Optimisation

Lecture 8 - Introduction to Non-Linear Programming

Fall semester - 2024

Dr. Eng. Valentin Leplat

Innopolis University

October 22, 2024

Outline

General Introduction

Why ?

Help to choose the **best** decision.

$$\left\{\begin{array}{cc} \text{Decision :} & \text{vector of variables } x \\ \text{Best :} & \text{objective function } f(x) \\ \text{Constraints :} & \text{feasible set } \mathcal{X} \end{array}\right\} \rightarrow \text{Optimization}$$

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

- Many applications in practice
- Efficient methods in practice
- Modelling and resolution of large-scale problems

Opening Remark and Credit

About more than 386 years ago.....In 1629, Fermat suggested the following:

- Give f, solve for x:
- $\lim_{d \to 0} \left[\frac{f(x+d)-f(x)}{d} \right] = 0$

...We can hardly expect to find a more general method to get the maximum or minimum points on a curve.....

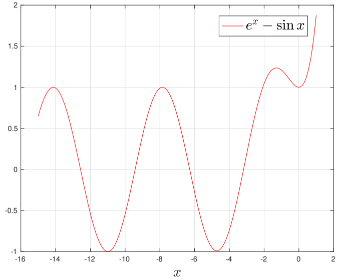
Why "numerical" optimization ?

- Say we want to find the minimum of :

$$f(x) := e^x + \cos x$$

- At some point: we want have to solve the following equation:

$$f'(x) = e^x - \sin x = 0$$



- Analytically: impossible
- Non-linear programming is dedicated to the study of numerical methods allowing to optimize functions.

Notations and formulation

Minimization of function $f : \mathbb{R}^n \to \mathbb{R}$ over a feasible set $\mathcal{X}$ writes as:

$$\begin{aligned} \min_{x} \quad & f(x) \\ \text{such that} \quad & x \in \mathcal{X} \end{aligned}$$

- The feasible set is usually defined with functional constraints:

$$\mathcal{X} = \{x \in \mathbb{R}^n | h_i(x) = 0 \quad \forall i \in \mathcal{E}, \quad h_j(x) = 0 \quad \forall j \in \mathcal{I}\}$$

- The constraints h_i, h_j are functions $h_i, h_j : \mathbb{R}^n \to \mathbb{R}$:
  - $h_i(x) = 0$ is an equality constraint
  - $h_j(x) \geqslant 0$ is an inequality constraint
- $\mathcal{E}$ and $\mathcal{I}$ are set of indices.

Notations and formulation

The general optimization problem can be therefore written as follows:

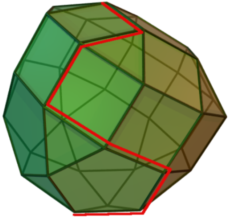
$$
\begin{aligned}
\min_{x} \quad & f(x) \\
\text{such that} \quad & h_i(x) = 0 \quad \forall i \in \mathcal{E} \\
& h_j(x) \geqslant 0 \quad \forall j \in \mathcal{I}
\end{aligned}
$$

Particular case: all the functions are linear !

- If all the functions $f(.), h_i(.), h_j(.)$ are linear, the general problem can be simplified as:

$$\begin{array}{ll} \min\limits_{x} & c^T x \\ \text{such that} & Ax \geqslant b \end{array}$$

- The constraints $Ax \geqslant b$ defines a feasible set that is a polyhedron (a bounded and non-empty polytope).
- The objective function $c^T x$ forms a translating hyperplane in space.



In most cases: an optimal solution will be one of the vertices of the polyhedron

Particular case: all the functions are linear !

Why the linear programming ?

- Because many problems can be modelled as linear programs
- Because there is a very efficient algorithm (the simplex algorithm) for solving these problems
- Because these problems have a very rich structure (properties of optimality, duality)

Particular case: all the functions are linear !

Why the linear programming ?

- Because many problems can be modelled as linear programs
- Because there is a very efficient algorithm (the simplex algorithm) for solving these problems
- Because these problems have a very rich structure (properties of optimality, duality)

Why studying non-linear programming ?

- Because some problems are impossible to model linearly
- Because the simplex algorithm for linear programming is inapplicable to non-linear problems

Recall of basics from analysis

Table of Contents

Blanket assumptions

This course: underlying space is the Euclidean space $\mathbb{R}^n$, a particular case of Hilbert space $\mathcal{X}$ of finite dimension n over the field $\mathbb{R}$ equipped with:

- an inner product $\langle ., . \rangle$, here we consider the dot product $\langle x, y \rangle = \sum_i^n x_i y_i = x^T y$ for $x, y \in \mathbb{R}^n$,
- induced norm $\|.\| = \sqrt{\langle ., . \rangle}$

Notation: x_i denotes the i-th component of x.

Gradient and Hessian matrix

Gradient

For a differentiable function f, the mapping $\nabla f(x) : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is called the gradient of f, and is defined as:

$$\nabla f(x) := \begin{pmatrix} \frac{\partial f(x)}{\partial x_1} \\ \cdots \\ \frac{\partial f(x)}{\partial x_n} \end{pmatrix}$$

It is the vector of all the partial derivatives, that gives the direction to follow to increase at most the function f at point x.

Gradient and Hessian matrix

Hessian

For a twice differentiable function f, the mapping $\nabla^2 f(x) : \mathbb{R}^n \rightarrow \mathbb{R}^{n\times n}$ is called the Hessian of f, and is defined as:

$$\nabla^2 f(x) := \begin{pmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n x_1} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{pmatrix}$$

Example: Compute the gradient and the Hessian of:

- $f(x_1, x_2) = \frac{1}{2}(x_1)^2 + 2(x_2)^2$
- $f(x_1, x_2, x_3) = e^{x_1} + (x_1)^2 x_3 - x_1 x_2 x_3$

{Positive (semi) definite, indefinite} matrix

Positive (Semi-) Definite Matrix

Let $A \in \mathbb{R}^{n \times n}$ be symmetric matrix.

