Optimization

Màster de Fonaments de Ciència de Dades

Lecture V. Constrained optimization

Gerard Gómez

Constrained optimization. Main results

Constrained optimization. Main results

Assume given a set $\mathcal{C} \subset \mathbb{R}^n$, and the real-valued functions

$$\begin{aligned} f : \mathcal{C} &\longrightarrow \mathbb{R}, \\ g_i : \mathcal{C} &\longrightarrow \mathbb{R}, \quad i = 1, \dots p \\ h_j : \mathcal{C} &\longrightarrow \mathbb{R}, \quad j = 1, \dots, m \end{aligned}$$

Constrained optimization. Main results

The **general constrained optimization problem** is defined by

$$\begin{array}{rll} \min & f(\boldsymbol{x}) & \\ \text{subject to:} & g_i(\boldsymbol{x}) \geq 0, & i = 1, \dots, p \\ & h_j(\boldsymbol{x}) = 0, & j = 1, \dots, m \ \text{ with } m < n \end{array}$$

The **Lagrangian** associated with the problem is defined as

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

The **feasible set** X is defined as the set of point fulfilling the constraints

$$X = \{\boldsymbol{x} \in \mathcal{C} \mid g_i(\boldsymbol{x}) \geq 0, i = 1, \dots, p \text{ and } h_j(\boldsymbol{x}) = 0, j = 1, \dots, m\}$$

Constrained optimization. Main results

The **equality constrained optimization problem** is defined by

$$\begin{array}{rl} \min & f(\boldsymbol{x}) \\ \text{subject to} & h_j(\boldsymbol{x}) = 0, \qquad j = 1, \dots, m \ \text{ with } m < n \end{array}$$

The **Lagrangian** associated with the problem is defined as

$$L(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) - \sum_{j=1}^{m} \lambda_i h_i(\boldsymbol{x})$$

The **feasible set** X is defined as the set of point fulfilling the constraints

$$X = \{\boldsymbol{x} \in \mathcal{C} \mid h_j(\boldsymbol{x}) = 0, j = 1, \dots, m\}$$

Constrained optimization. Main results

Goal:

Stablish the necessary and sufficient conditions to characterize the local extrema (maximum or minimum) of f

Constrained optimization. Main results

Theorem (Necessary conditions for the equality constrained problem)

Let f, $h_1, \ldots, h_m$ be real continuously differentiable functions on an open set C containing X

If:

1. $\boldsymbol{x}^* \in X \subset \mathbb{R}^n$ *is a solution of the equality constrained problem*
2. *at* $\boldsymbol{x} = \boldsymbol{x}^*$, *the Jacobian matrix*

$$\begin{pmatrix} \frac{\partial h_1}{\partial x_1} & \cdots\cdots & \frac{\partial h_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_m}{\partial x_1} & \cdots\cdots & \frac{\partial h_m}{\partial x_n} \end{pmatrix}$$

has rank m, this is: the constraint gradients $\nabla h_1(\boldsymbol{x}^*),\ldots,\nabla h_m(\boldsymbol{x}^*)$ *are linearly independent*

Then:
there exists a vector of multipliers $\boldsymbol{\lambda}^* = (\lambda_1^*, \ldots, \lambda_m^*)^T$ *such that*

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

(($\boldsymbol{x}^*, \boldsymbol{\lambda}^*$) is a stationary vector of the Lagrangian $L(\boldsymbol{x}, \boldsymbol{\lambda})$)

Constrained optimization. Main results

Theorem (Necessary conditions for the general constrained problem)

Let f, $h_1, \dots, h_m$ and $g_1, \dots, g_p$ be real continuously differentiable functions on an open set C containing the feasible set X

If:

1. $\boldsymbol{x}^* \in X \subset \mathbb{R}^n$ *is a solution of the constrained problem*
2. $$(Z^1(\boldsymbol{x}^*))' = (S(X, \boldsymbol{x}^*))'$$

Then:
there exist $\boldsymbol{\lambda}^* = (\lambda_1^*, \dots, \lambda_p^*)^T$ *and* $\boldsymbol{\mu}^* = (\mu_1^*, \dots, \mu_m^*)^T$ *such that*

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

(Karush–Kuhn–Tucker conditions)

Constrained optimization. Main results

Theorem (Sufficient conditions for the equality constrained problem)

Let f, h_1,...,h_m be twice continuously differentiable real-valued functions in $\mathbb{R}^n$

If:

- *there exist* $\boldsymbol{x}^* \in X \subset \mathbb{R}^n$, $\boldsymbol{\lambda}^* \in \mathbb{R}^m$ *such that*
  1. *The vector* $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ *is a stationary point of the Lagrangian*
  $$\nabla L(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = 0$$
  2. *For every* $\boldsymbol{z} \in \mathbb{R}^n$, $\boldsymbol{z} \neq 0$ *satisfying*
  $$(\nabla h_i(\boldsymbol{x}^*))^T \boldsymbol{z} = \boldsymbol{z}^T \nabla h_i(\boldsymbol{x}^*) = 0, \quad i = 1, ..., m$$
  it follows that
  $$\boldsymbol{z}^T \nabla^2_{\boldsymbol{xx}} L(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) \boldsymbol{z} > 0$$

Then:

$\boldsymbol{x}^*$ *is a strict local minimum of the equality constrained optimization problem*

Constrained optimization. Main results

Theorem (Sufficient conditionsfor the general constrained problem)

Let f, $g_1,\dots,g_p$, $h_1,\dots,h_m$ be twice continuously differentiable real-valued functions in $\mathbb{R}^n$, and $\mathbf{x}^*$ be a feasible point of the general constrained optimization problem

If there exist $\boldsymbol{x}^* \in X \subset \mathbb{R}^n$, $\boldsymbol{\lambda}^* \in \mathbb{R}^p$, $\boldsymbol{\mu}^* \in \mathbb{R}^m$ such that

1. They satisfy the Karush–Kuhn–Tucker conditions:

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

2. For every $\boldsymbol{z} \neq \mathbf{0}$, such that $\boldsymbol{z} \in \overline{Z}^1(\boldsymbol{x}^*)$ it follows that

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

Then, $\boldsymbol{x}^*$ is a strict local minimum of the general constrained optimization problem

Exercises

Exercise 7. To be delivered before 9-XI-2021 as: `Ex07-YourSurname.pdf`

Solve the two-dimensional problem

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

for all possible values of the scalars a and b

Exercise 8. To be delivered before 9-XI-2021 as: `Ex08-YourSurname.pdf`

Given a vector $\boldsymbol{y}$, consider the problem

$$\begin{array}{ll} \text{maximize} & \boldsymbol{y}^T \boldsymbol{x} \\ \text{subject to:} & \boldsymbol{x}^T Q \boldsymbol{x} \le 1 \end{array}$$

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

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

Equality constrained extrema

Equality constrained extrema

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

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

subject to the equality constraints

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

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 set of feasible points X determined by equations (1)

As we have already seen, Lagrange's method consists of transforming an equality constrained extremum problem into a problem of finding a stationary point $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ of the Lagrangian function

$$L(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) - \sum_{i=1}^{m} \lambda_i h_i(\boldsymbol{x})$$

Lagrange's method

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, find the maximum of

$$f(x, y) = 4xy$$

subject to the constraint

$$h(x, y) = \frac{x^2}{a^2} + \frac{y^2}{b^2} - 1 = 0$$

In this example

$$L(x, y, \lambda) = 4xy - \lambda \left(\frac{x^2}{a^2} + \frac{y^2}{b^2} - 1 \right)$$

Lagrange's method

Theorem (Necessary conditions)

Suppose that

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

are real-valued functions that satisfy:

- *They are all continuosly differentiable on a neighborhood around* $\mathbf{x}^*$ *of radius* ϵ*:* $N_\epsilon(\mathbf{x}^*) \subset \mathcal{C}$
- $\mathbf{x}^*$ *is a local minimum (or maximum) of* f *in* $N_\epsilon(\mathbf{x}^*)$
- *If* $\mathbf{x} \in N_\epsilon(\mathbf{x}^*)$*, then all the constraints are satisfied*

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

- *The Jacobian matrix*

$$\begin{pmatrix} \frac{\partial h_1}{\partial x_1} & \cdots & \frac{\partial h_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_m}{\partial x_1} & \cdots & \frac{\partial h_m}{\partial x_n} \end{pmatrix}$$

  has rank m*, this is: the constraint gradients* $\nabla h_1(\mathbf{x}^*)$*,...,*$\nabla h_m(\mathbf{x}^*)$ *are linearly independent*

Then, there exists a vector of multipliers $\boldsymbol{\lambda}^* = (\lambda_1^*, ..., \lambda_m^*)^T$ *such that* $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ *is a stationary vector of the Lagrangian*

$$\nabla L(\mathbf{x}^*, \boldsymbol{\lambda}^*) = 0$$

Lagrange's method. First order feasible variations

Definition

The subspace of **first order feasible variations at $\boldsymbol{x}^*$** *is defined by*

