Numerical optimization

Mines Nancy – Fall 2024

session 1 – general introduction

Lecturer: Julien Flamant (CNRS, CRAN)

julien.flamant@univ-lorraine.fr

Office 425, FST 1er cycle

Course material:

name-here

github.com/jflamant/mines-nancy-fall24-optimization

Outline of the course

36h of classes, including lectures, exercices and lab work (Python)

Course material in english, lecture given in french or english

Main results will be on lecture slides, but detailed computations, proofs and exercises solution will not, in general. Take notes!

Prerequisites: elementary differential calculus and linear algebra

Evaluation

- intermediate exams (40%)
- final exam (60%)

Lecturers Julien Flamant (CNRS) and Christophe Zhang (INRIA)
Contact info julien.flamant@univ-lorraine.fr

What to expect from this course

in short: an introduction to numerical optimization

Objectives

- learn the basic language of optimization;
- characterize objective functions and critical points
- solve least squares problems;
- learn about classical first and second-order descent algorithms
- first principles and algorithms in constrained optimization

hands-on approach with Python notebooks

Some useful references

Books

- D. Bertsekas (1999). *Nonlinear programming*. Second edition. Belmont, MA: Athena Scientific
- J. Nocedal and S. J. Wright (1999). *Numerical optimization*. Springer
- S. P. Boyd and L. Vandenberghe (2004). *Convex optimization*. Cambridge university press
- A. Beck (2017). *First-order methods in optimization*. SIAM

Online material (include lecture slides and notebooks)

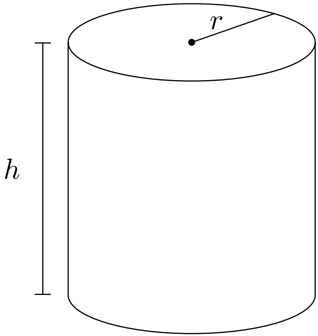
- MIT OpenCourseWare | Optimization (Boyd and Bertsekas courses)
- Numerical tours | Optimization

Outline

A first example

A can produce wants to optimize his consumption of raw material.

The only constraint is that their volume $V = 1\text{L}$ is fixed.



How to formulate this shape optimization problem?

quantity of raw material $\equiv$ area of cylinder $\times$ constant thickness

$\rightarrow$ how to minimize the area A of a cylinder with given volume V?

Formulation of the optimization problem

Expression of the area A (objective function or criterion)

$$A(r,h) = A_{\text{cylinder}}(r,h) + A_{\text{lids}}(r,h) = 2\pi rh + 2\pi r^2$$

Formulation of the optimization problem

Expression of the area A (objective function or criterion)

$$A(r,h) = A_{\text{cylinder}}(r,h) + A_{\text{lids}}(r,h) = 2\pi rh + 2\pi r^2$$

The dimensions r and h are the variables of the optimization problem

r and h must be non-negative → constraint $\quad r \geq 0, h \geq 0$

Formulation of the optimization problem

Expression of the area A (objective function or criterion)

$$A(r,h) = A_{\text{cylinder}}(r,h) + A_{\text{lids}}(r,h) = 2\pi rh + 2\pi r^2$$

The dimensions r and h are the variables of the optimization problem

r and h must be non-negative $\rightarrow$ constraint $\quad r \geq 0, h \geq 0$

The condition of constant volume is also a constraint.

$$V(r,h) = \pi r^2 h = V_0$$

Formulation of the optimization problem

Expression of the area A (objective function or criterion)

$$A(r,h) = A_{\text{cylinder}}(r,h) + A_{\text{lids}}(r,h) = 2\pi rh + 2\pi r^2$$

The dimensions r and h are the variables of the optimization problem

$$r \text{ and } h \text{ must be non-negative} \rightarrow \text{constraint} \quad r \geq 0, h \geq 0$$

The condition of constant volume is also a constraint.

$$V(r,h) = \pi r^2 h = V_0$$

The constrained optimization problem is finally written as

$$\text{Find } (\hat{r}, \hat{h}) \text{ such that } A(\hat{r}, \hat{h}) \text{ is minimal and } \begin{cases} r \geq 0 \text{ and } h \geq 0 \\ \pi r^2 h = V_0 \end{cases}$$

Simplification of the problem

We notice that h and r are related by the constraint $V_0 = \pi r^2 h$.
$\rightarrow$ The area $A(r, h)$ can be rewritten as a function of r only.

$$h(r) = \frac{V_0}{\pi r^2} \quad (h \geq 0 \quad \text{OK!})$$

Simplification of the problem

We notice that h and r are related by the constraint $V_0 = \pi r^2 h$.
$\rightarrow$ The area $A(r,h)$ can be rewritten as a function of r only.