- A is said to be Positive Definite:

$$\begin{aligned} A \succ 0 &\Leftrightarrow \lambda_i > 0 \quad \forall i \\ &\Leftrightarrow x^T A x > 0 \quad \forall x \in \mathbb{R}^n, \text{ and } x \neq 0 \end{aligned}$$

- A is said to be Semi-Positive Definite:

$$\begin{aligned} A \succcurlyeq 0 &\Leftrightarrow \lambda_i \geqslant 0 \quad \forall i \\ &\Leftrightarrow x^T A x \geqslant 0 \quad \forall x \in \mathbb{R}^n \end{aligned}$$

{Positive (semi) definite, indefinite} matrix

Positive (Semi-) Definite Matrix

Let $A \in \mathbb{R}^{n\times n}$ be symmetric matrix.

- A is said to be Positive Definite:

$$\begin{aligned} A \succ 0 &\Leftrightarrow \lambda_i > 0 \quad \forall i \\ &\Leftrightarrow x^T A x > 0 \quad \forall x \in \mathbb{R}^n, \text{ and } x \neq 0 \end{aligned}$$

- A is said to be Semi-Positive Definite:

$$\begin{aligned} A \succcurlyeq 0 &\Leftrightarrow \lambda_i \geqslant 0 \quad \forall i \\ &\Leftrightarrow x^T A x \geqslant 0 \quad \forall x \in \mathbb{R}^n \end{aligned}$$

Indefinite Matrix

A symmetric matrix $A \in \mathbb{R}^{n\times n}$ is said to be ”indefinite” if and only if it has at least one strictly positive eigenvalue and one strictly negative eigenvalue.

{Positive (semi) definite, indefinite} matrix

Examples:

1.
$$A = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix}$$

$\lambda_1 = 2,\ \lambda_2 = 2 - \sqrt{2},\ \lambda_3 = 2 + \sqrt{2}$
$x^T Ax = (x_1)^2 + (x_1 - x_2)^2 + (x_2 - x_3)^2 + (x_3)^2$

2.
$$A = \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix}$$

$\lambda_1 = 0,\ \lambda_2 = 3,\ \lambda_3 = 3$
$x^T Ax = (x_1 - x_2)^2 + (x_2 - x_3)^2 + (x_3 - x_1)^2$

Matrix notation for a quadratic form

A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be a quadratic form if it can be written in the form:

$$f(x) = \frac{1}{2}x^T Ax - b^T x + c,$$

where $A \in \mathbb{R}^{n\times n}$, $b \in \mathbb{R}^n$ and $c \in \mathbb{R}$. We then have:

- The gradient: $\nabla f(x) = Ax - b$,
- the Hessian matrix: $\nabla^2 f(x) = A$.

Matrix A can always be symmetric. If not, we have:

$$x^T Ax = \sum_i^n \sum_j^n a_{ij} x_i x_j = \sum_i^n \sum_j^n \frac{1}{2}(a_{ij} + a_{ji}) x_i x_j$$

Then, we build A' symmetric such that $a'_{ij} = a'_{ji} = \frac{1}{2}(a_{ij} + a_{ji})$, and $x^T Ax = x^T A' x$.

Matrix notation for a quadratic form

A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be a quadratic form if it can be written in the form:

$$f(x) = \frac{1}{2}x^T Ax - b^T x + c,$$

where $A \in \mathbb{R}^{n\times n}$, $b \in \mathbb{R}^n$ and $c \in \mathbb{R}$. We then have:

- The gradient: $\nabla f(x) = Ax - b$,
- the Hessian matrix: $\nabla^2 f(x) = A$.

Matrix A can always be symmetric. If not, we have:

$$x^T Ax = \sum_i^n \sum_j^n a_{ij}x_i x_j = \sum_i^n \sum_j^n \frac{1}{2}(a_{ij} + a_{ji})x_i x_j$$

Then, we build A' symmetric such that $a'_{ij} = a'_{ji} = \frac{1}{2}(a_{ij} + a_{ji})$, and $x^T Ax = x^T A' x$.
Exercise: how to write $f(x_1, x_2) = 3(x_1)^2 + 4x_1x_2 + 3(x_2)^2$?

Taylor expansion at order 1 and 2

- The Taylor expansion is an essential tool for characterizing the optimality. [1]
- for a real multivariate function $f : \mathbb{R}^n \to \mathbb{R}$, we have:

$$f(x) = P(x) + R(x)$$

- For a Taylor expansion built around a and limited at order k, the residual term $R(x) \sim (x-a)^{k+1}$
- When $x \to a$, then $R(x)$ becomes negligeable compared to $(x-a)^k$, we denote this by $R(x) = o(||x-a||^k)$ (notation of Landau).

[1] It is then required that the functions (objective and constraints) of the Problem to solve are enough differentiable !

