Continuous optimization
ENT305A

Elise Grosjean

Ensta-Paris
Institut Polytechnique de Paris
September 2025

Organisation

Organization:

- Class 1: lecture + programming exercises
- Class 2: lecture (1h 30) + programming exercises (2h)
- Class 3: lecture (1h 30) + programming exercises (2h)
- Class 4: programming exercises
- Class 5: programming exercise (1h 30) + exam (2h).

Contact me:

elise.grosjean@ensta-paris.fr

Course website:

https://grosjean1.github.io/teaching/ensta/

Main objectives

Skills to be developed:

- **Modelling** of practical situations as an *optimization problem*.
- Basic knowledge in optimization: **theory and numerics**.
- **Numerical resolution** of such problems with the help of AMPL (A Mathematical Programming Language) and python.

Pre-requisite:

- Programming: little (python)
- Maths: little (Topology & Differential calculus).

Main objectives

Skills to be developed:

- **Modelling** of practical situations as an *optimization problem*.
- Basic knowledge in optimization: **theory and numerics**.
- **Numerical resolution** of such problems with the help of AMPL (A Mathematical Programming Language) and python.

Pre-requisite:

- Programming: little (python)
- Maths: little (Topology & Differential calculus).

Classes of optimization problems

From the point of view of **applications**, one can distinguish four classes of optimization problems.

1. Economical problems
2. Physical problems
3. Inverse problems
4. Learning problems.

1. Economical problems.

Any practical situation involving

- a **cost** to be minimized, some revenue or performance index to be maximized
- **operational decisions** (production level in thermal power plants, amount of water flowing out from a hydropower plant, beginning and end of the maintenance of a nuclear power plant, etc.)
- constraints **bounding** the decisions (which are often non-negative!)
- **physical constraints** ("total production=demand", "variation of stock= input - output",...).

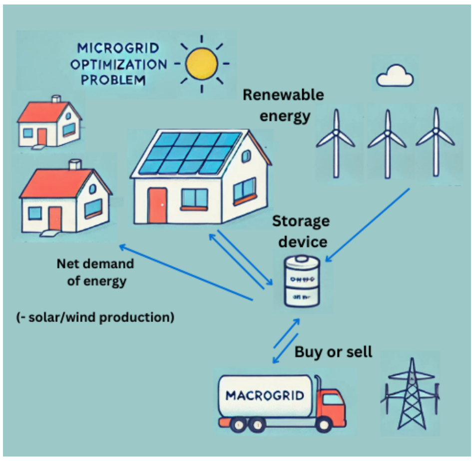
Optimization problem

Optimisation/decision variable :

- $x(s)$: state of charge of the battery.
- $a(s)$: amount of electricity *bought* on the network.
- $v(s)$: amount of energy *sold* on the network.

Parameters:

- $d(s)$: net demand of energy.
- $P_a(s)$: unitary *buying* price of energy at time s
- $P_v(s)$: unitary *selling* price of energy at time s
- $x_{\max}$: storage capacity of the battery.



Horizon: 24 hours.
Stepsize: 1 hour.
Optimization over $T = 24$ intervals.

Optimization problem

Optimisation/decision variable :

- $x(s)$: state of charge of the battery.
- $a(s)$: amount of electricity *bought* on the network.
- $v(s)$: amount of energy *sold* on the network.

Parameters:

- $d(s)$: net demand of energy.
- $P_a(s)$: unitary *buying* price of energy at time s
- $P_v(s)$: unitary *selling* price of energy at time s
- $x_{\max}$: storage capacity of the battery.

- Contraints:
  - $\forall s = 1, ..., T$,

    $$x(s+1) = x(s) - d(s) + a(s) - v(s)$$

  - $x(1) = 0$
  - $a(s) \geq 0, \ \forall s = 1, ..., T$
  - $v(s) \geq 0, \ \forall s = 1, ..., T$
  - $0 \leq x(s) \leq x_{\max}$, $\forall s = 1, \ldots T+1$.
- Cost function to be minimized:

$$J(x, a, v) = \sum_{s=1}^{T} \Big(P_a(s)a(s) - P_v(s)v(s) \Big)$$

Classes of optimization problems

2. Physical problems.

Some equilibrium problems in **physics** can be formulated as optimization problems, involving an energy to be minimized.

- Mechanical structures
- Electricity networks
- Gas networks

Some similar problems arise in **economics** and game theory:

- Traffic models on road networks.

Classes of optimization problems

3. Inverse problems

Context. A variable x must be identified, with the help of another variable y, related to x via a relation $y = F(x)$.

Examples:

- tissue regeneration with different parameters
- the epicenter x of an earthquake, given seismic measurements y.
- localization x of a crack in a mechanical structure, given displacements measurements y provided by captors
- temperature in the core of a nuclear plant, given external temperature measurements

Classes of optimization problems

The equation $y = F(x)$ (with unknown x)...

- may not have a solution (because of inaccurate measurements)
- may have several solutions (too few measurements).

Optimization is the solution! Consider

$$\inf_{x \in \mathcal{D}} \|y - F(x)\|^2, \quad \text{subject to: } x \in K,$$

where the constraints may model a priori knowledge on x.

What is an optimization problem?

Definition 1

An **optimization problem** is a mathematical expression of the form:

$$\inf_{x \in \mathcal{D}} f(x), \quad \text{subject to: } x \in K, \tag{P}$$

where:

- $\mathcal{D}$ is a set, called **domain** of f
- $f : \mathcal{D} \to \mathbb{R}$ is called **cost function** (or **objective** function)
- $K \subseteq \mathcal{D}$ is called **feasible set**.

In this class: $\mathcal{D} = \mathbb{R}^n$. **Unconstrained** optimization: $\mathcal{D} = K = \mathbb{R}^n$.