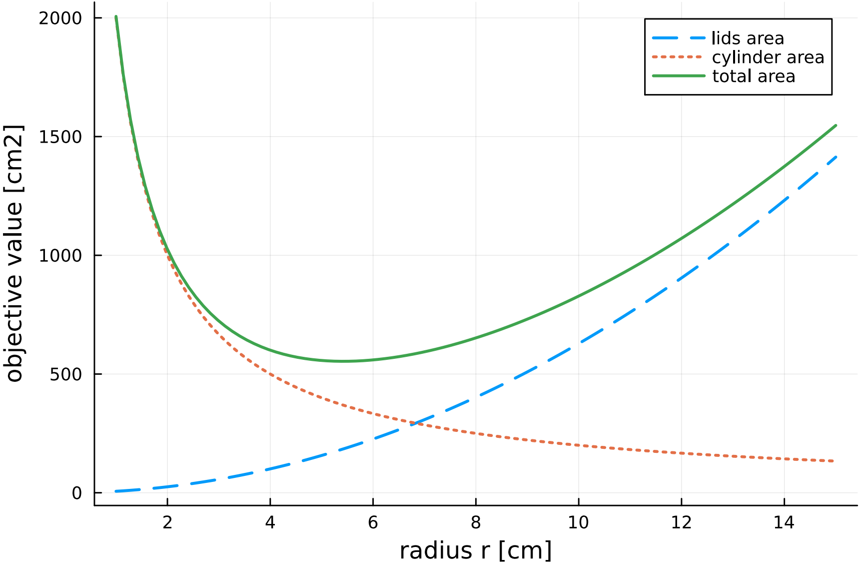
$$h(r) = \frac{V_0}{\pi r^2} \quad (h \geq 0 \quad \text{OK!})$$

By substituting this into the expression for $A(r,h)$, we then obtain

$$\begin{aligned} A(r,h) &= 2\pi r h + 2\pi r^2 \\ &= 2\pi r \frac{V_0}{\pi r^2} + 2\pi r^2 \\ &= \frac{2V_0}{r} + 2\pi r^2 \\ &:= \tilde{A}(r) \end{aligned}$$

The volume constraint is directly taken into account in the new objective function $\tilde{A}(r)$. We consider the constraint $r \geq 0$ when searching for the minimum of $\tilde{A}(r)$ over $\mathbb{R}^+$.

Objective function of the problem ($V_0 = 1\text{L}$)



Explicit solution

The derivative of the objective function vanishes at its minimum.

$$\frac{\mathrm{d}\tilde{A}(r)}{\mathrm{d}r} = 4\pi r - \frac{2V_0}{r^2}$$

We solve

$$\left.\frac{\mathrm{d}\tilde{A}(r)}{\mathrm{d}r}\right|_{r=\hat{r}} = 0 \Leftrightarrow 4\pi\hat{r} = \frac{2V_0}{\hat{r}^2}$$

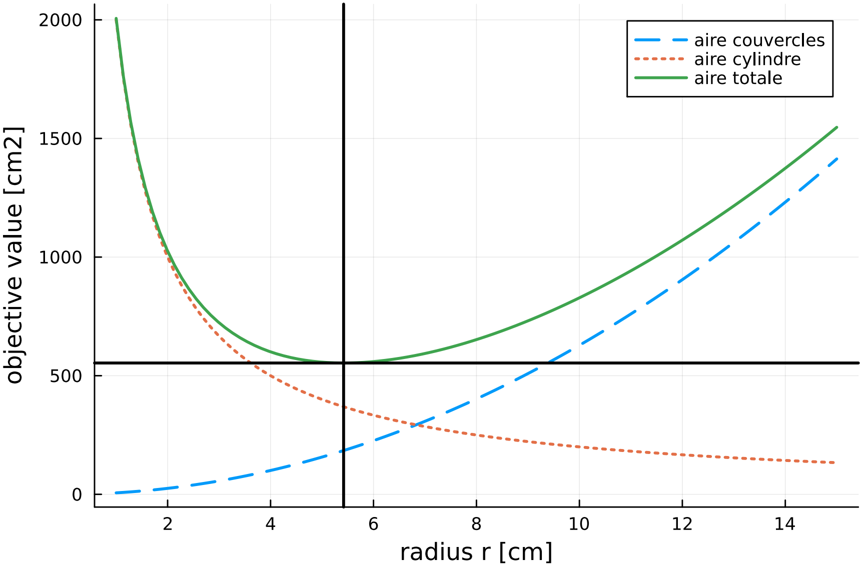
$$\Leftrightarrow \hat{r} = \left[\frac{V_0}{2\pi}\right]^{1/3}$$