Taylor expansion at order 1 and 2

- for a real multivariate function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$, we have:

$$f(x) = P(x) + R(x)$$

- Around the point a, we have:
  1. for $k = 1$:

$$f(x) = f(a) + (x-a)^T \nabla f(a) + o(||x-a||)$$
$$f(a+\alpha d) = f(a) + \alpha d^T \nabla f(a) + o(||\alpha d||)$$

Taylor expansion at order 1 and 2

- for a real multivariate function $f : \mathbb{R}^n \to \mathbb{R}$, we have:

$$f(x) = P(x) + R(x)$$

- Around the point a, we have:
  1. for $k = 1$:
$$f(x) = f(a) + (x-a)^T \nabla f(a) + o(||x-a||)$$
$$f(a+\alpha d) = f(a) + \alpha d^T \nabla f(a) + o(||\alpha d||)$$
  2. for $k = 2$:
$$f(x) = f(a) + (x-a)^T \nabla f(a) + \frac{1}{2}(x-a)^T \nabla^2 f(a)(x-a) + o(||x-a||^2)$$
$$f(a+\alpha d) = f(a) + \alpha d^T \nabla f(a) + \frac{\alpha^2}{2} d^T \nabla^2 f(a) d + o(||\alpha d||^2)$$

Details on the Landau notation $o(.)$

- When $x \to \infty$, $f(x) = o(g(x))$ means that f is dominated asymptotically by g, that is:

$$\forall \epsilon > 0, \exists K | \forall x > K : |f(x)| \leqslant \epsilon |g(x)|.$$

  In other words:

$$\lim_{x \to \infty} \frac{f(x)}{g(x)} = 0$$

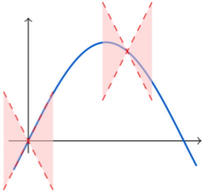
- When $x \to 0$, $f(x) = o(g(x))$ means that f tends to zero faster than $g(x)$. In other words:

$$\lim_{x \to 0} \frac{f(x)}{g(x)} = 0$$

For a Taylor expansion of order k, this means:

$$\lim_{x \to a} \frac{R(x)}{(||x - a||^k)} = 0$$

Continuity in the Lispschitz sense



A function f is Lispschitz continuous on its domain if there exists a constant $L > 0$ such that for any $x, y \in \mathrm{dom} f$, we have:

$$|f(x) - f(y)| \leqslant L||x - y||$$

Insights:

- Such a function is limited in how fast it can change, L is an upper bound to the maximum steepness of $f(x)$
- Stronger than continuous, weaker than continuously differentiable
  Example: $f(x) = |x|$ is Lipschitz continuous but not continuously differentiable (in 0).

Globally, we have no hope...

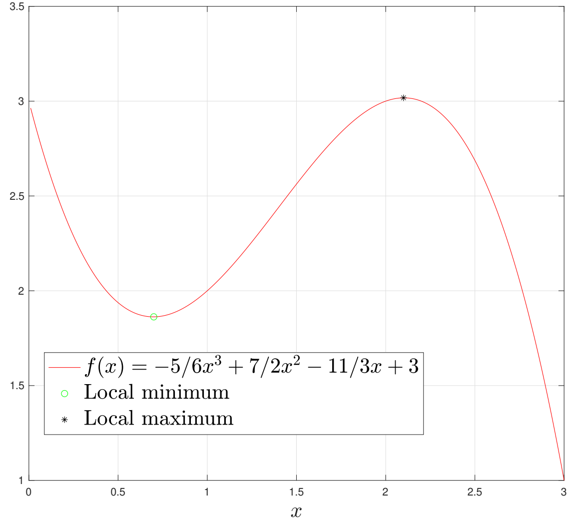
Table of Contents

Global minima and local minima

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

- $x^\star$ is a global minimum if and only if $x^\star \in \mathcal{X}$ and $f(x) \geqslant f(x^\star)$ for all $x \in \mathcal{X}$.
- $x^\star$ is a local minimum if and only if
  1. $x^\star \in \mathcal{X}$
  2. $\exists$ a neighborhood V around $x^\star$ such that $f(x) \geqslant f(x^\star)$ for all $x \in \mathcal{X} \cap V$.
- $x^\star$ is a strict local minimum if and only if
  1. $x^\star \in \mathcal{X}$
  2. $\exists$ a neighborhood V around $x^\star$ such that $f(x) > f(x^\star)$ for all $x \in \mathcal{X} \cap V$ and $x \neq x^\star$.

Global minima and local minima



A function with a global minimum difficult to find

- We show here that finding the global minimum for any problem is tough.
- **How ?**: we compute the minimum number of points at which we need to evaluate the objective function
- The example is illustrative, and we make some assumptions:
  1. the function $f : \mathbb{R}^n \to \mathbb{R}$ is Lipschitz continuous with constant L.
  2. We use only a method of order "zero" (we just use the objective function values at different points, no use of derivatives).
  3. The work space is the ball $([0,1]^n)$, that is $\mathbb{B}_n = \{x \in \mathbb{R}^n | 0 \leqslant x_i \leqslant 1, i = 1, ...n\}$
- The function is the worst possible one: for any x evaluated, the answer is the same: $f(x) = 0$.

A function with a global minimum difficult to find

- $p \geqslant 1$ is an integer input parameter
- Form p^n points $x^{(i_1,\ldots,i_n)} = (\frac{1}{2p} + \frac{i_1}{p}, \ldots, \frac{1}{2p} + \frac{i_n}{p})$ where $i_1 = 0, \ldots, p-1, \ldots, i_n = 0, \ldots, p-1$.
- We obtain a uniform grid where two "consecutive points" are separated by $1/p$ along each dimension.
- **Question**: *In order to have an error at most equal to ϵ ($|f(x) - f^\star| \leqslant \epsilon$), what is the number of points to evaluate ?*

