Constrained Optimization

K.Bouanane

May 10, 2023

Optimization with equality constraints

Introduction

Consider the problem

Introduction

Consider the problem

$$\begin{cases} \min\limits_{\mathbf{x}\in\mathbb{R}^n} J(\mathbf{x}) \\ \text{such that} \\ \theta_i(\mathbf{x}) = 0, i = 1..m \end{cases}$$

Introduction

Consider the problem

$$\begin{cases} \min_{\mathbf{x} \in \mathbb{R}^n} J(\mathbf{x}) \\ \text{such that} \\ \theta_i(\mathbf{x}) = 0, i = 1..m \end{cases}$$

that consists in finding $\mathbf{x} \in \mathbb{R}^n$ that minimizes J and verifies the constraints $\theta_i(\mathbf{x}) = 0, i = 1..m$ at the same time.

Example

Suppose a consumer consumes two goods, x and y and has utility function $u(x, y) = xy$.
He has a budget of $400\$$. The price of x is $P_x = 10$ and the price of y is $P_y = 20$.
We want to find his optimal consumption bundle.
Mathematically, we want to find a solution to the problem

$$\begin{cases} \max u(x, y) = xy \\ \text{ such that} \\ 10x + 20y = 400. \end{cases}$$

Lagrange multipliers and optimality conditions

Definitions and Notations

Let $J, \theta_1, \theta_2, \ldots, \theta_m : \mathbb{R}^n \longmapsto \mathbb{R}$ functions of class C^1, with $m \in \mathbb{N}^*$.

Definitions and Notations

Let $J, \theta_1, \theta_2, \ldots, \theta_m : \mathbb{R}^n \longmapsto \mathbb{R}$ functions of class C^1, with $m \in \mathbb{N}^*$.
We note

$$\widetilde{O} = \{u \in \mathbb{R}^n, \theta_i(u) = 0, i = 1, 2, \ldots, m\}.$$

Definitions and Notations

Let $J, \theta_1, \theta_2, \ldots, \theta_m : \mathbb{R}^n \longmapsto \mathbb{R}$ functions of class C^1, with $m \in \mathbb{N}^*$.
We note

$$\widetilde{O} = \{u \in \mathbb{R}^n, \theta_i(u) = 0, i = 1, 2, \ldots, m\}.$$

We say that $\widetilde{O}$ is a **variety** of $\mathbb{R}^n$.

Definitions and Notations

- If u is such that the family of vectors $\nabla\theta_i(u)_{i=1,\dots,m}$ forms a free system in $\mathbb{R}^n$, then u is called a **regular** point of $\widetilde{O}$.

Definitions and Notations

- If u is such that the family of vectors $\nabla\theta_i(u)_{i=1,\ldots,m}$ forms a free system in $\mathbb{R}^n$, then u is called a **regular** point of $\widetilde{O}$.
- The variety $\widetilde{O}$ is said to be **regular** if all points of $\widetilde{O}$ are regular.

Remark

An equivalent condition of regularity of $\widetilde{O}$ at a point u is that

Remark

An equivalent condition of regularity of $\widetilde{O}$ at a point u is that $\mathrm{rk}(\mathbf{B}) = m$,
where $\mathbf{B}$ is the Jacobian matrix given by

$$\mathbf{B}_{ij}(u) = \frac{\partial \theta_i}{\partial u_j}(u), i = 1, \ldots m, j = 1, \ldots, n.$$

Remark

An equivalent condition of regularity of $\widetilde{O}$ at a point u is that $\mathrm{rk}(\mathbf{B}) = m$,
where $\mathbf{B}$ is the Jacobian matrix given by

$$\mathbf{B}_{ij}(u) = \frac{\partial \theta_i}{\partial u_j}(u), i = 1, \ldots m, j = 1, \ldots, n.$$

And we have then necessarily $m \leq n$.

Example

Consider the variety

$$\widetilde{O} = \{(x,y) \in \mathbb{R}^2 \mid \exp(x) - \cos(y) = 0; x + y = 0\}$$

To test whether a point is regular or not we must first compute the Jacobian of $\Theta = (\theta_1, \theta_2)^\top$.