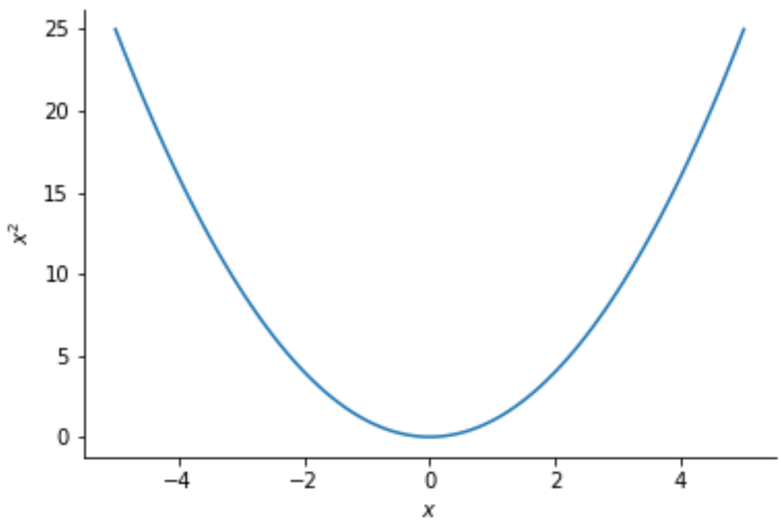
Straightforward adaptation of all results of the class to **maximization** problems, replacing f by $-f$.

Abbreviation: "subject to" $\rightsquigarrow$ "s.t.".

What is an optimization problem?

$$f : x \to x^2$$

$$\inf_{x \in \mathbb{R}} f(x), s.t. \ x \in [-5, 5]$$



What is an optimization problem?

Definition 2

- A point x is called **feasible** if $x \in K$.
- A feasible point $\bar{x}$ is called **(global) solution** (to problem P) if

$$f(x) \geq f(\bar{x}), \quad \text{for all } x \in K.$$

- If $\bar{x}$ is a global solution, then the real number $f(\bar{x})$ is called **value** of the optimization problem, it is denoted $\mathrm{val}(P)(val(P) = \alpha)$.

Example. The point $x = \pi$ is the solution of the problem

$$\inf_{x \in \mathbb{R}} \cos(x), \quad x \in [0, 2\pi].$$

What is an optimization problem?

Remarks.

- An optimization problem may **not** have a solution.
  Examples:

$$\inf_{x \in \mathbb{R}} e^x, \tag{P_1}$$

$$\inf_{x \in \mathbb{R}} x^3. \tag{P_2}$$

- The concept of **value** of an optimization problem can also be defined whether the problem has a solution or not, as an element of $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$. In particular:

$$\text{val}(P_1) = 0, \quad \text{val}(P_2) = -\infty.$$

What is an optimization problem?

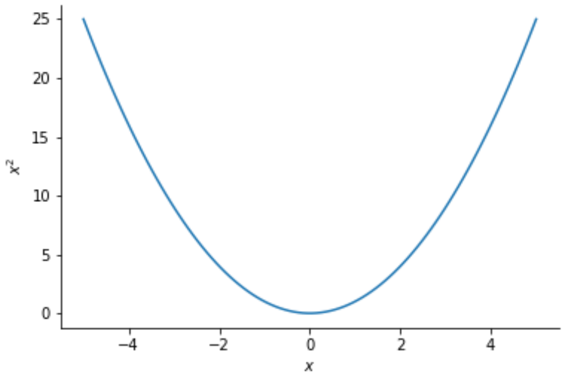
Definition 3

Let $\bar{x} \in K$. We call $\bar{x}$ a **local solution** to (P) if there exists $\varepsilon > 0$ such that the following holds true: for all $x \in K$,

$$\|x - \bar{x}\| \leq \varepsilon \Longrightarrow f(x) \geq f(\bar{x}).$$

Example: $\inf_{x\in\mathbb{R}} -x^2$, s.t. $x \in [-1, 2]$. Local solutions: -1 and 2.

$$\sup_{x\in\mathbb{R}} x^2, s.t. \; x \in [-1, 2]$$



What is an optimization problem?

Remarks.

- A global solution is also a local solution.
- The notion of local optimality does not depend on the norm, if K is a subset of a finite dimensional vector space.

What is an optimization problem?

Notation.
Let $\bar{B}(\bar{x}, \varepsilon)$ denote the closed ball of center $\bar{x}$ and radius ε.

Equivalent definition.
A feasible point $\bar{x}$ is a local solution to (P) if and only if there exists $\varepsilon > 0$ such that $\bar{x}$ is a **global** solution to the following **localized** problem:

$$\inf_{x \in \mathbb{R}^n} f(x), \quad x \in K \cap \bar{B}(\bar{x}, \varepsilon).$$

What is an optimization problem?

Constraints.

Most of the time, the feasible set K is described by

$$K = \left\{ x \in \mathbb{R}^n \,\middle|\, \begin{array}{ll} h_i(x) = 0, & \forall i \in \mathcal{E} \\ g_j(x) \leq 0, & \forall j \in \mathcal{I} \end{array} \right\},$$

where $h \colon \mathbb{R}^n \to \mathbb{R}^{m_1}$, $g : \mathbb{R}^n \to \mathbb{R}^{m_2}$.

We call the expressions

- $h_i(x) = 0$: **equality constraint**
- $g_j(x) \leq 0$: **inequality constraint**.

And to sum up the courses ...

	Necessary conditions	Sufficient conditions
Abstract formulation (exist.)		if ... then at least one solution
		if ... ... then at least one solution

	Necessary conditions	Sufficient conditions
No constraints $K = \mathbb{R}^d$ (opt.)	if ... ... then ...	if ... ... then ...
Affine constraints	...	... then global sol.
Non-linear constraints	...	... , ... then global sol.

And to sum up the courses ...

	Find a local solution
No constraints	methods
Affine constraints	
Non-linear constraints	

Existence of a solution

Theorem 4 (existence of extreme value (Weierstrass))

Assume the following:

- K *is* **non-empty** *and* **compact** *(i.e. closed and bounded)*
- f *is* **continuous** *on* K.