A function with a global minimum difficult to find

- Recall that for any x, anyway the function returns $f(x) = 0$.
- The idea is to tune p (and then know how many points p^n we need) such that the error $|f(x) - f^\star| = |0 - f^\star| = |f^\star| \leqslant \epsilon$
- To help us:
  1. The objective function $f(.)$ is Lipschitz continuous on $\mathbb{B}_n$

  $$|f(x) - f(y)| \leqslant L\|x - y\|_\infty, \forall x, y \in \mathbb{B}_n$$

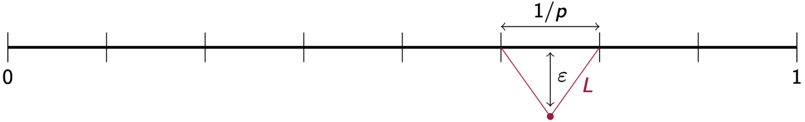
  2. It is clear that $\mathbb{B}_n = \bigcup_{(i)} \mathcal{B}_\infty(x^{(i)}, \frac{1}{2p})$ where $\mathcal{B}_\infty(x, r) = \{y \in \mathbb{R}^n | \|y - x\|_\infty \leqslant r\}$
  3. Let $x^\star$ be the global minimum of our problem. Then there exists an index $i^\star = (i_1^\star, ..., i_n^\star)$ such that

  $$\|x^{(i^\star)} - x^\star\|_\infty \leqslant \frac{1}{2p}$$

  4. The point $x^{(i^\star)}$ belongs to the grid, hence $f(x^{(i^\star)}) = 0$.
  5. To ease the notation, recall that $f^\star = f(x^\star)$

A function with a global minimum difficult to find

Example for $n = 1$:



- Since the slope of the function $f(.)$ is bounded by L, the lower bound for $f()$ somewhere inside a "box" is $-\frac{L}{2p}$, and we want that

$$-\frac{L}{2p} \geqslant -\epsilon \Leftrightarrow \frac{L}{2p} \leqslant \epsilon \tag{1}$$

- The choice $p = \frac{L}{2\epsilon}$ satisfies our goal $|f^\star| \leqslant \epsilon$, indeed, we have the following chain

$$|f(x^{(i^\star)}) - f(x^\star)| = |f^\star| \leqslant L\|x^{(i^\star)} - x^\star\|_\infty \leqslant \frac{L}{2p} = \epsilon \tag{2}$$

A function with a global minimum difficult to find

For a function with n variables, we find that the number of points should be:

$$p^n = \left(\frac{L}{2\epsilon}\right)^n$$

Example

- Let $n = 10$ variables and $L = 2$.
- We want an accuracy $\epsilon = 0.001$

$$\left(\frac{L}{2\epsilon}\right)^n = \left(\frac{2}{2 \times 0.001}\right)^{10} = 10^{30} \text{ evaluations of the function}$$

With 10^{12} evaluations/second, we need 32.000.000.000 years :)

- There is no hope to find a method able to find the global minimum of any Problem !
- A universal algorithm does not exist, it is anyway possible to solve efficiently many classes of problems with dedicated methods.

Optimality Conditions

Why optimal conditions in optimization?

$$\min_x f(x)$$

$$\text{s.t.} \quad \begin{aligned} h_i(x) &= 0 \text{ for } i \in \mathcal{E} \\ h_i(x) &\geqslant 0 \text{ for } i \in \mathcal{I} \end{aligned}$$

It's easy to check that a given solution is feasible, but how do you check that it's optimal?
→ conditions of optimality!

- A set of OCs is used to characterize optimal solutions
  - by using a system of equalities and/or inequalities (easy to check)
  - rather than a definition involving a quantifier $\forall$ or $\exists$ (difficult to check)
- According to the versions, the optimal conditions characterize
  - either local solutions (most often)
  - or only the global solutions (but this requires additional hypotheses on the on the problem; for example convexity)

Optimality conditions in optimization

Generally speaking, in an optimization problem, we're looking for $x^\star$.

If we define :

- P : Satisfy OCs.
- Q : Be an optimal solution

Ideally, it would be interesting to have $P \Leftrightarrow Q$, i.e. :

- satisfying the OCs is necessary ($Q \Rightarrow P$) :
  all optimal solutions satisfy OCs
- satisfying OCs is sufficient ($P \Rightarrow Q$) :
  all points verifying OCs are optimal solutions

Optimality conditions in optimization

Unfortunately, in most cases, OCs are either :

- necessary but not sufficient: among the points that satisfy the OCs, we find **all** optimal solutions, but also non-optimal solutions.
- sufficient but not necessary: among the points that satisfy the OCs, we only find optimal solutions, but some other optimal solutions do not satisfy the OCs.

Necessary conditions

In the unconstrained case, the admissible domain of the problem is $\mathcal{X} = \mathbb{R}^n$, hence :

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

The following conditions characterize local optimal solutions.

First-order necessary condition.

Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable.

$$x^\star \text{ local minimum of } f \Rightarrow \nabla f(x^\star) = 0.$$

Second-order necessary condition.

Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice differentiable.

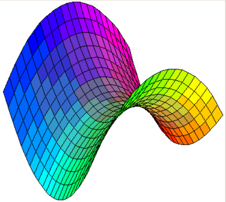
$$x^\star \text{ local minimum of } f \Rightarrow \nabla^2 f(x^\star) \succeq 0.$$

These conditions are necessary, but not sufficient: among the points verifying these conditions are minima, maxima and points that are neither.

Stationary points and saddle points

A stationary point (or critical point) of a function f is a point $x^\star$ satisfying the first-order optimality conditions.
In the unconstrained case, this means $\nabla f(x^\star) = 0$.

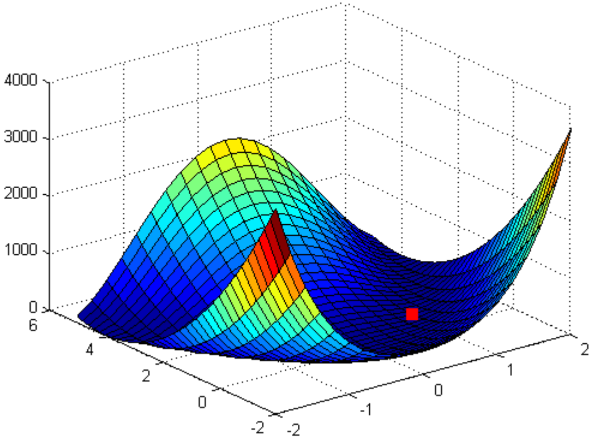
A saddle point of a function f is a stationary point that is neither a local minimum nor a local maximum. Intuitively, this means that there is a downward and an upward direction:



A sufficient condition for $x^\star$ to be a saddle point is that $\nabla^2 f(x^\star)$ has one strictly positive and one strictly negative eigenvalue.

