Optimization

Màster de Fonaments de Ciència de Dades

Gerard Gómez

Lecture IV. Constrained optimization. Lagrange multipliers

Equality constrained extrema

Consider the problem of finding the minimum (or maximum) of a real-valued function f with domain $\mathcal{C} \subset \mathbb{R}^n$

$$f : \mathcal{C} \longrightarrow \mathbb{R}$$

subject to the constraints

$$h_i(x) = 0, \quad i = 1, ..., m, \quad m < n \tag{1}$$

where each of the h_i is a real-valued function defined on $\mathcal{C}$. This is, the problem is to find an extremum of f in the region determined by the equations (1).

Example. *Find the area of the largest rectangle that can be inscribed in the ellipse*

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

this is

$$f(x, y) = 4xy \quad \text{and} \quad h(x, y) = \frac{x^2}{a^2} + \frac{y^2}{b^2} - 1 = 0$$

Lagrange's method

Lagrange's method consists of transforming an equality constrained extremum problem into a problem of finding a stationary point of the Lagrangian function

$$L(x,\lambda) = f(x) - \sum_{i=1}^{m} \lambda_i h_i(x), \quad \text{in the example:} \quad L = 4xy - \lambda\left(\frac{x^2}{a^2} + \frac{y^2}{b^2} - 1\right)$$

Theorem (Necessary conditions)
Suppose that

$$f : \mathcal{C} \longrightarrow \mathbb{R}, \quad \textit{and} \quad h_i : \mathcal{C} \longrightarrow \mathbb{R}, \quad i = 1, ..., m$$

are real-valued functions that satisfy:

- *They are all continuosly differentiable on a neighborhood* $N_\epsilon(x^*) \subset \mathcal{C}$
- x^* *is a local minimum (or maximum) of* f *in* $N_\epsilon(x^*)$
- *If* $x \in N_\epsilon(x^*)$, *then*
  $$h_i(x) = 0, \quad i = 1, ..., m$$
- *The Jacobian matrix* $(\partial h_i(x^*)/\partial x_j)$ *has rank* m.

Then, there exists a vector of multipliers $\lambda^* = (\lambda_1^*, ..., \lambda_m^*)^T$ *such that*

$$\nabla L(x^*, \lambda^*) = 0$$

Lagrange's method

There are two ways to interpret the equation

$$\nabla L(x^*, \lambda^*) = 0 \quad \Leftrightarrow \quad \nabla f(x^*) - \sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) = 0$$

1. The cost gradient $\nabla f(x^*)$ belongs to the subspace spanned by the constraint gradients at x^*
2. The cost gradient $\nabla f(x^*)$ is orthogonal to the subspace of **first order feasible variations**

$$V(x^*) = \{\Delta x \mid \nabla h_i(x^*)^T \Delta x = \Delta x^T \nabla h_i(x^*) = 0, \; i = 1, ..., m\}$$

   This is

$$\nabla f(x^*)^T \Delta x = 0 \quad \text{for all} \quad \Delta x \in V(x^*)$$

Remark: $V(x^*)$ is the subspace of variations Δx for which the point $x^* + \Delta x$ satisfies the constraint $h = 0$ up to the first order

$$h(x^* + \Delta x) = h(x^*) + \nabla h(x^*)^T \Delta x = \nabla h(x^*)^T \Delta x = 0$$

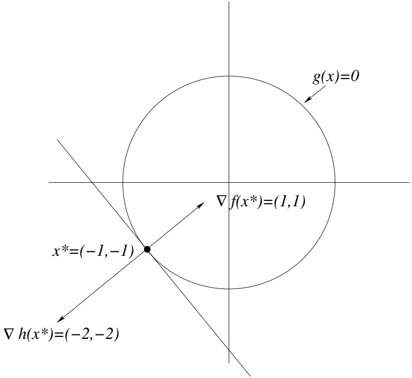
Lagrange necessary conditions. Example

Example.

$$\text{minimize} \quad x_1 + x_2,$$

$$\text{subject to:} \quad h(x_1, x_2) = 2 - x_1^2 - x_2^2 = 0$$

At the local minimum $x^* = (-1, -1)^T$, the cost gradient $\nabla f(x^*) = (1, 1)^T$ is normal to the constraint circle and is, therefore, collinear with the constraint gradient $\nabla h(x^*) = (-2, -2)^T$. The Lagrange multiplier is $\lambda = 1/2$



Feasible variations

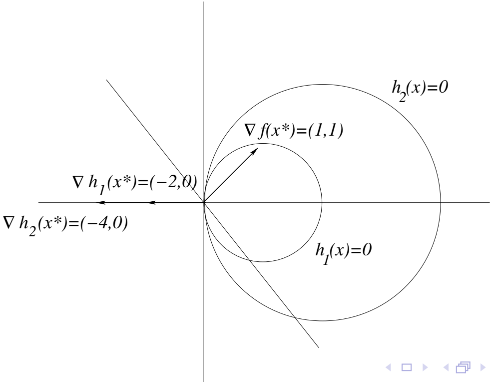
A feasibel variation x for which the gradients $\nabla h_1(x), \ldots, \nabla h_m(x)$ are linearly independent is called **regular**. For a local minimum that is not regular there may not exist Lagrange multipliers. In the next example we will have $m = n$ instead of $m < n$, but this is not relevant for what follows.

Example. Consider the problem of minimizing

$$f(x) = x_1 + x_2$$

subject to

$$h_1(x) = (x_1 - 1)^2 + x_2^2 - 1 = 0, \quad h_2(x) = (x_2 - 2)^2 + x_2^2 - 4 = 0$$



Example (cont.)

