Continuous optimization
ENT305A

Elise Grosjean

Ensta-Paris
Institut Polytechnique de Paris
September 2024

Organisation

Organization:

- Class 1: lecture
- 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

Main objectives

Skills to be developed:

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

Pre-requisite:

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

Main objectives

Skills to be developed:

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

Pre-requisite:

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

1 General introduction
- What is an optimization problem?
- Classes of problems
- Existence of a solution
- Derivatives

2 Methods for unconstrained optimization
- Optimality conditions
- Gradient methods
- Newton's method

3 Optimality conditions for constrained problems
- Linear constraints
- Non-linear constraints
- Sensitivity analysis

Infimum (and supremum)

Let $K \subset \mathbb{R}^d$ and $f : K \to \mathbb{R}$ be a numerical function.
By construction of real numbers, the set $\{f(x) | x \in K\}$ has an infimum $\boldsymbol{\alpha = \inf_{x \in K} \; f(x)}$.
It satisfies: $\alpha \leq f(x)$ for all $x \in K$.
Remark: $\alpha = -\infty$ is a possible value.
Characterisation of infimum:

- $\alpha = \inf_{x \in K} f(x) > -\infty$
  For all $\varepsilon > 0$, there exists $x \in K$ such that $f(x) < \alpha + \varepsilon$.
- $\alpha = \inf_{x \in K} f(x) = -\infty$
  For all $N > 0$, there exists $x \in K$ such that $f(x) < -N$.

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**.

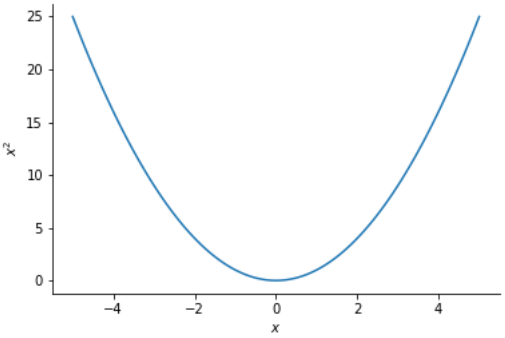
In this class: $\mathcal{D} = \mathbb{R}^n$. **Unconstrained** optimization: $\mathcal{D} = K = \mathbb{R}^n$.

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?

$$f : x \to x^2, x \in [-5, 5]$$



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 $\text{val}(P)$ $(val(P) = \alpha)$.

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.

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**.

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",...).

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:

- Cournot models with competing firms
- 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:

- 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.

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.

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.

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 $F: \mathbb{R}^n \to \mathbb{R}^m$ 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 $F\colon \mathbb{R}^n \to \mathbb{R}^m$ 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

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

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).*

Optimality conditions

Definition 12

The function f is said to be convex if

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y),$$

for all x and $y \in \mathbb{R}^n$ and for all $\lambda \in [0,1]$.

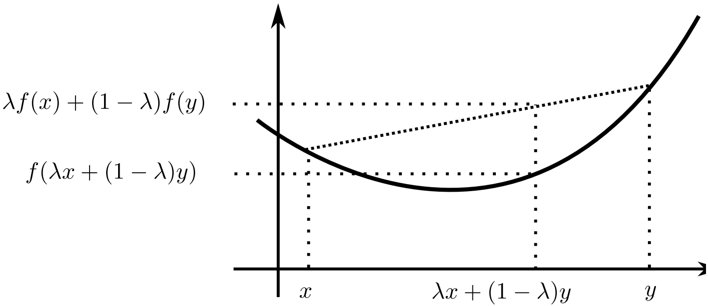
Theorem 13

- *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 14

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

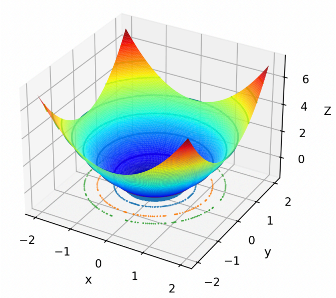
Solution.

