Optimisation

Lecture 9 - Optimality conditions in the constrained case

Fall semester - 2024

Dr. Eng. Valentin Leplat
Innopolis University
October 29, 2024

Outline

Summary of OC's in the unconstrained case

Summary

In the unconstrained case, the feasible set $\mathcal{X} = \mathbb{R}^n$, then:

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

First and second order necessary conditions

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

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

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

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

Sufficient conditions

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

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

For today's session

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

For today's session

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

What happens to the optimality conditions when we add

- equality constraints ?
- inequality constraints ?

For today's session

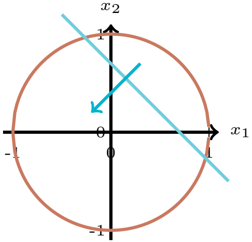
$$\min_{x} f(x)$$
$$\text{s.t.} \quad h_i(x) = 0 \text{ for } i \in \mathcal{E}$$
$$h_i(x) \leqslant 0 \text{ for } i \in \mathcal{I}$$

What happens to the optimality conditions when we add

- equality constraints ?
- inequality constraints ?

Example :

$$\min_{x_1, x_2} x_1 + x_2$$
$$\text{s.t.} \quad x_1^2 + x_2^2 = 1$$



For today's session

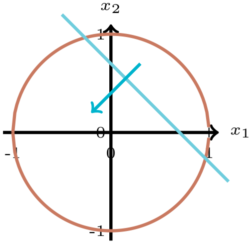
$$\begin{aligned} \min_{x} \; & f(x) \\ \text{s.t.} \quad & h_i(x) = 0 \text{ for } i \in \mathcal{E} \\ & h_i(x) \leqslant 0 \text{ for } i \in \mathcal{I} \end{aligned}$$

What happens to the optimality conditions when we add

- equality constraints ?
- inequality constraints ?

Example :

$$\begin{aligned} \min_{x_1, x_2} \; & x_1 + x_2 \\ \text{s.t.} \quad & x_1^2 + x_2^2 = 1 \end{aligned}$$



$(\frac{-1}{\sqrt{2}}, \frac{-1}{\sqrt{2}})$ is the optimal solution and, at this point, the gradient does not cancel out:

$$\nabla f \left(\frac{-1}{\sqrt{2}}, \frac{-1}{\sqrt{2}} \right) = (1, 1)^T \neq 0$$

Descent direction

Directional derivative

Let $f : \mathbb{R}^n \to \mathbb{R}$. The restriction of f at a point x in a direction $d \in \mathbb{R}^n$ is called directional function and is defined using a function $g : \mathbb{R} \to \mathbb{R}$ such that

$$g(\alpha) = f(x + \alpha d).$$

Directional derivative

Let $f : \mathbb{R}^n \to \mathbb{R}$. The restriction of f at a point x in a direction $d \in \mathbb{R}^n$ is called directional function and is defined using a function $g : \mathbb{R} \to \mathbb{R}$ such that

$$g(\alpha) = f(x + \alpha d).$$

Let $f : \mathbb{R}^n \to \mathbb{R}$. For a direction d of $\mathbb{R}^n$ (of unit length), the directional derivative of f at x is :

$$\frac{\partial f}{\partial d}(x) = \lim_{\alpha \to 0} \frac{f(x + \alpha d) - f(x)}{\alpha} = \lim_{\alpha \to 0} \frac{g(\alpha) - g(0)}{\alpha - 0} = g'(0).$$

The directional derivative gives information about the slope of f in the d direction. When d is a canonical e_i direction, we speak of partial derivative :

$$\frac{\partial f}{\partial x_i}(x) = \lim_{\alpha \to 0} \frac{f(x + \alpha e_i) - f(x)}{\alpha}.$$

Directional derivative

How to calculate a directional derivative?

The directional derivative is the dot product of the d direction and ∇f :

$$\frac{\partial f}{\partial d}(x) = d_1 \cdot \frac{\partial f}{\partial x_1}(x) + ... + d_n \cdot \frac{\partial f}{\partial x_n}(x) = \sum_{i=1}^{n} d_i \cdot \frac{\partial f}{\partial x_i}(x) = d^T \nabla f(x).$$

Directional derivative

How to calculate a directional derivative?

The directional derivative is the dot product of the d direction and ∇f :

$$\frac{\partial f}{\partial d}(x) = d_1 \cdot \frac{\partial f}{\partial x_1}(x) + ... + d_n \cdot \frac{\partial f}{\partial x_n}(x) = \sum_{i=1}^{n} d_i \cdot \frac{\partial f}{\partial x_i}(x) = d^T \nabla f(x).$$

Example.

Find the directional derivative of the function $f(x_1, x_2) = (x_1 + x_2)^4 - 2(x_1 + x_2)^2 + 1$ at $(\frac{1}{4}, \frac{1}{4})$ along direction $d = (1, 1)^T$.

Definition of a descent direction

Definition

Let $f : \mathbb{R}^n \to \mathbb{R}$, a direction d is called a *descent* direction at x if:

$$\frac{\partial f}{\partial d}(x) = d^T \nabla f(x) < 0$$

In other words, this means that the function f is locally decreasing along d

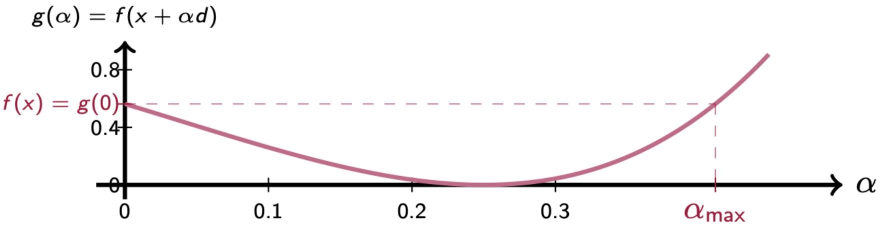
Definition of a descent direction

Theorem

Let $f : \mathbb{R}^n \to \mathbb{R}$, if d is a *descent* direction at x, then $\exists \alpha_{\max} > 0$:

$$f(x + \alpha d) < f(x), \quad \forall \alpha \in (0, \alpha_{\max}].$$

Illustration: $f(x_1, x_2) = (x_1 + x_2)^4 - 2(x_1 + x_2)^2 + 1$, $x = (\frac{1}{4}, \frac{1}{4})^T$ along direction $d = (1, 1)^T$:



Here: $\alpha_{\max} = 0.411438$

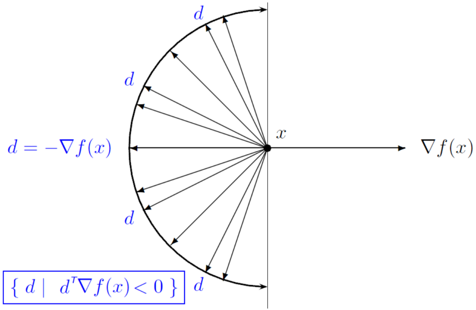
Steepest descent direction

How to find the steepest d ?

In other words: which direction d minimizes the directional derivative of f at x?
Equivalent to solve

$$\min_{d} \frac{\partial f}{\partial d}(x)$$

where $\frac{\partial f}{\partial d}(x) := d^T \nabla f(x)$. Without loss of generality, we may assume $||d||^2 = 1$.



OCs for problems with equality constraints only

Intuition for OC in the constrained case

OCs in the presence of constraints are based on the same intuition as in the unconstrained case: it's impossible to go down at a minimum of the problem.

Intuition for OC in the constrained case

OCs in the presence of constraints are based on the same intuition as in the unconstrained case: it's impossible to go down at a minimum of the problem.

For the minimization problem with a equality constraint only

$$\min_x f(x)$$
$$\text{s.t.} \quad h(x) = 0,$$

the point $x + d$ is better if $f(x + d) < f(x)$ and if $h(x + d) = 0$.

Intuition for OC in the constrained case

OCs in the presence of constraints are based on the same intuition as in the unconstrained case: it's impossible to go down at a minimum of the problem.

For the minimization problem with a equality constraint only

$$\min_x f(x)$$
$$\text{s.t.} \quad h(x) = 0,$$

the point $x + d$ is better if $f(x + d) < f(x)$ and if $h(x + d) = 0$.

- You need $h(x + d) = 0$. By Taylor : $h(x + d) \approx h(x) + d^T \nabla h(x)$, hence :

$$d^T \nabla h(x) = 0.$$

Intuition for OC in the constrained case

OCs in the presence of constraints are based on the same intuition as in the unconstrained case: it's impossible to go down at a minimum of the problem.

For the minimization problem with a equality constraint only

$$\min_x f(x)$$
$$\text{s.t.} \quad h(x) = 0,$$

the point $x + d$ is better if $f(x + d) < f(x)$ and if $h(x + d) = 0$.

- You need $h(x + d) = 0$. By Taylor : $h(x + d) \approx h(x) + d^T \nabla h(x)$, hence :
$$d^T \nabla h(x) = 0.$$
- A descent direction d improving the objective f verifies
$$d^T \nabla f(x) < 0.$$

For x extremum, a displacement d would have no component along ∇f, so $d^T \nabla f(x) = 0$.

Intuition for OC in the constrained case

OCs in the presence of constraints are based on the same intuition as in the unconstrained case: it's impossible to go down at a minimum of the problem.

For the minimization problem with a equality constraint only

$$\min_x f(x)$$
$$\text{s.t.} \quad h(x) = 0,$$

the point $x + d$ is better if $f(x + d) < f(x)$ and if $h(x + d) = 0$.

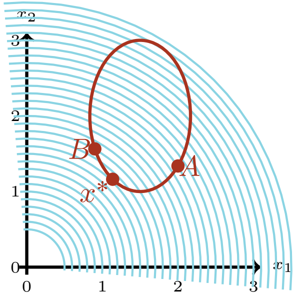
- You need $h(x + d) = 0$. By Taylor : $h(x + d) \approx h(x) + d^T \nabla h(x)$, hence :
$$d^T \nabla h(x) = 0.$$
- A descent direction d improving the objective f verifies
$$d^T \nabla f(x) < 0.$$

For x extremum, a displacement d would have no component along ∇f, so $d^T \nabla f(x) = 0$.
At the optimum, this means that the two vectors ∇f and ∇h are parallel, i.e. there is a real λ such that :

$$\nabla f = \lambda \nabla h.$$

Example

$$\min_{x_1,x_2} x_1^2 + x_2^2 \text{ such that } \frac{(x_1-1.5)^2}{4} + \frac{(x_2-2)^2}{9} - \frac{1}{9} = 0.$$



1. By representing level curves of the objective function, we can see that the gradient of this function is orthogonal to these curves and oriented towards those of higher levels.
2. By following the constraint from a point A to a point B, we move in a direction close to $-\nabla f$: we reduce the value of f while respecting the constraint.
3. At the x^* point, motion is orthogonal to ∇f. Then, the motion has a component in the direction of ∇f and the value of f increases again.
4. The minimum is therefore reached at x^*. At this point, motion is orthogonal to ∇h to stay on the curve, and orthogonal to ∇f since we are at the minimum of f: the two gradients are then collinear.

