Continuous optimization
ENT305A

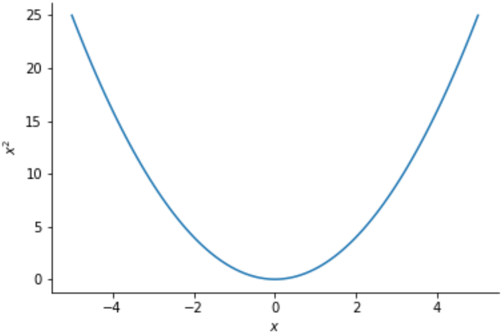
Elise Grosjean

Ensta-Paris
Institut Polytechnique de Paris
September 2025

What is an optimization problem?

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

$\mathcal{D} = K = \mathbb{R}$



What is an optimization problem?

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

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**.

Existence of a solution

Theorem 1 (existence of extreme value (Weierstrass))

Assume the following:

- *?*
- *?*

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

Existence of a solution

Theorem 2 (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.

Optimality conditions

Let us fix a continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ 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 3 (Necessary optimality condition)

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

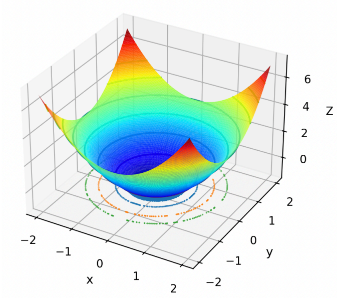
Remark. Stationarity is only a necessary condition!

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		
Non-linear constraints		then KKT=global sol.

Gradient methods



Our goal: solving **numerically** the problem

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

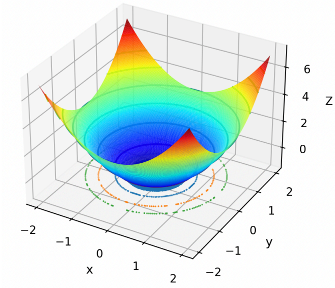
General idea: to compute a sequence $(x_k)_{k\in\mathbb{N}}$ such that

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

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

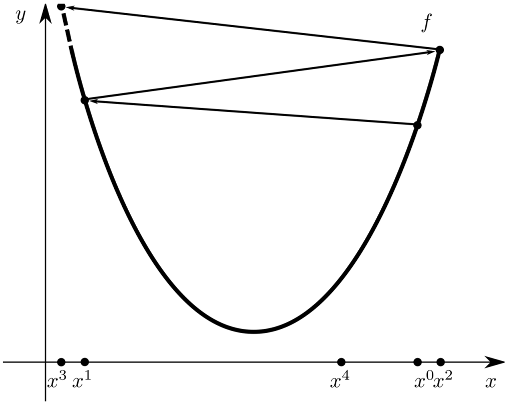
Gradient descent algorithm.

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

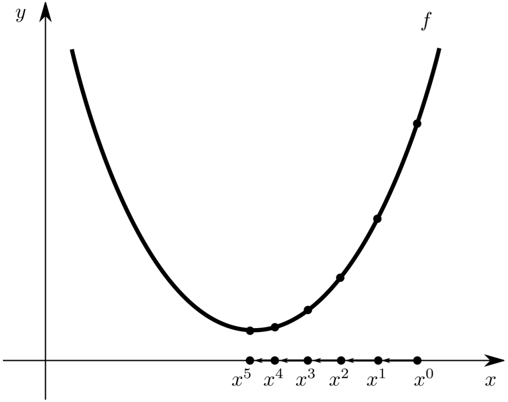
Gradient methods

On the choice of α_k.



Gradient methods

On the choice of α_k.



Gradient methods

On the choice of α_k.

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,

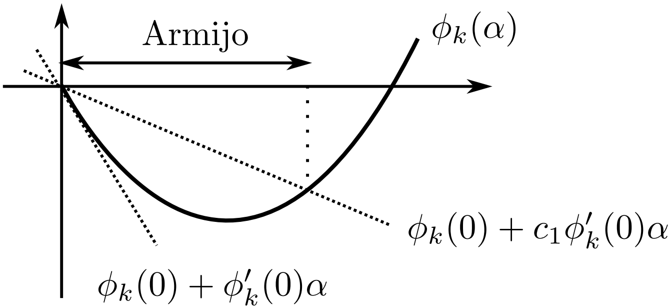
$$\phi_k'(0) = \langle \nabla f(x_k), d_k \rangle < 0.$$

Definition 4

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.