In summary, the solutions to the optimization problem are

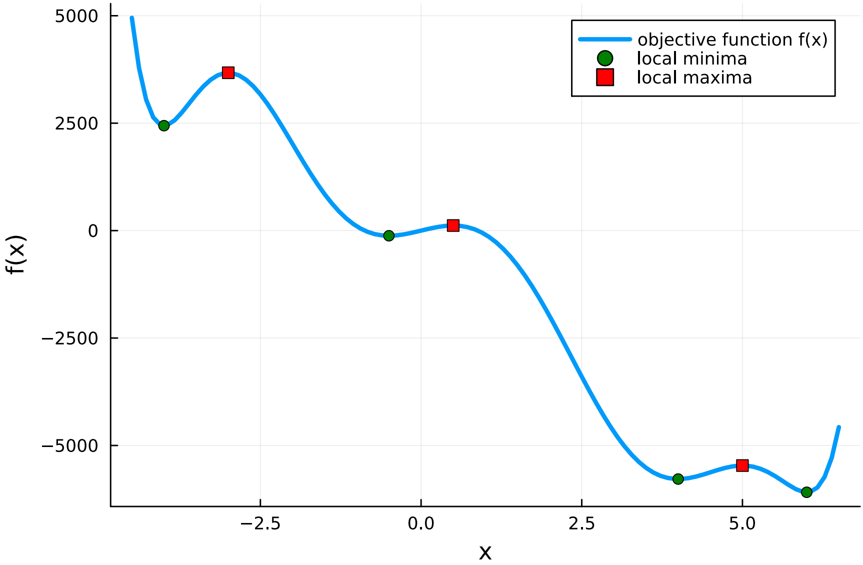
$$\hat{r} = \left[\frac{V_0}{2\pi}\right]^{1/3} \qquad \hat{h} = \frac{V_0}{\pi\hat{r}^2} = 2\hat{r}$$

In this simple case, the solution is explicit and does not require a resolution algorithm → this is not the case in general.

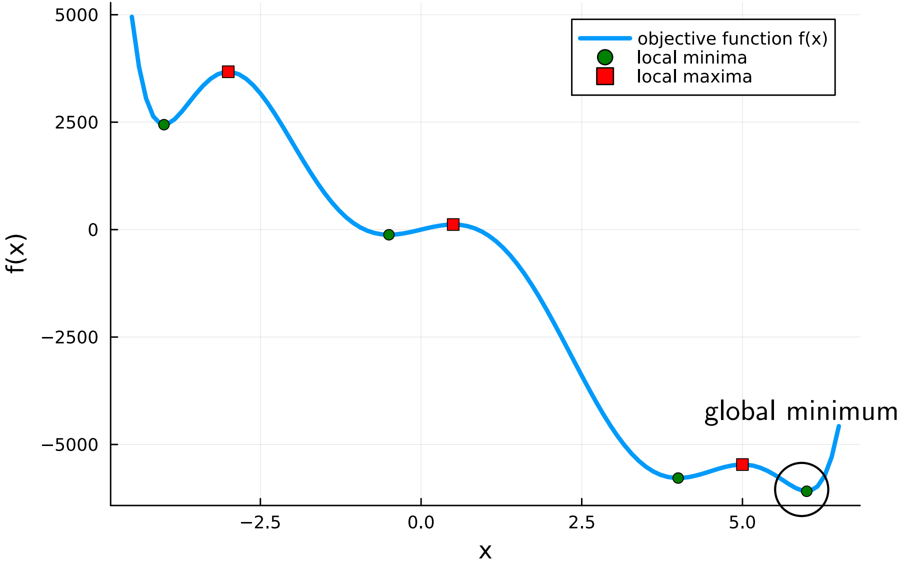
Objective function of the problem ($V_0 = 1$L) and solution