Lagrange multipliers

Minimization problem with equality constraints :

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

$$\text{s.t.} \quad h_i(x) = 0 \text{ for } i \in \mathcal{E}.$$

Lagrange multipliers

Minimization problem with equality constraints :

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

$$\text{s.t.} \quad h_i(x) = 0 \text{ for } i \in \mathcal{E}.$$

- For each constraint $h_i(x) = 0$, we associate a real variable λ_i called Lagrange multiplier (so we have a vector λ).

Lagrange multipliers

Minimization problem with equality constraints :

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

$$\text{s.t.} \quad h_i(x) = 0 \text{ for } i \in \mathcal{E}.$$

- For each constraint $h_i(x) = 0$, we associate a real variable λ_i called Lagrange multiplier (so we have a vector λ).
- We define a function named Lagrangian function depending on the variables x_j, $j = 1, \ldots, n$, and the multipliers λ_i :

$$\mathcal{L}(x, \lambda) = f(x) + \sum_{i \in \mathcal{E}} \lambda_i h_i(x).$$

Lagrange multipliers

Minimization problem with equality constraints :

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

$$\text{s.t.} \quad h_i(x) = 0 \text{ for } i \in \mathcal{E}.$$

- For each constraint $h_i(x) = 0$, we associate a real variable λ_i called Lagrange multiplier (so we have a vector λ).
- We define a function named Lagrangian function depending on the variables x_j, $j = 1, ..., n$, and the multipliers λ_i :

$$\mathcal{L}(x, \lambda) = f(x) + \sum_{i \in \mathcal{E}} \lambda_i h_i(x).$$

The gradient of $\mathcal{L}(x, \lambda)$ w.r.t. x is :

$$\nabla_x \mathcal{L}(x, \lambda) = \nabla_x f(x) + \sum_{i \in \mathcal{E}} \lambda_i \nabla_x h_i(x).$$

Lagrange multipliers

Minimization problem with equality constraints :

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

$$\text{s.t.} \quad h_i(x) = 0 \text{ for } i \in \mathcal{E}.$$

- For each constraint $h_i(x) = 0$, we associate a real variable λ_i called Lagrange multiplier (so we have a vector λ).
- We define a function named Lagrangian function depending on the variables x_j, $j = 1, ..., n$, and the multipliers λ_i :

$$\mathcal{L}(x, \lambda) = f(x) + \sum_{i\in\mathcal{E}} \lambda_i h_i(x).$$

The gradient of $\mathcal{L}(x, \lambda)$ w.r.t. x is :

$$\nabla_x \mathcal{L}(x, \lambda) = \nabla_x f(x) + \sum_{i\in\mathcal{E}} \lambda_i \nabla_x h_i(x).$$

The gradient of $\mathcal{L}(x, \lambda)$ w.r.t. λ_i is :

$$\nabla_{\lambda_i} \mathcal{L}(x, \lambda) = h_i(x).$$

Linear independence of gradients

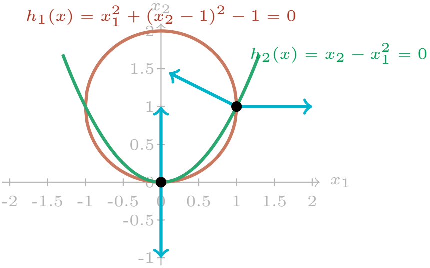
Optimality conditions in constrained cases often involve a condition known as the linear independence of constraint gradients condition, abbreviated LICQ.

LICQ condition at a point $x^\star$.

The LICQ condition is satisfied at a point $x^\star$ if, at this point, the set of active constraint gradients is linearly independent.

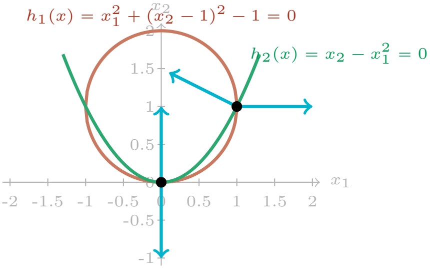
Linear independence of gradients - example

Example.



Linear independence of gradients - example

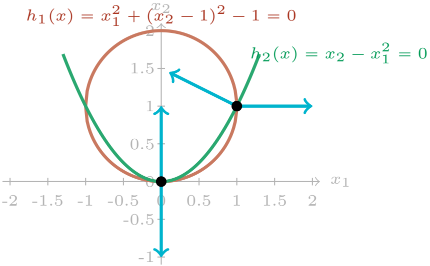
Example.



- At $x^\star = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, we have $\nabla h_1(x^\star) = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$ and $\nabla h_2(x^\star) = \begin{pmatrix} -2 \\ 1 \end{pmatrix}$

Linear independence of gradients - example

Example.



- At $x^\star = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, we have $\nabla h_1(x^\star) = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$ and $\nabla h_2(x^\star) = \begin{pmatrix} -2 \\ 1 \end{pmatrix}$
- At $x^\star = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, we have $\nabla h_1(x^\star) = \begin{pmatrix} 0 \\ -2 \end{pmatrix}$ and $\nabla h_2(x^\star) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$

Lagrangian-based optimality conditions

First-order necessary conditions

If x^* is a local minimum and the LICQ condition is satisfied in x^*, then there exists a vector λ^* (which is unique) such that :

$$\nabla_x \mathcal{L}(x^*, \lambda^*) = 0.$$

Lagrangian-based optimality conditions - Interpretations I

- This is a **first-order** condition

Lagrangian-based optimality conditions - Interpretations I

- This is a **first-order** condition
- In the absence of constraint ($\mathcal{E} = \emptyset$), we find $\nabla f(x^*) = 0$.