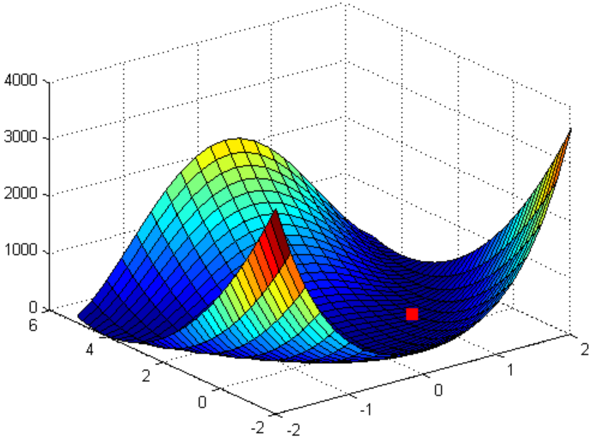
Necessary conditions: illustration 1

$f(x_1, x_2) = (1 - x_1)^2 + 100(x_2 - x_1^2)^2$ has a local minimum in $x^\star = (1, 1)^T$



Necessary conditions: illustration 1

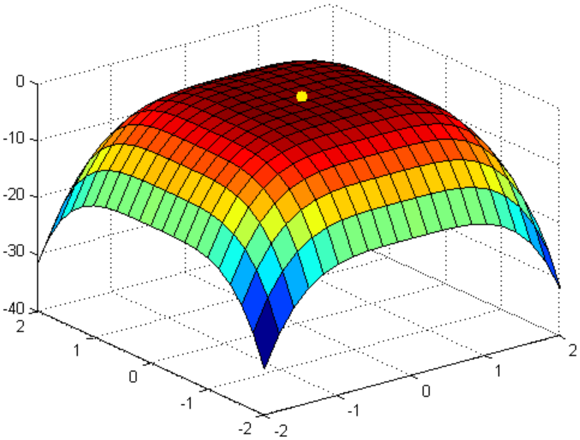
$$f(x_1, x_2) = (1 - x_1)^2 + 100(x_2 - x_1^2)^2 \text{ has a local minimum in } x^\star = (1, 1)^T$$



We have $\nabla f(x^\star) = 0$ and $\nabla^2 f(x^\star) \succeq 0$.

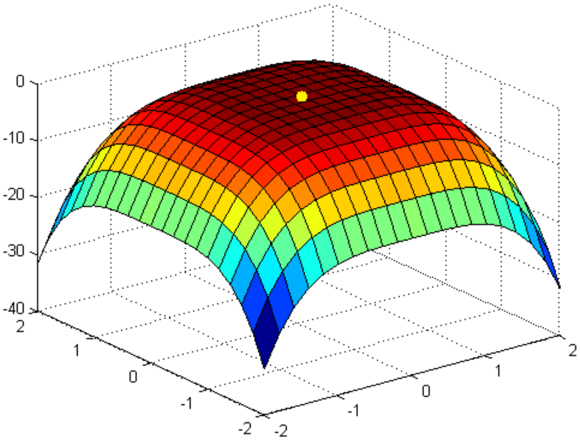
Necessary conditions: illustration 2

$$f(x_1, x_2) = -x_1^4 - x_2^4$$



Necessary conditions: illustration 2

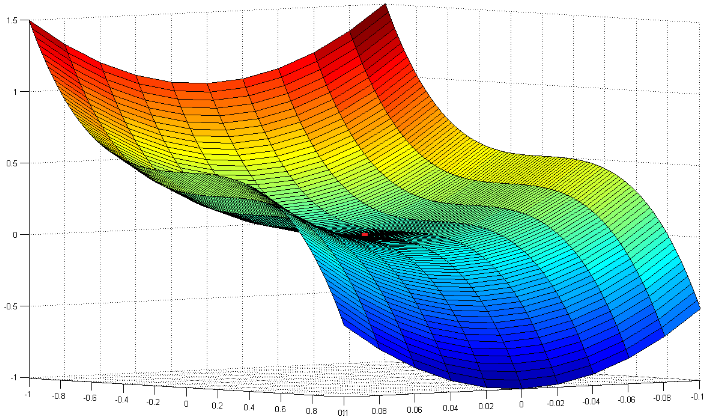
$$f(x_1, x_2) = -x_1^4 - x_2^4$$



For $x^\star = (0,0)^T$, we have $\nabla f(x^\star) = 0$ and $\nabla^2 f(x^\star) \succeq 0$, and yet $x^\star$ is not a local minimum (it's even a maximum in this case).

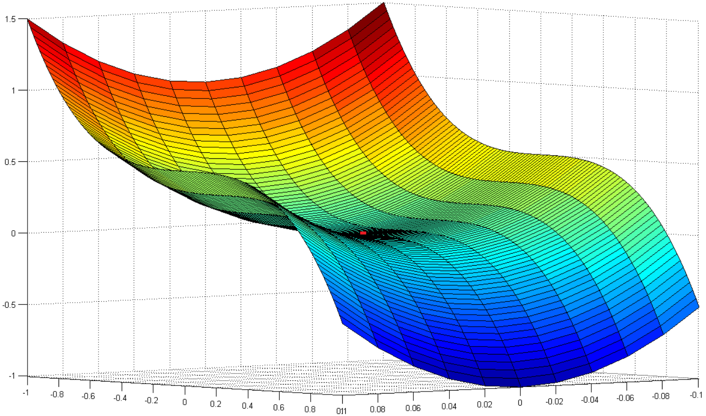
Necessary conditions: illustration 3

$$f(x_1, x_2) = 50x_1^2 - x_2^3$$



Necessary conditions: illustration 3

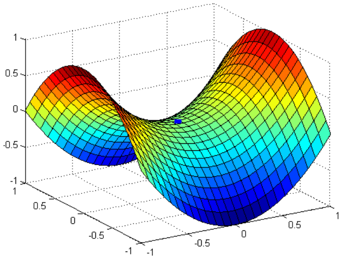
$$f(x_1, x_2) = 50x_1^2 - x_2^3$$



For $x^\star = (0, 0)^T$, we have $\nabla f(x^\star) = 0$ and $\nabla^2 f(x^\star) \succeq 0$, yet $x^\star$ is not a local minimum (it's a saddle point in this case).