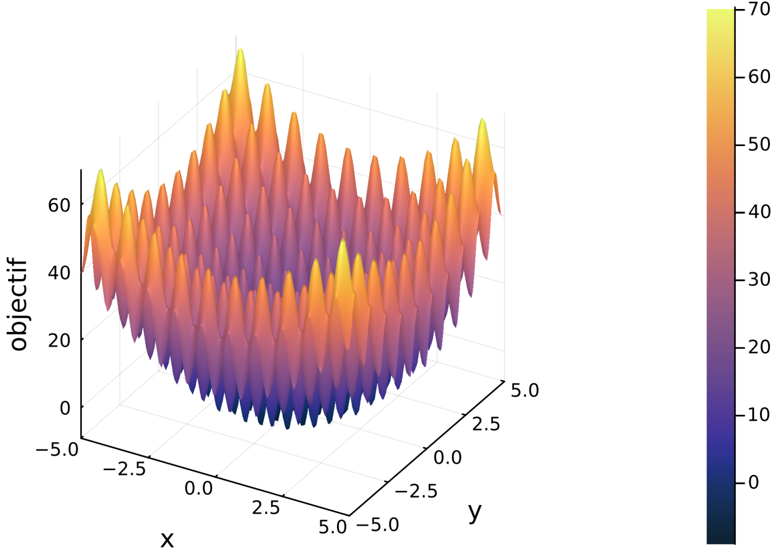
It's not always that simple... example in 1D



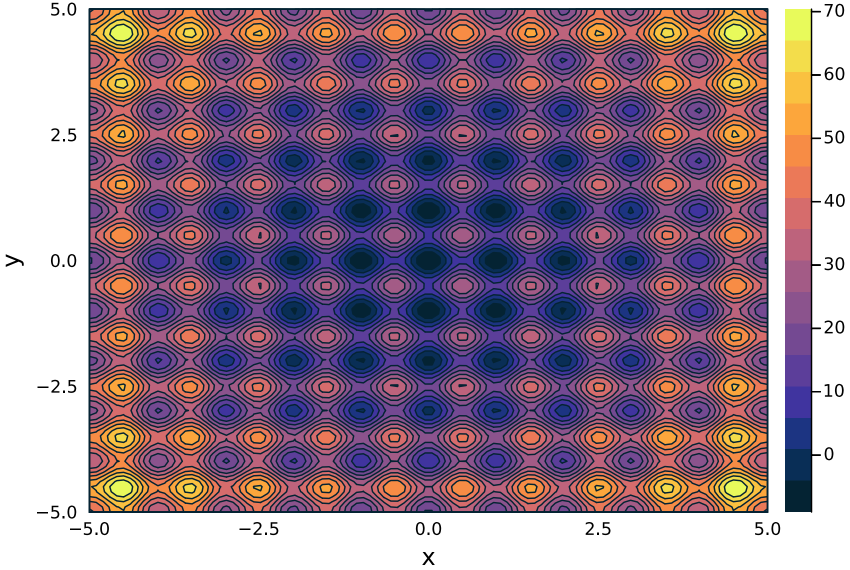
It's not always that simple... example in 1D