- 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

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

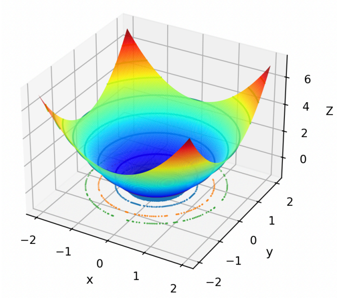
General idea: to compute a sequence $(x_k)_{k\in\mathbb{N}}$ such that

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

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

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

Gradient methods

Definition 15

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

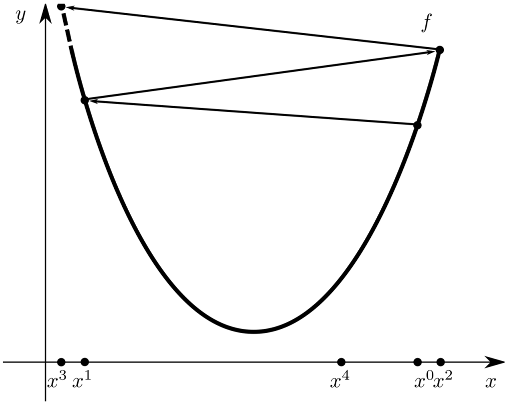
Gradient descent algorithm.

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

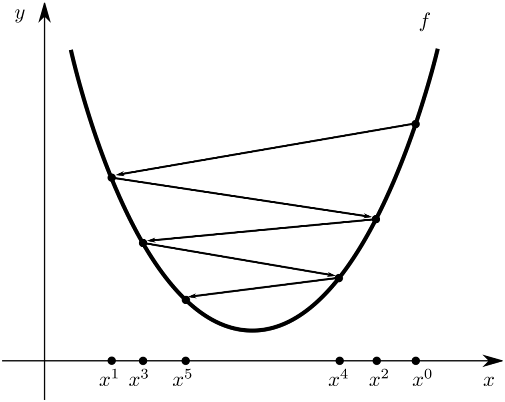
Gradient methods

On the choice of α_k.



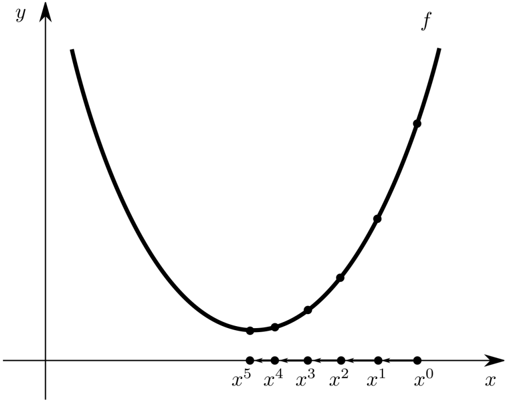
Gradient methods

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

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

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,

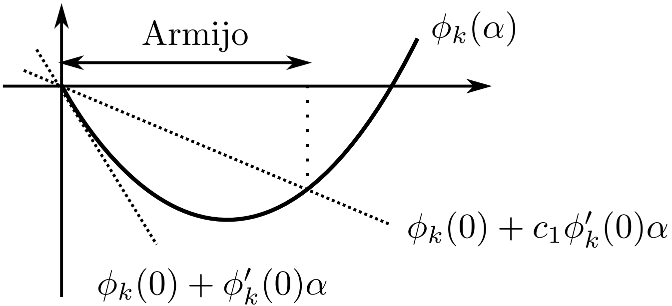
$$\phi_k'(0) = \langle \nabla f(x_k), d_k \rangle < 0.$$

Definition 16

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.

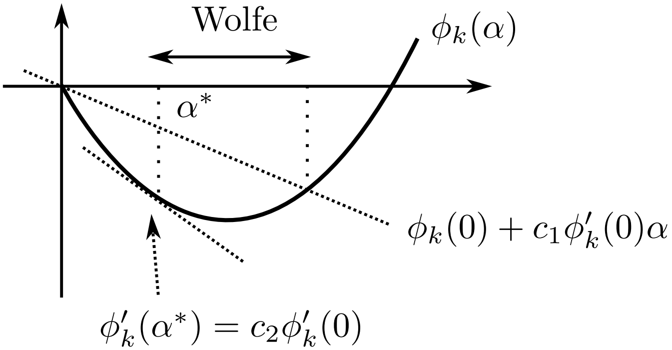
Gradient methods

Definition 17

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!**

Linear equality constraints

- We investigate in this section the problem

$$\inf_{x\in\mathbb{R}^n} f(x), \quad \text{s.t.} \begin{cases} h_i(x) = 0, & \forall i \in \mathcal{E} \\ g_j(x) \leq 0, & \forall j \in \mathcal{I}. \end{cases} \tag{P}$$

- Let $x \in \mathbb{R}^n$ be feasible. Let $j \in \mathcal{I}$. We say that
  - the inequality constraint j is **active** if $g_j(x) = 0$
  - the inequality constraint j is **inactive** if $g_j(x) < 0$.
- *Remark*. All results of the section are true if $\mathcal{E} = \emptyset$ or $\mathcal{I} = \emptyset$.