Then the optimization problem (P) *has (at least) one solution.*

Remarks. If $K = \{x \in \mathbb{R}^n \mid h_i(x) = 0, \forall i \in \mathcal{E}, g_j(x) \leq 0, \forall j \in \mathcal{I}\}$, where h_i, g_j are continuous, then K is closed. In practical exercises, it is not necessary to justify the continuity of h_i or g_j.

one example

Exercise. Consider for $x \in \mathbb{R}^n$, the problem

$$\min < c, x >, \ \text{s.t.} ||x||_2 \leq 1, \ x \geq 0.$$

Prove that the problem is well posed and provide an obvious global solution if $c \geq 0$.

one example

Solution. The constraint set $K = \{x \in \mathbb{R}^n : \ x \geq 0, ||x||_2 \leq 1\}$ is clearly bounded, since it is contained in the closed unit ball. It is also closed in $\mathbb{R}^n$ by continuity of the functions involved. Hence, it is compact, and since the objective function is continuous, the problem is well-posed. In the case where $c \geq 0$, we have $< c, x > \geq 0$ for all $x \in K$. Now, for $x = 0 \in K$, the objective function takes the value 0. Therefore 0 is a global minimizer of the problem.

Existence of a solution

Definition 5

We say that $f : \mathbb{R}^d \to \overline{\mathbb{R}}$ is **coercive** if the following holds: for any sequence $(x_k)_{k\in\mathbb{N}}$ in $\mathbb{R}^d$,

$$\|x_k\| \to \infty \Longrightarrow f(x_k) \to +\infty.$$

Remark. The definition is independent of the norm.

Existence of a solution

Exercise. Consider

$$f : (x, y) \in \mathbb{R}^2 \mapsto x^4 - 2xy + 2y^2.$$

Prove that f is coercive on $\mathbb{R}^2$.

Existence of a solution

Solution. We have $x^4 \geq 2x^2 - 1$, since

$$0 \leq (x^2 - 1)^2 = x^4 - 2x^2 + 1.$$

Therefore

$$\begin{aligned} f(x, y) &\geq 2x^2 - 1 - 2xy + 2y^2 \\ &= (x^2 + y^2) - 1 + (x - y)^2 \\ &\geq \|(x, y)\|^2 - 1 \underset{\|(x,y)\| \to \infty}{\longrightarrow} \infty, \end{aligned}$$

where $\| \cdot \|$ denotes the Euclidean norm. Thus f is coercive. *CQFD (or "Areuh-areuh")*

Existence of a solution

Lemma 6

Assume the following:

- K *is* **non-empty** *and* **closed**
- f *is* **continuous** *on* K
- f *is* **coercive** *on* K.

Then the optimization problem (P) *has (at least) one solution.*

Existence of a solution

Elements of proof.
Fix $x_0 \in K$.
If f is coercive, then f goes large when x moves away from x_0, thus there exists a radius R_{x_0} such that for all x located outside the ball B centered on x_0 of radius R_{x_0}, $f(x) \geq f(x_0)$.

By Weierstrass extreme value theorem, there is a global minimizer x^* on the closed ball B.
x^* being minimizer within a ball, we have $f(x^*) \leq f(x)$ for any x in B.

In particular for x_0, thus $f(x^*) \leq f(x)$, for all $\|x - x_0\| \geq R_{x_0}$. So x^* is a global minimum.

Derivatives

Definition 7

A function $F\colon \mathbb{R}^n \to \mathbb{R}^m$ is called **differentiable** at $\bar{x}$ if for all $i = 1, \ldots m$, for all $j = 1, \ldots, n$, the function

$$x \in \mathbb{R} \mapsto F_i(\bar{x}_1, \ldots, \bar{x}_{j-1}, x, \bar{x}_{j+1}, \ldots) \in \mathbb{R}$$

is differentiable. Its derivative at $\bar{x}_j$ is called **partial derivative** of F, it is denoted $\frac{\partial F_i}{\partial x_j}(\bar{x})$.

The matrix

$$DF(\bar{x}) = \left(\frac{\partial F_i}{\partial x_j}(\bar{x}) \right)_{\substack{i=1,\ldots,m \\ j=1,\ldots,n}} \in \mathbb{R}^{m \times n}$$

is called **Jacobian** matrix.

Derivatives

- The function F is said to be continuously differentiable if the Jacobian $DF\colon x \in \mathbb{R}^n \to \mathbb{R}^{m\times n}$ is continuous.
- If F is continuously differentiable, then we have the **first order Taylor expansion**

$$F(x+\delta x) = F(x) + DF(x)\delta x + o(\|\delta x\|).$$

- **Chain rule**. Let $F\colon \mathbb{R}^n \to \mathbb{R}^m$ and let $G\colon \mathbb{R}^p \to \mathbb{R}^n$ be continuously differentiable functions. Let $H = F \circ G$ (that is, $H(x) = F(G(x))$). Then

$$DH(x) = DF(G(x))DG(x), \quad \text{for all } x \in \mathbb{R}^p.$$

Derivatives

Definition 8

Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable at $x \in \mathbb{R}^n$. We call **gradient** of f (at x) the column vector:

$$\nabla f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(x) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x) \end{pmatrix} = Df(x)^\top .$$

Derivatives

Definition 9

The function $F: \mathbb{R}^n \to \mathbb{R}^m$ is said to be **twice differentiable** if it is differentiable and DF is differentiable.

We denote: $\frac{\partial^2 F_i}{\partial x_j \partial x_k}(x) = \frac{\partial}{\partial x_j}\left(\frac{\partial F}{\partial x_k}\right)(x)$.

If $m = 1$, the matrix

$$D^2 F(x) = \left(\frac{\partial^2 F}{\partial x_j \partial x_k}(x)\right)_{\substack{j=1,\dots,n \\ k=1,\dots,n}}$$

is called **Hessian** matrix. It is symmetric if F is twice continuously differentiable.