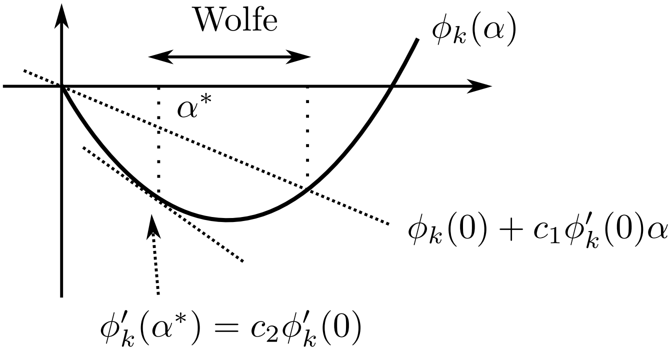
Gradient methods

Definition 5

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.

- 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.

1. Reminders
   - Optimization problem
   - Existence of a solution
   - Gradient methods
   - Newton's method

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

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.

- 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!**

1. Reminders
   - Optimization problem
   - Existence of a solution
   - Gradient methods
   - Newton's method

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

Linear equality constraints

- We investigate in this section the problem

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

- 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 $g_j(x) < 0$.

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 6

Assume that h and g are affine, that it to say, there exists $A \in \mathbb{R}^{m_2 \times n}$ *and* $b \in \mathbb{R}_2^m$ *such that*

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

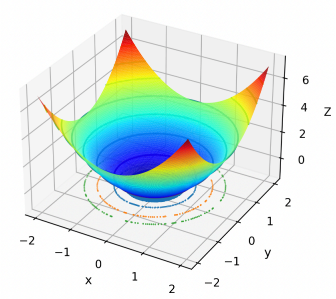
Let $\bar{x}$ *be a* **local solution** *to* (P).

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:*

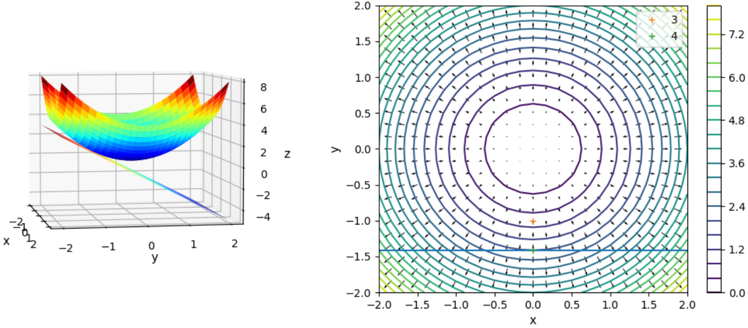
1. **Stationarity** *condition:* $\nabla_x L(\bar{x}, \mu, \lambda) = 0$.
2. **Sign** *condition: for all* $j \in \mathcal{I}$, $\lambda_j \geq 0$.
3. **Complementarity** *condition: for all* $j \in \mathcal{I}$, $g_j(\bar{x}) < 0 \Longrightarrow \lambda_j = 0$.

Linear constraints

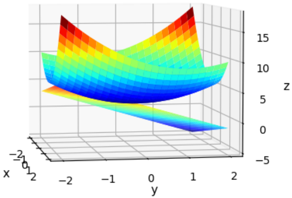
Illustration.