Linear constraints

- Let $f\colon \mathbb{R}^n \to \mathbb{R}$ and let $h\colon \mathbb{R}^n \to \mathbb{R}^{m_1}$ and $g\colon \mathbb{R}^n \to \mathbb{R}^{m_2}$ be two continuously differentiable functions.

- Let the **Lagrangian** $L\colon \mathbb{R}^n \times \mathbb{R}^{m_1} \times \mathbb{R}^{m_2} \to \mathbb{R}$ be defined by

$$\begin{aligned} L(x,\mu,\lambda) &= f(x) + \langle \mu, h(x)\rangle + \langle \lambda, g(x)\rangle \\ &= f(x) + \sum_{i=1}^{m_1} \mu_i h_i(x) + \sum_{j=1}^{m_2} \lambda_j g_j(x). \end{aligned}$$

The variables μ, λ are referred to as **dual variables**.

Linear equality constraints

Theorem 18

Assume that h and g are affine, that it to say, there exists $A \in \mathbb{R}^{m_2 \times n}$ and $b \in \mathbb{R}^m_2$ such that

$$g(x) = Ax + b.$$

Let $\bar{x}$ be a **local solution** *to (P).*

Then there exists $(\mu, \lambda) \in \mathbb{R}^{m1} \times \mathbb{R}^{m_2}$ such that the following three conditions, referred to as Karush-Kuhn-Tucker (KKT) conditions, are satisfied:

1. **Stationarity** *condition:* $\nabla_x L(\bar{x}, \mu, \lambda) = 0$.
2. **Sign** *condition: for all $j \in \mathcal{I}$,* $\lambda_j \geq 0$.
3. **Complementarity** *condition: for all $j \in \mathcal{I}$,* $g_j(\bar{x}) < 0 \Longrightarrow \lambda_j = 0$.

Linear equality constraints

Remarks.

- A dual variable (μ, λ) satisfying the KKT conditions is called **Lagrange multiplier** (associated with $\bar{x}$).
- Further assumptions are required to have uniqueness of (μ, λ).
- If $\mathcal{I} = \emptyset$, then the sign condition and the complementarity conditions are trivially satisfied.
- The theorem allows to have $m \geq n$.

Lagrangian formulation

Main ideas of **lagrangian formulation**:
Consider the primal problem (P): $\inf_{x\in\mathbb{R}^n} f(x)$ s.t.: $g(x) \leq 0$
(with $g : \mathbb{R}^n \to \mathbb{R}^m$).
Let p^* be the optimal cost of (P), and L the associated Lagrangian formulation. Then $p^* = \inf_{x\in\mathbb{R}^n} \sup_{\lambda\in\mathbb{R}^m} L(x,\lambda)$. Indeed,

$$\sup_{\lambda\in\mathbb{R}^m} L(x,\lambda) = \sup_{\lambda\in\mathbb{R}^m} \Big(f(x) + \sum_{j=1}^{m} \lambda_j g_j(x)\Big),$$

$$= \begin{cases} f(x) & \text{if } g_j(x) \leq 0,\ \forall j = 1,\dots,m\ (\lambda_j^* = 0) \\ \infty & \text{otherwise (if } x \text{ not feasible and } g_j(x) > 0), \end{cases}$$

Economic interpretation

Example with $\mathcal{E} = \emptyset$.

Aim: minimizing a cost under some constraints (e.g. warehouse space):

$$\min_{g_j(x) \leq 0} f(x)$$

Optimal cost : p^*. Lagrangian: $L(x, \lambda) = f(x) + \sum_{j=1}^{m} \lambda_j g_j(x)$.

- We relax the constraints while paying an additional cost linear in the constraints (e.g. by renting an extra space at a price λ_1 (in €$/m^2$)).
- $\lambda_j \geq 0$. Suppose $\lambda_j > 0$,
  - then if $g_j(x) < 0$: the cost is reduced (the company rents parts of its own space).
  - and if $g_j(x) > 0$: constraints are violated (the company pays for extra space).
- We can define an optimal cost, called **the dual function**, depending on the price λ. The optimal dual value is d^* : the optimal cost under the less favorable set of prices.
- We always have $d^* \leq p^*$.
- Strong duality: $d^* = p^*$. Then the company has no advantage to pay for an extra space (or to receive extra payment for renting parts of its own space).