Derivatives

Exercise.
Calculate the gradient and the Hessian of the function

$$f : x \in \mathbb{R}^n \mapsto \frac{1}{2}\langle x, Ax\rangle + \langle b, x\rangle,$$

where $A \in \mathbb{R}^{n\times n}$ and $b \in \mathbb{R}^n$.

Derivatives

Solution. We have

$$f(x) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}x_i x_j + \sum_{i=1}^{n} b_i x_i.$$

Therefore,

$$\begin{aligned}\frac{\partial f}{\partial x_k}(x) &= \frac{1}{2}\sum_{j=1}^{n} A_{kj}x_j + \frac{1}{2}\sum_{i=1}^{n} A_{ik}x_i + b_k \\ &= \frac{1}{2}(Ax)_k + \frac{1}{2}(A^\top x)_k + b_k.\end{aligned}$$

Therefore,

$$\nabla f(x) = \frac{1}{2}(A + \bar{A}^\top)x + b.$$

Hessian: $D^2 f(x) = \frac{1}{2}(A + A^\top)$.

Optimality conditions

Let us fix a continuously differentiable function $f\colon \mathbb{R}^n \to \mathbb{R}$ for the whole section. Let us consider

$$\inf_{x\in\mathbb{R}^n} f(x) \qquad (P)$$

The function f is said to be **stationary** at $x \in \mathbb{R}^n$ if $\nabla f(x) = 0$.

Theorem 10 (Necessary optimality condition)

Let $\bar{x} \in \mathbb{R}^n$ be a **local solution** *of (P). Then, f is* **stationary** *at $\bar{x}$.*

Remark. Stationarity is only a necessary condition!

Optimality conditions

Theorem 11

Assume that f is twice continuously differentiable. Let $\bar{x}$ be a **stationary** *point.*

- **Necessary** *condition.*
  If $\bar{x}$ is a **local solution** *of (P), then $D^2 f(\bar{x})$ is* **positive semi-definite**, *that is to say,*

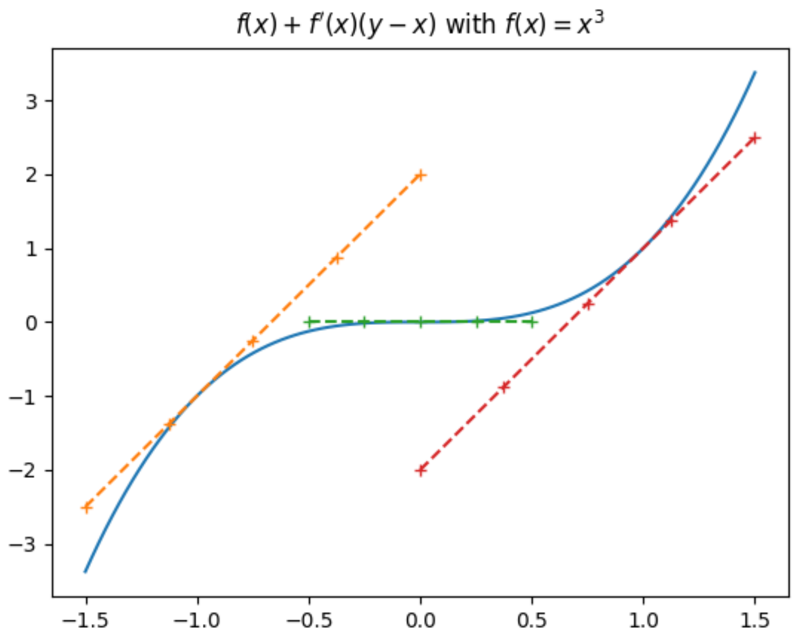
$$\langle h, D^2 f(\bar{x}) h \rangle \geq 0, \quad \text{for all } h \in \mathbb{R}^n.$$

- **Sufficient** *condition.*
  If $D^2 f(\bar{x})$ is **positive definite**, *that is to say if*

$$\langle h, D^2 f(\bar{x}) h \rangle > 0, \quad \text{for all } h \in \mathbb{R}^n \backslash \{0\} \ ,$$

  then $\bar{x}$ is a **local solution** *of (P).*

illustration



Optimality conditions

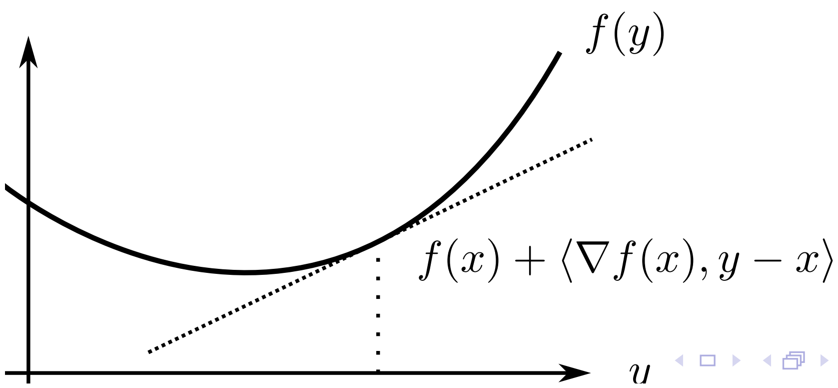
Theorem 12

- *The function f is convex if and only if*

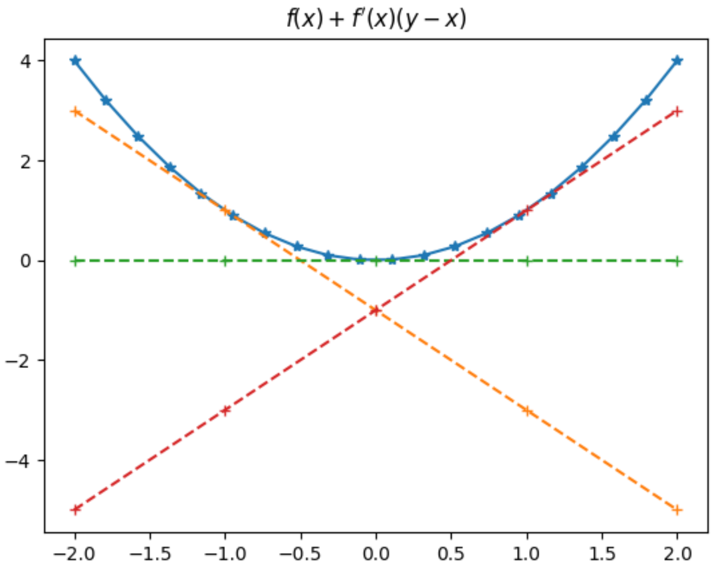
$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle,$$

for all x and $y \in \mathbb{R}^n$.