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

Note that $V(\boldsymbol{x}^*)$ is the subspace of variations $\Delta\boldsymbol{x}$ for which the point $\boldsymbol{x}^* + \Delta\boldsymbol{x}$ satisfies the constraint

$$h(\boldsymbol{x}) = 0$$

up to the first order:

$$h(\boldsymbol{x}^* + \Delta\boldsymbol{x}) \approx h(\boldsymbol{x}^*) + \nabla h(\boldsymbol{x}^*)^T \Delta\boldsymbol{x} = \nabla h(\boldsymbol{x}^*)^T \Delta\boldsymbol{x} = 0$$

Lagrange's method. First order feasible variations

There are two ways to interpret the necessary condition given by the equation

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

1. The gradient of the cost function $\nabla f(\boldsymbol{x}^*)$ belongs to the subspace spanned by the gradients of the constraints $\nabla h_i(\boldsymbol{x}^*)$ at $\boldsymbol{x}^*$

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

2. $\nabla f(\boldsymbol{x}^*)$ is orthogonal to the subspace of first order feasible variations at $\boldsymbol{x}^*$ $(\nabla h_i(\boldsymbol{x}^*)^T \Delta \boldsymbol{x} = 0)$, this is

$$\text{If} \quad \Delta \boldsymbol{x} \in V(\boldsymbol{x}^*) \quad \text{then} \quad \nabla f(\boldsymbol{x}^*)^T \Delta \boldsymbol{x} = \sum_{i=1}^{m} \lambda_i \nabla h_i(\boldsymbol{x}^*)^T \Delta \boldsymbol{x} = 0$$

Lagrange necessary conditions

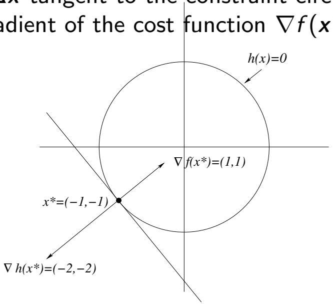
Example

$$\begin{array}{rl} \text{minimize} & f(x,y) = x + y \\ \text{subject to} & h(x,y) = 2 - x^2 - y^2 = 0 \end{array}$$

At the local minimum $\boldsymbol{x}^* = (-1,-1)^T$, the first order feasible variations $\Delta\boldsymbol{x}$ that must satisfy

$$\nabla h_i(\boldsymbol{x}^*)^T \Delta\boldsymbol{x} = \Delta\boldsymbol{x}^T \nabla h_i(\boldsymbol{x}^*) = 0$$

are the displacements $\Delta\boldsymbol{x}$ tangent to the constraint circle at $\boldsymbol{x}^*$, and are also perpendicular to the gradient of the cost function $\nabla f(\boldsymbol{x}^*) = (1,1)^T$



In this example, the gradient of the cost function $\nabla f(\boldsymbol{x}^*) = (1,1)^T$ is also collinear with the gradient of the constraint $\nabla h(\boldsymbol{x}^*) = (-2,-2)^T$

$$(1,1)^T = \nabla f(\boldsymbol{x}^*) = \lambda \nabla h(\boldsymbol{x}^*) = (1/2)(-2,-2)^T$$

Feasible variations

Definition

A point $\mathbf{x}$ *for which* $h_1(\mathbf{x}) = 0, \ldots, h_m(\mathbf{x}) = 0$ *(feasible point) and such that the gradients* $\nabla h_1(\mathbf{x}), \ldots, \nabla h_m(\mathbf{x})$ **are linearly independent** *is called* **regular**

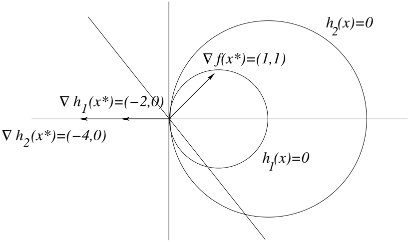
Remark: For a local minimum that is not regular there may not exist Lagrange multipliers

Example. Consider the problem of minimizing

$$f(\mathbf{x}) = x + y$$

subject to

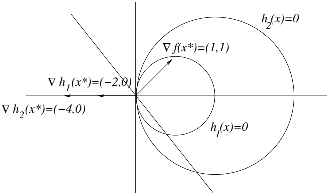
$$h_1(\mathbf{x}) = (x-1)^2 + y^2 - 1 = 0, \quad h_2(\mathbf{x}) = (x-2)^2 + y^2 - 4 = 0$$



Note that in this example we have $m = n$ instead of $m < n$, but this is not relevant for what follows

Example (cont.)

- At the local minimum of $f = x + y$, $\boldsymbol{x}^* = (0,0)^T$ (the only feasible point), the cost gradient $\nabla f(\boldsymbol{x}^*) = (1,1)^T$ cannot be expressed as a linear combination of $\nabla h_1(\boldsymbol{x}^*) = (-2,0)^T$ and $\nabla h_2(\boldsymbol{x}^*) = (-4,0)^T$



Thus, the Lagrange multiplier condition

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

cannot hold for any λ_1^* and λ_2^*

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

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

has dimension 1, that is larger than the one of the **true set of feasible variations** $\{\Delta \boldsymbol{x} = (0,0)^T\}$

Lagrange's method

Theorem (Sufficient conditions).

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

1. *The vector $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ is a stationary point of the Lagrangian*

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

2. *For every $\boldsymbol{z} \in \mathbb{R}^n$, $\boldsymbol{z} \neq 0$ satisfying*

$$(\nabla h_i(\boldsymbol{x}^*))^T \boldsymbol{z} = \boldsymbol{z}^T \nabla h_i(\boldsymbol{x}^*) = 0, \quad i = 1, \dots, m$$

($\boldsymbol{z}$ is a feasible first order variation) it follows that

$$\boldsymbol{z}^T \nabla^2_{xx} L(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) \boldsymbol{z} > 0$$

Then, f has a strict local minimum at $\boldsymbol{x}^*$ subject to $h_i(\boldsymbol{x}) = 0$, $i = 1, \dots, m$ (similar for a maximum if $\boldsymbol{z}^T \nabla^2_{xx} L(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) \boldsymbol{z} < 0$)

We will see the proof of both theorems (necessary and sufficient conditions) later, when we also consider inequality constraints

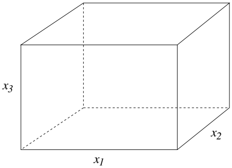
Sufficient conditions

Example

Consider the problem

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

this is, minimize the surface area of a rectangular parallelepiped P subject to the sum of the edge lengths of P being equal to 3



Since

$$L(\boldsymbol{x}, \lambda) = -(x_1x_2 + x_2x_3 + x_1x_3) - \lambda(x_1 + x_2 + x_3 - 3)$$

the necessary conditions ($\nabla L(\boldsymbol{x}^*, \lambda^*) = 0$) are

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

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

Sufficient conditions. Example (cont.)

The subspace of first order feasible variations V is

$$V = \{\boldsymbol{z} \mid \boldsymbol{z}^T \nabla h(\boldsymbol{x}^*) = 0\} = \left\{ \boldsymbol{z} \mid (z_1, z_2, z_3) \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = 0 \right\} = \{\boldsymbol{z} \mid z_1 + z_2 + z_3 = 0\}$$

The Hessian of the Lagrangian
$L(\boldsymbol{x}, \lambda) = -(x_1 x_2 + x_2 x_3 + x_1 x_3) - \lambda(x_1 + x_2 + x_3 - 3)$ is

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

We have for all $\boldsymbol{z} \in V$ with $\boldsymbol{z} \neq 0$, that

$$\boldsymbol{z}^T \nabla^2_{\boldsymbol{xx}} L(\boldsymbol{x}^*, \lambda^*) \boldsymbol{z} = -z_1(z_2 + z_3) - z_2(z_1 + z_3) - z_3(z_1 + z_2) = z_1^2 + z_2^2 + z_3^2 > 0$$

hence, the sufficient conditions for a strict local minimum

$$\boldsymbol{z}^T \nabla^2_{\boldsymbol{xx}} L(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) \boldsymbol{z} > 0$$

are satisfied

Inequality constrained extrema

First-order necessary conditions for inequality constrained extrema

We begin with the **first-order** (involving only first derivatives) necessary conditions

- Consider the general problem (P) defined by

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

  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 in D satisfying the constraints defined by (2)

$$X = \{\boldsymbol{x} \in D \mid g_i(\boldsymbol{x}) \geq 0,\ i = 1, ..., p;\ h_j(\boldsymbol{x}) = 0,\ j = 1, ..., m\}$$

  If $\boldsymbol{x} \in X$, we say that $\boldsymbol{x}$ is a feasible point

First-order necessary conditions for inequality constrained extrema

- A point $\boldsymbol{x}^* \in X$ is said to be a local minimum of problem (P), if there exist $\delta > 0$ such that

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

  where $N_\delta(\boldsymbol{x}^*)$ is the neighbourhood of radius δ centred at $\boldsymbol{x}^*$

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

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

  then $\boldsymbol{x}^*$ is said to be a global minimum of problem (P)