Linear constraints

Exercise.

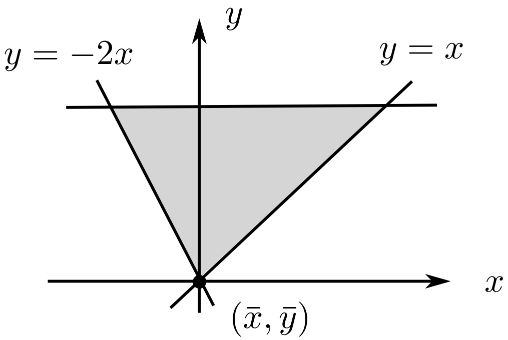
Consider the problem

$$\inf_{(x,y)\in\mathbb{R}^2} f(x,y) := y, \quad \text{s.t.} \begin{cases} g_1(x,y) := & -2x - y & \leq 0, \\ g_2(x,y) := & x - y & \leq 0, \\ g_3(x,y) := & y - 3 & \leq 0. \end{cases}$$

- Draw the feasible set and find (geometrically) the solution.
- Verify that the KKT conditions are satisfied.

Linear constraints

Solution.



Linear constraints

- Solution to the problem: $(\bar{x}, \bar{y}) = 0$.
- Let $\lambda \in \mathbb{R}^3$ be the associated Lagrange multiplier. Necessarily $\lambda_3 = 0$, since $g_3(\bar{x}, \bar{y}) < 0$, by complementarity.
- Lagrangian:

$$L(x, y, \lambda) = y - \lambda_1(2x + y) - \lambda_2(-x + y).$$

- The stationarity condition yields:

$$0 = \frac{\partial L}{\partial x}(0,0) = -2\lambda_1 + \lambda_2$$
$$0 = \frac{\partial L}{\partial y}(0,0) = 1 - \lambda_1 - \lambda_2.$$

Linear constraints

- This linear system has a unique solution

$$\lambda_1 = 1/3 \geq 0 \quad \lambda_2 = 2/3 \geq 0.$$

The sign condition is satisfied.

Linear constraints

Example 1. Case of **one equality constraint**:

$$m = 1, \quad \mathcal{E} = \{1\}, \quad \mathcal{I} = \emptyset.$$

The matrix A is a row vector, let $q = A^\top$.

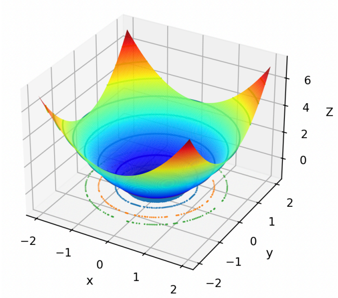
Proof of KKT conditions.

- Geometrically, we understand that $\nabla f(\bar{x})$ and q are **colinear**.
- Let $\mu \in \mathbb{R}$ be such that $\nabla f(\bar{x}) = \mu q$.
- We have:

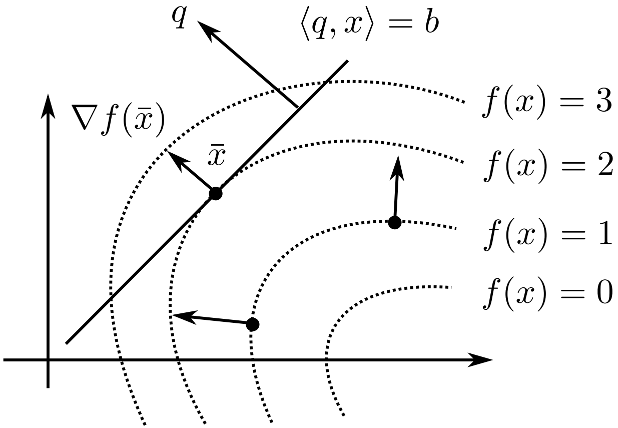
$$\nabla_x L(\bar{x}, \mu) = \nabla f(\bar{x}) + \mu \nabla g(\bar{x}) = \nabla f(\bar{x}) + \mu q = 0.$$

Linear constraints

Illustration.