- The point (0,0) is not regular,
- the variety $\widetilde{O}$ is not regular.

Lagrange Multipliers Theorem

Theorem 1

Let u^* be a regular point of $\widetilde{\mathrm{O}}$ such that u^* is a local extremum of J on $\widetilde{\mathrm{O}}$ (a point of local minimum or a point of local maximum).

Theorem 1

Let u^* be a regular point of $\widetilde{\mathrm{O}}$ such that u^* is a local extremum of J on $\widetilde{\mathrm{O}}$ (a point of local minimum or a point of local maximum).
Then there exist $\lambda_1^*, \lambda_2^*, \ldots \lambda_m^* \in \mathbb{R}^m$, called ***Lagrange multipliers****, such that*

Theorem 1

Let u^* be a regular point of $\widetilde{O}$ such that u^* is a local extremum of J on $\widetilde{O}$ (a point of local minimum or a point of local maximum).
*Then there exist $\lambda_1^*, \lambda_2^*, \dots \lambda_m^* \in \mathbb{R}^m$, called **Lagrange multipliers**, such that*

$$\nabla J(u^*) = \sum_{i=1}^{m} \lambda_i^* \nabla \theta_i(u^*). \tag{1}$$

Remarks

- The system (1) gives necessary optimality conditions **first order necessary conditions of optimality.**

Remarks

- The system (1) gives necessary optimality conditions **first order necessary conditions of optimality**.
- These conditions are, generally, not sufficient. There are n equations with $n+m$ unknowns.

Remarks

- The system (1) gives necessary optimality conditions **first order necessary conditions of optimality**.
- These conditions are, generally, not sufficient. There are n equations with $n+m$ unknowns.
- But the fact that $u^* \in \widetilde{O}$ still gives us m equations $\theta_i(u^*) = 0$, $i = 1, \ldots, m$, which gives us a total of $n+m$ equations with $n+m$ unknowns.

We define **the Lagrangian**

$$L(u, \lambda) = J(u) - \lambda^T \theta(u)$$

with $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_m)^T$.

We define **the Lagrangian**

$$L(u,\lambda) = J(u) - \lambda^T \theta(u)$$

with $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_m)^T$.
The conditions can then be written as follows

$$\begin{cases} \nabla_u L(u,\lambda) = 0 \in \mathbb{R}^n \\ \nabla_\lambda L(u,\lambda) = 0 \in \mathbb{R}^m. \end{cases}$$

Example

$$\begin{cases} \min f(x,y) = x^4 + y^4 \\ (x,y) \in R^2; \text{ such that } x^2 + y^2 = 1. \end{cases} \tag{2}$$

We write the Lagrangian

$$L(x, y, \lambda) = x^4 + y^4 - \lambda(x^2 + y^2 - 1).$$

Using the Lagrange multipliers theorem, we have

We write the Lagrangian

$$L(x, y, \lambda) = x^4 + y^4 - \lambda(x^2 + y^2 - 1).$$

Using the Lagrange multipliers theorem, we have

$$\begin{cases} x(x^2 - \lambda) = 0 \\ y(y^2 - \lambda) = 0 \\ x^2 + y^2 - 1 = 0 \end{cases}$$

Geometric interpretation of Lagrange multipliers

Consider a surface S not passing through the origin. We propose to determine the points u^* of S closest to the origin.

Consider a surface S not passing through the origin. We propose to determine the points u^* of S closest to the origin.
The function to minimize here is the distance function

$$J(u) = \|u\|,$$

Consider a surface S not passing through the origin. We propose to determine the points u^* of S closest to the origin.
The function to minimize here is the distance function

$$J(u) = \|u\|,$$

The set of constraints $\widetilde{O}$ is the surface itself since we must indeed have $u^* \in S$.
We are therefore dealing with a constrained problem

$$\min_{u \in S} \quad J(u).$$

Let us denote by r the distance of a point u from the origin.

Let us denote by r the distance of a point u from the origin. Then, a point is at distance r from the origin if it satisfies: $\|u\| = r$, that is, if it lies on the sphere of radius r centered in the origin.