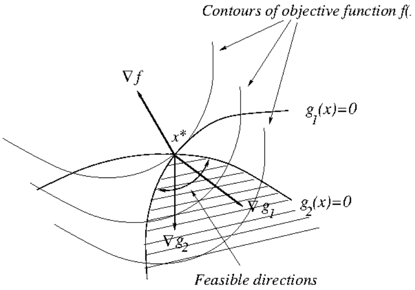
- Note that every point $\boldsymbol{x} \in N_\delta(\boldsymbol{x}^*)$ can be written as $\boldsymbol{x}^* + \boldsymbol{z}$, where $\boldsymbol{z} \neq 0$ if and only if $\boldsymbol{x} \neq \boldsymbol{x}^*$

Feasible directions

Definition

A vector $\boldsymbol{z} \neq 0$ *is called a* **feasible direction** *from* $\boldsymbol{x}^*$ *if there exist* $\delta > 0$ *such that*

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



- We are interested in feasible directions since

If $\boldsymbol{x}^*$ *is a local minimum of problem* (P)*, and if* $\boldsymbol{z}$ *is a feasible direction for* $\boldsymbol{x}^*$*, then* $f(\boldsymbol{x}^* + \theta \boldsymbol{z}) \geq f(\boldsymbol{x}^*)$*, if* $\theta > 0$ *is small enough*

Feasible directions characterization

Recall that one set of constraints is given by $g_i(\boldsymbol{x}) \geq 0$, for $i = 1, ..., p$

Define the set of index $I(\boldsymbol{x}^*)$ as:

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

Lemma

If $\boldsymbol{z}$ is a feasible direction, then

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

(the angle between $\boldsymbol{z}$ and $\nabla g_i(\boldsymbol{x}^*)$ is in $[-90^o, 90^o]$)

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

$$\boldsymbol{z}^T \nabla g_k(\boldsymbol{x}^*) < 0$$

Then, since $k \in I(\boldsymbol{x}^*)$, we can write

$$g_k(\boldsymbol{x}^* + \theta \boldsymbol{z}) = g_k(\boldsymbol{x}^*) + \theta \boldsymbol{z}^T \nabla g_k(\boldsymbol{x}^*) + \theta \epsilon_k(\theta) = \theta \boldsymbol{z}^T \nabla g_k(\boldsymbol{x}^*) + \theta \epsilon_k(\theta)$$

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

If θ is small enough, and since we have assumed that $\boldsymbol{z}^T \nabla g_k(\boldsymbol{x}^*) + \epsilon_k(\theta) < 0$, it follows that $g_k(\boldsymbol{x}^* + \theta \boldsymbol{z}) < 0$ for all $\theta > 0$ small enough, contradicting the fact that $\boldsymbol{z}$ is a feasible direction vector from $\boldsymbol{x}^*$ ($\boldsymbol{x}^* + \theta \boldsymbol{z} \in X \cap N_{\delta_1}(\boldsymbol{x}^*)$). So the claim is true

Feasible directions characterization

For the equality constraints defined by $h_j(\boldsymbol{x}) = 0$, for $j = 1, \ldots, m$, the following lemma holds.

Lemma

If $\boldsymbol{z}$ is a certain feasible direction, then

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

The proof is similar to the one of the previous lemma

Feasible directions characterization

- **Define**

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

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

- Note that $0 \in Z^1(\boldsymbol{x}^*)$, so $Z^1(\boldsymbol{x}^*) \neq \emptyset$

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

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

- **Define**

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

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

Constrained optimization. Summary of definitions

- The **first order feasible variations at x^***, Δx are defined as

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

  and satisfy the constraint in the linear approximation: $h(\mathbf{x}^* + \Delta \mathbf{x}) \approx 0$

- The necessary condition for equality constrained problems implies that **the gradient of the cost function $\nabla f(\mathbf{x}^*)$ is orthogonal to** $V(\mathbf{x}^*)$, since

$$\nabla f(\mathbf{x}^*) = \sum_{i=1}^{m} \lambda_i^* \nabla h_i(\mathbf{x}^*) \quad \Rightarrow \quad \nabla f(\mathbf{x}^*)^T \Delta \mathbf{x} = \sum_{i=1}^{m} \lambda_i \nabla h_i(\mathbf{x}^*)^T \Delta \mathbf{x} = 0$$

Constrained optimization. Summary of definitions

Feasible directions characterization

- Given $\boldsymbol{x}^*$ (not necessarily the solution of problem (P)), define the following sets

$$\begin{aligned} I(\boldsymbol{x}^*) &= \{i \mid g_i(\boldsymbol{x}^*) = 0\} \\ Z^1(\boldsymbol{x}^*) &= \left\{\boldsymbol{z} \mid \boldsymbol{z}^T \nabla g_i(\boldsymbol{x}^*) \geq 0, i \in I(\boldsymbol{x}^*);\ \boldsymbol{z}^T \nabla h_j(\boldsymbol{x}^*) = 0, j = 1, ..., m\right\} \neq \emptyset \\ Z^2(\boldsymbol{x}^*) &= \left\{\boldsymbol{z} \mid \boldsymbol{z}^T \nabla f(\boldsymbol{x}^*) < 0\right\} \end{aligned}$$

- Iif $\boldsymbol{z}$ is a certain feasible direction, we have

$$\boldsymbol{z}^T \nabla g_i(\boldsymbol{x}^*) \geq 0, \quad i \in I(\boldsymbol{x}^*) \quad \text{and} \quad \boldsymbol{z}^T \nabla h_j(\boldsymbol{x}^*) = 0, \quad j = 1, ...m$$

- If $\boldsymbol{z}$ is a feasible direction for $\boldsymbol{x}^*$, then $\boldsymbol{z} \in Z^1(\boldsymbol{x}^*)$, but it may happen that $\boldsymbol{z} \in Z^1(\boldsymbol{x}^*)$ without being a feasible direction

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

Necessary conditions "candidates"

Definition

The Lagrangian associated with problem (P) *is defined as*

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

The following Theorem gives **candidate conditions to become the necessary conditions** for $\boldsymbol{x}^0$ to be the solution of problem (P)

Necessary conditions "candidates"

Theorem

Given $\boldsymbol{x}^0 \in X$, *then* $Z^1(\boldsymbol{x}^0) \cap Z^2(\boldsymbol{x}^0) = \emptyset$ **if and only if** *there exist vectors* $\boldsymbol{\lambda}^0$, $\boldsymbol{\mu}^0$ *such that*

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

((3), (4) and (5) are called Lagrange conditions)

Remarks:

- Recall that if $\boldsymbol{z}$ is a feasible direction for $\boldsymbol{x}^0$ then $\boldsymbol{z} \in Z^1(\boldsymbol{x}^0)$
- Recall that if $\boldsymbol{z} \in Z^2(\boldsymbol{x}^0)$ then the function f decreases along $\boldsymbol{z}$
- From the above two remarks it follows that the condition $Z^1(\boldsymbol{x}^0) \cap Z^2(\boldsymbol{x}^0) = \emptyset$ implies that **there are no feasible directions at** x^0 **along which** f **decreases**

Necessary conditions "candidates"*

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

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

we have

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

this is, if $\boldsymbol{z} \in Z^1(\boldsymbol{x}^0)$, then $\boldsymbol{z} \notin Z^2(\boldsymbol{x}^0)$
We can write (7) as

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

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

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

Letting $\lambda_i^0 = 0$ for $i \notin I(\boldsymbol{x}^0)$, $\mu_j^0 = \mu_j^1 - \mu_j^2$, we conclude that $Z^1(\boldsymbol{x}^0) \cap Z^2(\boldsymbol{x}^0) = \emptyset$ 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 $\boldsymbol{x}^0$ to be the solution $\boldsymbol{x}^*$ of problem (P)
- According to them, we must guarantee that $Z^1(\boldsymbol{x}^*) \cap Z^2(\boldsymbol{x}^*) = \emptyset$ when $\boldsymbol{x}^*$ is a solution of (P). This condition (that will be characterized later) ensures that f can not decrease along any feasible direction
- For most problems $Z^1(\boldsymbol{x}^*) \cap Z^2(\boldsymbol{x}^*) = \emptyset$, and then the Lagrange conditions (3), (4) and (5) hold at $\boldsymbol{x}^*$
- Unfortunately, we can not state that if $\boldsymbol{x}^0$ is a solution of problem (P) and $Z^1(\boldsymbol{x}^0) \cap Z^2(\boldsymbol{x}^0) = \emptyset$, then the Lagrange conditions are satisfied, as we will see in the next example

Example

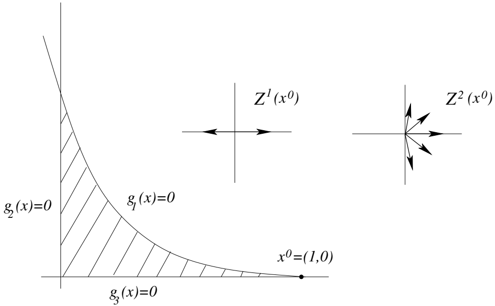
Example: Consider $\boldsymbol{x} = (x_1, x_2)^T \in \mathbb{R}^2$, $f(\boldsymbol{x}) = -x_1$ with the following constraints:

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

that define the feasible set X. The feasible point $\boldsymbol{x}^0 = (1,0)^T$ is the solution of the problem

$$\max_X x_1 = \min_X (-x_1)$$

Let's see that $Z^1(\boldsymbol{x}^*) \cap Z^2(\boldsymbol{x}^*) \neq \emptyset$

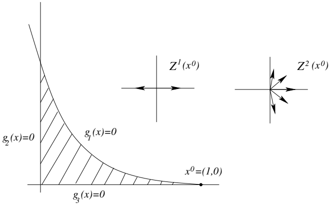


Example (cont.)