- *If f is twice differentiable, then f is convex if and only if $D^2 f(x)$ is symmetric positive semi-definite for all $x \in \mathbb{R}^n$.*



Optimality conditions



Optimality conditions

Theorem 13

*Assume that f is **convex**. Let $\bar{x}$ be a **stationary** point of f. Then it is a **global solution** of (P).*

Proof. For all $x \in \mathbb{R}^n$, we have

$$f(x) \geq f(\bar{x}) + \langle \nabla f(\bar{x}), x - \bar{x} \rangle = f(\bar{x}).$$

Optimality conditions

Exercise.
Let $A \in \mathbb{R}^{n\times n}$ be symmetric positive definite and let $b \in \mathbb{R}^n$. Let

$$f : x \in \mathbb{R}^n \mapsto \frac{1}{2}\langle x, Ax\rangle + \langle b, x\rangle.$$

Prove that

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

has a unique solution.

Optimality conditions

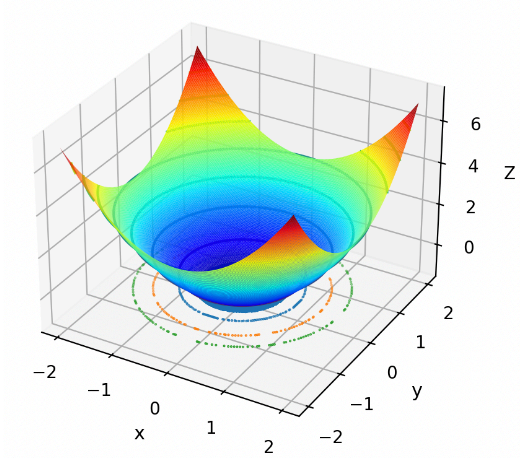
Solution.

- We have $\nabla f(x) = Ax + b$ and $\nabla^2 f(x) = A$. Since A is symmetric positive definite, thus symmetric positive semi-definite, the function f is convex.
- For a convex function, a point is a solution if and only if it is a stationary point. Thus it suffices to prove the existence and uniqueness of a stationary point.
- We have

$$\begin{aligned} x \text{ is stationary} &\iff \nabla f(x) = 0 \\ &\iff Ax + b = 0 \\ &\iff x = -A^{-1}b. \end{aligned}$$

  Therefore there is a unique stationary point, which concludes the proof.

Gradient methods



Our goal: solving **numerically** the problem

$$\inf_{x\in\mathbb{R}^n} f(x). \qquad (P)$$

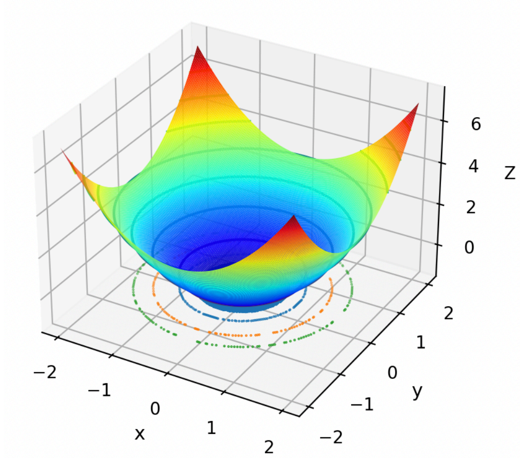
General idea: to compute a sequence $(x_k)_{k\in\mathbb{N}}$ such that

$$f(x_{k+1}) \le f(x_k), \quad \forall k \in \mathbb{N},$$

the inequality being strict if $\nabla f(x_k) \neq 0$. $\to$ **Iterative** method.

How to compute x_{k+1}?

Gradient methods



Our goal: solving **numerically** the problem

$$\inf_{x\in\mathbb{R}^n} f(x). \qquad (P)$$

General idea: to compute a sequence $(x_k)_{k\in\mathbb{N}}$ such that

$$f(x_{k+1}) \leq f(x_k), \quad \forall k \in \mathbb{N},$$

the inequality being strict if $\nabla f(x_k) \neq 0$. → **Iterative** method.
How to compute x_{k+1}?

Gradient methods

Main idea of gradient methods.

Let $x_k \in \mathbb{R}^n$. Let d_k be a descent direction at x_k. Let $\alpha > 0$. Then

$$f(x_k + \alpha d_k) = f(x_k) + \alpha \underbrace{\langle \nabla f(x_k), d_k \rangle}_{<0} + o(\alpha).$$