Let us denote by r the distance of a point u from the origin.
Then, a point is at distance r from the origin if it satisfies:
$\|u\| = r$, that is, if it lies on the sphere of radius r centered in the origin.
Let's start at $r = 0$ and then increase the value of r.

Let us denote by r the distance of a point u from the origin.
Then, a point is at distance r from the origin if it satisfies:
$\|u\| = r$, that is, if it lies on the sphere of radius r centered in the origin.
Let's start at $r = 0$ and then increase the value of r.
At some point, this level surface will touch S, each point of contact u^* being then a point we are looking for.

Let us denote by r the distance of a point u from the origin.
Then, a point is at distance r from the origin if it satisfies:
$\|u\| = r$, that is, if it lies on the sphere of radius r centered in the origin.
Let's start at $r = 0$ and then increase the value of r.
At some point, this level surface will touch S, each point of contact u^* being then a point we are looking for.
To determine the coordinates of the contact point, we assume that S is described by a Cartesian equation: $\theta(u) = 0$.

If now S has a tangent plane at a point of contact, this tangent plane must also be tangent to the level surface.

If now S has a tangent plane at a point of contact, this tangent plane must also be tangent to the level surface.
Therefore, the gradient of the surface $\theta(u) = 0$ must be parallel to the gradient of the contact surface $J(u) = r$.

If now S has a tangent plane at a point of contact, this tangent plane must also be tangent to the level surface.
Therefore, the gradient of the surface $\theta(u) = 0$ must be parallel to the gradient of the contact surface $J(u) = r$.
This means that there exists a scalar λ such that $\nabla J = \lambda \nabla \theta$, at any contact point.

If now S has a tangent plane at a point of contact, this tangent plane must also be tangent to the level surface.
Therefore, the gradient of the surface $\theta(u) = 0$ must be parallel to the gradient of the contact surface $J(u) = r$.
This means that there exists a scalar λ such that $\nabla J = \lambda \nabla \theta$, at any contact point.
We can see that this last equation corresponds well to the Lagrange multiplier approach, λ being the multiplier.

Second order optimality conditions

Theorem 2

Let u^* be a regular point of the variety $\widetilde{O} = \{u \mid \theta(u) = 0\}$. We suppose that θ and J are of class C^2. If u^* is a local minimum point of J on $\widetilde{O}$, then there exists a vector $\lambda^* \in \mathbb{R}^m$ such that

$$\nabla L(u^*, \lambda^*) = \nabla J(u^*) - \sum_{i=1}^{m} \lambda_i^* \nabla \theta_i(u^*) = 0,$$

and moreover $\nabla^2_{uu} L(u^*, \lambda^*)$ is positive semidefinite on $\ker \nabla \theta(u^*)$.

Example

We want to minimize $f(u) = x^2 + y^2 + z^2$ in the set

$$\widetilde{O} = \{u = (x, y, z) \in \mathbb{R}^3, x + y + 3z - 2 = 0, 5x + 2y + z - 5 = 0\}.$$

We define then the Lagrangian

$$L(u, \lambda) = x^2 + y^2 + z^2 - \lambda_1(x + y + 3z - 2) - \lambda_2(5x + 2y + z - 5).$$

with $\lambda = (\lambda_1, \lambda_2)^T$.

After solving the system

$$\begin{cases} \frac{\partial L}{\partial u}(u, \lambda) = 0 \\ \frac{\partial L}{\partial \lambda}(u, \lambda) = 0 \end{cases}$$

After solving the system

$$\begin{cases} \frac{\partial L}{\partial u}(u, \lambda) = 0 \\ \frac{\partial L}{\partial \lambda}(u, \lambda) = 0 \end{cases}$$

We obtain the unique critical point $u^* = [0.8043; 0.3478; 0.2826]^T$ and $\lambda^* = -[0.0870; 0.3044]^T$. However we don't know the nature of this point.

Example

So we calculate $L_{uu}(u^*, \lambda^*)$. We have

$$L_{uu}(u^*, \lambda^*) = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

The matrix $L_{uu}(u^*, \lambda^*)$ is positive definite, so u^* is indeed a point of the global minimum of f on $\widetilde{O}$.