We can easily verify that

$$I(\boldsymbol{x}^0) = I((1,0)^T) = \{1,3\}, \quad \nabla g_1(\boldsymbol{x}^0) = (0,-1)^T, \quad \nabla g_3(\boldsymbol{x}^0)) = (0,1)^T$$

$$Z^1(\boldsymbol{x}^0) = \left\{\boldsymbol{z} \in \mathbb{R}^2 \mid \boldsymbol{z}^T \nabla g_i(\boldsymbol{x}^0) \geq 0, i \in I(\boldsymbol{x}^0)\right\} = \left\{\boldsymbol{z} = (z_1, z_2)^T \mid z_2 = 0\right\}$$



But at $\boldsymbol{x}^0$

$$Z^2(\boldsymbol{x}^0) = \left\{\boldsymbol{z} \in \mathbb{R}^2 \mid \boldsymbol{z}^T \nabla f(\boldsymbol{x}^0) < 0\right\} = \left\{\boldsymbol{z} = (z_1, z_2)^T \mid z_1 > 0\right\}$$

and

$$Z^1(\boldsymbol{x}^*) \cap Z^2(\boldsymbol{x}^*) = \left\{\boldsymbol{z} \in \mathbb{R}^2 \mid z_1 > 0, \ z_2 = 0\right\} \neq \emptyset$$

hence, due to the above Theorem, there exist no $\boldsymbol{\lambda}^0$ satisfying Lagrange conditions (3), (4) and (5)

Two technical results

Farkas Lemma and the Theorem of the Alternative

Farkas Lemma

Lemma

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

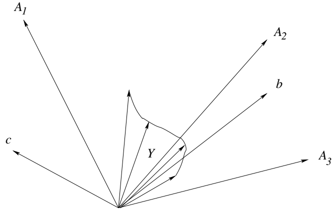
$$(b_1 \cdots b_n) \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \geq 0, \ \forall y \in \mathbb{R}^n \text{ s.t. } A\boldsymbol{y} \geq 0 \Leftrightarrow$$

$$\Leftrightarrow \exists \boldsymbol{\rho} \in \mathbb{R}^m, \ \boldsymbol{\rho} \geq 0, \ \text{s.t.} \begin{pmatrix} a_{11} & \cdots & a_{m1} \\ \vdots & & \vdots \\ a_{1n} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} \rho_1 \\ \vdots \\ \rho_m \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$$

Farkas Lemma. Geometric interpretation

$$(b_1 \cdots b_n)\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \geq 0,\ \forall y \in \mathbb{R}^n \text{ s.t. } A\boldsymbol{y} \geq 0 \Leftrightarrow \exists \boldsymbol{\rho} \in \mathbb{R}^m,\ \boldsymbol{\rho} \geq 0,\ \text{s.t.} \begin{pmatrix} a_{11} & \cdots & a_{m1} \\ \vdots & & \vdots \\ a_{1n} & \cdots & a_{mn} \end{pmatrix}\begin{pmatrix} \rho_1 \\ \vdots \\ \rho_m \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$$

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



The set $Y = \{\boldsymbol{y} \mid A\boldsymbol{y} \geq \boldsymbol{0}\}$ consists of all the vectors $\boldsymbol{y} \in \mathbb{R}^2$ that make an acute angle with every row of A

The Lemma states that $\boldsymbol{b}$ makes an acute angle with every $\boldsymbol{y} \in Y$ if and only if $\boldsymbol{b}$ can be expressed as a nonnegative linear combibation of the rows of A

In the figure, $\boldsymbol{b}$ satisfies these conditions and $\boldsymbol{c}$ does not

Theorem of the Alternative

Theorem

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

$$A\boldsymbol{x} < 0$$

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

$$\boldsymbol{u}^T A = 0, \quad \boldsymbol{u} \geq 0$$

but never both

Theorem of the Alternative. Proof

Proof: Assume that there exist $\boldsymbol{x}$ and $\boldsymbol{u}$ such that both

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

are satisfied. Then we have $\boldsymbol{u}^T A\boldsymbol{x} < 0$, and $\boldsymbol{u}^T A\boldsymbol{x} = 0$ simultaneously, which is a contradiction

Assume now that there exist no $\boldsymbol{x}$ satisfying the first condition ($A\boldsymbol{x} < 0$), and let us see that we can find $\boldsymbol{u}$ that satisfies the second condition of the Theorem. The assumption means that we cannot find a negative number $w < 0$ satisfying

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

for every $\boldsymbol{x} \in \mathbb{R}^n$, where A_i is the ith-row of A. This is, if for $i = 1, ..., m$, and $\forall \boldsymbol{x} \in \mathbb{R}^n$

$$A_i\boldsymbol{x} \leq w \quad \Leftrightarrow \quad w - A_i\boldsymbol{x} \geq 0, \qquad \text{then } w \geq 0$$

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

Theorem of the Alternative. Proof (cont.)

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

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

then

$$w = \boldsymbol{b}^T \boldsymbol{y} \geq 0$$

According to Farkas lemma, there exists a m-vector $\boldsymbol{u} \geq 0$, such that

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

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 $\boldsymbol{u}$ that satisfies the second condition of the Theorem of the Alternative

$\square$

Weak necessary optimality conditions

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

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

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

where λ_0 is an additional parameter

Weak necessary optimality conditions

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

$$\begin{array}{rll} \min & f(x_1, ..., x_n) & \\ \text{subject to} & g_i(x_1, ..., x_n) \geq 0, \qquad & i = 1, ..., p \end{array} \qquad (P)$$

Remark: The equality constraints become inequality constraints according to:

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

Weak necessary optimality conditions

Theorem

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

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

Proof

Proof: We must proof that the necessary conditions for $\boldsymbol{x}^*$ to be the solution of problem (P), is the existence of a vector $\boldsymbol{\lambda}^*$ satisfying (11), (12) and (13)

If $g_i(\boldsymbol{x}^*) > 0$ for all i (the point $\boldsymbol{x}^*$ is in the interior of the feasible set X), then $I(\boldsymbol{x}^*) = \{i \mid g_i(\boldsymbol{x}^*) = 0\} = \emptyset$. Choose $\lambda_0^* = 1$, $\lambda_1^* = \lambda_2^* = \ldots = \lambda_p^* = 0$ and the conditions (11), (12) and (13) hold since $\nabla f(\boldsymbol{x}^*) = 0$

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

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

we cannot have

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

This follows from the following: According to Taylor's formula, we can see that if there exists $\boldsymbol{z}$ satisfying (14), then we can find a sufficiently small δ such that if $0 < \theta < \delta$, then $\boldsymbol{x} = \boldsymbol{x}^* + \theta \boldsymbol{z}$ satisfies

$$g_i(\boldsymbol{x}) = g_i(\boldsymbol{x}^*) + \theta \boldsymbol{z}^T \nabla g_i(\boldsymbol{x}^*) + O_2$$

and, since $g_i(\boldsymbol{x}^*) = 0$ we get

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

for all $0 < \theta < \delta$, that is, $\boldsymbol{x}$ is a feasible point

Proof (cont.)

Now, if (15) also holds, then

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

contradicting that $\boldsymbol{x}^*$ is a minimum

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

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

has no solution. Using the matrix A with rows equal to $\nabla f(\boldsymbol{x}^*)$ and $-\nabla g_i(\boldsymbol{x}^*)$:

$$A = \begin{pmatrix} \nabla f(\boldsymbol{x}^*) \\ -\nabla g_{i_1}(\boldsymbol{x}^*) \\ \cdots \\ -\nabla g_{i_r}(\boldsymbol{x}^*) \end{pmatrix}$$

the above system of inequalities, which has no solution, can be written as $A\boldsymbol{z} < 0$ According to the Theorem of the Alternative, we get that there exists a nonzero vector $\boldsymbol{\lambda}^* \geq 0$, such that

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

Proof (cont.)*

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

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

and clearly

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

$\square$

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 $\boldsymbol{x}^*$ is a solution of problem (P), then there exist $\boldsymbol{\lambda}^* = (\lambda_0^*, \lambda_1^*, \ldots, \lambda_p^*)^T$ and $\boldsymbol{\mu}^* = (\mu_1^*, \ldots, \mu_m^*)^T$ such that:

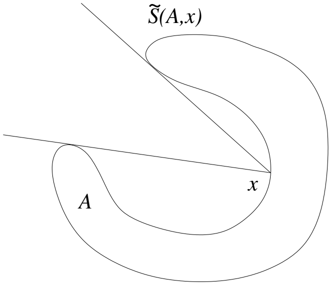
$$
\begin{aligned}
\nabla_x \tilde{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) &= \lambda_0^* \nabla f(\boldsymbol{x}^*) - \sum_{i=1}^{p} \lambda_i^* \nabla g_i(\boldsymbol{x}^*) - \sum_{j=1}^{m} \mu_j^* \nabla h_j(\boldsymbol{x}^*) = 0 \\
\lambda_i^* g_i(\boldsymbol{x}^*) &= 0,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