It can be seen that at the local minimum $x^* = (0,0)^T$ (the only feasible solution), the cost gradient $\nabla f(x^*) = (1,1)^T$ cannot be expressed as a linear combination of the constraints gradients $\nabla h_1(x^*) = (-2,0)^T$ and $\nabla h_2(x^*) = (-4,0)^T$. Thus, the Lagrange multiplier condition

$$\nabla f(x^*) - \lambda_1^* \nabla h_1(x^*) - \lambda_2^* \nabla h_2(x^*) = 0$$

cannot hold for any λ_1^* and λ_2^*

The difficulty here is that the subspace of first order feasible variations

$$V(x^*) = \{\Delta x \mid \nabla h_1(x^*)^T \Delta x = 0,\ \nabla h_2(x^*)^T \Delta x = 0\}$$

which is $\{\Delta x = (0, x_2)^T\}$, has larger dimension than the true set of feasible variations $\{\Delta x = (0,0)^T\}$

The Penalty approach

In the penalty approach, the original constrained problem is approximated by an unconstrained optimization problem that involves a penalty for violation of the constraints.

For $k = 1, 2, \ldots$, we introduce the cost function

$$F^k(x) = f(x) + \frac{k}{2}\|h(x)\|^2 + \frac{\alpha}{2}\|x - x^*\|^2$$

where x^* is the local minimum of the constrained problem and α some positive scalar.

- The term $(k/2)\|h(x)\|^2$ imposes a penalty for violating the constraint $h(x) = 0$
- The term $(\alpha/2)\|x - x^*\|^2$ is introduced for technical proof-related reasons (to ensure that x^* is a strict local minimum of the function $f(x) + (\alpha/2)\|x - x^*\|^2$ subject to $h(x) = 0$)

It can be shown that if x^k is an optimal solution of

$$\begin{array}{rl} \text{minimize} & F^k(x), \\ \text{subject to:} & x \in S = \{x \mid \|x - x^*\| \leq \epsilon\} \end{array}$$

that exists because S is compact, then the sequence $\{x^k\}$ converges to x^*

Theorem (Sufficient conditions).
Let f, h_1,...,h_m be twice continuously differentiable real-valued functions in $\mathbb{R}^n$. If there exist vectors $x^* \in \mathbb{R}^n$, $\lambda^* \in \mathbb{R}^m$ such that

$$\nabla L(x^*, \lambda^*) = 0,$$

and for every $z \in \mathbb{R}^n$, $z \neq 0$ satisfying

$$z^T \nabla h_i(x^*) = 0, \quad i = 1, ..., m \qquad \textit{(z is a feasible variation)}$$

it follows that

$$z^T \nabla_x^2 L(x^*, \lambda^*) z > 0,$$

then, f has a strict local minimum at x^* subject to $h_i(x) = 0$, $i = 1, ..., m$ (similar for a maximum).

Sufficient conditions

Example. Consider the problem (P)

$$\begin{array}{rl} \text{minimize} & -(x_1x_2 + x_2x_3 + x_1x_3), \\ \text{subject to:} & x_1 + x_2 + x_3 = 3 \end{array}$$

If x_1, x_2 and x_3 represent the length, width and height of a rectangular parallelepiped P, respectively, the problem can be interpreted as maximizing the surface area of P subject to the sum of the edge lengths of P being equal to 3.

$$L = -(x_1x_2 + x_2x_3 + x_1x_3) - \lambda(x_1 + x_2 + x_3 - 3)$$

The first order necessary conditions are

$$\begin{array}{rcl} -x_2^* - x_3^* - \lambda^* & = & 0 \\ -x_1^* - x_3^* - \lambda^* & = & 0 \\ -x_1^* - x_2^* - \lambda^* & = & 0 \\ x_1^* + x_2^* + x_3^* & = & 3 \end{array}$$

which have the unique solution $x_1^* = x_2^* = x_3^* = 1$, $\lambda^* = -2$.

Example (cont.)

The Hessian of the Lagrangian is

$$\nabla^2_{xx} L(x^*, \lambda^*) = \begin{pmatrix} 0 & -1 & -1 \\ -1 & 0 & -1 \\ -1 & -1 & 0 \end{pmatrix}$$

We have for all $y \in V$

$$V = \{y \mid y^T \nabla h(x^*) = 0\} = \{y \mid (y_1, y_2, y_3)(1, 1, 1)^T = 0\} = \{y \mid y_1 + y_2 + y_3 = 0\}$$

with $y \neq 0$ that

$$y^T \nabla^2_{xx} L(x^*, \lambda^*) y = -y_1(y_2 + y_3) - y_2(y_1 + y_3) - y_3(y_1 + y_2) = y_1^2 + y_2^2 + y_3^2 > 0$$

hence, the sufficient conditions for a strict local minimum are satisfied

First-order necessary conditions for inequality constrained extrema

- We begin deriving first-order necessary conditions for inequality and equality constrained extremum problems involving only first derivatives.
- Consider the general problem (P) defined by

$$\begin{array}{rlr} \min & f(x_1, ..., x_n) & \\ \text{subject to:} & g_i(x_1, ..., x_n) \geq 0, & i = 1, ..., p \\ & h_j(x_1, ..., x_n) = 0, & j = 1, ..., m \end{array} \qquad (2)$$

  the functions f, g_i, h_j are assumed to be defined and continuously differentiable on some open set $D \subset \mathbb{R}^n$.

- Let $X \subset D$ denote the **feasible set for problem** (P) this is, the set of all points $x \in D$ satisfying (2). If $x \in X$, we say that x is a **feasible point**
- A point $x^* \in X$ is said to be a **local minimum of problem** (P), or a local solution of (P), if there exist $\delta > 0$ such that

$$f(x) \geq f(x^*), \quad \forall x \in X \cap N_\delta(x^*).$$

- If this condition holds for all $x \in X$

$$f(x) \geq f(x^*), \quad \forall x \in X$$

  then x^* is said to be a **global minimum** of problem (P).

Feasible directions

- Every point $x \in N_\delta(x^*)$ can be written as $x^* + z$, where $z \neq 0$ if and only of $x \neq x^*$.
- A vector $z \neq 0$ is called a **feasible direction** from x^* if there exist $\delta_1 > 0$ such that

$$x^* + \theta z \in X \cap N_{\delta_1}(x^*) \quad \text{for all} \quad 0 \leq \theta < \delta_1 / \|z\|$$

- Feasible directions are important in optimization algorithms. For the moment, we are interested in them for the simple reason that:
  - If x^* is a local minimum of problem (P), and
  - if z is a feasible direction for x^*,
  - then $f(x^* + \theta z) \geq f(x^*)$, if $\theta > 0$ is small enough.