Optimization under inequality constraints

Introduction and notations

Le us consider the functions $J, \theta_1, \theta_2, \ldots \theta_m : \mathbb{R}^n \mapsto \mathbb{R}$ of class C^1, with $m \in \mathbb{N}^*$.

Le us consider the functions $J, \theta_1, \theta_2, \dots \theta_m : \mathbb{R}^n \mapsto \mathbb{R}$ of class C^1, with $m \in \mathbb{N}^*$.

We denote:

$$O = \{x \in \mathbb{R}^n, \theta_i(x) \leq 0, \forall i = 1, 2, \dots m\}$$

Le us consider the functions $J, \theta_1, \theta_2, \dots \theta_m : \mathbb{R}^n \mapsto \mathbb{R}$ of class C^1, with $m \in \mathbb{N}^*$.
We denote:

$$O = \{x \in \mathbb{R}^n, \theta_i(x) \leq 0, \forall i = 1, 2, \dots m\}$$

We are interested in solving the minimization problem:

$$\min_{x \in O} J(x). \tag{3}$$

Hypothesis: We will assume that the functions θ_i are such that $O \neq \emptyset$ and also that O is not reduced to a single point or a finite number of points (in which case the minimization problem (3) is trivial!)

Hypothesis: We will assume that the functions θ_i are such that $O \neq \emptyset$ and also that O is not reduced to a single point or a finite number of points (in which case the minimization problem (3) is trivial!)

We will give the conditions of optimality similar to the system

$$\nabla J(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla \theta_i(x^*) = 0$$

but for inequality constraints.

- We say that the constraint $\theta_i(u) \leq 0$ is **active** at the point $v \in O$, if $\theta_i(v) = 0$

- We say that the constraint $\theta_i(u) \leq 0$ is **active** at the point $v \in O$, if $\theta_i(v) = 0$
- We then define the set $I(v) = \{i \in 1, 2 \ldots, m, \theta_i(v) = 0\}$

- We say that the constraint $\theta_i(u) \leq 0$ is **active** at the point $v \in O$, if $\theta_i(v) = 0$
- We then define the set $I(v) = \{i \in 1, 2 \ldots, m, \theta_i(v) = 0\}$
- and the variety (defined only if $I(v) \neq \emptyset$).
  $\widetilde{O}(v) = \{u \in \mathbb{R}^n, \theta_i(u) = 0, \forall i \in I(v)\}$ (note that $v \in \widetilde{O}(v)$).

- We say that the constraint $\theta_i(u) \leq 0$ is **active** at the point $v \in O$, if $\theta_i(v) = 0$
- We then define the set $I(v) = \{i \in 1, 2 \ldots, m, \theta_i(v) = 0\}$
- and the variety (defined only if $I(v) \neq \emptyset$).
$\widetilde{O}(v) = \{u \in \mathbb{R}^n, \theta_i(u) = 0, \forall i \in I(v)\}$ (note that $v \in \widetilde{O}(v)$).
- We will say that $v \in O$ is a regular **point** of O if either $I(v) = \emptyset$ or v is a regular point of the variety $\widetilde{O}(v)$.

Example

Consider the set O defined by the constraints

$$\begin{cases} x^2 - y^2 \leq 0 \\ -\exp(x) + y^3 - 2 \leq 0 \end{cases}$$

Example

Consider the set O defined by the constraints

$$\begin{cases} x^2 - y^2 \leq 0 \\ -\exp(x) + y^3 - 2 \leq 0 \end{cases}$$

for $v = (1,1)$, we have $I(v) = \{1\}$ and $\widetilde{O}(v) = \{(x,y) \in \mathbb{R}^2 \mid x^2 - y^2 = 0\}$
$v = (1,1)$ is a regular point of $\widetilde{O}(v)$ and thus is a regular point of O.

First order optimality conditions: Karush-kuhn-Tucker multipliers

First order optimality conditions: Karush-kuhn-Tucker multipliers

Theorem 3

Let x^* be a regular point of O.