It's not always that simple... example in 2D

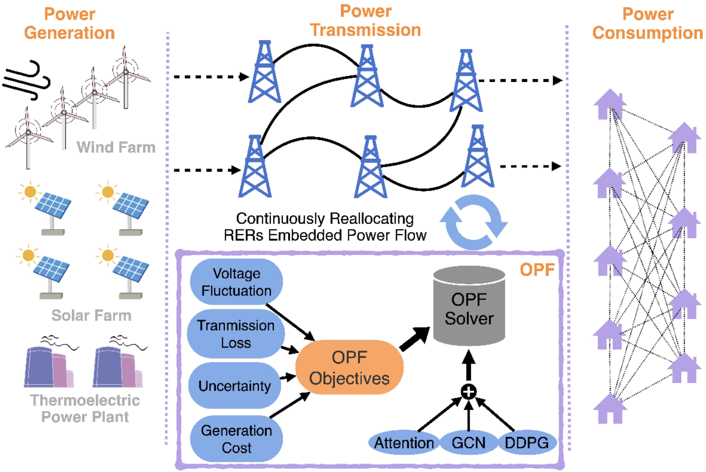


It's not always that simple... example in 2D



Outline

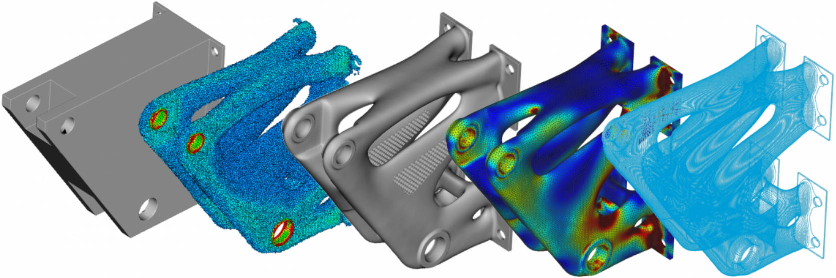
Energy chain optimization



often a multi-objective problem: maximize system efficiency + minimize costs, etc.

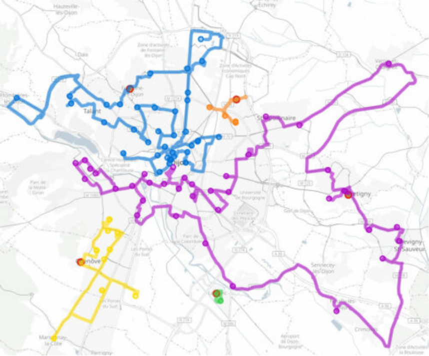
variables of different types: discrete + continuous

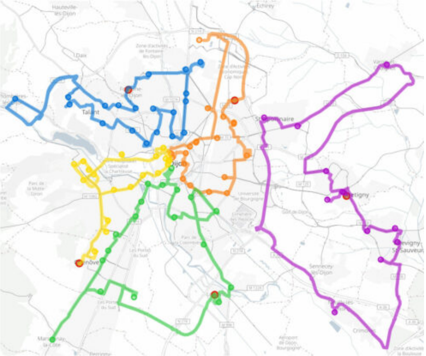
Shape optimization



objective: strongly non-linear, based on physics/mechanics
variables: several billion positions of 3D mesh points

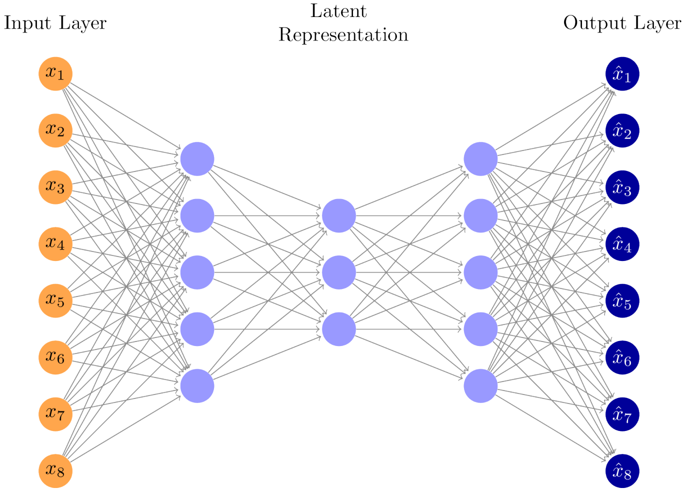
Route or path optimization





objective: minimize the distance/time of each path/journey/yield/etc.,
often multi-objective
generally discrete variables

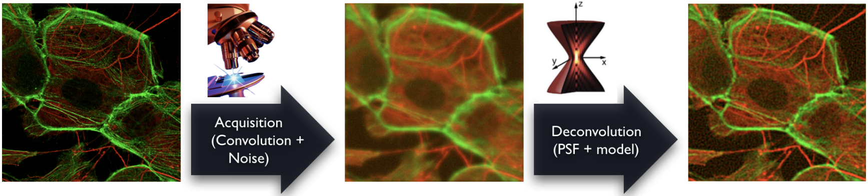
Learning and neural networks



objective: minimize the error between input and output

$\sim 10^6$ to 10^9 optimization variables

Image processing: deconvolution and denoising



Inverse problem: reconstruct an image X from the measurements
$Y = \mathrm{PSF} * X + \text{noise}$