Feasible directions characterization

Characterization of the feasible directions in terms of the constraint functions g_i and h_i.

Define

$$I(x^*) = \{i \mid g_i(x^*) = 0\}.$$

Lemma

If z is a certain feasible direction, we must have

$$z^T \nabla g_i(x^*) \geq 0 \quad \textit{for all} \quad i \in I(x^*)$$

Proof: Assume that for a certain $k \in I(x^*)$ and a certain feasible direction z that:

$$z^T \nabla g_k(x^*) < 0 \quad \text{(the angle is greater than } 90^o\text{)}$$

then, we can write

$$g_k(x^* + \theta z) = g_k(x^*) + \theta z^T \nabla g_k(x^*) + \theta \epsilon_k(\theta) = \theta z^T \nabla g_k(x^*) + \theta \epsilon_k(\theta),$$

with $\theta > 0$, and where $\epsilon_k(\theta)$ tends to zero as $\theta \to 0$.

If θ is small enough, then $z^T \nabla g_k(x^*) + \epsilon_k(\theta) < 0$, so $g_k(x^* + \theta z) < 0$ for all $\theta > 0$ small enough, contradicting the fact that z is a feasible direction vector from x^*.

Similar reasoning can be applied to show that

$$z^T \nabla h_j(x^*) = 0 \quad \text{for} \quad j = 1, ..., m$$

Feasible directions characterization

- Define

$$Z^1(x^*) = \left\{ z \mid z^T \nabla g_i(x^*) \geq 0, i \in I(x^*) \; ; \; z^T \nabla h_j(x^*) = 0, j = 1, ..., m \right\}.$$

According to what it has been said, if z is a feasible direction for x^*, then $z \in Z^1(x^*)$, but it may happen that $z \in Z^1(x^*)$ without being a feasible direction.

- A set $K \subset \mathbb{R}^n$ is called a **cone** if $x \in K \Rightarrow \alpha x \in K$ for all $\alpha \geq 0$.

  The set $Z^1(x^*)$ is clearly a cone, and is also called the **linearizing cone of** X at x^*, since it is generated by linearizing the constraint functions at x^*.

- Define

$$Z^2(x^*) = \left\{ z \mid z^T \nabla f(x^*) < 0 \right\}.$$

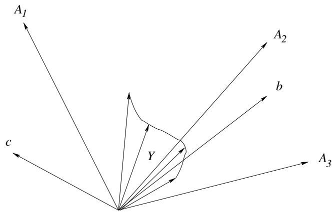
If $z \in Z^2(x^*)$, using Taylor's formula, it can be shown that there exist a point $x = x^* + \theta z$, sufficiently close to x^*, such that $f(x^*) > f(x)$, this is, $Z^2(x^*)$ is formed by the directions along which the function f decreases.

Farkas Lemma

Lemma

Let A be a given $m \times n$ real matrix and $b \in \mathbb{R}^n$ a given vector. The inequality $b^T y \geq 0$ holds for all vectors y satisfying $Ay \geq 0$ if and only if there exists a vector $\rho \in \mathbb{R}^m$, $\rho \geq 0$, such that $A^T \rho = b$.

Interpretation: Let A be a 3×2 matrix and A_1, A_2, $A_3 \in \mathbb{R}^2$ the rows of A.



The set $Y = \{y \mid Ay \geq 0\}$ consists of all the vectors $y \in \mathbb{R}^2$ that make an acute angle with every row of A. The Lemma states that b makes an acute angle with every $y \in Y$ if and only if b can be expressed as a nonnegative linear combibation of the rows of A. In the figure, b satisfies these conditions and c does not.

Necessary conditions "candidates"

As in the case of equality constraints, we define the Lagrangian associated with problem (P) as

$$L(x, \lambda, \mu) = f(x) - \sum_{i=1}^{p} \lambda_i g_i(x) - \sum_{j=1}^{m} \mu_j h_j(x)$$

The following Theorem holds.

Theorem

Assume that $x^0 \in X$, then $Z^1(x^0) \cap Z^2(x^0) = \emptyset$ if and only if there exist vectors λ^0, μ^0 such that

$$\begin{aligned}
\nabla_x L(x^0, \lambda^0, \mu^0) &= \nabla f(x^0) - \sum_{i=1}^{p} \lambda_i^0 \nabla g_i(x^0) - \sum_{j=1}^{m} \mu_j^0 \nabla h_j(x^0) = 0, && (3)\\
\lambda_i^0 g_i(x^0) &= 0, \quad i = 1, \ldots, p && (4)\\
\lambda_i^0 &\geq 0. && (5)
\end{aligned}$$

(Lagrange conditions)

The condition $Z^1(x^0) \cap Z^2(x^0) = \emptyset$ implies that there are no feasible directions along which f decreases.

Necessary conditions "candidates"

Proof: The $Z^1(x^0)$ is never empty, since $0 \in Z^1(x^0)$.
The condition $Z^1(x^0) \cap Z^2(x^0) = \emptyset$ holds if and only if for every z satisfying

$$\begin{aligned} z^T \nabla g_i(x^0) &\geq 0, \quad i \in I(x^0), &(6)\\ z^T \nabla h_j(x^0) &= 0, \quad j = 1, ..., m, &(7) \end{aligned}$$

we have

$$z^T \nabla f(x^0) \geq 0. \tag{8}$$

We can write (7) as

$$\begin{aligned} z^T \nabla h_j(x^0) &\geq 0, \quad j = 1, ..., m &(9)\\ z^T [-\nabla h_j(x^0)] &\geq 0, \quad j = 1, ..., m &(10) \end{aligned}$$

From Farkas Lemma, it follows that (8) holds for all vectors z satisfying (6), (9) and (10) if and only if there exist vectors $\lambda^0 \geq 0$, $\mu^1 \geq 0$, $\mu^2 \geq 0$ such that

$$\nabla f(x^0) = \sum_{i \in I(x^0)} \lambda_i^0 \nabla g_i(x^0) + \sum_{j=1}^{m} (\mu_j^1 - \mu_j^2) \nabla h_j(x^0).$$

Letting $\lambda_i^0 = 0$ for $i \notin I(x^0)$, $\mu_j^0 = \mu_j^1 - \mu_j^2$, we conclude that
$Z^1(x^0) \cap Z^2(x^0) = \emptyset$ if and only if (3), (4) and (5) hold. $\square$

Some remarks

- The Lagrange conditions of the above Theorem are the natural candidates to become the necessary conditions for x^0 to be the solution x^* of problem (P) if we can guarantee that $Z^1(x^*) \cap Z^2(x^*) = \emptyset$, when x^* is a solution of (P). This condition ensures that f can not decrease along any feasible direction.