First order optimality conditions: Karush-kuhn-Tucker multipliers

Theorem 3

Let x^* be a regular point of O.
If x^* is a point of local minimum of J on O, then there exists $p_1^*, p_2^*, \dots p_m^* \in [0, \infty[$, called ***Karush-Kuhn-Tucker multipliers****, such that:*

First order optimality conditions: Karush-kuhn-Tucker multipliers

Theorem 3

Let x^* be a regular point of O.
If x^* is a point of local minimum of J on O, then there exists $p_1^*, p_2^*, \dots p_m^* \in [0, \infty[$, called ***Karush-Kuhn-Tucker multipliers****, such that:*

$$\nabla J(x^*) + \sum_{i=1}^{m} p_i^* \nabla \theta_i(x^*) = 0 \tag{4}$$

$$p_i^* \theta_i(x^*) = 0, \forall i = 1, 2, \dots, m. \tag{5}$$

Example

Consider the problem

$$\begin{cases} \min f(x,y) = x^2 + y^2 \\ s.c \\ -x - y \leq -1 \end{cases}$$

If we consider the set

$$O = \{x \in \mathbb{R}^n, \theta_i \leq 0, i = 1, \ldots, m_1, \varphi_j(x) = 0, j = 1, \ldots, m_2\}$$

with $m_2 \geq 1$, then we can consider O as given only by inequality constraints by writing :

If we consider the set

$$O = \{x \in \mathbb{R}^n, \theta_i \leq 0, i = 1, \ldots, m_1, \varphi_j(x) = 0, j = 1, \ldots, m_2\}$$

with $m_2 \geq 1$, then we can consider O as given only by inequality constraints by writing :

$$O = \{x \in \mathbb{R}^n, \theta_i(x) \leq 0, i = 1, \ldots, m_1, \\ \varphi_j(x) \leq 0, j = 1, \ldots, m_2, -\varphi_j(x) \leq 0, j = 1, \ldots, m_2\} \quad (6)$$

If we consider the set

$$O = \{x \in \mathbb{R}^n, \theta_i \leq 0, i = 1, \ldots, m_1, \varphi_j(x) = 0, j = 1, \ldots, m_2\}$$

with $m_2 \geq 1$, then we can consider O as given only by inequality constraints by writing :

$$\begin{aligned} O = \{x \in \mathbb{R}^n, \theta_i(x) \leq 0, i = 1, \ldots, m_1, \\ \varphi_j(x) \leq 0, j = 1, \ldots, m_2, -\varphi_j(x) \leq 0, j = 1, \ldots, m_2\} \end{aligned} \quad (6)$$

It is easy to see that no point x of O is regular, because on the one hand the two constraints $\varphi_j(x) \leq 0$ and $-\varphi_j(x) \leq 0$ are active and on the other hand, no family of vectors which contains both $\nabla\varphi_j(x)$ and $\nabla\varphi_j(x)$ can be independent.

Conclusion:

The regularity assumption is not satisfied for any $x \in O$ if O has at least one equality constraint artificially transformed into two inequality constraints.

Example

The set $O = \{x \in \mathbb{R}^2, x_1^2 - x_2 \le 0, x_2 - x_1 - 4 = 0\}$ can be written as

$$O = \{x \in \mathbb{R}^2, x_1^2 - x_2 \le 0, x_2 - x_1 - 4 \le 0, -x_2 + x_1 + 4 \le 0\}.$$

Example

The set $O = \{x \in \mathbb{R}^2, x_1^2 - x_2 \leq 0, x_2 - x_1 - 4 = 0\}$ can be written as

$$O = \{x \in \mathbb{R}^2, x_1^2 - x_2 \leq 0, x_2 - x_1 - 4 \leq 0, -x_2 + x_1 + 4 \leq 0\}.$$

By introducing the functions: $\theta_1, \theta_2, \theta_3 : \mathbb{R}^2 \to \mathbb{R}$ given by

$$\theta_1(x) = x_1^2 - x_2, \theta_2(x) = x_2 - x_1 - 4, \theta_3(x) = -x_2 + x_1 + 4$$

we can write O in the standard form

$$O = \{x \in \mathbb{R}^2, \theta_i(x) \leq 0, i = 1, 2, 3\}$$

Example

Example

It is clear then that for any $x \in O$, we have $I(x) = \{2, 3\}$.