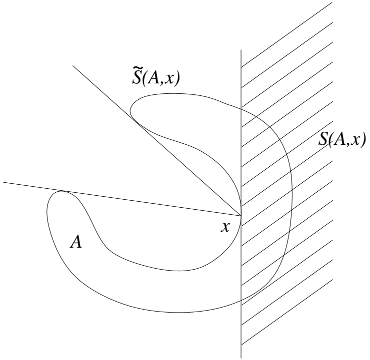
More definitions: the cone and the closed cone of tangents

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

Define the **cone of tangents** of the set A at $\boldsymbol{x} \in A$, $\tilde{S}(A, \boldsymbol{x})$, as the intersection of all closed cones containing the set $\{\boldsymbol{a} - \boldsymbol{x} \mid \boldsymbol{a} \in A\}$, this is

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



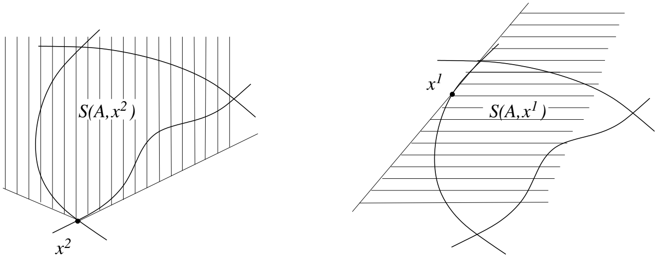


The cone and the closed cone of tangents

Define the **closed cone of tangents** of the set A at $\boldsymbol{x} \in A$, $S(A, \boldsymbol{x})$ as

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

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



The following lemma characterizes $S(A, \boldsymbol{x})$

The closed cone of tangents. Characterization

Lemma

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

$$\mathbf{z} \in S(A, \mathbf{x}) \quad \Leftrightarrow \quad \exists \{\mathbf{x}^k\} \text{ and } \{\alpha^k \geq 0\} \text{ such that } \{\mathbf{x}^k\} \to \mathbf{x},\ \{\alpha^k(\mathbf{x}^k - \mathbf{x})\} \to \mathbf{z}$$

Proof: Assume that $\mathbf{z} \in S(A, \mathbf{x})$. Then $\mathbf{z} \in \tilde{S}(A \cap N_{1/k}(\mathbf{x}), \mathbf{x})$ for $k = 1, 2, ...,$ and, by definition:

$$\tilde{S}(A \cap N_{1/k}(\mathbf{x}), \mathbf{x}) = \mathrm{cl}\{\alpha(\mathbf{y} - \mathbf{x}) \mid \alpha \geq 0, \mathbf{y} \in A \cap N_{1/k}(\mathbf{x})\}, \quad k = 1, 2, \dots \quad (16)$$

where cl denotes the closure operation of sets in $\mathbb{R}^n$

Choose any sequence of positive numbers $\{\epsilon^k\} \to 0$, and consider the vectors $\mathbf{z}(\epsilon^k) \in \{\alpha(\mathbf{y} - \mathbf{x}) \mid \alpha \geq 0, \mathbf{y} \in A \cap N_{1/k}(\mathbf{x})\}$ such that

$$\|\mathbf{z}(\epsilon^k) - \mathbf{z}\| \leq \epsilon^k \quad (17)$$

Due to the condition (16), the points $\mathbf{z}(\epsilon^k)$ can be written as

$$\mathbf{z}(\epsilon^k) = \alpha(\epsilon^k)(\mathbf{y}(\epsilon^k) - \mathbf{x}), \quad \alpha(\epsilon^k) \geq 0, \quad \mathbf{y}(\epsilon^k) \in A \cap N_{1/k}(\mathbf{x}) \quad (18)$$

The closed cone of tangents. Characterization (cont.)

Letting $k = 1, 2...$ we generate a sequence of vectors $\boldsymbol{y}(\epsilon^1)$, $\boldsymbol{y}(\epsilon^2), \ldots$ that is contained in A and converges to $\boldsymbol{x}$, and a sequence of nonnegative numbers $\alpha(\epsilon^1)$, $\alpha(\epsilon^2), \ldots$ such that, according to (17) and (18), the sequence $\{\alpha(\epsilon^k)(\boldsymbol{y}(\epsilon^k) - \boldsymbol{x})\}$ converges to $\boldsymbol{z}$

Conversely, suppose that there exist a sequence of vectors $\{\boldsymbol{x}^k\} \subset A$ converging to $\boldsymbol{x}$ and a sequence of nonnegative numbers $\{\alpha^k\}$ such that $\{\alpha^k(\boldsymbol{x}^k - \boldsymbol{x})\}$ converges to $\boldsymbol{z}$. Let p be any natural number. Then, there exists a natural number K such that $k \geq K$ implies $\boldsymbol{x}^k \in A \cap N_{1/p}(\boldsymbol{x})$, so

$$\alpha^k(\boldsymbol{x}^k - \boldsymbol{x}) \in \tilde{S}(A \cap N_{1/p}(\boldsymbol{x})), \quad k \geq K$$

and, since $\tilde{S}$ is closed

$$z \in \tilde{S}(A \cap N_{1/p}(\boldsymbol{x}))$$

Since this last expression holds for any natural number p, it follows that

$$z \in \bigcap_{p \geq 1} \tilde{S}(A \cap N_{1/p}(x)) = S(A, x)$$

$\square$

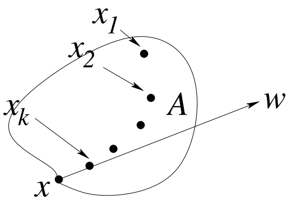
The closed cone of tangents (alternative description)

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

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

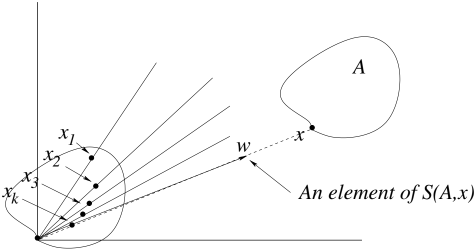
$$\lim_{k \to \infty} \frac{\boldsymbol{x}^k - \boldsymbol{x}}{\|\boldsymbol{x}^k - \boldsymbol{x}\|} = \boldsymbol{w}$$

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



The closed cone of tangents (alternative description)

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



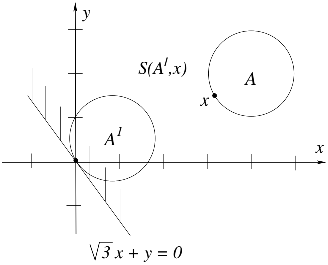
The closed cone of tangents. Example

Example: Consider the closed ball A with center at $(4, 2)$ and radius 1:

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

Let us find the cone of tangents of A at the boundary point

$$\boldsymbol{x} = (4 - \sqrt{3}/2, 3/2)^T$$



First we translate A to the origin, obtaining the ball

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

Taking sequences of points $\{\boldsymbol{x}^k\}$ 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 circle A^1 at the origin

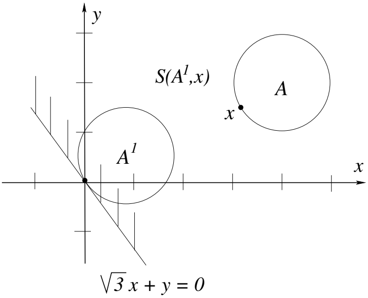
The closed cone of tangents. Example

The tangent line to the circle at the origin 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

$$S(A^1, \boldsymbol{x}) = \{(x_1, x_2) \,|\, \sqrt{3}x_1 + x_2 \geq 0\}$$

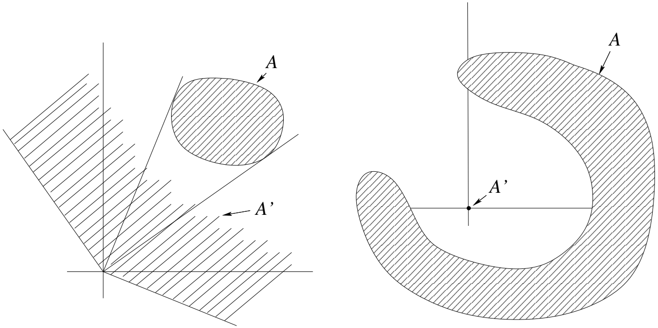


Positively normal cones

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

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

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



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

We have defined the positively normal cone to a set $A \subset \mathbb{R}^n$ as

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

so, the positively normal cone of $Z^1(\boldsymbol{x}^0)$ is

$$(Z^1(\boldsymbol{x}^0))' = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{z}^T \boldsymbol{x} \geq 0, \ \forall \boldsymbol{z} \in Z^1(\boldsymbol{x}^0)\}$$