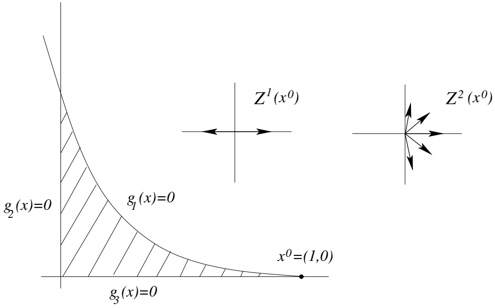
- For most problems $Z^1(x^*) \cap Z^2(x^*) = \emptyset$ and the Lagrange conditions (3), (4) and (5) hold at x^*; however, this is not always the case as the following example shows.

Example

Example: Consider the following constraints in $\mathbb{R}^2$:

$$\begin{aligned} g_1(x) &= (1-x_1)^3 - x_2 \geq 0, \\ g_2(x) &= x_1 \geq 0, \\ g_3(x) &= x_2 \geq 0, \end{aligned}$$

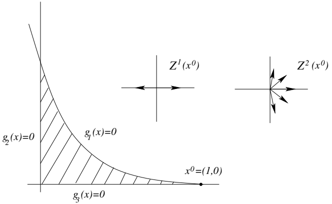
that define the feasible set X.



Example

The point $x^0 = (1,0)$ is feasible, and we can easily verify that $I(x^0) = \{1,3\}$ and that $\nabla g_1(x^0) = (0,-1)^T$, $\nabla g_3(x^0)) = (0,1)^T$, so

$$Z^1(x^0) = \{(z_1, z_2) \mid z_2 = 0\}.$$



Letting $f(x) = -x_1$, we can see that x^0 is a solution of the problem

$$\min_x -x_1,$$

subject to the above constraints
At this point

$$Z^2(x^0) = \{(z_1, z_2) \mid z_1 > 0\},$$

an $Z^1(x^*) \cap Z^2(x^*)$ is nonempty, hence there exist no λ^0 satisfying conditions (3), (4) and (5).

Weak necessary optimality conditions

It is possible te derive weak necessary conditions for optimality without requiring the set $Z^1(x^*) \cap Z^2(x^*)$ to be empty at the solution.

Let the **weak Lagrangian** $\tilde{L}$ be defined by

$$\tilde{L}(x, \lambda, \mu) = \lambda_0 f(x) - \sum_{i=1}^{p} \lambda_i g_i(x) - \sum_{j=1}^{m} \mu_j h_j(x),$$

where λ_0 is an additional parameter.

To proof necessary conditions for equality and inequality constrained problems we need the following result, called a **"theorem of the alternative"**

Theorem

Let A be an $m \times n$ real matrix. Then either there exists an $x \in \mathbb{R}^n$ such that

$$Ax < 0,$$

or there exists a nonzero vector $u \in \mathbb{R}^m$, $u \neq 0$ such that

$$u^T A = 0, \quad u \geq 0$$

but never both

Theorem

Proof: Assume that there exist x and u such that both

$$Ax < 0, \quad \text{and} \quad u^T A = 0, \quad u \geq 0$$

are satisfied. Then we have $u^T Ax < 0, \quad$ and $\quad u^T Ax = 0$ simultaneously, a contradiction.

Assume now that there exist no x satisfying the first condition ($Ax < 0$). This means that we cannot find a negative number $w < 0$ satisfying

$$(Ax)_i = A_i x = \sum_{j=1}^{n} a_{ij} x_j \leq w, \quad i = 1, ..., m$$

for every $x \in \mathbb{R}^n$ where A_i is the ith-row of A. This is, if

$$A_i x \leq w \Leftrightarrow w - A_i x \geq 0, \; i = 1, ..., m \quad \forall x \in \mathbb{R}^n, \text{ then } w \geq 0.$$

Take $y = (w, x)^T$, $b = (1, 0, ..., 0)^T \in \mathbb{R}^{n+1}$, $e = (1, ..., 1)^T \in \mathbb{R}^m$, and $\tilde{A} = (e \,|\, -A)$.

Using this notattion, what we have stablished is that: if for any $y = (w, x)^T$ the following inequality is fulfilled

$$w - A_i x = (\tilde{A} y)_i \geq 0, \quad i = 1, ..., m, \quad \Leftrightarrow \quad \tilde{A} y \geq 0,$$

then

$$w = b^T y \geq 0.$$

Proof (cont.)

According to Farkas lemma, there exists an m vector $u \geq 0$, such that

$$\tilde{A}^T u = \begin{pmatrix} 1 & \dots & 1 \\ & -A^T & \end{pmatrix} \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix} = b = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

so

$$\sum_{i=1}^{m} u_i = 1, \quad \sum_{i=1}^{m} u_i a_{ij} = 0, \; j = 1, ..., n$$

hence, we have found u that satisfies the second condition of the theorem.