Example

It is clear then that for any $x \in O$, we have $I(x) = \{2, 3\}$.
On the other hand, $\nabla\theta_2(x) = (-1, 1)^T$ and $\nabla\theta_3(x) = (1, -1)^T = -\nabla\theta_2(x)$.
So no family of vectors that includes $\nabla\theta_2(x)$ and $\nabla\theta_3(x)$ can be independent, so no $x \in O$ is regular.

To remedy this kind of drawback, we weaken the regularity hypothesis of the Theorem 3.

To remedy this kind of drawback, we weaken the regularity hypothesis of the Theorem 3.

We say that the constraints of O are qualified in $v \in O$ if

To remedy this kind of drawback, we weaken the regularity hypothesis of the Theorem 3.

We say that the constraints of O are qualified in $v \in O$ if

1. Either $I(v) = \emptyset$
2. Or there exists $w \in \mathbb{R}^n$ such that $\forall i \in I(v)$ we have,

$$< \nabla\theta_i(v), w > \leq 0,$$

   with

$$< \nabla\theta_i(v), w > < 0 \text{ si } \theta_i \text{ is non-affine.}$$

Remarks

1. If θ_i is affine for all $i \in I(v)$ then the constraints of O are qualified in v.
2. In particular if the functions θ_i are affine for all $i = 1, \dots, m$, then the constraints are qualified at any point of O.
3. As a consequence, the vector w "points" to the set O (so $\exists t_0 > 0$ such that $v + tw \in O, \forall t \in [0, t_0]$)

We have then

If a point $v \in O$ is regular then the constraints of O are qualified in v.

In other words, the qualification condition is **weaker** than the regularity condition.

We then have the following result, stronger than Theorem 3

Theorem 4

Let $x^* \in O$ such that the constraints of O are qualified in x^*. If x^* is a local minimum point of J on O then there exist $p_1^*, p_2^*, \dots p_m^* \in [0, +\infty[$ (KKT multipliers), such that:

$$\nabla J(x^*) + \sum_{i=1}^{m} p_i^* \nabla \theta_i(x^*) = 0 \tag{7}$$

$$p_i^* \theta_i(x^*) = 0, \forall i = 1, 2, \dots, m. \tag{8}$$

Remark

The statement of this last result is obtained from the statement of Theorem 3, by replacing the hypothesis "x^* regular point of O" by the weaker hypothesis "the constraints of O are qualified in x^*".

The following result gives us a sufficient condition for qualification at any point of O:

Proposition 1

Suppose that the functions $\theta_1, \theta_2, \ldots, \theta_m$ are convex. Let us also suppose that

- either all functions $\theta_1, \theta_2, \ldots, \theta_m$ are affine

- or there exists $y \in \mathbb{R}^n$ such that for all $i = 1, \ldots, m$, we have

$$\theta_i(y) \leq 0$$
$$\theta_i(y) < 0 \text{ if } \theta_i \text{ is not affine.}$$

Then for all $x \in O$, constraints of O are qualified in x.

Example 1

$$O_1 = \{x \in \mathbb{R}^2, x_2 \geq x_1^2, x_2 - x_1 = 4\}.$$

It is clear that we can write O_1 in the form

$$O_1 = \{x \in \mathbb{R}^2, \theta_i(x) \leq 0, i = 1, 2, 3\}$$

with $\theta_1, \theta_2, \theta_3 : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$\theta_1(x) = x_1^2 - x_2, \theta_2(x) = x_2 - x_1 - 4, \theta_3(x) = -x_2 + x_1 + 4.$$

We observe that θ_1, θ_2 and θ_3 are convex, with in addition θ_2 and θ_3 affine.

Taking $y = (0, 4)^T$, we have

$$\theta_1(y) < 0, \theta_2(y) = \theta_3(y) = 0$$

Then the assumptions of the proposition 1 are satisfied. So all the points of O_1 are qualified

Example 2

$$O_2 = \{x \in \mathbb{R}^2, x_2 = x_1^2, x_2 - x_1 \leq 4\}.$$

We can write

$$O_2 = \{x \in \mathbb{R}^2, \theta_i(x) \leq 0, i = 1, 2, 3\}$$

with $\theta_1(x) = x_1^2 - x_2, \theta_2(x) = -x_1^2 + x_2, \theta_3(x) = x_2 - x_1 - 4$.
Since θ_2 is not convex, Proposition 1 does not apply (this does not automatically imply that the constraints are not qualified!)