Objective: minimize the error between the measurements and the direct model
image X with several megapixels $\rightarrow$ as many optimization variables

Outline

General form of an optimization problem

$$\min_{\mathbf{x} \in \mathbb{X}^N} f(\mathbf{x}) \text{ such that } \mathbf{x} \in \Omega$$

Vocabulary

- $\mathbf{x} = (x_1, x_2, \ldots, x_N)^\top \in \mathbb{X}^N$ is the vector of variables or parameters of the optimization problem

General form of an optimization problem

$$\min_{\mathbf{x} \in \mathbb{X}^N} f(\mathbf{x}) \text{ such that } \mathbf{x} \in \Omega$$

Vocabulary

- $\mathbf{x} = (x_1, x_2, \ldots, x_N)^\top \in \mathbb{X}^N$ is the vector of variables or parameters of the optimization problem
- N is the number of variables or the dimension of the problem

General form of an optimization problem

$$\min_{\mathbf{x} \in \mathbb{X}^N} f(\mathbf{x}) \text{ such that } \mathbf{x} \in \Omega$$

Vocabulary

- $\mathbf{x} = (x_1, x_2, \ldots, x_N)^\top \in \mathbb{X}^N$ is the vector of variables or parameters of the optimization problem
- N is the number of variables or the dimension of the problem
- the function $f : \mathbb{X}^N \to \mathbb{R}$ is the objective function, also called the cost function or criterion

General form of an optimization problem

$$\min_{\mathbf{x}\in\mathbb{X}^N} f(\mathbf{x}) \text{ such that } \mathbf{x}\in\Omega$$

Vocabulary

- $\mathbf{x} = (x_1, x_2, \dots, x_N)^\top \in \mathbb{X}^N$ is the vector of variables or parameters of the optimization problem
- N is the number of variables or the dimension of the problem
- the function $f : \mathbb{X}^N \to \mathbb{R}$ is the objective function, also called the cost function or criterion
- the set $\Omega \subseteq \mathbb{X}^N$ is the constraint or feasible set. Often, Ω takes the form

  $$\Omega = \{\mathbf{x} \mid \mathbf{h}(\mathbf{x}) = \mathbf{0},\ \mathbf{g}(\mathbf{x}) \leq \mathbf{0}\}$$

  i.e., it is defined in terms of *functional constraints*.

General form of an optimization problem

$$\min_{\mathbf{x} \in \mathbb{X}^N} f(\mathbf{x}) \text{ such that } \mathbf{x} \in \Omega$$

Nomenclature

- The space $\mathbb{X}$ can be:
  - finite or discrete: discrete (or combinatorial) optimization
  - continuous $\mathbb{X} = \mathbb{R}$: continuous optimization

General form of an optimization problem

$$\min_{\mathbf{x} \in \mathbb{X}^N} f(\mathbf{x}) \text{ such that } \mathbf{x} \in \Omega$$

Nomenclature

- The space $\mathbb{X}$ can be:
  - finite or discrete: discrete (or combinatorial) optimization
  - continuous $\mathbb{X} = \mathbb{R}$: continuous optimization
- if no constraints, $\Omega = \mathbb{X}^N$: unconstrained optimization

General form of an optimization problem

$$\min_{\mathbf{x} \in \mathbb{X}^N} f(\mathbf{x}) \text{ such that } \mathbf{x} \in \Omega$$

Nomenclature

- The space $\mathbb{X}$ can be:
  - finite or discrete: discrete (or combinatorial) optimization
  - continuous $\mathbb{X} = \mathbb{R}$: continuous optimization
- if no constraints, $\Omega = \mathbb{X}^N$: unconstrained optimization
- with constraints $\Omega \subset \mathbb{X}^N$: constrained optimization

General form of an optimization problem

$$\min_{\mathbf{x}\in\mathbb{X}^N} f(\mathbf{x}) \text{ such that } \mathbf{x}\in\Omega$$

Nomenclature

- The space $\mathbb{X}$ can be:
    - finite or discrete: discrete (or combinatorial) optimization
    - continuous $\mathbb{X}=\mathbb{R}$: continuous optimization
- if no constraints, $\Omega=\mathbb{X}^N$: unconstrained optimization
- with constraints $\Omega\subset\mathbb{X}^N$: constrained optimization
- the nature of f defines specific problems:
    - f linear: linear optimization
    - f quadratic: quadratic optimization
    - f convex and convex constraints: convex optimization
    - f non-differentiable, non-linear, etc.