Lemma

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

$$\nabla f(\boldsymbol{x}^0) \in (Z^1(\boldsymbol{x}^0))'$$

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

Lemma

Assume that $\boldsymbol{x}^0$ is a solution of problem (P). Then

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

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

Proof: We must show that $\boldsymbol{z}^T \nabla f(\boldsymbol{x}^0) \geq 0$ for every $\boldsymbol{z} \in S(X, \boldsymbol{x}^0)$

Let $\boldsymbol{z} \in S(X, \boldsymbol{x}^0)$. According to the previous characterization Lemma of the tangent cone (see page 33), there exists a sequence $\{\boldsymbol{x}^k\} \subset X$ converging to $\boldsymbol{x}^0$ and a sequence of nonnegative numbers $\{\alpha^k\}$ such that $\{\alpha^k(\boldsymbol{x}^k - \boldsymbol{x}^0)\}$ converges to $\boldsymbol{z}$

Since f is differentiable at $\boldsymbol{x}^0$, we can write

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

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

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

Cones of tangents and positively normal cones (cont.)*

Since $\boldsymbol{x}^k \in X$, and $\boldsymbol{x}^0$ is a local minimum ($f(\boldsymbol{x}^k) - f(\boldsymbol{x}^0) \geq 0$ if k is large enough), it follows that, by letting $k \to \infty$, the term $\epsilon \|\alpha^k (x^k - x^0)\|$ in the above equation

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

goes to 0, and $\alpha^k (f(\boldsymbol{x}^k) - f(\boldsymbol{x}^0))$ converges to a non-negative limit. Thus

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

That is

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

$\square$

The Karush-Kuhn-Tucker necessary optimality conditions

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

Theorem

Let $\boldsymbol{x}^*$ be a solution of problem (P) and suppose that

$$(Z^1(\boldsymbol{x}^*))' = (S(X, \boldsymbol{x}^*))' \tag{19}$$

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

$$\begin{aligned} \nabla f(\boldsymbol{x}^*) - \sum_{i=1}^{p} \lambda_i^* \nabla g_i(\boldsymbol{x}^*) - \sum_{j=1}^{m} \mu_j^* \nabla h_j(\boldsymbol{x}^*) &= 0 && (20) \\ \lambda_i^* g_i(\boldsymbol{x}^*) &= 0, \quad i = 1, ..., p && (21) \\ \boldsymbol{\lambda}^* &\geq 0. && (22) \end{aligned}$$

(Karush–Kuhn–Tucker conditions)

Proof: Suppose that $\boldsymbol{x}^*$ is a solution of (P). According to a previous Lemma, $\nabla f(\boldsymbol{x}^*) \in (S(X, \boldsymbol{x}^*))'$. If $(Z^1(\boldsymbol{x}^*))' = (S(X, \boldsymbol{x}^*))'$, then $\nabla f(\boldsymbol{x}^*) \in (Z^1(\boldsymbol{x}^*))'$, and we have already seen that then $Z^1(\boldsymbol{x}^*) \cap Z^2(\boldsymbol{x}^*)$ is empty (see page 48). According to the characterization theorem of the condition $Z^1(\boldsymbol{x}^*) \cap Z^2(\boldsymbol{x}^*) = \emptyset$ (see page 26), conditions (20), (21) and (22) hold □

Essentially, what the above theorem says is that the condition

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

is a sufficient condition for the existence of the multipliers $\boldsymbol{\lambda}^*$ and $\boldsymbol{\mu}^*$ satisfying conditions (20), (21) and (22).

Notice that if

$$Z^1(\boldsymbol{x}^*) = S(X, \boldsymbol{x}^*)$$

at a solution point $\boldsymbol{x}^*$ of problem (P) implies the hypotheses of the last theorem

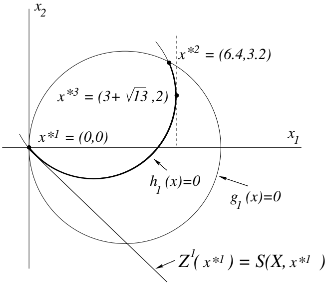
The Karush-Kuhn-Tucker necessary optimality conditions

Example: Consider the following problem

$$\min f(\boldsymbol{x}) = x_1$$

subject to

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



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

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

The Karush-Kuhn-Tucker necessary optimality conditions (cont.)

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

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

hence $Z^1(\boldsymbol{x}^{*1}) \cap Z^2(\boldsymbol{x}^{*1}) = \emptyset$. The Karush–Kuhn–Tucker conditions (20), (21) and (22) are satisfied for $\lambda_1^* = 1/8$ and $\mu_1^* = 0$

At the second point

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

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

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

It turns out that at $\boldsymbol{x}^{*3} = (3 + \sqrt{13}, 2)$ the Karush–Kuhn–Tucker necessary conditions also hold. At this point $Z^1(\boldsymbol{x}^{*3}) \cap Z^2(\boldsymbol{x}^{*3}) = \emptyset$ and the corresponding multipliers are $\lambda_1^* = 0$ and $\mu_1^* = \sqrt{13}/26$

From the above figure is clear that $\boldsymbol{x}^{*3}$ is not a solution of our problem but is a solution of

$$\max f(\boldsymbol{x}) = x_1$$

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 Karush–Kuhn–Tucker conditions; later we will give the sufficient contions for optimality

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

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

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

Recall that $Z^1(\boldsymbol{x})$ is

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

Second-order optimality conditions

Definition: The second-order constraint qualification is said to hold at $\boldsymbol{x}^0 \in X$ if for each $\boldsymbol{z} \in \hat{Z}^1(\boldsymbol{x}^0)$ there is a twice differentiable function

$$\alpha : [0, \epsilon] \subset \mathbb{R} \longrightarrow \mathbb{R}^n$$

such that

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

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

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

for some positive $\lambda > 0$

Since $\hat{Z}^1(\boldsymbol{x}^*)$ is a cone, we can always assume that $\lambda = 1$

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

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

Second-order optimality conditions theorem

Theorem

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

- *If there exist $\boldsymbol{\lambda}^* = (\lambda_1^*, ..., \lambda_p^*)^T$ and $\boldsymbol{\mu}^* = (\mu_1^*, ..., \mu_m^*)^T$ satisfying the Karush–Kuhn–Tucker conditions (20), (21) and (22):*

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

  and

- *If for every $\boldsymbol{z} \neq 0$ such that $\boldsymbol{z} \in \hat{Z}^1(\boldsymbol{x}^*)$, it follows that*

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

then $\boldsymbol{x}^*$ is a strict local minimum of problem (P)

Second-order optimality conditions theorem*

Proof: Let $\boldsymbol{z} \neq 0$ such that $\boldsymbol{z} \in \hat{Z}^1(\boldsymbol{x}^*)$ and $\alpha(t)$ the function that appears in the second order constraint qualification; that is

$$\alpha(0) = \boldsymbol{x}^*, \quad d\alpha(0)/dt = \boldsymbol{z}$$

Let $d^2\alpha(0)/dt^2 = \boldsymbol{w}$. From the second order conditions and the chain rule it follows that for $i \in I(\boldsymbol{x}^*)$

$$\frac{dg_i(\alpha(0))}{dt} = \boldsymbol{z}^T \nabla g_i(\boldsymbol{x}^*) \quad \Rightarrow$$

$$\frac{d^2 g_i(\alpha(0))}{dt^2} = \boldsymbol{z}^T \nabla^2 g_i(\boldsymbol{x}^*)\boldsymbol{z} + \boldsymbol{w}^T \nabla g_i(\boldsymbol{x}^*) = 0, \quad i \in I(\boldsymbol{x}^*) \tag{24}$$

$$\frac{dh_j(\alpha(0))}{dt} = \boldsymbol{z}^T \nabla h_j(\boldsymbol{x}^*) \quad \Rightarrow$$

$$\frac{d^2 h_j(\alpha(0))}{dt^2} = z^T \nabla^2 h_j(\boldsymbol{x}^*)\boldsymbol{z} + \boldsymbol{w}^T \nabla h_j(\boldsymbol{x}^*) = 0, \quad j = 1, ..., p \tag{25}$$

From condition (20), $\nabla_x L(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = 0$, and the definition of $\hat{Z}^1(\boldsymbol{x}^*)$, we have

$$\frac{df(\alpha(0))}{dt} \; = \; \boldsymbol{z}^T \nabla f(\boldsymbol{x}^*) \; = \; \boldsymbol{z}^T \left[\sum_{i=1}^{m} \lambda_i^* \nabla g_i(\boldsymbol{x}^*) + \sum_{j=1}^{p} \mu_j^* \nabla h_j(\boldsymbol{x}^*) \right] = 0$$

Second-order optimality conditions theorem (cont.)*

Since $\boldsymbol{x}^*$ is a local minimum, and $df(\alpha(0))/dt = 0$, it follows that $d^2 f(\alpha(0))/dt^2 \geq 0$, this is

$$\frac{d^2 f(\alpha(0))}{dt^2} = \boldsymbol{z}^T \nabla^2 f(\boldsymbol{x}^*) \boldsymbol{z} + \boldsymbol{w}^T \nabla f(\boldsymbol{x}^*) \geq 0 \tag{26}$$