Let $x \in O_2$. It is clear that $\{1,2\} \subset I(x)$. To have the qualification, there should exist $w \in \mathbb{R}^n$ such that

$$< \nabla\theta_1(x), w >< 0 \text{ and } < \nabla\theta_2(x), w >< 0$$

but this is impossible because $\nabla\theta_2(x) = -\nabla\theta_1(x)$. So no point of O_2 is qualified.

Example

$$\min_{u_1,u_2} \frac{1}{2}(u_1 - 1)^2 + \frac{1}{2}(u_2 - 2)^2,$$

such that $u_1 - u_2 = 1, u_1 + u_2 \leq 2, u_1 \geq 0, u_2 \geq 0.$

Projected Gradient Method

Necessary condition for local minimum

Directional derivative

Consider a function $f : \mathbb{R}^n \to \mathbb{R}$ of class C^1. Let h be a vector from

The directional derivative of f at a point u in the direction of a vector h is defined by

$$\nabla_h f(u) = \lim_{t \to 0} \frac{f(u + th) - f(u)}{t}.$$

It is a measure of the rate of change of the function f at u in the direction of h.

Directional derivative

Consider a function $f : \mathbb{R}^n \to \mathbb{R}$ of class C^1. Let h be a vector from

The directional derivative of f at a point u in the direction of a vector h is defined by

$$\nabla_h f(u) = \lim_{t \to 0} \frac{f(u+th) - f(u)}{t}.$$

It is a measure of the rate of change of the function f at u in the direction of h. It is also defined as the dot product of the gradient of f at the point u and h. More formally

$$\nabla_h f(u) = \langle \nabla f, h \rangle$$

Necessary condition

Theorem 5

Let $\Omega \subset \mathbb{R}^n$ *a set,* $U \subset \Omega$ *a convex set and* $f : \Omega \mapsto \mathbb{R}$ *a function of* C^1*. Let* $u^* \in U$ *be a local minimum of* f *onto* U*. Then*

$$\langle \nabla f(u^*), u - u^* \rangle \geq 0, \forall u \in U \tag{9}$$

Theorem 5 provide us with a necessary condition for minimum existence in case of constrained minimization problems.

Proof

Since U is convex, then $tu + (1-t)u^* = u^* + t(u-u^*) \in U$
Therefore, $f(u^* + t(u-u^*)) \geq f(u^*)$. Consequently,

$$f(u^* + t(u-u^*)) - f(u^*) \geq 0$$

by dividing by t and considering the limit when $t \to 0$, we obtain

$$\lim_{t\to 0} \frac{f(u^* + t(u-u^*)) - f(u^*)}{t} = \langle \nabla f(u^*), u - u^* \rangle \geq 0$$

Orthogonal projection onto a convex set

Consider a non empty convex set $U \in \mathbb{R}^n$. We call orthogonal projection of $v \notin U$ onto U, every point u that is solution of the problem

$$\min_{u \in U} \|u - v\|^2 \tag{10}$$

This solution is denoted $v^* = P_U(v)$.

Theorem 6

Let U be a non empty convex set from $\mathbb{R}^n$. Then the problem 10 has a unique solution.

Example

Consider the set $U = \mathbb{R}^2_+$ and $v = (2, -2)$. The projection of v onto U is the optimal solution of the problem

$$\min_{u \geq 0} \|u - v\|^2 = \min_{u \geq 0} [(u_1 - 2)^2 + (u_2 + 2)^2]$$

Which has a unique solution $P_U(2, -2) = (2, 0)$.