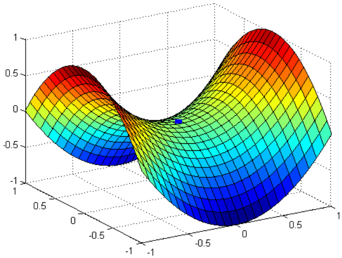
Necessary conditions: illustration 4

$$f(x_1, x_2) = x_1^2 - x_2^2$$



Necessary conditions: illustration 4

$$f(x_1, x_2) = x_1^2 - x_2^2$$



For $x^\star = (0, 0)^T$:

- we have $\nabla f(x^\star) = 0$ ($x^\star$ is therefore a stationary point) but the necessary conditions for second-order optimality are not satisfied: $\nabla^2 f(x^\star)$ is undefined.

  By contrapositive, we can therefore conclude that $x^\star$ is not a local minimum.

- Since $\nabla^2 f(x^\star)$ has a positive and a negative eigenvalue, this is sufficient to conclude that $x^\star$ is a saddle point.

Sufficient conditions for an unconstrained problem

Sufficient conditions.

Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice differentiable.

$$\text{If } \nabla f(x^\star) = 0 \text{ and } \nabla^2 f(x^\star) \succ 0, \text{ then } x^\star \text{ is a strict local minimum of } f.$$

Sufficient conditions for an unconstrained problem

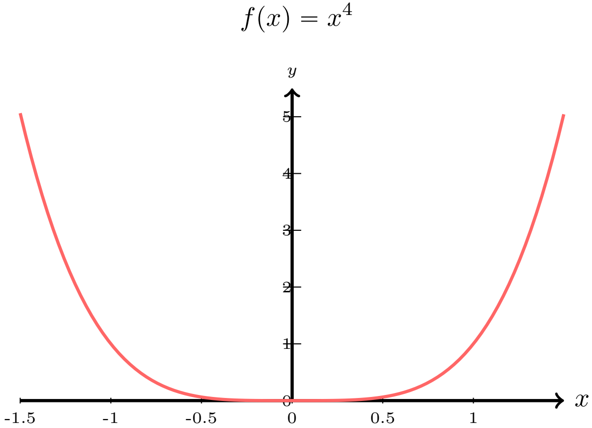
Sufficient conditions.

Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice differentiable.

$$\text{If } \nabla f(x^\star) = 0 \text{ and } \nabla^2 f(x^\star) \succ 0, \text{ then } x^\star \text{ is a strict local minimum of } f.$$

- These sufficient conditions guarantee a stronger result than the necessary conditions: the local minimum is strict.
- These sufficient conditions are not necessary: a point $x^\star$ can be a strict local minimum and still not satisfy these conditions (see example below).

Sufficient conditions: illustration



For $x^\star = 0$, we have $f'(x^\star) = 0$ ($x^\star$ is therefore a stationary point) and $f''(x^\star) = 0$.
The sufficient conditions for optimality are not satisfied, yet $x^\star$ is a strict local minimum.

Summary

First and second order necessary conditions

If f is differentiable in a neighborhood of an $x^\star$ point

$$x^\star \text{ local minimum} \Rightarrow \nabla f(x^\star) = 0$$

If f is twice differentiable in a neighbourhood of $x^\star$

$$x^\star \text{ local minimum} \Rightarrow \nabla^2 f(x^\star) \succeq 0$$

Sufficient conditions

If f is twice differentiable in a neighbourhood of a point $x^\star$.

$$\nabla f(x^\star) = 0 \text{ and } \nabla^2 f(x^\star) \succ 0 \Rightarrow x^* \text{ strict local minimum}$$

How to use of optimality conditions ?

Recipe

1. Compute $\nabla f(x)$ and $\nabla^2 f(x)$.
2. Identify stationary points.
3. Eliminate those that do not meet the necessary optimality conditions.
4. Identify those that verify the sufficient conditions for optimality.

How to use of optimality conditions ?

Recipe

1. Compute $\nabla f(x)$ and $\nabla^2 f(x)$.
2. Identify stationary points.
3. Eliminate those that do not meet the necessary optimality conditions.
4. Identify those that verify the sufficient conditions for optimality.

(points 3 and 4 do not cover all possible cases: there may always be a stationary point $x^\star$ with $\nabla^2 f(x^\star)$ possessing at least one zero eigenvalue, in which case nothing can be concluded for such a point.)

How to use of optimality conditions ?

Recipe

1. Compute $\nabla f(x)$ and $\nabla^2 f(x)$.
2. Identify stationary points.
3. Eliminate those that do not meet the necessary optimality conditions.
4. Identify those that verify the sufficient conditions for optimality.

(points 3 and 4 do not cover all possible cases: there may always be a stationary point $x^\star$ with $\nabla^2 f(x^\star)$ possessing at least one zero eigenvalue, in which case nothing can be concluded for such a point.)