General form of an optimization problem

$$\min_{\mathbf{x} \in \mathbb{X}^N} f(\mathbf{x}) \text{ such that } \mathbf{x} \in \Omega$$

Nomenclature

- The space $\mathbb{X}$ can be:
  - finite or discrete: discrete (or combinatorial) optimization
  - continuous $\mathbb{X} = \mathbb{R}$: continuous optimization
- if no constraints, $\Omega = \mathbb{X}^N$: unconstrained optimization
- with constraints $\Omega \subset \mathbb{X}^N$: constrained optimization
- the nature of f defines specific problems:
  - f linear: linear optimization
  - f quadratic: quadratic optimization
  - f convex and convex constraints: convex optimization
  - f non-differentiable, non-linear, etc.

each category has its own resolution algorithms

Scope of this course

continuous optimization with and without constraints

$$\min_{\mathbf{x} \in \mathbb{R}^N} f(\mathbf{x}) \text{ such that } \mathbf{x} \in \Omega$$

typical assumption: $f : \mathbb{R}^N \to \mathbb{R}$ is $\mathcal{C}^1$ (or at least, differentiable)

Scope of this course

continuous optimization with and without constraints

$$\min_{\mathbf{x}\in\mathbb{R}^N} f(\mathbf{x}) \text{ such that } \mathbf{x}\in\Omega$$

typical assumption: $f:\mathbb{R}^N\to\mathbb{R}$ is $\mathcal{C}^1$ (or at least, differentiable)

Note: minimization vs. maximization

There are also maximization problems (e.g., maximizing the return of a stock portfolio)

$$\max_{\mathbf{x}\in\mathbb{R}^N} g(\mathbf{x}) \text{ such that } \mathbf{x}\in\Omega$$

This problem is equivalent to the previous problem for $f(\mathbf{x}) = -g(\mathbf{x})$.

$\rightarrow$ In practice, we always reduce the problem to a minimization problem.

Typology of solutions

$$\min_{\mathbf{x} \in \mathbb{R}^N} f(\mathbf{x}) \text{ such that } \mathbf{x} \in \Omega$$

solving this optimization problem means finding $\mathbf{x}^\star \in \Omega$ (i.e. satisfying the constraints) that minimizes the function f.

An optimization problem can have:

- no solution e.g. $\min_{x \in \mathbb{R}} f(x) := -x^2$
- a unique solution
- a finite number of solutions
- an infinite number of solutions

when it exists, we call a solution $\mathbf{x}^\star$ a minimizer of f over Ω

Global and local minimizers

$$\min_{\mathbf{x}\in\mathbb{R}^N} f(\mathbf{x}) \text{ such that } \mathbf{x}\in\Omega$$

Definition (Global minimizer)

A point $\mathbf{x}^\star \in \Omega$ is a *global minimizer* of f over Ω if $f(\mathbf{x}) \geq f(\mathbf{x}^\star)$ for all $\mathbf{x}\in\Omega\setminus\{\mathbf{x}^\star\}$.

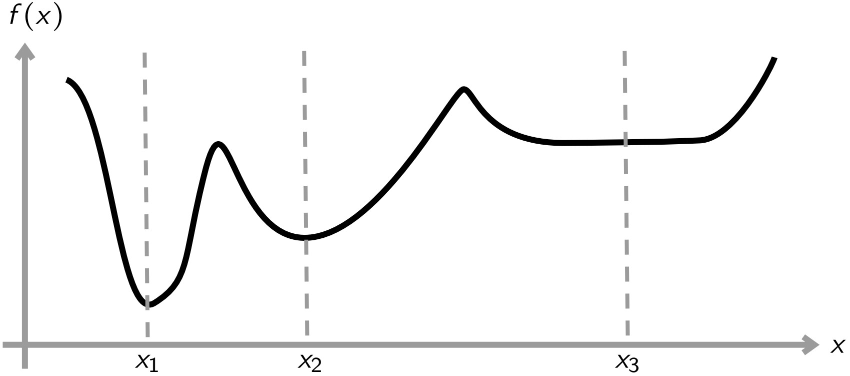
Definition (Local minimizer)