More generally, The projection of $v = (v_1, v_2, \ldots, v_n) \in \mathbb{R}^n$ onto $U = \mathbb{R}^n_+$ is defined by

$$v_i^* = \begin{cases} v_i & \text{if } v_i \geq 0 \\ 0 & \text{if } v_i < 0 \end{cases}$$

The projection of $v = (v_1, v_2, \dots, v_n) \in \mathbb{R}^n$ onto $U = [m_1, p_1] \times [m_2, p_2] \times \ldots \times [m_n, p_n]$ is given by

$$v_i^* = \begin{cases} p_i & \text{if } v_i > p_i \\ v_i & \text{if } m_i \leq v_i \leq p_i \\ m_i & \text{if } v_i < m_i \end{cases}$$

for all $i = 1, 2, \dots, n$.

Theorem 7

Let U be a non empty convex set from $\mathbb{R}^n$. Then $z = P_U(v)$ if and only if $(v-z)^\top(u-z) \leq 0, \quad u \in U$.

Proof

We have, $z = P_U(v)$ if and only if z is the optimal solution of the minimization problem

$$\begin{cases} \min g(u) = \|u - v\|^2 \\ u \in U \end{cases}$$

Since U is convex, we have

$$\forall t \in [0,1], tu + (1-t)z = z + t(u - z) \in U$$

Hence

$$
\begin{aligned}
g(z+t(u-z))-g(z) &\geq 0 \\
\Rightarrow \frac{g(z+t(u-z))-g(z)}{t} &\geq 0 \\
\Rightarrow \lim_{t\to 0}\frac{g(z+t(u-z))-g(z)}{t} &\geq 0 \\
\Rightarrow \nabla_{(u-z)}g(z) &\geq 0 \\
\Rightarrow \langle \nabla g(z), u-z\rangle &\geq 0 \\
\Rightarrow \langle 2(z-v), u-z\rangle &\geq 0 \\
\Rightarrow \langle v-z, u-z\rangle &\leq 0
\end{aligned}
$$

Theorem 8

Let U be a convex set and $f : U \to \mathbb{R}$ a differentiable function, and $\alpha > 0$. Then u^* is a stationary point of the problem

$$(P)\begin{cases} \min f(u) \\ s.t \\ u \in U \end{cases}$$

if and only if

$$u^* = P_U(u^* - \alpha \nabla f(u^*)) \tag{11}$$

Proof

Using Theorem 7, we have $u^* = P_U(u^* - \alpha\nabla f(u^*))$ if and only if
$(u^* - \alpha\nabla f(u^*)) - u^*)^\top (u - u^*) \leq 0, \quad u \in U$.
This means that $\nabla f(u^*)^\top (u - u^*) \geq 0, \quad u \in U$.
According to Theorem 5, u^* is a stationary point of f onto U.

Projected Gradient Method

Let U be a convex set and $f : U \to \mathbb{R}$ a differentiable function, and $\alpha > 0$. We want to solve the problem

$$(P) \begin{cases} \min f(u) \\ s.t \\ u \in U \end{cases}$$

The projected gradient method consists of an iterative process to find $u^* = P_U(u^* - \alpha \nabla f(u^*))$ a stationary point of f onto U.

The process is described as follows:

$$\begin{cases} u^{(0)} \in U \\ u^{(k+1)} = P_U(u^{(k)} - \alpha \nabla f(u^{(k)})) \end{cases}$$

The stopping criteria is $\|u^{(k+1)} - u^{(k)}\| \leq \epsilon$.

Theorem 9

Let U be a convex set and $f : U \to \mathbb{R}$ a differentiable convex lower bounded function. We assume that

$$\exists L > 0, \forall u, v \in U, \|\nabla f(u) - \nabla f(v)\| \leq L\|u - v\|$$

Let $(u^{(k)})$ be the sequence obtained via the process

$$\begin{cases} u^{(0)} \in U \\ u^{(k+1)} = P_U(u^{(k)} - \alpha \nabla f(u^{(k)})) \end{cases}$$

If $0 < \alpha < \dfrac{2}{L}$ and $(u^{(k)})$ converge to u^* then u^* is a global minimum of f in U.

Example

$$(P)\begin{cases} \min f(u) = u_1^2 + 2u_2^2 + u_1 u_2 \\ s.t \\ 1 \leq u_1 \leq 2 \\ 0 \leq u_2 \leq 1 \end{cases}$$