$\square$

Weak necessary optimality conditions

We consider problem (P) when there are no equality constraints $h_i(x) = 0$, $i = 1, ..., m$, this is:

$$\min f(x), \quad \text{subject to} \quad g_i(x) \geq 0, \quad i = 1, ..., p,$$

Remark: The equality constraints become inequality constraints according to:

$$\begin{aligned} h_j(x) &= g_{p+j}(x) \geq 0, \quad j = 1, ..., m \\ -h_j(x) &= g_{p+m+j}(x) \geq 0, \quad j = 1, ..., m. \end{aligned}$$

Theorem

Let $f, g_1, ..., g_m$ be real continuously differentiable functions on an open set containing X. If x^* is a solution of problem (P), then there exist $\lambda^* = (\lambda_0^*, \lambda_1^*, ..., \lambda_p^*)^T$ such that

$$\begin{aligned} \nabla_x \tilde{L}(x^*, \lambda^*) &= \lambda_0^* \nabla f(x^*) - \sum_{i=1}^{p} \lambda_i^* \nabla g_i(x^*) = 0, && (11) \\ \lambda_i^* g_i(x^*) &= 0, \quad i = 1, ..., p && (12) \\ \lambda^* &\neq 0, \quad \lambda^* \geq 0 && (13) \end{aligned}$$

Theorem

Proof: We must proof that the necessary conditions for x^* to be the solution of problem (P), are the existence of a vector λ^* satisfying (16), (16) and (16).

If $g_i(x^*) > 0$ for all i (the point x^* is in the interior of the feasible set X), then $I(x^*) = \{i \mid g_i(x^*) = 0\} = \emptyset$. Choose $\lambda_0^* = 1$, $\lambda_1^* = \lambda_2^* = ... = \lambda_p^* = 0$ and the conditions (16), (16) and (16) hold since $\nabla f(x^*) = 0$.

Suppose now that $I(x^*) \neq \emptyset$. Then, for every z satisfying

$$z^T \nabla g_i(x^*) > 0, \quad i \in I(x^*) \tag{14}$$

we cannot have

$$z^T \nabla f(x^*) < 0. \tag{15}$$

This follows from the following: according to Taylor formula, we can see that if there exists z satisfying (14), then we can find a sufficiently small δ such that $x = x^* + \theta z$ satisfies

$$g_i(x) = g_i(x^*) + \theta z^T \nabla g_i(x^*) + O_2,$$

and, since $g_i(x^*) = 0$ we get

$$g_i(x) > 0, \quad \text{if} \quad i \in I(x^*),$$

for all $0 < \theta < \delta$, that is, x is a feasible point.
If (15) also holds, then

$$f(x) = f(x^*) + \theta z^T \nabla f(x^*) + O_2 < f(x^*),$$

contradicting that x^* is a minimum.

Proof (cont.)

Thus, the system of inequalities (14) and (15), that can also be written as

$$\begin{aligned} z^T[-\nabla g_i(x^*)] &< 0, \quad i \in I(x^*), \\ z^T \nabla f(x^*) &< 0, \end{aligned}$$

has no solution. According to the Theorem of the Alternative, and taking as matrix A one with rows equal to $\nabla f(x^*)$ and $-\nabla g_i(x^*)$, we get that there exists a nonzero vector $\lambda^* \geq 0$, such that

$$(\lambda^*)^T A = A^T \lambda^* = \lambda_0^* \nabla f(x^*) + \sum_{i \in I(x^*)} \lambda_i^* [-\nabla g_i(x^*)] = 0.$$

Letting $\lambda_i^* = 0$ for $i \notin I(x^*)$, we can write this equation as

$$\lambda_0^* \nabla f(x^*) - \sum_{i=1}^{p} \lambda_i^* \nabla g_i(x^*) = 0,$$

and clearly

$$\lambda_i^* g_i(x^*) = 0, \quad i = 1, ..., p.$$

$\square$

Weak necessary optimality conditions

If we dont want to transform the equality constraints into inequalities, the following theorem also holds.

Theorem

Let f, $h_1, \ldots, h_m$ and $g_1, \ldots, g_p$ be real continuously differentiable functions on an open set containing X. If x^* is a solution of problem (P), then there exist $\lambda^* = (\lambda_0^*, \lambda_1^*, \ldots, \lambda_p^*)^T$ and $\mu^* = (\mu_1^*, \ldots, \mu_m^*)^T$ such that

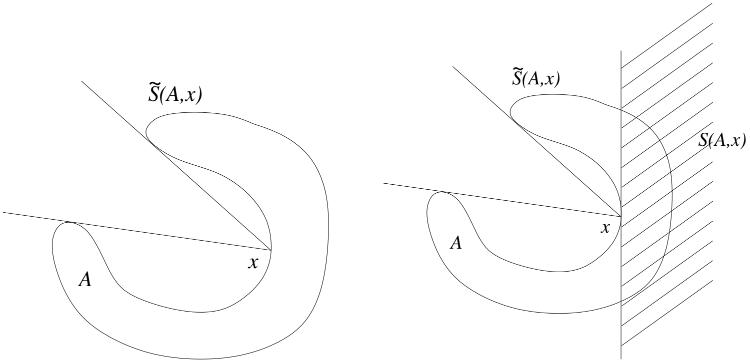
$$
\begin{aligned}
\nabla_x \tilde{L}(x^*, \lambda^*) &= \lambda_0^* \nabla f(x^*) - \sum_{i=1}^{p} \lambda_i^* \nabla g_i(x^*) - \sum_{j=1}^{m} \mu_j^* \nabla h_j(x^*) = 0, \\
\lambda_i^* g_i(x^*) &= 0, \quad i = 1, \ldots, p \\
(\lambda^*, \mu^*) &\neq 0, \quad \lambda^* \geq 0
\end{aligned}
$$

The closed cone of tangents

Let $x \in A \subset \mathbb{R}^n$, where A is a nonempty set.

Denote by $\tilde{S}(A,x)$ the intersection of all closed cones containing the set $\{a - x \mid a \in A\}$, this is

$$\tilde{S}(A,x) = \{\alpha(a - x) \mid \alpha \geq 0, a \in A\}.$$

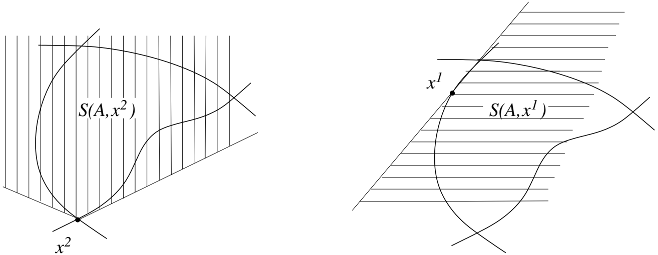


The closed cone of tangents

The *closed cone of tangents* of the set A at x, $S(A,x)$, is defined as

$$S(A,x) = \cap_{k=1}^{\infty} \tilde{S}(A \cap N_{1/k}(x), x),$$

where $N_{1/k}(x)$ is a spherical neighborhood of x with radius $1/k$, $k \in \mathbb{N}$.