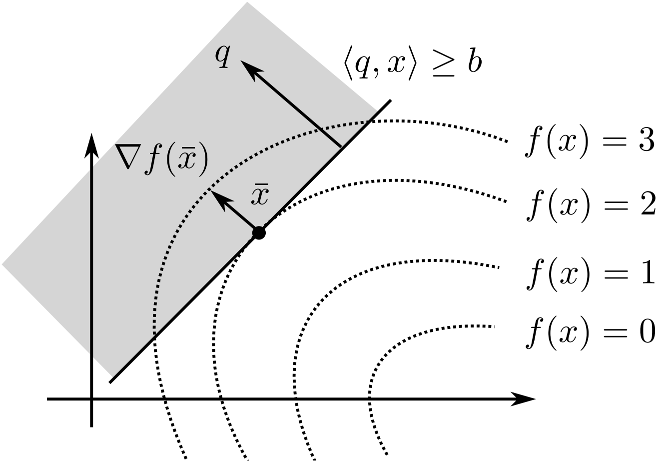


Linear constraints



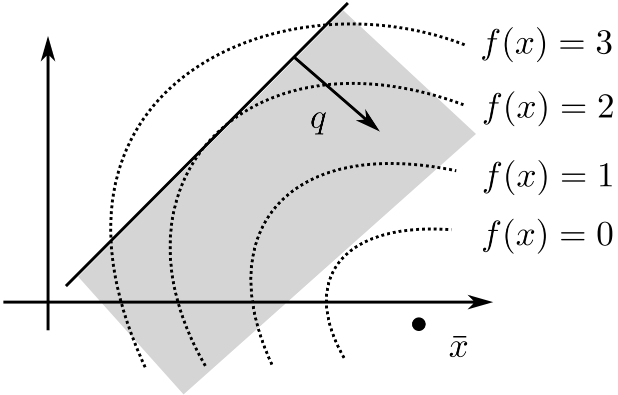
Linear constraints

Example 2(a). Case of **one (active) inequality equality constraint**:



Linear constraints

Example 2(b). Case of **one (inactive) inequality equality constraint**:



Linear constraints

Example 3. Case of m equality constraints ($\mathcal{I} = \emptyset$).

Proof.

- Let $\varepsilon > 0$ be given by the definition of a local solution.
  Let $h \in \text{Ker}(A)$ (that is $Ah = 0$).
  For all $\theta \in \mathbb{R}$, let $x_\theta = \bar{x} + \theta h$.

- For all $\theta \in \mathbb{R}$, x_θ is **feasible:**

$$g(x_\theta) = Ax_\theta + b = A\bar{x} + b + \theta Ah = 0.$$

- For all $\theta \in [0, \varepsilon/\|h\|]$, we have $\|x_\theta - \bar{x}\| \leq \varepsilon$ and thus

$$f(x_\theta) \geq f(\bar{x}).$$

Linear constraints

- We deduce that

$$0 \leq \lim_{\theta \searrow 0} \frac{f(\bar{x} + \theta h) - f(\bar{x})}{\theta} = \langle \nabla f(\bar{x}), h \rangle.$$

- Since $h \in \text{Ker}(A)$, we also have $-h \in \text{Ker}(A)$. Therefore,

$$0 \leq \langle \nabla f(\bar{x}), -h \rangle$$

  and therefore $\langle \nabla f(\bar{x}), h \rangle = 0$.
- We deduce that

$$\nabla f(\bar{x}) \in (\text{Ker}(A))^{\perp} = \text{Im}(A^{\top}),$$

  that is, there exists $\mu \in \mathbb{R}^m$ such that $\nabla f(\bar{x}) = A^{\top} \mu$.
- We have $\nabla_x L(\bar{x}, \mu) = \nabla f(\bar{x}) - A^{\top} \mu = 0$.

Non-linear constraints

Definition 19

Let $\bar{x}$ be a feasible point. Let the set of **active inequality constraints** $\mathcal{I}_0(\bar{x})$ be defined by

$$\mathcal{I}_0(\bar{x}) = \{j \in \mathcal{I} \,|\, g_j(\bar{x}) = 0\}.$$

We say that the **Linear Independence Qualification Condition (LICQ)** holds at $\bar{x}$, if the following set of vectors is linearly independent:

$$\{\nabla k_l(\bar{x})\}_{l \in \mathcal{E} \cup \mathcal{I}_0(\bar{x})},$$

where

$$k_l(\bar{x}) = \begin{cases} h_l(\bar{x}) \text{ if } l \in \mathcal{E} \\ g_l(\bar{x}) \text{ if } l \in \mathcal{I}_0. \end{cases}$$

Non-linear constraints

Theorem 20

Let $\bar{x}$ be a **local solution** *to (P). Assume that the LICQ holds at $\bar{x}$. Then there exists a unique (μ, λ) such that the* **KKT conditions** *are satisfied.*

Remarks.