Lagrangian-based optimality conditions - Interpretations I

- This is a **first-order** condition
- In the absence of constraint ($\mathcal{E} = \emptyset$), we find $\nabla f(x^*) = 0$.
- This is a **necessary** condition that characterizes only **certain local** solutions:
  1. This condition is not sufficient: it can also be satisfied by maxima or saddle points.
  2. For a local minimum where the independence of the gradients of the constraints is not satisfied, the existence of λ_i allowing the optimality condition to be satisfied is not guaranteed.
  3. The condition does not distinguish between global and local solutions

Lagrangian-based optimality conditions - Interpretations I

- This is a **first-order** condition
- In the absence of constraint ($\mathcal{E} = \emptyset$), we find $\nabla f(x^*) = 0$.
- This is a **necessary** condition that characterizes only **certain local** solutions:
  1. This condition is not sufficient: it can also be satisfied by maxima or saddle points.
  2. For a local minimum where the independence of the gradients of the constraints is not satisfied, the existence of λ_i allowing the optimality condition to be satisfied is not guaranteed.
  3. The condition does not distinguish between global and local solutions
- The condition $\nabla_\lambda \mathcal{L}(x^*, \lambda^*) = 0$ is also necessary, and guarantees that x^* is feasible (satisfies the constraints):

$$\nabla_\lambda \mathcal{L}(x^*, \lambda^*) = 0 \Leftrightarrow h_i(x^*) = 0.$$

Lagrangian-based optimality conditions - Interpretations II

- This condition can also be written as $\nabla f(x^*) + \sum_{i \in \mathcal{E}} \lambda_i^* \nabla h_i(x^*) = 0$, where :

$$\nabla f(x^*) = -\sum_{i \in \mathcal{E}} \lambda_i^* \nabla h_i(x^*).$$

  The gradient of f can therefore be non-zero at an optimal solution, but it is then a linear combination of the gradients of the $h_i(x)$ constraints.

Lagrangian-based optimality conditions - Interpretations II

- This condition can also be written as $\nabla f(x^*) + \sum_{i \in \mathcal{E}} \lambda_i^* \nabla h_i(x^*) = 0$, where :

$$\nabla f(x^*) = -\sum_{i \in \mathcal{E}} \lambda_i^* \nabla h_i(x^*).$$

  The gradient of f can therefore be non-zero at an optimal solution, but it is then a linear combination of the gradients of the $h_i(x)$ constraints.

- Geometrically, each $\nabla_x h_i(x^*)$ points in a direction perpendicular to the equation surface $h_i(x) = 0$.

Lagrangian-based optimality conditions - Interpretations II

- This condition can also be written as $\nabla f(x^*) + \sum_{i\in\mathcal{E}} \lambda_i^* \nabla h_i(x^*) = 0$, where :

$$\nabla f(x^*) = -\sum_{i\in\mathcal{E}} \lambda_i^* \nabla h_i(x^*).$$

  The gradient of f can therefore be non-zero at an optimal solution, but it is then a linear combination of the gradients of the $h_i(x)$ constraints.
- Geometrically, each $\nabla_x h_i(x^*)$ points in a direction perpendicular to the equation surface $h_i(x) = 0$.
- This means that the gradient points in a linear combination of directions **normal** to the surfaces defined by the constraints $h_i(x) = 0$, whose intersection is the feasible domain/set.

Linear independence of gradients (LICQ)

If we wish to identify all the local minima of the problem

$$\min_{x \in \mathbb{R}^n} f(x) \text{ such that } h_i(x) = 0 \text{ for } i \in \mathcal{E},$$

Linear independence of gradients (LICQ)

If we wish to identify all the local minima of the problem

$$\min_{x \in \mathbb{R}^n} f(x) \text{ such that } h_i(x) = 0 \text{ for } i \in \mathcal{E},$$

not only must all feasible points be considered (x^*, λ^*) which satisfy the necessary optimality condition

$$\nabla_x \mathcal{L}(x^*, \lambda^*) = 0,$$

Linear independence of gradients (LICQ)

If we wish to identify all the local minima of the problem

$$\min_{x \in \mathbb{R}^n} f(x) \text{ such that } h_i(x) = 0 \text{ for } i \in \mathcal{E},$$

not only must all feasible points be considered (x^*, λ^*) which satisfy the necessary optimality condition

$$\nabla_x \mathcal{L}(x^*, \lambda^*) = 0,$$

but also take into account x^* points where the constraint gradients are not independent. :

$$\{\nabla h_i(x^*)\}_{i \in \mathcal{E}} \text{ is linearly dependent.}$$

Linear independence of gradients (LICQ)

If we wish to identify all the local minima of the problem

$$\min_{x \in \mathbb{R}^n} f(x) \text{ such that } h_i(x) = 0 \text{ for } i \in \mathcal{E},$$

not only must all feasible points be considered (x^*, λ^*) which satisfy the necessary optimality condition

$$\nabla_x \mathcal{L}(x^*, \lambda^*) = 0,$$

but also take into account x^* points where the constraint gradients are not independent. :

$$\{\nabla h_i(x^*)\}_{i \in \mathcal{E}} \text{ is linearly dependent.}$$

If one of the gradients cancels out, the whole set of gradients is automatically linearly dependent. :

$$\exists k \in \mathcal{E} \mid \nabla h_k(x^*) = 0 \Rightarrow \{\nabla h_i(x^*)\}_{i \in \mathcal{E}} \text{ is linearly dependent.}$$

Summary - for equality constraints only

First-order necessary conditions

If f and h_i are differentiable in a neighborhood of x^*

$$x^* \text{ local minimum and } \{\nabla h_i(x^*)\}_{i\in\mathcal{E}} \text{ lin. indep.} \Rightarrow \nabla_x \mathcal{L}(x^*, \lambda^*) = 0.$$

Full procedure

Minimization problem with equality constraints :

$$\min_{x \in \mathbb{R}^n} f(x) \text{ s.t. } h_i(x) = 0 \text{ for } i \in \mathcal{E}.$$