The following lemma characterizes $S(A,x)$.

Lemma

A vector z is contained in $S(A,x)$ if and only if there exists a sequence of vectors $\{x^k\} \subset A$ converging to x and a sequence of nonnegative numbers $\{\alpha^k\}$ such that the sequence $\{\alpha^k(x^k - x)\}$ converges to z.

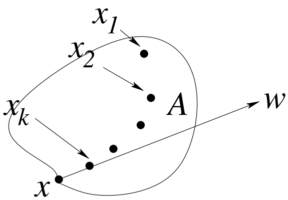
The closed cone of tangents

With the aid of this lemma, it is possible to give alternative descriptions of $S(A, x)$.

- First observe that the vector $w = 0$ is always in $S(A, x)$ for every A and x.
- Let w be a unit vector, and suppose that there exists a sequence of points $\{x^k\} \subset A$ such that: $x^k \to x$, $x^k \neq x$ and

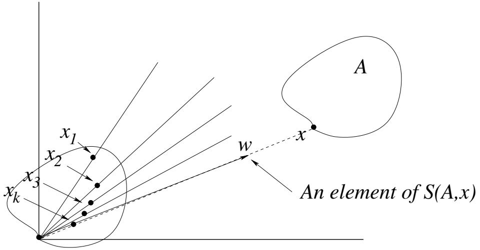
$$\lim_{k\to\infty} \frac{x^k - x}{\|x^k - x\|} = w.$$

  This is, a sequence of vectors $\{x^k\}$ converging to x in the direction of w.
- The cone of tangents of the set A at x contains all the vectors that are nonnegative multiples of the w obtained by this method.



The closed cone of tangents (second description)

- Translate the set A by substracting x from each of its elements
- Let $\{x^k\}$ be a sequence of the translated set, $x^k \neq 0$, converging to the origin.
- Construct a sequence of half-lines from the origin and passing through x^k.
- These half-lines tend to a half-line that will be a member of $S(A, x)$.
- The union of all the half-lines formed by taking all such sequences will then be the cone of tangents of A at x.



The closed cone of tangents. Example

Example: Let

$$A = \{(x_1, x_2) \mid (x_1 - 4)^2 + (x_2 - 2)^2 \le 1\}.$$

Let us find the cone of tangents of A at the boundary point $x = (4 - \sqrt{3}/2, 3/2)$.

First we translate A to the origin, obtaining the ball

$$A^1 = \{(x_1, x_2) \mid (x_1 - \sqrt{3}/2)^2 + (x_2 - 1/2)^2 \le 1\}.$$

Taking sequences of points $\{x^k\}$ on the boundary of A^1 converging to the origin we generate sequences of half-lines converging to a line that is actually the tangent line to the curve at the origin. This line satisfies

$$\sqrt{3}x_1 + x_2 = 0.$$

Repeating this process for all sequences in the interior of A^1 converging to the origin, we get the cone of tangents of A^1 at 0 as

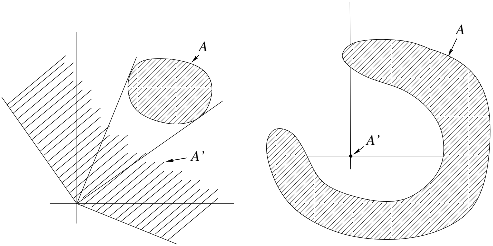
$$S(A^1, x) = \{(x_1, x_2) \mid \sqrt{3}x_1 + x_2 \ge 0\}.$$

Positively normal cones

The next notion is the **positively normal cone** to a set $A \subset \mathbb{R}^n$ and that will be denoted by A'. It is defined by

$$A' = \{x \in \mathbb{R}^n \mid x^T y \geq 0,\ \forall y \in A\},$$

This is, it consists of all vectors $x \in \mathbb{R}^n$ that make an angle less or equal to 90^o with all $y \in A$.



An important property of normal cones is the following: given two sets $A_1 \subset \mathbb{R}^n$, $A_2 \subset \mathbb{R}^n$, then

$$A_1 \subset A_2 \quad \Longrightarrow \quad A_2' \subset A_1'.$$

Cones of tangents and positively normal cones

Cones of tangents and positively normal cones play a central role in stablishing strong optimality conditions.

Lemma

Let $x^0 \in X$. The set $Z^1(x^0) \cap Z^2(x^0)$ is empty if and only if

$$\nabla f(x^0) \in (Z^1(x^0))'.$$

Proof: The set $Z^1(x^0) \cap Z^2(x^0)$ is empty if and only if for all $z \in Z^1(x^0)$ we have $z^T \nabla f(x^0) \geq 0$. Then, $\nabla f(x^0)$ is contained in the positively normal cone of $Z^1(x^0)$ that is $(Z^1(x^0))'$. $\square$

Lemma

Assume that x^0 is a solution of problem (P). Then

$$\nabla f(x^0) \in (S(X, x^0))'.$$

Remark: $(S(X, x^0))'$ is the positively normal cone of the closed tangent cone of the feasible set X at the point x^0.