- Many available **variants** of this theorem in the literature, with different qualification conditions.
- At a numerical level, a solution that does not satisfy the LICQ is hard to compute.

Non-linear constraints

Example 4.

Consider the problem

$$\inf_{x \in \mathbb{R}} x, \quad \text{subject to: } x^2 \leq 0.$$

Unique feasible point: $\bar{x} = 0$, thus the solution.

Lagrangian:

$$L(x, \lambda) = x + \lambda x^2.$$

At zero:

$$\nabla_x L(0, \lambda) = 1 + 2\lambda\bar{x} = 1 \neq 0.$$

The LICQ is not satisfied, since $\nabla g_1(0) = 0$.

Non-linear constraints

Theorem 21

Assume that

- f *is* **convex**
- *for all* $i \in \mathcal{E}$, *the map* $x \mapsto h_i(x)$ *is* **affine**
- *for all* $j \in \mathcal{I}$, *the map* $x \mapsto g_j(x)$ *is* **convex**.

Then any feasible point $\bar{x}$ *satisfying the* **KKT conditions** *is a* **global solution** *to the problem.*

Remark. The result holds whether the LICQ holds or not at $\bar{x}$.

Exercise

Exercise. Consider the function $f\colon (x,y) \in \mathbb{R}^2 \mapsto \exp(x+y^2) + y + x^2$.

1. Prove that f is coercive.
2. Calcule $\nabla f(x,y)$ and $\nabla^2 f(x,y)$.
3. We recall that a symmetric matrix of size 2 of the form $\begin{pmatrix} a & b \\ b & c \end{pmatrix}$ is positive semidefinite if and only if $a+c \geq 0$ and $ac - b^2 \geq 0$. Using this fact, prove that f is convex.
4. We consider the following problem:

$$\inf_{(x,y)\in\mathbb{R}^2} f(x,y), \quad \text{subject to: } \begin{cases} -x-y \leq 0 \\ -x-2 \leq 0. \end{cases} \qquad (\mathcal{P})$$

   Verify that $(0,0)$ is feasible and satisfies the KKT conditions.
5. Is the point $(0,0)$ a global solution to problem $(\mathcal{P})$?

Exercise

Solution.

1. We use the inequality: $\exp(z) \geq 1 + z$, which yields:

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

Exercise

2. It holds:

$$\frac{\partial f}{\partial x} = \exp(x+y^2) + 2x, \qquad \frac{\partial f}{\partial y} = 2y\exp(x+y^2) + 1.$$

Therefore, $\nabla f(x,y) = \begin{pmatrix} \exp(x+y^2)+2x \\ 2y\exp(x+y^2)+1 \end{pmatrix}$.

We also have

$$\frac{\partial^2 f}{\partial x^2} = \exp(x+y^2) + 2, \qquad \frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x} = 2y\exp(x+y^2),$$

$$\frac{\partial^2 f}{\partial y^2} = 2\exp(x+y^2) + 4y^2\exp(x+y^2).$$

Thus, $D^2 f(x,y) = \begin{pmatrix} \exp(x+y^2)+2 & 2y\exp(x+y^2) \\ 2y\exp(x+y^2) & (2+4y^2)\exp(x+y^2) \end{pmatrix}$.

Exercise

3. Proof of positive definiteness of $D^2 f$. It holds:

$$a + c = (3 + 4y^2)\exp(x + y^2) + 2 \geq 0$$

and

$$ac - b^2 = 2\exp(2x + 2y^2) + 4(1 + 2y^2)\exp(x + y^2) \geq 0.$$

It follows that $D^2 f(x, y)$ is positive semidefinite, for all (x, y). Therefore f is a convex function.