1. Write the Lagrangian function $\mathcal{L}(x, \lambda) = f(x) + \sum_{i \in \mathcal{E}} \lambda_i h_i(x)$.

Full procedure

Minimization problem with equality constraints :

$$\min_{x \in \mathbb{R}^n} f(x) \text{ s.t. } h_i(x) = 0 \text{ for } i \in \mathcal{E}.$$

1. Write the Lagrangian function $\mathcal{L}(x, \lambda) = f(x) + \sum_{i \in \mathcal{E}} \lambda_i h_i(x)$.
2. Identify all the solutions (x^*, λ^*) satisfying

$$\begin{pmatrix} \nabla_x \mathcal{L}(x^*, \lambda^*) \\ \nabla_\lambda \mathcal{L}(x^*, \lambda^*) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Full procedure

Minimization problem with equality constraints :

$$\min_{x \in \mathbb{R}^n} f(x) \text{ s.t. } h_i(x) = 0 \text{ for } i \in \mathcal{E}.$$

1. Write the Lagrangian function $\mathcal{L}(x, \lambda) = f(x) + \sum_{i \in \mathcal{E}} \lambda_i h_i(x)$.
2. Identify all the solutions (x^*, λ^*) satisfying

$$\begin{pmatrix} \nabla_x \mathcal{L}(x^*, \lambda^*) \\ \nabla_\lambda \mathcal{L}(x^*, \lambda^*) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

3. Add to this the feasible solutions where the gradients of the constraints are linearly dependent.

Full procedure

Minimization problem with equality constraints :

$$\min_{x \in \mathbb{R}^n} f(x) \text{ s.t. } h_i(x) = 0 \text{ for } i \in \mathcal{E}.$$

1. Write the Lagrangian function $\mathcal{L}(x, \lambda) = f(x) + \sum_{i \in \mathcal{E}} \lambda_i h_i(x)$.
2. Identify all the solutions (x^*, λ^*) satisfying

$$\begin{pmatrix} \nabla_x \mathcal{L}(x^*, \lambda^*) \\ \nabla_\lambda \mathcal{L}(x^*, \lambda^*) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

3. Add to this the feasible solutions where the gradients of the constraints are linearly dependent.
   - We have thus identified a set of feasible points including all local minima and possibly other points that are not local minima.

Full procedure

Minimization problem with equality constraints :

$$\min_{x \in \mathbb{R}^n} f(x) \text{ s.t. } h_i(x) = 0 \text{ for } i \in \mathcal{E}.$$

1. Write the Lagrangian function $\mathcal{L}(x, \lambda) = f(x) + \sum_{i \in \mathcal{E}} \lambda_i h_i(x)$.
2. Identify all the solutions (x^*, λ^*) satisfying

$$\begin{pmatrix} \nabla_x \mathcal{L}(x^*, \lambda^*) \\ \nabla_\lambda \mathcal{L}(x^*, \lambda^*) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

3. Add to this the feasible solutions where the gradients of the constraints are linearly dependent.
   - We have thus identified a set of feasible points including all local minima and possibly other points that are not local minima.
   - Any global minimum is necessarily found among these local minima

Exercise.

$$\min x_1^2 x_2 \text{ such that } x_1^2 + x_2^2 = 3.$$

Additional comments

- In certain situations, rather than using Lagrange multipliers multipliers, it may be advantageous to use constraints to eliminate variables by substitution, and eventually obtain a constraint-free problem with fewer variables.
- There is also a **second**-order version of the conditions
  – for instance, the necessary condition $\nabla^2 f(x^*) \succeq 0$ becomes

$$w^T \nabla_{xx} \mathcal{L}(x^*, \lambda^*) w \geqslant 0 \;\forall w \text{ such that } \nabla h_i(x^*)^T w = 0 \;\forall i \in \mathcal{E}$$

  – whereas the sufficient conditions $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*) \succ 0$ become:

$$\begin{aligned} &\nabla_x \mathcal{L}(x^*, \lambda^*) = 0 \text{ with } h_i(x^*) \;\forall i \in \mathcal{E} \\ &\text{and } w^T \nabla_{xx} \mathcal{L}(x^*, \lambda^*) w > 0 \;\forall w \neq 0 \text{ such that } \nabla h_i(x^*)^T w = 0 \;\forall i \in \mathcal{E} \end{aligned}$$

OCs for problems with equality and inequality constraints

Description

We wish to generalize the optimality conditions for the following problem:

Minimization problem with equality and inequality constraints. :

$$
\begin{aligned}
&\min_{x\in\mathbb{R}^n} f(x) \\
\text{s.t.} \quad & h_i(x) = 0 \text{ for } i \in \mathcal{E} \\
& h_i(x) \leqslant 0 \text{ for } i \in \mathcal{I}
\end{aligned}
$$

Description

We wish to generalize the optimality conditions for the following problem:

Minimization problem with equality and inequality constraints. :

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \; & f(x) \\ \text{s.t.} \quad & h_i(x) = 0 \text{ for } i \in \mathcal{E} \\ & h_i(x) \leqslant 0 \text{ for } i \in \mathcal{I} \end{aligned}$$

Active constraints and gradients independence.

Description

We wish to generalize the optimality conditions for the following problem:

Minimization problem with equality and inequality constraints. :

$$\begin{aligned} &\min_{x\in\mathbb{R}^n} f(x) \\ \text{s.t.} \quad & h_i(x) = 0 \text{ for } i \in \mathcal{E} \\ & h_i(x) \leqslant 0 \text{ for } i \in \mathcal{I} \end{aligned}$$

Active constraints and gradients independence.