Therefore, if α is small enough,

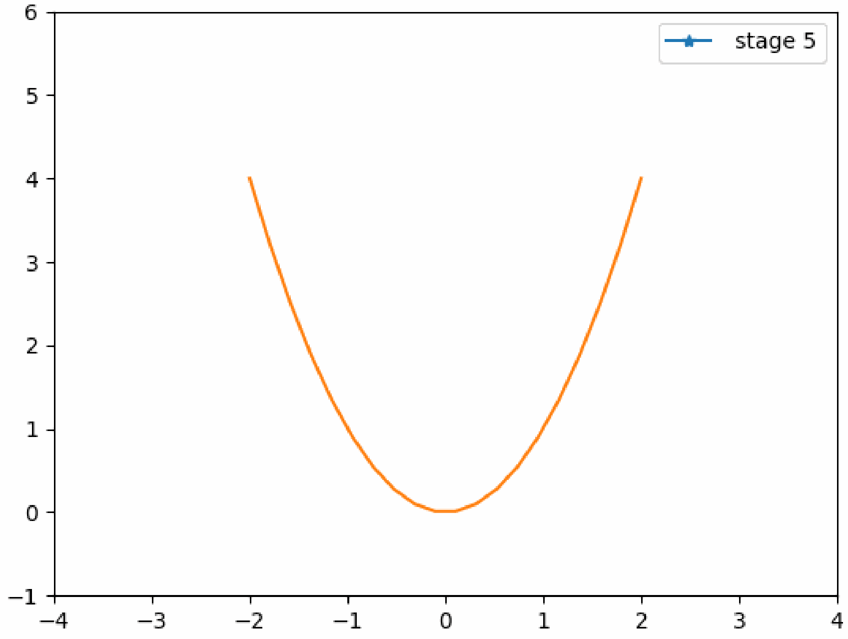
$$f(x_k + \alpha d_k) < f(x_k).$$

We can set

$$x_{k+1} = x_k + \alpha d_k.$$

illustration

Descent gradient with $\alpha = 0.25$



Gradient methods

Definition 14

Let $x \in \mathbb{R}^n$ and let $d \in \mathbb{R}^n$. The vector d is called **descent direction** if

$$\langle \nabla f(x), d \rangle < 0.$$

Remark. If $\nabla f(x) \neq 0$, then $d = -\nabla f(x)$ is a descent direction. Indeed,

$$\langle \nabla f(x), d \rangle = -\langle \nabla f(x), \nabla f(x) \rangle = -\|\nabla f(x)\|^2 < 0.$$

Gradient methods

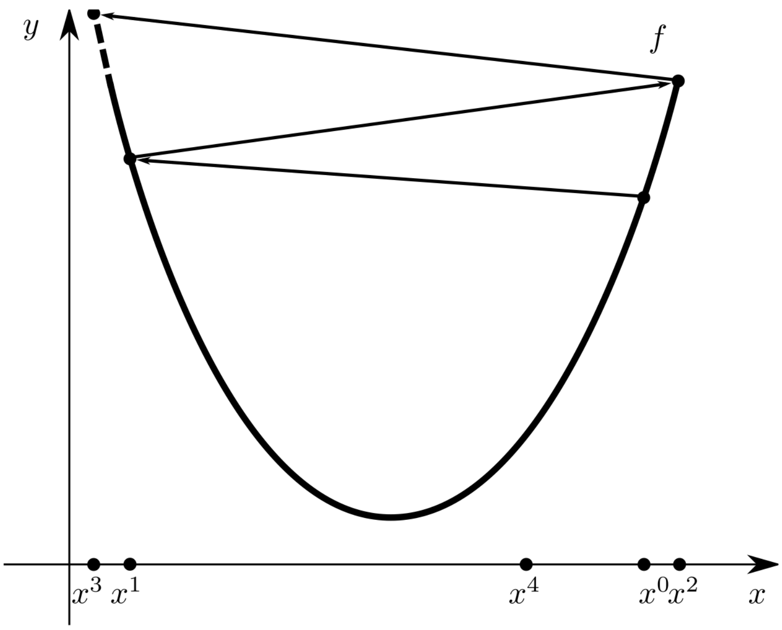
Gradient descent algorithm.

1. Input: $x_0 \in \mathbb{R}^n$, $\varepsilon > 0$.
2. Set $k = 0$.
3. While $\|\nabla f(x_k)\| \geq \varepsilon$, do
   (a) Find a descent direction d_k.
   (b) Find $\alpha_k > 0$ such that $f(x_k + \alpha_k d_k) < f(x_k)$.
   (c) Set $x_{k+1} = x_k + \alpha_k d_k$.
   (d) Set $k = k + 1$.
4. Output: x_k.

Remark. Step (b) is crucial; it is called **line search**.
The real α_k is called **stepsize**.
Exercice: Code the gradient descent algorithm

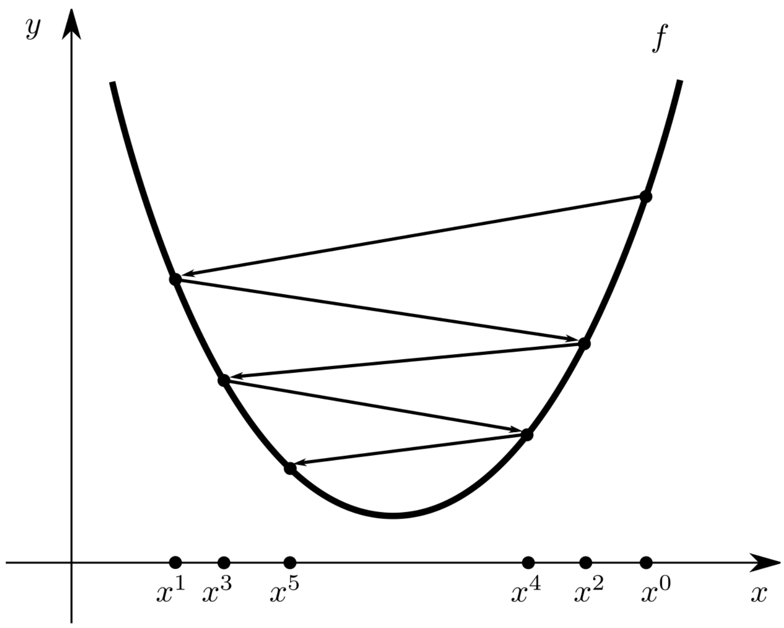
Gradient methods

On the choice of α_k.