Summary exercise.
Calculate the stationary points of the following function and characterize them, if possible, using the necessary and sufficient conditions for optimality.

$$f(x_1, x_2) = x_1 x_2^2 - 12x_2 - 6x_1 x_2 - 4x_1 + \frac{x_1^3}{3}$$

What about global solutions ?

- If a function has a global minimum, it is necessarily found among the local minima.
  $\rightarrow$ identify all local minima and select among them the one or more possessing the smallest value of the objective function

What about global solutions ?

- If a function has a global minimum, it is necessarily found among the local minima.
  $\rightarrow$ identify all local minima and select among them the one or more possessing the smallest value of the objective function
- However, a function does not necessarily have a global minimum.

What about global solutions ?

- If a function has a global minimum, it is necessarily found among the local minima.
  $\rightarrow$ identify all local minima and select among them the one or more possessing the smallest value of the objective function
- However, a function does not necessarily have a global minimum.
  - A function can be unbounded ($\rightarrow -\infty$)

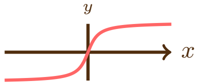
What about global solutions ?

- If a function has a global minimum, it is necessarily found among the local minima.

  $\rightarrow$ identify all local minima and select among them the one or more possessing the smallest value of the objective function

- However, a function does not necessarily have a global minimum.
  - A function can be unbounded ($\rightarrow -\infty$)
  - Even a bounded function may not have a global minimum :

$$f(x) = \arctan x \text{ or } f(x, y) = e^{-x^2-y^2}$$

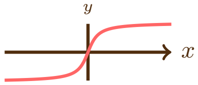


What about global solutions ?

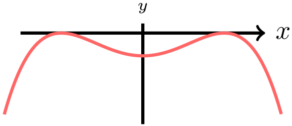
- If a function has a global minimum, it is necessarily found among the local minima.

  $\rightarrow$ identify all local minima and select among them the one or more possessing the smallest value of the objective function
- However, a function does not necessarily have a global minimum.
  - A function can be unbounded ($\rightarrow -\infty$)
  - Even a bounded function may not have a global minimum :

$$f(x) = \arctan x \text{ or } f(x, y) = e^{-x^2 - y^2}$$



- Caution : the smallest local minimum is not automatically a global minimum, as in the case of $f(x) = -(x^2 - 1)^2$:



A first condition for a global minimum

- In the best case: the conditions described above can only guarantee the local optimality of stationary points, since they only exploit local information: the values of the gradient and the Hessian matrix at a given point.
- The conditions that ensure global optimality must use global information.
  For example, in the result below, the Hessian matrix is positive semidefinite over the entire domain.

Condition for global optimality.
Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice differentiable.

$$\text{If } \nabla f(x^\star) = 0 \text{ and } \nabla^2 f(x) \succeq 0 \quad \forall x \in \mathbb{R}^n, \text{ then } x^\star \text{ is a global minimum of } f.$$

Example

Example Let $f(x) = x_1^2 + x_2^2 + x_3^2 + x_1x_2 + x_1x_3 + x_2x_3 + (x_1^2 + x_2^2 + x_3^2)^2$. We have :

$$\nabla f(x) = \begin{pmatrix} 2x_1 + x_2 + x_3 + 4x_1(x_1^2 + x_2^2 + x_3^2) \\ 2x_2 + x_1 + x_3 + 4x_2(x_1^2 + x_2^2 + x_3^2) \\ 2x_3 + x_1 + x_1 + 4x_3(x_1^2 + x_2^2 + x_3^2) \end{pmatrix}, \text{ and}$$

$$\nabla^2 f(x) = \begin{pmatrix} 2 + 4(x_1^2 + x_2^2 + x_3^2) + 8x_1^2 & 1 + 8x_1x_2 & 1 + 8x_1x_3 \\ 1 + 8x_1x_2 & 2 + 4(x_1^2 + x_2^2 + x_3^2) + 8x_2^2 & 1 + 8x_2x_3 \\ 1 + 8x_1x_3 & 1 + 8x_2x_3 & 2 + 4(x_1^2 + x_2^2 + x_3^2) + 8x_3^2 \end{pmatrix}$$

$x^\star = (0, 0, 0)$ is a stationary point. If we show that $\nabla^2 f(x) \succeq 0 \ \forall x \in \mathbb{R}^3$, it implies $x^\star$ is global minimum. First, let's note that $\nabla^2 f(x) = A + B(x) + C(x)$ with

$$A = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}, \ B(x) = 4(x_1^2 + x_2^2 + x_3^2)\mathbf{I}_3, \text{ and } C(x) = \begin{pmatrix} 8x_1^2 & 8x_1x_2 & 8x_1x_3 \\ 8x_1x_3 & 8x_2^2 & 8x_2x_3 \\ 8x_1x_3 & 8x_2x_3 & 8x_3^2 \end{pmatrix}.$$

A and B are positive semidefinites because they are dominant diagonals. C can be written as $8xx^T$ and is therefore also positive semidefinite. Hence $\nabla^2 f(x) \succeq 0 \ \forall x \in \mathbb{R}^3$, and then $x^\star$ is global minimum.

Application to quadratic problems

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

$$f(x) = \frac{1}{2}x^T Ax - b^T x + c,$$

where $A \in \mathbb{R}^{n\times n}$ is symmetric matrix, $b \in \mathbb{R}^n$ a vector and c a scalar.

Application to quadratic problems

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

$$f(x) = \frac{1}{2}x^T Ax - b^T x + c,$$

where $A \in \mathbb{R}^{n\times n}$ is symmetric matrix, $b \in \mathbb{R}^n$ a vector and c a scalar.
We have :

- $\nabla f(x) = Ax - b$. The stationary points of f are therefore the solutions of the linear system $Ax = b$ (if they exist).
- $\nabla^2 f(x) = A$. The Hessian matrix is the same $\forall x \in \mathbb{R}^n$.