Multiplying (24) and (25) by the corresponding multipliers, substracting from (26) and using the Karush–Kuhn–Tucker conditions (20), we get de desired inequality (23) $\square$

Sufficient optimality conditions

Denote by $\overline{I}(\boldsymbol{x}^*)$ the set of indices i for which $g_i(\boldsymbol{x}^*) = 0$ and the Karush–Kuhn–Tucker conditions (20), (21) and (22) are satisfied by $\lambda_i^* > 0$

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

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

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

The following theorem gives sufficient optimality conditions

Sufficient optimality conditions

Theorem

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

$$\begin{aligned}
\nabla_{\mathbf{x}} L(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) &= \nabla f(\mathbf{x}^*) - \sum_{i=1}^{p} \lambda_i^* \nabla g_i(\mathbf{x}^*) - \sum_{j=1}^{m} \mu_j^* \nabla h_j(\mathbf{x}^*) = 0 \quad (27)\\
\lambda_i^* g_i(\mathbf{x}^*) &= 0, \quad i = 1, ..., p \quad (28)\\
\boldsymbol{\lambda}^* &\geq 0 \quad (29)
\end{aligned}$$

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

$$\mathbf{z}^T \left[\nabla^2 f(\mathbf{x}^*) - \sum_{i=1}^{p} \lambda_i^* \nabla^2 g_i(\mathbf{x}^*) - \sum_{j=1}^{m} \mu_j^* \nabla^2 h_j(\mathbf{x}^*) \right] \mathbf{z} = \mathbf{z}^T \nabla_{\mathbf{x}}^2 L(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \mathbf{z} > 0 \quad (30)$$

then, $\mathbf{x}^*$ is a strict local minimum of problem (P)

Sufficient optimality conditions (cont.)*

Proof: Assume that the conditions (27), (28), (29) and (30) hold, and that $\boldsymbol{x}^*$ is not a strict local minimum. Then, there exists a sequence $\{\boldsymbol{z}^k\}$ of feasible points, $\boldsymbol{z}^k \neq \boldsymbol{x}^*$, convergent to $\boldsymbol{x}^*$, such that for each $\boldsymbol{z}^k$

$$f(\boldsymbol{x}^*) \geq f(\boldsymbol{z}^k) \tag{31}$$

Let $\boldsymbol{z}^k = \boldsymbol{x}^* + \theta^k \boldsymbol{y}^k$, with $\theta^k > 0$ and $\|\boldsymbol{y}^k\| = 1$. Without loss of generality, assume that the sequence $\{(\theta^k, \boldsymbol{y}^k)\}$ converges to $(0, \overline{\boldsymbol{y}})$, where $\|\overline{\boldsymbol{y}}\| = 1$. Since the points $\boldsymbol{z}^k$ are feasible

$$\begin{aligned} g_i(\boldsymbol{z}^k) - g_i(\boldsymbol{x}^*) &= \theta^k (\boldsymbol{y}^k)^T \nabla g_i(\boldsymbol{x}^* + \eta_i^k \theta^k \boldsymbol{y}^k) \geq 0, \quad i \in I(\boldsymbol{x}^*) && (32) \\ h_j(\boldsymbol{z}^k) - h_j(\boldsymbol{x}^*) &= \theta^k (\boldsymbol{y}^k)^T \nabla h_j(\boldsymbol{x}^* + \overline{\eta}_j^k \theta^k \boldsymbol{y}^k) = 0, \quad j = 1, ..., p && (33) \end{aligned}$$

and from (31)

$$f(\boldsymbol{z}^k) - f(\boldsymbol{x}^*) = \theta^k (\boldsymbol{y}^k)^T \nabla f(\boldsymbol{x}^* + \eta^k \theta^k \boldsymbol{y}^k) \leq 0 \tag{34}$$

where η^k, η_i^k and $\overline{\eta}_j^k$ are numbers between 0 and 1. Dividing (32), (33) and (34) by $\theta^k > 0$, and taking limits, we get

$$\begin{aligned} \overline{\boldsymbol{y}}^T \nabla g_i(\boldsymbol{x}^*) &\geq 0, \quad i \in I(\boldsymbol{x}^*) && (35) \\ \overline{\boldsymbol{y}}^T \nabla h_j(\boldsymbol{x}^*) &= 0, \quad j = 1, ..., p && (36) \\ \overline{\boldsymbol{y}}^T \nabla f(\boldsymbol{x}^*) &\leq 0 && (37) \end{aligned}$$

Sufficient optimality conditions (cont.)*

Assume now that (35) holds with a strict inequality for some $i \in \bar{I}(\boldsymbol{x}^*)$. Then, from (27), (35) and (36) we get

$$\bar{\boldsymbol{y}}^T \nabla f(\boldsymbol{x}^*) = \sum_{i=1}^{m} \lambda_i^* \bar{\boldsymbol{y}}^T \nabla g_i(\boldsymbol{x}^*) + \sum_{j=1}^{p} \mu_j^* \bar{\boldsymbol{y}}^T \nabla h_j(\boldsymbol{x}^*) > 0$$

contradicting (37). Therefore $\bar{\boldsymbol{y}}^T \nabla g_i(\boldsymbol{x}^*) = 0$ for all $i \in \bar{I}(\boldsymbol{x}^*)$, and so $\bar{y} \in \bar{Z}^1(\boldsymbol{x}^*)$. From Taylor's formula we obtain

$$\begin{aligned} g_i(\boldsymbol{z}^k) &= g_i(\boldsymbol{x}^*) + \theta^k (\boldsymbol{y}^k)^T \nabla g_i(\boldsymbol{x}^*) && (38) \\ &\quad + \frac{1}{2}(\theta^k)^2 (\boldsymbol{y}^k)^T [\nabla^2 g_i(\boldsymbol{x}^* + \xi_i^k \theta^k \boldsymbol{y}^k)] \boldsymbol{y}^k \geq 0, \quad i = 1, \dots, m \\ h_j(\boldsymbol{z}^k) &= h_j(\boldsymbol{x}^*) + \theta^k (\boldsymbol{y}^k)^T \nabla h_j(\boldsymbol{x}^*) && (39) \\ &\quad + \frac{1}{2}(\theta^k)^2 (\boldsymbol{y}^k)^T [\nabla^2 h_j(\boldsymbol{x}^* + \bar{\xi}_j^k \theta^k \boldsymbol{y}^k)] \boldsymbol{y}^k = 0, \quad j = 1, \dots, p \\ f(\boldsymbol{z}^k) - f(\boldsymbol{x}^*) &= \theta^k (\boldsymbol{y}^k)^T \nabla f(\boldsymbol{x}^*) && (40) \\ &\quad + \frac{1}{2}(\theta^k)^2 (\boldsymbol{y}^k)^T [\nabla^2 f(\boldsymbol{x}^* + \xi^k \theta^k \boldsymbol{y}^k)] \boldsymbol{y}^k \leq 0 \end{aligned}$$

where ξ^k, ξ_i^k and $\bar{\xi}_j^k$ are again numbers between 0 and 1

Sufficient optimality conditions (cont.)*

Multiplying (38) and (39) by λ_i^* and μ_j^*, respectively, and substracting from (40), we obtain

$$\theta^k (\boldsymbol{y}^k)^T \left\{ \nabla f(\boldsymbol{x}^*) - \sum_{i=1}^m \lambda_i^* \nabla g_i(\boldsymbol{x}^*) - \sum_{j=1}^p \mu_j^* \nabla h_j(\boldsymbol{x}^*) \right\}$$

$$+\tfrac{1}{2}(\theta^k)^2 (\boldsymbol{y}^k)^T \left[\nabla^2 f(\boldsymbol{x}^* + \xi^k \theta^k \boldsymbol{y}^k) - \sum_{i=1}^m \lambda_i^* \nabla^2 g_i(\boldsymbol{x}^* + \xi_i^k \theta^k \boldsymbol{y}^k) - \right.$$

$$\left. \sum_{j=1}^p \mu_j^* \nabla^2 h_j(\boldsymbol{x}^* + \overline{\xi}_j^k \theta^k \boldsymbol{y}^k) \right] \boldsymbol{y}^k \leq 0$$

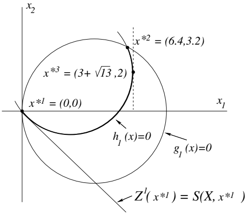
Since (27), the expression in braces (in the first line) vanishes. Dividing the remaining portion by $(\theta^k)^2/2$ and taking limits, we obtain

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

Since $\overline{\boldsymbol{y}} \neq 0$ and $\overline{\boldsymbol{y}} \in \overline{Z}^1(\boldsymbol{x}^*)$, it follows that this last inequality contradicts (30)
$\square$

The Karush-Kuhn-Tucker necessary optimality conditions

Example: Consider again the problem min $f(\boldsymbol{x}) = x_1$ 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 $\boldsymbol{x}^{*1}$ we have that

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

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

At $\boldsymbol{x}^{*3}$, however

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

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

Saddel points of the Lagrangian

Another type of **optimality conditions** is related to the Lagrangian and is **expressed in tems of its saddle points**.