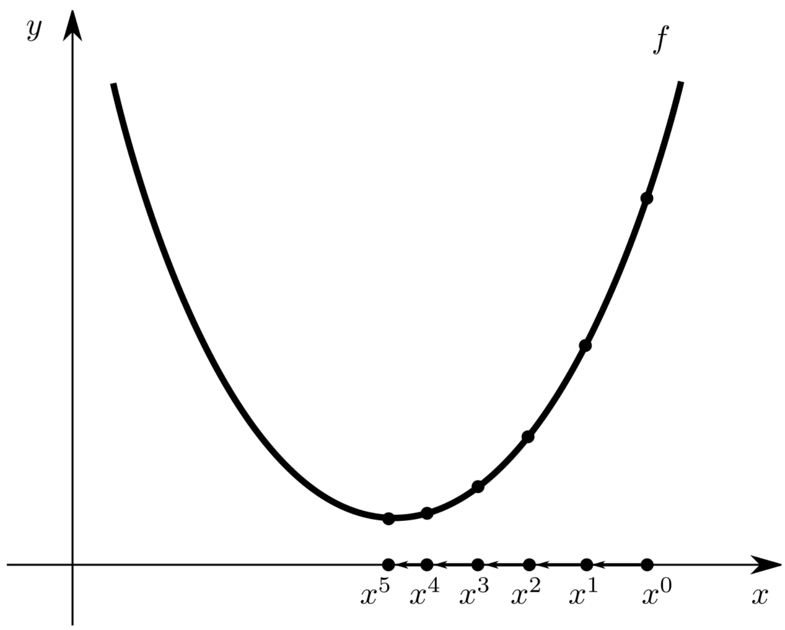
Gradient methods

On the choice of α_k.



Gradient methods

On the choice of α_k.



Gradient methods

On the choice of α_k.

Let us fix $x_k \in \mathbb{R}^n$. Let us define

$$\phi_k : \alpha \in \mathbb{R} \mapsto f(x_k + \alpha d_k).$$

The condition $f(x_k + \alpha_k d_k) < f(x_k)$ is equivalent to

$$\phi_k(\alpha_k) < \phi_k(0).$$

A natural idea: define α_k as a solution to

$$\inf_{\alpha \geq 0} \phi_k(\alpha).$$

Minimizing ϕ_k would take too much time! A **compromise** must be found between simplicity of computation and quality of α.

Gradient methods

Observation. Recall that $\phi_k(\alpha) = f(x_k + \alpha d_k)$. We have

$$\phi_k'(\alpha) = \langle \nabla f(x_k + \alpha d_k), d_k \rangle.$$

In particular, since d_k is a descent direction,

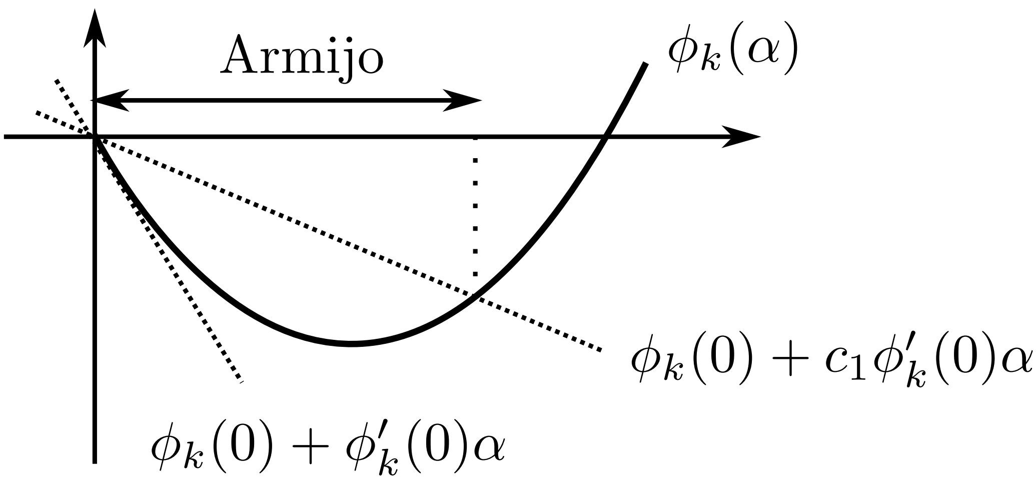
$$\phi_k'(0) = \langle \nabla f(x_k), d_k \rangle < 0.$$

Definition 15

Let us fix $0 < c_1 < 1$. We say that α satisfies **Armijo's rule** if

$$\phi_k(\alpha) \leq \phi_k(0) + c_1 \phi_k'(0)\alpha.$$

Gradient methods



Gradient methods

Backstepping algorithm for Armijo's rule

1. Input: $c_1 \in (0, 1)$, $\beta > 0$, and $\gamma \in (0, 1)$.
2. Set $\alpha = \beta$.
3. While α does not satisfy Armijo's rule,
   - Set $\alpha = \gamma\alpha$.
4. Output α.