- A constraint i is said to be active at x^* if $h_i(x^*) = 0$.

Description

We wish to generalize the optimality conditions for the following problem:

Minimization problem with equality and inequality constraints. :

$$\min_{x \in \mathbb{R}^n} f(x)$$
$$\text{s.t.} \quad h_i(x) = 0 \text{ for } i \in \mathcal{E}$$
$$h_i(x) \leqslant 0 \text{ for } i \in \mathcal{I}$$

Active constraints and gradients independence.

- A constraint i is said to be active at x^* if $h_i(x^*) = 0$.
- When x is feasible, equality constraints are active, but not necessarily inequality constraints. For x^*, the set of active constraints is defined by :

$$\mathcal{A}(x^*) = \mathcal{E} \cup \{i \in \mathcal{I} : h_i(x^*) = 0\}.$$

Description

We wish to generalize the optimality conditions for the following problem:

Minimization problem with equality and inequality constraints. :

$$\min_{x \in \mathbb{R}^n} f(x)$$
$$\text{s.t.} \quad h_i(x) = 0 \text{ for } i \in \mathcal{E}$$
$$h_i(x) \leqslant 0 \text{ for } i \in \mathcal{I}$$

Active constraints and gradients independence.

- A constraint i is said to be active at x^* if $h_i(x^*) = 0$.
- When x is feasible, equality constraints are active, but not necessarily inequality constraints. For x^*, the set of active constraints is defined by :

$$\mathcal{A}(x^*) = \mathcal{E} \cup \{i \in \mathcal{I} : h_i(x^*) = 0\}.$$

- A feasible point x satisfies the condition LICQ if and only if

$$\{\nabla h_i(x)\}_{i \in \mathcal{A}(x)} \text{ is linearly independent.}$$

Description

We wish to generalize the optimality conditions for the following problem:

Minimization problem with equality and inequality constraints. :

$$\min_{x \in \mathbb{R}^n} f(x)$$
$$\text{s.t.} \quad h_i(x) = 0 \text{ for } i \in \mathcal{E}$$
$$h_i(x) \leqslant 0 \text{ for } i \in \mathcal{I}$$

Active constraints and gradients independence.

- A constraint i is said to be active at x^* if $h_i(x^*) = 0$.
- When x is feasible, equality constraints are active, but not necessarily inequality constraints. For x^*, the set of active constraints is defined by :

$$\mathcal{A}(x^*) = \mathcal{E} \cup \{i \in \mathcal{I} : h_i(x^*) = 0\}.$$

- A feasible point x satisfies the condition LICQ if and only if

$$\{\nabla h_i(x)\}_{i \in \mathcal{A}(x)} \text{ is linearly independent.}$$

Lagrangian function.

$$\mathcal{L}(x, \lambda) = f(x) + \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i h_i(x).$$

Karush-Kuhn-Tucker necessary optimality conditions

First-order necessary conditions (KKT conditions)

If x^* is a local minimum and the set of gradients of the constraints $\{\nabla h_i(x^*)\}_{i \in \mathcal{A}(x^*)}$ is linearly independent, then there exists a vector λ^* (which is unique) such that :

$$\begin{aligned} h_i(x^*) &= 0 \;\; \forall i \in \mathcal{E} \\ h_i(x^*) &\leqslant 0 \;\; \forall i \in \mathcal{I} \end{aligned}$$

Karush-Kuhn-Tucker necessary optimality conditions

First-order necessary conditions (KKT conditions)

If x^* is a local minimum and the set of gradients of the constraints $\{\nabla h_i(x^*)\}_{i \in \mathcal{A}(x^*)}$ is linearly independent, then there exists a vector λ^* (which is unique) such that :

$$
\begin{aligned}
h_i(x^*) &= 0 \;\; \forall i \in \mathcal{E} \\
h_i(x^*) &\leqslant 0 \;\; \forall i \in \mathcal{I} \\
\nabla_x \mathcal{L}(x^*, \lambda^*) &= 0 \\
\lambda_i^* &\geqslant 0 \;\; \forall i \in \mathcal{I} \\
\lambda_i^* h_i(x^*) &= 0 \;\; \forall i \in \mathcal{I}
\end{aligned}
$$

Terminology

Definition (KKT point)

A feasible point x^* is called a **KKT point** if there exist a vector λ^* such that

$$\nabla_x \mathcal{L}(x^*, \lambda^*) = 0, \quad \lambda_i^* \geqslant 0 \ \ \forall i \in \mathcal{I}, \quad \lambda_i^* h_i(x^*) = 0 \ \ \forall i \in \mathcal{I}$$

Definition (regularity)

A feasible point x^* is called **regular** if the set $\{\nabla h_i(x)\}_{i \in \mathcal{A}(x)}$ is linearly independent.

- The KKT theorem states that a necessary local optimality condition of a regular point is that it is a KKT point.
- The additional requirement of regularity is not required in linearly constrained problems in which no such assumption is needed.

Karush-Kuhn-Tucker necessary optimality conditions

- It's a set of conditions that are necessary but not sufficient: the conditions can also be met by other points.

[1] as the complementarity conditions impose $\lambda_i^* = 0$ for any inactive constraint i.

Karush-Kuhn-Tucker necessary optimality conditions

- It's a set of conditions that are necessary but not sufficient: the conditions can also be met by other points.
- The complementarity conditions $\lambda_i^* h_i(x^*) = 0$ are equivalent to imposing $\lambda_i^* = 0$ for any inactive i constraint.

[1] as the complementarity conditions impose $\lambda_i^* = 0$ for any inactive constraint i.