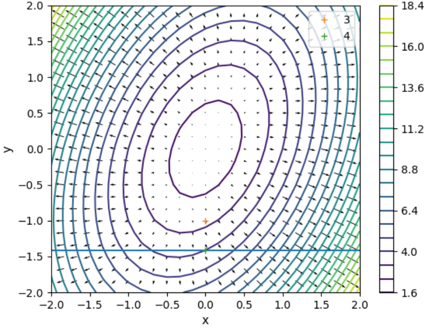


KKT stationarity

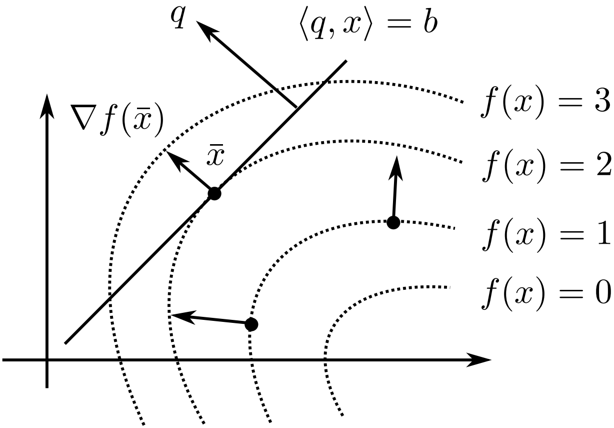


KKT stationarity



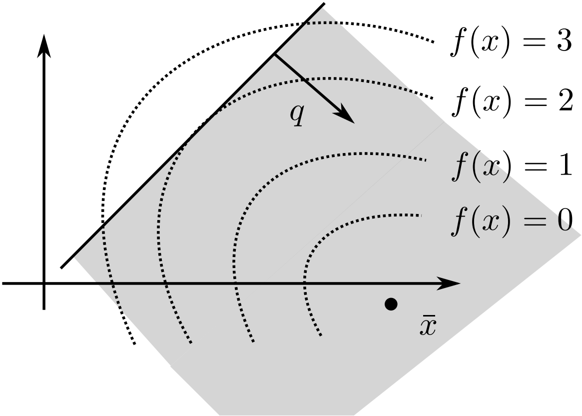


Linear constraints



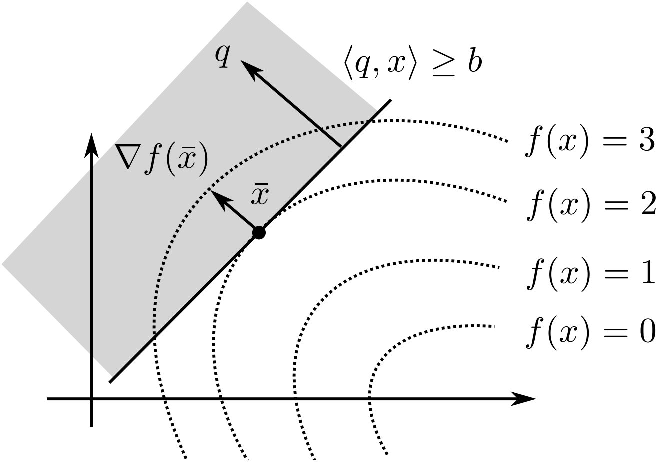
Linear constraints

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



Linear constraints

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



1. Reminders
   - Optimization problem
   - Existence of a solution
   - Gradient methods
   - Newton's method

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

Non-linear constraints

Definition 7

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 indepedent:

$$\{\nabla h_i(\bar{x})\}_{i \in \mathcal{E}} \cup \{\nabla g_j(\bar{x})\}_{j \in \mathcal{I}_0(\bar{x})}$$

Non-linear constraints

Non-linear constraints

Theorem 8

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.

- 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 9

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. *Indication: Use* $\exp(z) \geq 1+z$
2. Compute $\nabla f(x,y)$ 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:
$$\inf_{(x,y)\in\mathbb{R}^2} f(x,y), \quad \text{subject to: } \begin{cases} -x-y \leq 0 \\ -x-2 \leq 0. \end{cases} \tag{$\mathcal{P}$}$$
Verify that $(0,0)$ is feasible and satisfies the KKT conditions.
5. Is the point $(0,0)$ a global solution to problem $(\mathcal{P})$?

Non-linear constraints

Exercise.

Consider:

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

- Show that $\bar{x} = (1, 1)$ is feasible
- Verify that the LICQ and the KKT conditions hold at $\bar{x} = (1, 1)$.
- Prove that $\bar{x} = (1, 1)$ is a global solution.

1 Reminders
- Optimization problem
- Existence of a solution
- Gradient methods
- Newton's method

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

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

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

- A variation δy_i in the i-th constraint generates a variation of the optimal cost of $\lambda_i \delta y_i$.

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:

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

   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.

Temporary page!

LaTeX was unable to guess the total number of pages correctly.
there was some unprocessed data that should have been added
the final page this extra page has been added to receive it.
If you rerun the document (without altering it) this surplus page
will go away, because LaTeX now knows how many pages to expe
for this document.