Proof: We must show that $z^T \nabla f(x^0) \geq 0$ for every $z \in S(X, x^0)$. Let $z \in S(X, x^0)$. According to the previous characterization lemma of the tangent cone, there exists a sequence $\{x^k\} \in X$ converging to x^0 and a sequence of nonnegative numbers $\{\alpha^k\}$ such that $\{\alpha^k(x^k - x^0)\}$ converges to z. If f is differentiable at x^0, we can write

$$f(x^k) = f(x^0) + (x^k - x^0)^T \nabla f(x^0) + \epsilon \|x^k - x^0\|,$$

where ϵ tends to zero as $k \to \infty$. Hence

$$\alpha^k(f(x^k) - f(x^0)) = (\alpha^k(x^k - x^0))^T \nabla f(x^0) + \epsilon \|\alpha^k(x^k - x^0)\|.$$

Cones of tangents and positively normal cones (cont.)

$$\alpha^k(f(x^k) - f(x^0)) = (\alpha^k(x^k - x^0))^T \nabla f(x^0) + \epsilon \|\alpha^k(x^k - x^0)\|.$$

Since $x^k \in X$ and x^0 is a local minimum, it follows that, by letting $k \to \infty$, the term $\epsilon\|\alpha^k(x^k - x^0)\| \to 0$, and $\alpha^k(f(x^k) - f(x^0))$ converges to a nonnegative limit z. Thus

$$\lim_{k\to\infty} (\alpha^k(x^k - x^0))^T \nabla f(x^0) = z^T \nabla f(x^0) \geq 0,$$

That is

$$\nabla f(x^0) \in (S(X, x^0))'$$

$\square$

The Kuhn-Tucker necessary optimality conditions

The (generalized) Kuhn-Tucker necessary conditions for optimality are given by the following theorem.

Theorem

Let x^* be a solution of problem (P) and suppose that

$$(Z^1(x^*))' = (S(X, x^*))'. \tag{16}$$

Then, there exist $\lambda^* = (\lambda_1^*, ..., \lambda_p^*)^T$ and $\mu^* = (\mu_1^*, ..., \mu_m^*)^T$ such that

$$\begin{aligned} \nabla f(x^*) - \sum_{i=1}^{p} \lambda_i^* \nabla g_i(x^*) - \sum_{j=1}^{m} \mu_j^* \nabla h_j(x^*) &= 0, && (17) \\ \lambda_i^* g_i(x^*) &= 0, \quad i = 1, ..., p, && (18) \\ \lambda^* &\geq 0. && (19) \end{aligned}$$

(Kuhn–Tucker conditions).

Proof: Suppose that x^* is a solution of (P). According to a previous Lemma, $\nabla f(x^*) \in (S(X, x^*))'$. If $(Z^1(x^*))' = (S(X, x^*))'$, then $\nabla f(x^*) \in (Z^1(x^*))'$. We have already seen that then $Z^1(x^*) \cap Z^2(x^*)$ is empty and, according to the necessary optimality conditions theorem already seen, conditions (17), (18) and (19) hold $\square$

The Kuhn-Tucker necessary optimality conditions

Essentially, what the above theorem says is that the condition

$$(Z^1(x^*))' = (S(X, x^*))'$$

is a sufficient condition for the existence of the multipliers λ^* and μ^* satisfying conditions (17), (18) and (19).

Notice that if

$$Z^1(x^*) = S(X, x^*),$$

at a solution point x^* of problem (P), then the hypotheses of the last theorem are fulfilled.

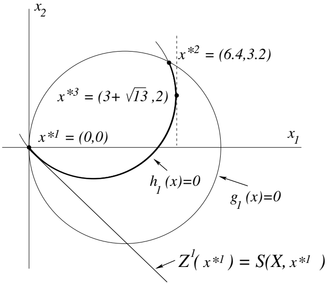
The Kuhn-Tucker necessary optimality conditions

Example: Consider the following problem

$$\min f(x) = x_1,$$

subject to

$$g_1(x) = 16 - (x_1 - 4)^2 - x_2^2 \geq 0, \quad h_1(x) = (x_1 - 3)^2 + (x_2 - 2)^2 - 13 = 0.$$



From the figure it follows that f has local minima at $x^{*1} = (0,0)$ and $x^{*2} = (32/5, 16/5)$. At both points, $I(x^{*1}) = I(x^{*2}) = \{1\}$. At the first point $\nabla g_1(x^{*1}) = (8,0)^T$, $\nabla h_1(x^{*1}) = (-6,-4)^T$, so

$$\begin{aligned} Z^1(x^{*1}) &= \{z \mid z^T \nabla g_1(x^{*1}) \geq 0,\ z^T \nabla h_1(x^{*1}) = 0\} \\ &= \{(z_1, z_2) \mid z_1 \geq 0, z_2 = -(3/2)z_1\}, \end{aligned}$$

The Kuhn-Tucker necessary optimality conditions (cont.)

It can be verified that the set $Z^1(x^{*1})$ is also $S(X, x^{*1})$. Now

$$Z^2(x^{*1}) = \{z \mid z^T \nabla f(x^{*1}) < 0\} = \{(z_1, z_2) \mid z_1 < 0\},$$

hence $Z^1(x^{*1}) \cap Z^2(x^{*1}) = \emptyset$. The Kuhn–Tucker conditions (17), (18) and (19) are satisfied for $\lambda_1^* = 1/8$ and $\mu_1^* = 0$. At the second point

$$Z^1(x^{*2}) = \{(z_1, z_2) \mid z_1 \geq 0, z_2 = -(17/6)z_1\},$$

$$Z^2(x^{*2}) = \{(z_1, z_2) \mid z_1 < 0\},$$

and again $Z^1(x^{*2}) \cap Z^2(x^{*2}) = \emptyset$. At this point $\lambda_1^* = 3/40$ i $\mu_1^* = 1/5$.

It turns out that at $x^{*3} = (3 + \sqrt{13}, 2)$ the Kuhn–Tucker necessary conditions also hold. At this point $Z^1(x^{*3}) \cap Z^2(x^{*3}) = \emptyset$ and the corresponding multipliers are $\lambda_1^* = 0$ and $\mu_1^* = \sqrt{13}/26$. From the Figure is clear that x^{*3} is not a solution of our problem but is a solution of

$$\max f(x) = x_1,$$

with the same constraints.

Second-order optimality conditions

Let us see optimality conditions for problem (P) that involve second derivatives.

We begin with the second-order necessary conditions that complement the above Kuhn–Tucker conditions; later we will give the sufficient contions for optimality.

In what follows all the functions f, $g_1, \ldots, g_p$, $h_1, \ldots, h_m$ will be twice continuously differentiable.