Exercise

4. Feasibility of $(0,0)$: we have $0+0 \geq 0$ and $0+2>0$.
   KKT conditions. Lagrangian:

$$L(x,y,\lambda_1,\lambda_2) = \exp(x+y^2)+y+x^2-\lambda_1(x+y)-\lambda_2(x+2).$$

Therefore,

$$\frac{\partial L}{\partial x}(0,0,\lambda_1,\lambda_2) = 1-\lambda_1-\lambda_2, \qquad \frac{\partial L}{\partial y}(0,0) = 1-\lambda_1.$$

Taking $\lambda_1 = 1$ and $\lambda_2 = 0$, we have:

1. Stationarity: $\frac{\partial L}{\partial x}(0,0,1,0) = \frac{\partial L}{\partial y}(0,0,1,0) = 0$.
2. Sign condition: $\lambda_1 \geq 0$, $\lambda_2 \geq 0$.
3. Complementarity: the second constraint is inactive and the corresponding Lagrange multiplier is null.

Exercise

5. We have the following:
   - The cost function is convex.
   - The functions $-(x+y)$ and $-(x+2)$ are convex.
   - The point $(0,0)$ is feasible and satisfies the KKT conditions.

   Therefore $(0,0)$ is a global solution.

Non-linear constraints

Exercise.

Consider:

$$\inf_{x\in\mathbb{R}^2} f(x) := -x_1 - x_2, \quad \text{s.t.} \begin{cases} g_1(x) = & x_1^2 + 2x_2^2 - 3 & \leq 0 \\ g_2(x) = & x_1 - 1 & \leq 0. \end{cases}$$

- Draw the feasible set and prove the existence of a solution.
- Verify that the LICQ at the KKT conditions hold at $\bar{x} = (1, 1)$.

Non-linear constraints

Verification of the LICQ.

$$\nabla g_1(\bar{x}) = \begin{pmatrix} -2\bar{x}_1 \\ -4\bar{x}_2 \end{pmatrix} = \begin{pmatrix} -2 \\ -4 \end{pmatrix} \quad \text{and} \quad \nabla g_2(\bar{x}) = \begin{pmatrix} -1 \\ 0 \end{pmatrix}.$$

We have: $\mathcal{E} = \emptyset$, $\mathcal{I}_0(\bar{x}) = \{1, 2\}$. The vectors $\nabla g_1(\bar{x})$ and $\nabla g_2(\bar{x})$ are linearly independant, since

$$\det \begin{pmatrix} -2 & -4 \\ -1 & 0 \end{pmatrix} = -4 \neq 0.$$

Thus the LICQ is satisfied at $\bar{x}$.

Non-linear constraints

KKT conditions.

- Lagrangian:
  $L(x, \lambda) = (-x_1 - x_2) - \lambda_1(-x_1^2 - 2x_2^2 + 3) - \lambda_2(-x_1 + 1)$.
- Stationarity condition:

$$\begin{pmatrix} -1 \\ -1 \end{pmatrix} + \begin{pmatrix} 2\bar{x}_1 \\ 4\bar{x}_2 \end{pmatrix} \lambda_1 + \begin{pmatrix} 1 \\ 0 \end{pmatrix} \lambda_2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

  It is satisfied at $\bar{x}$ with $\lambda_1 = 1/4 \geq 0$ and $\lambda_2 = 1/2 \geq 0$.
- The sign condition is satisfied.
- The complementarity condition is satisfied (all inequality constraints are active).

Sensitivity analysis

- Consider the family of optimization problems

$$\inf_{x \in \mathbb{R}^n} f(x), \quad \text{s.t.} \begin{cases} h_i(x) = y_i, & \forall i \in \mathcal{E}, \\ g_j(x) \leq y_j, & \forall j \in \mathcal{I}, \end{cases} \qquad (P(y))$$

  parametrized by the vector $y \in \mathbb{R}^m$.

- Let the **value function** V be defined by

$$V(y) = \text{val}(P(y)).$$

Sensitivity analysis

Theorem 22

Assume that for some $\bar{y}$, the problem $(P(\bar{y}))$ has a solution $\bar{x}$ satisfying the KKT conditions. Let λ denote the corresponding Lagrange multiplier.
Then, under some technical assumptions, V is **differentiable** *at $\bar{y}$ and*

$$\nabla V(\bar{y}) = \lambda.$$

Interpretation. A variation δy_i in the *i*-th constraint generates a variation of the optimal cost of $\lambda_i \delta y_i$ (as a **first approximation**).

Sensitivity analysis

Exercise.