A point $\mathbf{x}^\star \in \Omega$ is a *local minimizer* of f over Ω if there exists $\varepsilon > 0$ such that $f(\mathbf{x}) \geq f(\mathbf{x}^\star)$ for all $\mathbf{x}\in\Omega\setminus\{\mathbf{x}^\star\}$ and $\|\mathbf{x}-\mathbf{x}^\star\| < \varepsilon$.

Remarks

- a global minimizer is a local minimizer, but the converse is not in general
- Replace "$\geq$" by "$>$" in definitions: strict (local, global) minimizer
- if $\mathbf{x}^\star$ is a (local, global) minimizer, then the value $f(\mathbf{x}^\star)$ is a (local, global) minimum

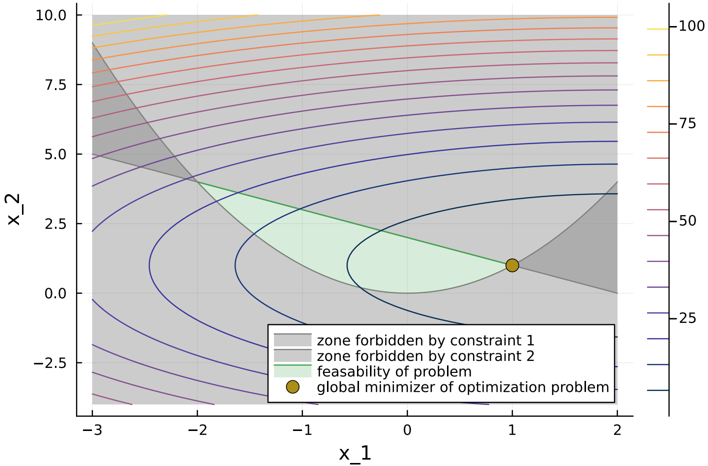
Example: unconstrained optimization in 1D



- $\mathbf{x}_1$ is a strict global minimizer
- $\mathbf{x}_2$ is a strict local minimizer
- $\mathbf{x}_3$ is a local minimizer

Example: constrained optimization in 2D

$$\min_{[x_1, x_2]^\top \in \mathbb{R}^2} (x_1 - 2)^2 + (x_2 - 1)^2 \text{ such that } \begin{cases} x_1^2 - x_2 \leq 0 \\ x_1 + x_2 \leq 2 \end{cases}$$



About notations [disgression]

Many names in the literature

- mathematical programming,
- mathematical optimization,
- numerical optimization,
- optimization

they all mean the field of study of *optimization problems*

About notations [disgression]

In the literature, you'll find many different equivalent notations for the same optimization problem

$$\min_{\mathbf{x} \in \mathbb{X}^N} f(\mathbf{x}) \text{ such that } \mathbf{x} \in \Omega$$

can be simply written as

$$\min_{\mathbf{x} \in \Omega} f(\mathbf{x})$$

or as

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{x} \in \Omega \end{aligned}$$

sometimes, you'll find the notation

$$\arg\min_{\mathbf{x} \in \Omega} f(\mathbf{x})$$

to highlight our interest in finding *minimizers* of the problem.

Outline

Recap and take-home message

$$\min_{\mathbf{x} \in \mathbb{X}^N} f(\mathbf{x}) \text{ such that } \mathbf{x} \in \Omega$$

Which questions do we ask in optimization?

- Existence and uniqueness of a solution $\mathbf{x}^\star$
- Characterization of solutions: necessary or sufficient conditions for $\mathbf{x}^\star$ to be a solution
- Design an efficient way to compute $\mathbf{x}^\star$ numerically.
  - by a closed-form solution (not very often)
  - usually, by an algorithm which produces a sequence

$$\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(n)}, \ldots \xrightarrow[n \to \infty]{} \mathbf{x}^\star$$

- Characterize the properties of this algorithm: rate of convergence, numerical complexity, etc.