Karush-Kuhn-Tucker necessary optimality conditions

- It's a set of conditions that are necessary but not sufficient: the conditions can also be met by other points.
- The complementarity conditions $\lambda_i^* h_i(x^*) = 0$ are equivalent to imposing $\lambda_i^* = 0$ for any inactive i constraint.
- The first condition $\nabla_x \mathcal{L}(x^*, \lambda^*) = 0$ can be written as $\nabla f(x^*) = -\sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i^* \nabla h_i(x^*)$ which can be simplified using the complementarity conditions[1] in :

$$\nabla f(x^*) = -\sum_{i \in \mathcal{A}(x^*)} \lambda_i^* \nabla h_i(x^*) \text{ and } \lambda_i^* \geqslant 0 \ \forall i \in \mathcal{I}.$$

This indicates that the gradient of f can therefore be non-zero in an optimal solution, but it is then a linear combination of the gradients of the active constraints h_i.

[1] as the complementarity conditions impose $\lambda_i^* = 0$ for any inactive constraint i.

Linear independence of gradients (LICQ) (idem)

If we wish to identify all the local minima of the problem

$$\min_{x \in \mathbb{R}^n} f(x) \text{ such that } h_i(x) = 0 \text{ for } i \in \mathcal{E} \text{ and } h_i(x) \leqslant 0 \text{ for } i \in \mathcal{I},$$

we must not only consider all feasible points (x^*, λ^*) that satisfy the necessary optimality conditions

$$\nabla_x \mathcal{L}(x^*, \lambda^*) = 0$$

$$\lambda_i^* \geqslant 0 \text{ for all } i \in \mathcal{I}$$

$$\lambda_i^* h_i(x^*) = 0 \text{ for all } i \in \mathcal{I}$$

but also take into account x^* points where the gradients of the active constraints are not independent. :

$$\{\nabla h_i(x^*)\}_{i \in \mathcal{A}(x^*)} \text{ is linearly dependent.}$$

Summary

First-order necessary conditions

If f and h_i are differentiable in a neighborhood of a point x^*

$$x^* \text{ local min. and } \{\nabla h_i(x^*)\}_{i \in \mathcal{A}(x^*)} \text{ lin. ind.} \Rightarrow \begin{cases} \nabla_x \mathcal{L}(x^*, \lambda^*) = 0 & \\ \lambda_i^* \geqslant 0 & \forall i \in \mathcal{I} \\ \lambda_i^* h_i(x^*) = 0 & \forall i \in \mathcal{I} \end{cases}$$

How to use optimal conditions

Minimization problem with equality and inequality constraints :

$$\min_{x \in \mathbb{R}^n} f(x) \text{ s.t. } h_i(x) = 0 \text{ for } i \in \mathcal{E} \text{ and } h_i(x) \leqslant 0 \text{ for } i \in \mathcal{I}.$$

1. Write the Lagrangian function $\mathcal{L}(x, \lambda) = f(x) + \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i h_i(x)$.

How to use optimal conditions

Minimization problem with equality and inequality constraints :

$$\min_{x \in \mathbb{R}^n} f(x) \text{ s.t. } h_i(x) = 0 \text{ for } i \in \mathcal{E} \text{ and } h_i(x) \leqslant 0 \text{ for } i \in \mathcal{I}.$$

1. Write the Lagrangian function $\mathcal{L}(x, \lambda) = f(x) + \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i h_i(x)$.
2. Identify all solutions (x^*, λ^*) satisfying KKT conditions.

How to use optimal conditions

Minimization problem with equality and inequality constraints :

$$\min_{x \in \mathbb{R}^n} f(x) \text{ s.t. } h_i(x) = 0 \text{ for } i \in \mathcal{E} \text{ and } h_i(x) \leqslant 0 \text{ for } i \in \mathcal{I}.$$

1. Write the Lagrangian function $\mathcal{L}(x, \lambda) = f(x) + \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i h_i(x)$.
2. Identify all solutions (x^*, λ^*) satisfying KKT conditions.
3. Add to this the feasible solutions where the gradients of the constraints are linearly dependent.
   (We have thus identified a set of feasible points including all local minima and possibly other points that are not local minima).

Extreme value theorem

Global minimum? Extreme value theorem

Extreme Value Theorem: f reaches its bounds

Let $f : [a, b] \subset \mathbb{R} \to \mathbb{R}$ continuous on $[a, b]$, then

$$f([a, b]) = \left[\min_{x \in [a,b]} f(x), \max_{x \in [a,b]} f(x) \right].$$

Extreme value theorem (*Weierstrass theorem*)

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuous function defined on a $X \subseteq \mathbb{R}^n$ compact (and non-empty) set, then there exists a global minimum of f on X.

Compact set = set *closed and bounded*.

Global minimum? Extreme value theorem

Extreme Value Theorem: f reaches its bounds

Let $f : [a, b] \subset \mathbb{R} \to \mathbb{R}$ continuous on $[a, b]$, then

$$f([a, b]) = \left[\min_{x \in [a,b]} f(x), \max_{x \in [a,b]} f(x) \right].$$

Extreme value theorem (*Weierstrass theorem*)

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuous function defined on a $X \subseteq \mathbb{R}^n$ compact (and non-empty) set, then there exists a global minimum of f on X.

Compact set = set *closed and bounded*. Examples :

- $\|x\| \leqslant R$ (ball),
- $\|x\| = R$,
- $l \leqslant x \leqslant u$ (box)

Global minimum? Extreme value theorem

Extreme Value Theorem: f reaches its bounds

Let $f : [a, b] \subset \mathbb{R} \to \mathbb{R}$ continuous on $[a, b]$, then

$$f([a, b]) = \left[\min_{x \in [a,b]} f(x), \max_{x \in [a,b]} f(x) \right].$$

Extreme value theorem (*Weierstrass theorem*)

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuous function defined on a $X \subseteq \mathbb{R}^n$ compact (and non-empty) set, then there exists a global minimum of f on X.