A company decides to rent an engine over d days. The engine can be used to produce two different objects. The two objects are not produced simultaneously. Let x_1 and x_2 denote the times dedicated to the production of each object. The resulting benefits (in k€) are given by:

$$\frac{x_1}{1+x_1} \quad \text{and} \quad \frac{x_2}{4+x_2}.$$

Sensitivity analysis

1. Formulate the problem as a minimization problem.
2. Justify the existence of a solution.
3. Write the KKT conditions. What is the unit of the dual variable?
4. Verify that $\bar{x} = (4, 6)$ satisfies the KKT conditions for $d = 10$ days. Is it a global solution to the problem?
5. The renting cost of the engine is 70€/day. Is it of interest for the company to rent the engine for a longer time?

Sensitivity analysis

1. Problem:

$$\inf_{x \in \mathbb{R}^2} -\frac{x_1}{1+x_1} - \frac{x_2}{4+x_2}, \quad \text{s.t.} \begin{cases} x_1 + x_2 = d \\ -x_1 \leq 0 \\ -x_2 \leq 0 \end{cases}$$

2. The feasible set is obviously compact and non-empty and the cost function is continuous. Therefore, there exists a solution.

Sensitivity analysis

3. Let $\bar{x}$ be a solution. Let $\mu \in \mathbb{R}^2$ and $\lambda \in \mathbb{R}^2$ be the associated Lagrange multipliers. Lagrangian:

$$L(x, \mu, \lambda) = -\frac{x_1}{1 + x_1} - \frac{x_2}{4 + x_2} - \mu(x_1 + x_2 - d) - \lambda_1 x_1 - \lambda_2 x_2.$$

KKT conditions:

- Stationarity:

$$-\frac{1}{(1 + \bar{x}_1)^2} - \mu - \lambda_1 = 0, \qquad -\frac{4}{(4 + \bar{x}_2)^2} - \mu - \lambda_2 = 0.$$

- Sign condition: $\lambda_1 \geq 0$, $\lambda_2 \geq 0$.
- Complementarity: $\bar{x}_1 > 0 \Rightarrow \lambda_1 = 0$, $\bar{x}_2 > 0 \Rightarrow \lambda_2 = 0$.

- Units: $[\mu] = [\lambda_1] = [\lambda_2] =$k€/day.

Sensitivity analysis

4. Let μ, λ be such that the KKT conditions hold true. By complementarity condition, we necessarily have $\lambda_1 = \lambda_2 = 0$. The stationarity condition holds true with

$$\mu = -\frac{1}{(1+\bar{x}_1)^2} = -\frac{4}{(4+\bar{x}_2)^2} = -\frac{1}{25} = -0.04.$$

The sign condition trivially holds true since the inequality constraints are inactive. Lagrangian:

$$L(x,\mu,\lambda) = -\frac{x_1}{1+x_1} - \frac{x_2}{4+x_2} + 0.04(x_1+x_2-d).$$

If $x_1 + x_2 > d$, the cost associated to constraints is increased, otherwise decreased (company rents the engine for the $d - x_1 - x_2$ remaining days).

The point $\bar{x}$ is feasible and satisfies the KKT conditions. We have affine constraints and a convex cost function, therefore, the KKT conditions are sufficient. The point $\bar{x}$ is a global solution.

Sensitivity analysis

5. d is fixed.

   Increasing the renting time of y days will generate a variation of cost of μy (approximately), that is, an augmentation of the benefit of 40€/day (less the renting price). It corresponds to the benefit that the company can have from another firm for renting the engine. Thus, the cost will corresponds to:

$$c(\overline{x}, \mu, \lambda) = -\frac{4}{1+4} - \frac{6}{4+6} - 0.04y + 0.07y = -1.4 + 0.03y.$$

   It would be of interest for the company to reduce the renting time.

And to sum up the courses ...

	Necessary conditions	Sufficient conditions
Abstract formulation (exist.)		if K compact, $f \in C^0(K)$ then at least one solution
		if K closed, $f \in C^0(K)$, coercive then at least one solution

	Necessary conditions	Sufficient conditions
No constraints $K = \mathbb{R}^d$ (opt.)	if $\overline{x}$ local sol., $f \in C^2(K)$ then, $D^2 f(\overline{x})$ is positive semi-def.	if $f \in C^2(K)$, $\nabla f(\overline{x}) = 0$, $D^2 f(\overline{x})$ positive def. then $\overline{x}$ local sol.
Affine constraints	$\overline{x}$ local sol. then KKT	f convex, then KKT=global sol.
Non-linear constraints	$\overline{x}$ local sol., LICQ then KKT	f convex, h affine, g convex, then KKT=global sol.