Let $x \in X$, we define the following modification of the set $Z^1(x)$:

$$\hat{Z}^1(x) = \{z \mid z^T \nabla g_i(x) = 0,\ i \in I(x),\ z^T \nabla h_j(x) = 0,\ j = 1, \ldots, m\}.$$

Second-order optimality conditions

Definition: The second-order constraint qualification is said to hold at $x^0 \in X$ if for each $z \in \hat{Z}^1(x^0)$ there is a twice differentiable function $\alpha : [0, \epsilon] \subset \mathbb{R} \longrightarrow \mathbb{R}^n$ such that

$$\begin{aligned} \alpha(0) &= x^0, & \\ g_i(\alpha(t)) &= 0, & i \in I(x^0), \\ h_j(\alpha(t)) &= 0, & j = 1, ..., m, \end{aligned}$$

for $0 \leq t \leq \epsilon$ $(\alpha(t) \in X)$ and

$$\frac{d\alpha(0)}{dt} = \lambda z,$$

for some positive $\lambda > 0$.

The above conditions mean that every $z \in \hat{Z}^1(x^0)$, $z \neq 0$, is tangent to a twice differentiable arc α contained in the boundary of X.

It can be shown that if $\nabla g_i(x)$, $i \in I(x)$, $\nabla h_j(x)$, $j = 1, ..., p$ are linearly independent, then the second-order constraint qualification hold at $x \in X$. restricció de segon ordre a $x \in X$.

Second-order optimality conditions theorem

Theorem

Let x^* be feasible for problem (P) that holds the second-order constraint qualification.

- *If there exist $\lambda^* = (\lambda_1^*, ..., \lambda_p^*)$ and $\mu^* = (\mu_1^*, ..., \mu_m^*)$ satisfying the Kuhn–Tucker conditions (17), (18) and (19), and*
- *if for every $z \neq 0$ such that $z \in \hat{Z}^1(x^*)$, it follows that*

$$z^T \left[\nabla^2 f(x^*) - \sum_{i=1}^{p} \lambda_i^* \nabla^2 g_i(x^*) - \sum_{j=1}^{m} \mu_j^* \nabla^2 h_j(x^*) \right] z > 0.$$

then x^* is a strict local minimum of problem (P).

Sufficient optimality conditions

Denote by $\bar{I}(x^*)$ the set of indices i for which $g_i(x^*) = 0$ and the Kuhn–Tucker conditions (17), (18) and (19) are satisfied by $\lambda_i^* > 0$.

Clearly $\bar{I}(x^*) \subset I(x^*)$. Let

$$\begin{aligned} \overline{Z}^1(x^*) \quad = \quad \{z \mid\ & z^T \nabla g_i(x^*) = 0,\ i \in \bar{I}(x^*), \\ & z^T \nabla g_i(x^*) \geq 0,\ i \in I(x^*), \\ & z^T \nabla h_j(x^*) = 0,\ j = 1, \ldots, m\}. \end{aligned}$$

Note that $\overline{Z}^1(x^*) \subset Z^1(x^*)$.

The following theorem gives sufficient optimality conditions

Sufficient optimality conditions

Theorem

Let x^* be a feasible point for problem (P). If there exist $\lambda^* = (\lambda_1^*, \ldots, \lambda_p^*)$, $\mu^* = (\mu_1^*, \ldots, \mu_m^*)$ satisfying

$$\nabla_x L(x^*, \lambda^*, \mu^*) = \nabla f(x^*) - \sum_{i=1}^{p} \lambda_i^* \nabla g_i(x^*) - \sum_{j=1}^{m} \mu_j^* \nabla h_j(x^*) = 0 \quad (20)$$

$$\lambda_i^* g_i(x^*) = 0, \quad i = 1, \ldots, p \quad (21)$$

$$\lambda^* \geq 0 \quad (22)$$

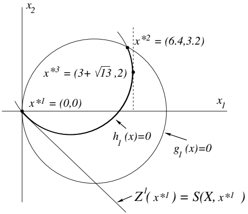
and for every $z \neq 0$, such that $z \in \overline{Z}^1(x^*)$ it follows that

$$z^T \left[\nabla^2 f(x^*) - \sum_{i=1}^{p} \lambda_i^* \nabla^2 g_i(x^*) - \sum_{j=1}^{m} \mu_j^* \nabla^2 h_j(x^*) \right] z = z^T \nabla_x^2 L(x^*, \lambda^*, \mu^*) z > 0, \quad (23)$$

then, x^* is a strict local minimum of problem (P).

The Kuhn-Tucker necessary optimality conditions

Example: Consider again the problem min $f(x) = x_1$ of the figure



We have seen that there are (at least) three points satisfying the necessary conditions for optimality. Let us check the sufficient conditions.
At x^{*1} we have that

$$\overline{Z}^1(x^{*1}) = \{0\},$$

and there are no vectors $z \neq 0$ such that $z \in \overline{Z}^1(x^{*1})$, so the sufficient conditions of the theorem are trivially satisfied. It can be seen that these conditions also hold at x^{*2}. At x^{*3}, however

$$\overline{Z}^1(x^{*3}) = \{(z_1, z_2) \mid z_1 = 0\},$$

an the quadratic form that appears in the Theorem is $-(\sqrt{13}/13)z^T z$, which is negative for all $z \neq 0$. Thus x^{*3} does not satisfy the sufficient conditions.

?

Exercises

Exercise 5. Solve the two-dimensional problem

$$\begin{aligned} \text{minimize} \quad & (x-a)^2 + (x-b)^2 + xy \\ \text{subject to:} \quad & 0 \le x \le 1, \quad 0 \le y \le 1 \end{aligned}$$

for all possible values of the scalars a and b.

Exercise 6. Given a vector y, consider the problem

$$\begin{aligned} \text{maximize} \quad & y^T x \\ \text{subject to:} \quad & x^T Q x \le 1 \end{aligned}$$

where Q is a positive definite symmetric matrix. Show that the optimal value is $\sqrt{y^T Q^{-1} y}$ and use this fact to establish the inequality

$$(x^T y)^2 \le (x^T Q\, x)(y^T Q^{-1} y)$$