Application to quadratic problems

Application of OCs :

Application to quadratic problems

Application of OCs :

- By the contrapositive of NC1, if the system $Ax = b$ has no solution, then there is no local minimum.

Application to quadratic problems

Application of OCs :

- By the contrapositive of NC1, if the system $Ax = b$ has no solution, then there is no local minimum.
- By the contrapositive of NC2, if A is not positive semidefinite, then there is no local minimum.

Application to quadratic problems

Application of OCs :

- By the contrapositive of NC1, if the system $Ax = b$ has no solution, then there is no local minimum.
- By the contrapositive of NC2, if A is not positive semidefinite, then there is no local minimum.
- By the SC of local optimality (see slide 42), if A is positive definite, then the solution of the system $Ax = b$ is the unique global minimum of f.

Application to quadratic problems

Application of OCs :

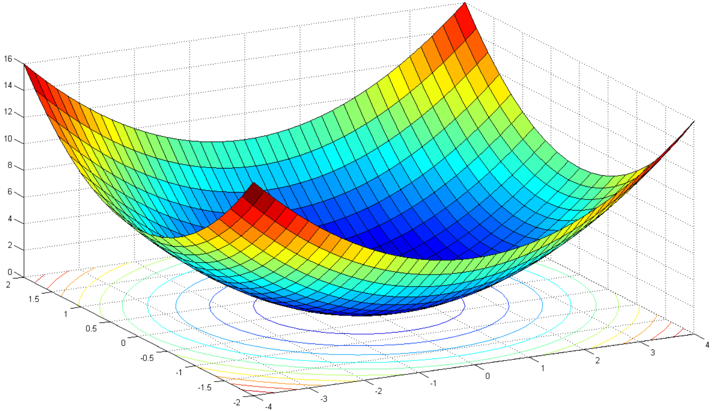
- By the contrapositive of NC1, if the system $Ax = b$ has no solution, then there is no local minimum.
- By the contrapositive of NC2, if A is not positive semidefinite, then there is no local minimum.
- By the SC of local optimality (see slide 42), if A is positive definite, then the solution of the system $Ax = b$ is the unique global minimum of f.
- By the SC of global optimality (see slide 47), if A is positive semidefinite, then the solutions of the system $Ax = b$ are global minima.

Application to quadratic problems: illustration 1

A positive definite :

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix}$$

$$\lambda_1 = 1 \text{ and } \lambda_2 = 4$$

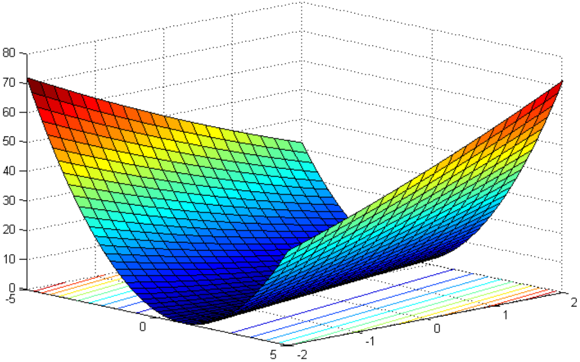


Application to quadratic problems: illustration 2

A positive semidefinite :

$$A = \begin{pmatrix} 4 & 2 \\ 2 & 1 \end{pmatrix}$$

$$\lambda_1 = 0 \text{ and } \lambda_2 = 5$$

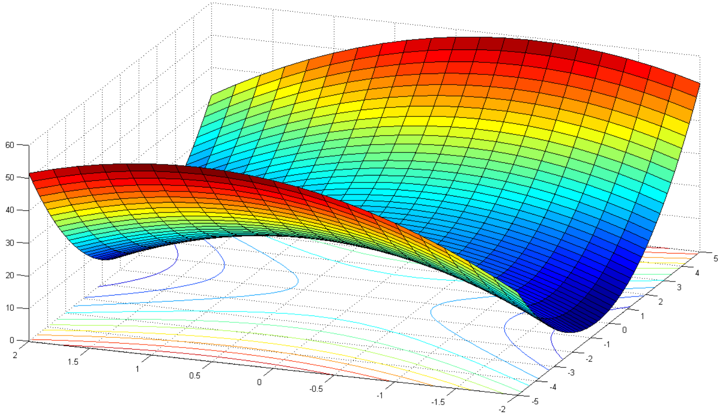


Application to quadratic problems: illustration 3

A indefinite :

$$A = \begin{pmatrix} 3 & -1 \\ -1 & -8 \end{pmatrix}$$

$$\lambda_1 = -8.09 \text{ and } \lambda_2 = 3.09$$



Application to quadratic problems: illustration 4

How can we interpret the quadratic function defined by :

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \text{ and } b = \begin{pmatrix} 1 \\ -1 \end{pmatrix},$$

(the eigenvalues of A are 0 and 2 but the system $Ax = b$ has no solution).

Conclusions

Summary

We have seen:

- the definition of *global* and *local* minima.
- The complexity bound for global optimization (using 0-order oracle).
- Definition of a *stationary* point, and a *saddle* point.
- Optimality conditions:
    1. Necessary conditions: first and second order.
    2. Sufficient conditions.
- How to use the OCs to identify local minima ?
- Condition for Global optimality.
- The quadratic case.

Preparations for the next lecture

- Review the lecture;
- The lab focuses on computing the stationary points, the local and potential global minima of functions $f : \mathbb{R}^n \to \mathbb{R}$ with $n = 2, 3$ (unconstrained case).
- Optimality conditions (necessary and sufficient) must be mastered before moving on to the constrained cases.

Goodbye, So Soon

THANKS FOR THE ATTENTION

- v.leplat@innopolis.ru
- sites.google.com/view/valentinleplat/