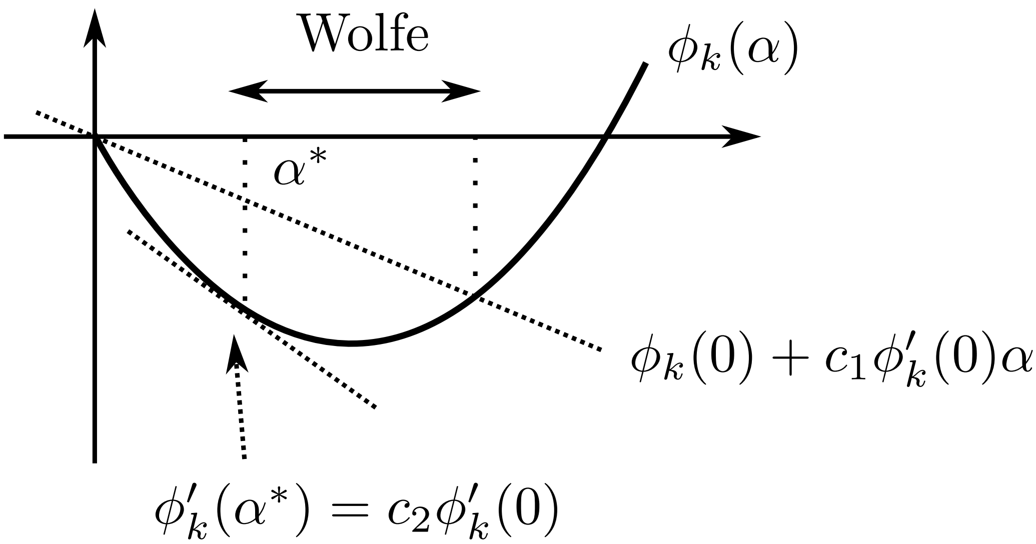
Gradient methods

Definition 16

Let $0 < c_1 < c_2 < 1$. We say that $\alpha > 0$ satisfies **Wolfe's rule** if

$$\phi_k(\alpha) < \phi_k(0) + c_1 \phi_k'(0)\alpha \quad \text{and} \quad \phi_k'(\alpha) \geq c_2 \phi'(0).$$

Gradient methods



Gradient methods

Bisection method for Wolfe's rule

1. Input: $c_1 \in (0,1)$, $c_2 \in (c_1,1)$, $\beta > 0$, $\alpha_{min}, \alpha_{max}$.
2. Set $\alpha = \beta$.

While Wolfe's rule not satified:

1. if α does not satisfy Armijo's rule :
   - Set $\alpha_{max} = \alpha$
   - $\alpha = 0.5(\alpha_{min} + \alpha_{max})$
2. if α satisfies Armijo's rule and $\phi'_k(\alpha) < c2\phi'_k(0)$, do
   - Set $\alpha_{min} = \alpha$
   - $\alpha = 0.5(\alpha_{min} + \alpha_{max})$
3. Output: α.

Gradient methods

General comments on theoretical results from literature.

- The algorithms for the computation of stepsizes satisfying Armijo and Wolfe's rules converge in **finitely many iterations** (under non-restrictive assumptions).

- Without convexity assumption on f, very little can be said about the convergence of the sequence $(x_k)_{k\in\mathbb{N}}$. Typical results ensure that any accumulation point is stationary.

- In practice: $(x_k)_{k\in\mathbb{N}}$ "usually" **converges to a local solution**. Thus a good **initialization** (that is the choice of x_0) is crucial.

- In general, **slow** convergence.

Newton's method

Main idea.

Originally, Newton's method aims at solving non-linear equations of the form

$$F(x) = 0,$$

where $F\colon \mathbb{R}^n \to \mathbb{R}^n$ is a given continuously differentiable function. It is an iterative method, generating a sequence $(x_k)_{k\in\mathbb{N}}$. Given x_k, we have

$$F(x) \approx F(x_k) + DF(x_k)(x - x_k).$$

Thus we look x_{k+1} as the solution to the linear equation

$$F(x_k) + DF(x_k)(x - x_k) = 0$$

that is, $x_{k+1} = x_k - DF(x_k)^{-1}F(x_k)$.

Newton's method

Remarks.

- If there exists $\bar{x}$ such that $F(\bar{x}) = 0$ and $DF(\bar{x})$ is regular, then for x_0 close enough to $\bar{x}$, the sequence $(x_k)_{k\in\mathbb{N}}$ is well-posed and converges "**quickly**" to $\bar{x}$.
- On the other hand, if x_0 is far away from $\bar{x}$, there is **no guaranty** of convergence.

Back to problem (P). Assume that f is continuously twice differentiable. Apply Newton's method with $F(x) = \nabla f(x)$ so as to solve $\nabla f(x) = 0$. Update formula:

$$x_{k+1} = x_k - D^2 f(x_k)^{-1} \nabla f(x_k).$$

The difficulties mentioned above are still relevant.

Newton's method

Optimization with Newton's method.

- Newton's formula can be written in the form:

$$x_{k+1} = x_k + \alpha_k d_k,$$

where

$$\alpha_k = 1 \quad \text{and} \quad d_k = -D^2 f(x_k)^{-1} \nabla f(x_k).$$

- If $D^2 f(x_k)$ is positive definite (and $\nabla f(x_k) \neq 0$), then $D^2 f(x_k)^{-1}$ is also positive definite, and therefore d_k is descent direction:

$$\langle \nabla f(x_k), d_k \rangle = -\langle \nabla f(x_k), D^2 f(x_k)^{-1} \nabla f(x_k) \rangle < 0.$$

Newton's method

Globalised Newton's method.

1. Input: $x_0 \in \mathbb{R}^n$, $\varepsilon > 0$, a linesearch rule (Armijo, Wolfe,...).
2. Set $k = 0$.
3. While $\|\nabla f(x_k)\| \geq \varepsilon$, do
   (a) If $-D^2 f(x_k)^{-1} \nabla f(x_k)$ is computable and is a descent direction, set $d_k = -D^2 f(x_k)^{-1} \nabla f(x_k)$, otherwise set $d_k = -\nabla f(x_k)$.
   (b) If $\alpha = 1$ satisfies the linesearch rule, then set $\alpha_k = 1$. Otherwise, find α_k with an appropriate method.
   (c) Set $x_{k+1} = x_k + \alpha_k d_k$.
   (d) Set $k = k + 1$.
4. Output: x_k.

Newton's method

Comments.

- Under non-restrictive assumptions, the globalized method converges, whatever the initial condition. Convergence is fast.
- The numerical computation of $D^2 f(x_k)$ may be **very time consuming** and may generate storage issues because of n^2 figures in general).
- **Quasi-Newton** methods construct a sequence of positive definite matrices H_k such that $H_k \approx D^2 f(x_k)^{-1}$. The matrix H_k can be stored efficiently (with $O(n)$ figures). Then $d_k = -H_k \nabla f(x_k)$ is a descent direction. Good speed of convergence is achieved. $\rightarrow$ **The ideal compromise!**