Let Φ be a real function defined in $D \times E \subset \mathbb{R}^n \times \mathbb{R}^m$:

$$\begin{array}{rccc} \Phi: & D \times E & \longrightarrow & \mathbb{R} \\ & (\boldsymbol{x}, \boldsymbol{y}) & \longrightarrow & \Phi(\boldsymbol{x}, \boldsymbol{y}). \end{array}$$

A point $(\overline{\boldsymbol{x}}, \overline{\boldsymbol{y}})$ is said to be a **saddle point** of Φ if:

$$\Phi(\overline{\boldsymbol{x}}, \boldsymbol{y}) \leq \Phi(\overline{\boldsymbol{x}}, \overline{\boldsymbol{y}}) \leq \Phi(\boldsymbol{x}, \overline{\boldsymbol{y}}), \quad \forall (\boldsymbol{x}, \boldsymbol{y}) \in D \times E.$$

Analogously to the nonlinear problem, there is a **saddle point problem** that can be stated as follows:

The saddle point problem (S)

Problem (S)**:** *Find* $\overline{\boldsymbol{x}} \in \mathbb{R}^n$, $\overline{\boldsymbol{\lambda}} \in \mathbb{R}^m$, $\overline{\boldsymbol{\lambda}} \geq 0$, $\overline{\boldsymbol{\mu}} \in \mathbb{R}^p$ *such that* $(\overline{\boldsymbol{x}}, \overline{\boldsymbol{\lambda}}, \overline{\boldsymbol{\mu}})$ *is a saddle point of the Lagrangian*

$$L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\boldsymbol{x}) - \sum_{i=1}^{m} \boldsymbol{\lambda}_i g_i(\boldsymbol{x}) - \sum_{j=1}^{p} \boldsymbol{\mu}_j h_j(\boldsymbol{x}).$$

That is

$$L(\overline{\boldsymbol{x}}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \leq L(\overline{\boldsymbol{x}}, \overline{\boldsymbol{\lambda}}, \overline{\boldsymbol{\mu}} \leq L(\boldsymbol{x}, \overline{\boldsymbol{\lambda}}, \overline{\boldsymbol{\mu}})$$

for every $\boldsymbol{x} \in \mathbb{R}^n$, $\boldsymbol{\lambda} \in \mathbb{R}^m$, $\boldsymbol{\lambda} \geq 0$, *and* $\boldsymbol{\mu} \in \mathbb{R}^p$.

A one-sided relation between a saddle point of the Lagrangian and a solution of problem (P)[1] *is given by the following theorem:*

[1] Problem (P): $\min f(\boldsymbol{x})$ subject to $g_i(\boldsymbol{x}) \geq 0$, $i = 1, \ldots, m$, and $h_j(\boldsymbol{x}) = 0$, $j = 1, \ldots, p$.

The saddle point problem (S)

Theorem (Sufficient condition of optimality for (P)).

If $(\overline{\boldsymbol{x}}, \overline{\boldsymbol{\lambda}}, \overline{\boldsymbol{\mu}})$ is a solution of problem (S), then $\overline{\boldsymbol{x}}$ is a solution of problem (P).

Proof Suppose that $(\overline{\boldsymbol{x}}, \overline{\boldsymbol{\lambda}}, \overline{\boldsymbol{\mu}})$ is a solution of problem (S). Then, for all $\boldsymbol{x} \in \mathbb{R}^n$, $\boldsymbol{\lambda} \in \mathbb{R}^m$, $\boldsymbol{\lambda} \geq 0$, and $\boldsymbol{\mu} \in \mathbb{R}^p$:

$$f(\overline{\boldsymbol{x}}) - \sum_{i=1}^{m} \boldsymbol{\lambda}_i g_i(\overline{\boldsymbol{x}}) - \sum_{j=1}^{p} \boldsymbol{\mu}_j h_j(\overline{\boldsymbol{x}}) \leq f(\overline{\boldsymbol{x}}) - \sum_{i=1}^{m} \overline{\boldsymbol{\lambda}}_i g_i(\overline{\boldsymbol{x}}) - \sum_{j=1}^{p} \overline{\boldsymbol{\mu}}_j h_j(\overline{\boldsymbol{x}}) \leq$$

$$\leq f(\boldsymbol{x}) - \sum_{i=1}^{m} \overline{\boldsymbol{\lambda}}_i g_i(\boldsymbol{x}) - \sum_{j=1}^{p} \overline{\boldsymbol{\mu}}_j h_j(\boldsymbol{x}).$$

Rearranging the first inequality, we obtain

$$\sum_{i=1}^{m} (\overline{\boldsymbol{\lambda}}_i - \boldsymbol{\lambda}_i) g_i(\overline{\boldsymbol{x}}) + \sum_{j=1}^{p} (\overline{\boldsymbol{\mu}}_j - \boldsymbol{\mu}_j) h_j(\overline{\boldsymbol{x}}) \leq 0, \tag{41}$$

for all $\boldsymbol{\lambda} \in \mathbb{R}^m$, $\boldsymbol{\lambda} \geq 0$, and $\boldsymbol{\mu} \in \mathbb{R}^p$.

Proof of the Theorem (cont. 1)

Suppose now that that for a certain k $(1 \leq k \leq p)$ $h_k(\overline{\mathbf{x}}) > 0$.
Letting

$$\boldsymbol{\lambda}_i = \overline{\boldsymbol{\lambda}}_i, \ i = 1, ..., m,$$

$$\boldsymbol{\mu}_j = \overline{\boldsymbol{\mu}}_j, \ j = 1, ..., p, \ j \neq k,$$

$$\boldsymbol{\mu}_k = \overline{\boldsymbol{\mu}}_k - 1,$$

we get a contradiction to (41)

If $h_k(\overline{\mathbf{x}}) < 0$ for some k, we can choose an appropiate $\boldsymbol{\mu}$ that results in a similar contradiction. Thus $h_j(\overline{\mathbf{x}}) = 0, j = 1, ..., p$.

Now set $\overline{\boldsymbol{\mu}} = \boldsymbol{\mu}$ and let $\boldsymbol{\lambda}_1 = \overline{\boldsymbol{\lambda}}_1 + 1$, $\boldsymbol{\lambda}_i = \overline{\boldsymbol{\lambda}}_i$, $i = 2, ..., m$, then we obtain $g_1(\overline{\mathbf{x}}) \geq 0$.

If $\boldsymbol{\lambda}_2 = \overline{\boldsymbol{\lambda}}_2 + 1$, $\boldsymbol{\lambda}_i = \overline{\boldsymbol{\lambda}}_i$, $i = 1, 3, ..., m$, we obtain $g_2(\overline{\mathbf{x}}) \geq 0$.

Repeating this process for all i we obtain $g_i(\overline{\mathbf{x}}) \geq 0, \ i = 1, ..., m$.

As a consequence, $\overline{\mathbf{x}}$ is a feasible point for problem (P).

Proof of the Theorem (cont. 2)

Next let $\boldsymbol{\lambda} = 0$. Then, by the first inequality of (41) we have

$$0 \leq -\sum_{i=1}^{m} \overline{\boldsymbol{\lambda}}_i g_i(\overline{\boldsymbol{x}}).$$

But $\overline{\boldsymbol{\lambda}}_i \geq 0$ and $g_i(\overline{\boldsymbol{x}}) \geq 0$ for $i = 1, \ldots, m$, therefore

$$\sum_{i=1}^{m} \overline{\boldsymbol{\lambda}}_i g_i(\overline{\boldsymbol{x}}) = 0,$$

and so $\overline{\boldsymbol{\lambda}}_i g_i(\overline{\boldsymbol{x}}) = 0$ for all i.
Consider the second inequality of (41). From the preceding arguments we get

$$f(\overline{\boldsymbol{x}}) \leq f(\boldsymbol{x}) - \sum_{i=1}^{m} \overline{\boldsymbol{\lambda}}_i g_i(\boldsymbol{x}) - \sum_{j=1}^{p} \overline{\boldsymbol{\mu}}_j h_j(\boldsymbol{x}).$$

If $\boldsymbol{x}$ is feasible for (P), then $g_i(\boldsymbol{x}) \geq 0$, $h_j(\boldsymbol{x}) = 0$, thus

$$f(\overline{\boldsymbol{x}}) \leq f(\boldsymbol{x}),$$

and $\overline{\boldsymbol{x}}$ is a solution of (P).

Example

Consider the following problem:

$$\min f(x) = x, \text{ such that } -(x^2) \geq 0, \quad x \in \mathbb{R},$$

whose optimal solution is $x^* = 0$.

The corresponding saddle point problem of the Lagrangian is to find $\lambda^* \geq 0$ such that

$$x^* + \lambda(x^*)^2 \leq x^* + \lambda^*(x^*)^2 \leq x + \lambda^* x^2,$$

for all $x \in \mathbb{R}$, or, equivalently

$$0 \leq x + \lambda^* x^2.$$

Clearly, λ^* cannot vanish, but for any $\lambda^* > 0$ we can choose $x > -1/\lambda^*$, and (41) will not hold. Thus, there exist no λ^* such that (x^*, λ^*) will be a saddel point.