Compact set = set *closed and bounded*. Examples :

- $\|x\| \leqslant R$ (ball),
- $\|x\| = R$,
- $l \leqslant x \leqslant u$ (box)

The theorem implies that, for the following optimization problem,

$$\min f(x) \text{ such that } x \in X,$$

where the constraints form a compact, non-empty feasible set X and f is continuous on X, then f reaches its global minimum at a point $x^* \in X$.

Can the result be generalized if the X domain is not compact?

Can the result be generalized if the X domain is not compact?

Example. $\min f(x_1, x_2) = x_1^2 + x_2^2 \quad \text{such that} \quad x_1 + x_2 = 1.$

The feasible set is not compact in this example...

Can the result be generalized if the X domain is not compact?

Example. $\min f(x_1, x_2) = x_1^2 + x_2^2$ such that $x_1 + x_2 = 1.$

The feasible set is not compact in this example...

However, it is clear that the solution $(1, 0)$ is feasible and $f(1, 0) = 1.$

Can the result be generalized if the X domain is not compact?

Example. $\min f(x_1, x_2) = x_1^2 + x_2^2 \quad \text{such that} \quad x_1 + x_2 = 1.$

The feasible set is not compact in this example...

However, it is clear that the solution $(1, 0)$ is feasible and $f(1, 0) = 1$.
It is therefore equivalent to solve :

$$\min f(x_1, x_2) = x_1^2 + x_2^2 \quad \text{such that} \quad x_1 + x_2 = 1 \text{ et } x_1^2 + x_2^2 \leqslant 1.$$

Can the result be generalized if the X domain is not compact?

Example. $\min f(x_1, x_2) = x_1^2 + x_2^2 \quad \text{such that} \quad x_1 + x_2 = 1.$

The feasible set is not compact in this example...

However, it is clear that the solution $(1, 0)$ is feasible and $f(1, 0) = 1$.
It is therefore equivalent to solve :

$$\min f(x_1, x_2) = x_1^2 + x_2^2 \quad \text{such that} \quad x_1 + x_2 = 1 \text{ et } x_1^2 + x_2^2 \leqslant 1.$$

Since f is continuous, and the admissible set formed by the constraints is compact and non-empty, this means that a global minimum of this problem exists among the feasible points.

Can the result be generalized if the X domain is not compact?

Example. $\min f(x_1, x_2) = x_1^2 + x_2^2$ such that $x_1 + x_2 = 1$.

The feasible set is not compact in this example...

However, it is clear that the solution $(1, 0)$ is feasible and $f(1, 0) = 1$.
It is therefore equivalent to solve :

$$\min f(x_1, x_2) = x_1^2 + x_2^2 \quad \text{such that} \quad x_1 + x_2 = 1 \text{ et } x_1^2 + x_2^2 \leqslant 1.$$

Since f is continuous, and the admissible set formed by the constraints is compact and non-empty, this means that a global minimum of this problem exists among the feasible points.

Corollary. For f continuous, if there exists $\bar{x} \in X$ such that $f(\bar{x})$ is finite, if $\{x \mid f(x) \leqslant f(\bar{x})\} \cap X$ is compact, then f reaches its global min. on X.

Examples

The problems

$$\min_{x_1,x_2} x_1 + x_2 \quad \text{such that} \quad x_1^2 + x_2^2 = 1,$$

and

$$\min_{x_1,x_2} x_1 + x_2 \quad \text{such that} \quad x_1^2 + x_2^2 \leqslant 1,$$

Examples

The problems

$$\min_{x_1,x_2} x_1 + x_2 \quad \text{such that} \quad x_1^2 + x_2^2 = 1,$$

and

$$\min_{x_1,x_2} x_1 + x_2 \quad \text{such that} \quad x_1^2 + x_2^2 \leqslant 1,$$

have a compact domain and are composed of continuous functions.
So they reach their global minimum.

Examples

The problems

$$\min_{x_1,x_2} x_1 + x_2 \quad \text{such that} \quad x_1^2 + x_2^2 = 1,$$

and

$$\min_{x_1,x_2} x_1 + x_2 \quad \text{such that} \quad x_1^2 + x_2^2 \leqslant 1,$$

have a compact domain and are composed of continuous functions.
So they reach their global minimum.

This implies that the global minimum must be found among candidate points, i.e. among stationary points and non-LICQ points.

Conclusions

Summary

We have seen:

- what is a descent direction d at x for $f : \mathbb{R}^n \to \mathbb{R}$: $d^T \nabla f(x) < 0$, hence function *locally*[2] decreases along d.
- First-order necessary conditions for a minimization problem with equality constraints.
- The notion of **linear independence** for the set of constraint gradients (LICQ).
- The KKT conditions (first-order necessary conditions) for minimization problems with **equality** and **inequality** constraints.

[2] in a neighborhood of x

Preparations for the next lecture

- Review the lecture;
- **Exercises.**
  1. $$\min -0.1(x_1 - 4)^2 + x_2^2 \text{ such that } x_1^2 + x_2^2 \leqslant 1.$$
  2. $$\min 4x_1^2 + x_2^2 - x_1 - 2x_2 \text{ such that } 2x_1 + x_2 \leqslant 1, x_1^2 \leqslant 1.$$
  3. Constrained Least Squares:
     $$\begin{aligned} &\min_x \|Ax - b\|_2^2 \\ &\text{s.t.} \quad \|x\|_2^2 \leqslant \alpha \end{aligned}$$
     where $A \in \mathbb{R}^{m \times n}$ has full column rank, $b \in \mathbb{R}^m$ and $\alpha > 0$.

Goodbye, So Soon

THANKS FOR THE ATTENTION

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