

Optimization

Màster de Fonaments de Ciència de Dades

Lecture II. Unconstrained and constrained optimization with equalities. Optimality conditions

Main issues in Optimization

1. **Characterization** of extrema (maxima/minima)
 - ▶ Necessary conditions
 - ▶ Sufficient conditions
 - ▶ Lagrange multiplier theory
 - ▶ The Karush-Kuhn-Tucker theory
2. Iterative **algorithms** for the computation of the extrema
 - ▶ Iterative descent
 - ▶ Approximation methods
 - ▶ Dual and primal-dual methods

Characterization of minima. Local and global minima

Let a function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$. A point $\mathbf{x}^* \in \mathbb{R}^n$ is a:

- ▶ **LOCAL minimum** of f if:

there is an $\epsilon > 0$ such that $f(\mathbf{x}^*) \leq f(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^n$ when $\|\mathbf{x} - \mathbf{x}^*\| \leq \epsilon$.

- ▶ **STRICT LOCAL minimum** of f if:

there is an $\epsilon > 0$ such that $f(\mathbf{x}^*) < f(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{x}^*\}$ when $\|\mathbf{x} - \mathbf{x}^*\| \leq \epsilon$.

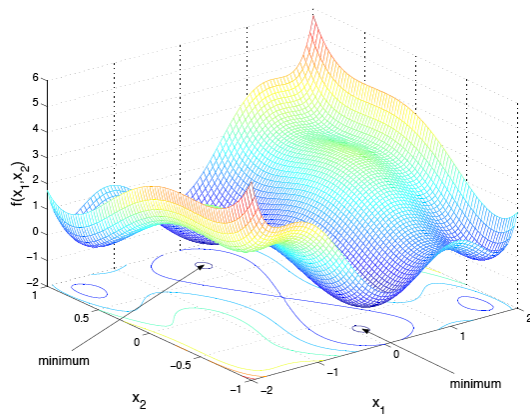
- ▶ **GLOBAL minimum** of f if:

$f(\mathbf{x}^*) \leq f(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^n$.

- ▶ **STRICT GLOBAL minimum** of f if:

$f(\mathbf{x}^*) < f(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{x}^*\}$.

Local and global minima



The function $f(x_1, x_2) = x_1^2(4 - 2.1x_1^2 + \frac{1}{3}x_1^4) + x_1x_2 + x_2^2(-4 + 4x_2^2)$ has two global minima, $(0.089, -0.717)$ and $(-0.0898, 0.717)$ and four local minima.

Derivatives and notation

Recall that if $x \in \mathcal{C} \subset \mathbb{R}^n$ is a point where the real function

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

is **differentiable** then:

- ▶ We define the **gradient** of f at x as the vector $\nabla f(x)$ given by:

$$\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \dots, \frac{\partial f(x)}{\partial x_n} \right)^T.$$

- ▶ If f is twice continuously differentiable at x we define the **Hessian** matrix of f at x as the $n \times n$ symmetric matrix $\nabla^2 f(x)$ given by:

$$\nabla^2 f(x) = \left(\frac{\partial^2 f(x)}{\partial x_i \partial x_j} \right), \quad i, j = 1, \dots, n.$$

- ▶ The **directional derivative** $D_u f(x)$ of the function f , at point $x \in \mathcal{C}$, in the direction $u \in \mathbb{R}^n$ ($\|u\| = 1$) is defined as

$$D_u f(x) = \lim_{\lambda \rightarrow 0} \frac{f(x + \lambda u) - f(x)}{\lambda} = \nabla f(x) \cdot u = \nabla f(x)^T u$$

Necessary and sufficient conditions for extrema

One must be careful with necessary and sufficient conditions.

Theorem : Let $c \in (a; b)$ and f be a real valued **continuous** function at c . If for some $\delta > 0$, f is increasing on $(c - \delta, c)$ and decreasing on $(c, c + \delta)$, then f has a local maximum at c .

Proof: Choose any x_1 and x such that $c - \delta < x_1 < x < c$. Then $f(x_1) \leq f(x)$ and by the continuity of f at c we have:

$$f(x_1) \leq \lim_{x \rightarrow c^-} f(x) = f(c).$$

Similarly, if $c < x < x_2 < c + \delta$, then

$$f(x_2) \leq \lim_{x \rightarrow c^+} f(x) = f(c).$$

This proves the result.

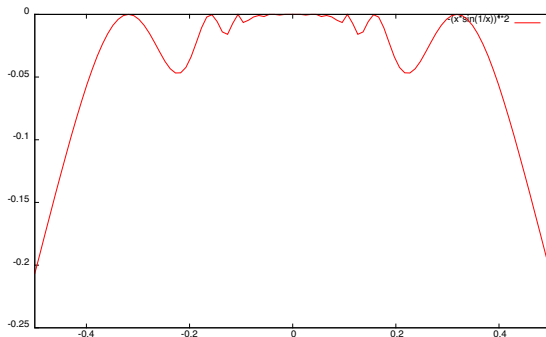
Remark: This Theorem gives a sufficient condition of maximum but not a necessary one.

Necessary and sufficient conditions for extrema

Remark: The converse of the above Theorem is not true, i.e.: If f is continuous at c and f has a local maximum at c , then f need not be increasing on $(c - \delta, c)$ or decreasing on $(c, c + \delta)$ **for any** $\delta > 0$.

Take, for example, $c = 0$ and

$$f(x) = -\left(x \sin \frac{1}{x}\right)^2 \quad \text{if } x \neq 0, \quad \text{and} \quad f(0) = 0.$$



Necessary and sufficient conditions for extrema

Theorem (Necessary condition)

Let $f : \mathcal{C} \rightarrow \mathbb{R}$ and \mathbf{x}^* an interior point of $\mathcal{C} \subset \mathbb{R}^n$ at which f has a local minimum (or a local maximum). If f is **differentiable** at \mathbf{x}^* then:

$$\nabla f(\mathbf{x}^*) = 0.$$

Proof. As \mathbf{x}^* is a local minimum, one has:

$$f(\mathbf{x}^*) \leq f(\mathbf{x}^* + \lambda \mathbf{s}) \quad \text{for all } \mathbf{s} \in \mathbb{R}^n, \|\mathbf{s}\| = 1 \quad \text{and } \lambda \in \mathbb{R} \text{ small enough.}$$

Fix \mathbf{s} and define $F(\lambda) = f(\mathbf{x}^* + \lambda \mathbf{s})$, which is continuous at $\lambda = 0$. Then the above inequality becomes

$$F(0) \leq F(\lambda), \quad \forall |\lambda| < \delta.$$

From the Mean Value Theorem, we have

$$F(\lambda) = F(0) + F'(\theta\lambda)\lambda,$$

where $\theta \in [0, 1]$.

Now...

Necessary and sufficient conditions for extrema

Now...

– If $F'(0) > 0$, then, by the continuity assumptions, there exists an $\epsilon > 0$ such that:

$$F'(\theta\lambda) > 0, \quad \forall \theta \in [0, 1], \text{ and } \forall \lambda \text{ s.t. } |\lambda| < \epsilon.$$

Hence, we can find a $\lambda < 0$, such that $|\lambda| < \delta$, and, since:

$$F(\lambda) = F(0) + F'(\theta\lambda)\lambda \Rightarrow F(0) > F(\lambda), \quad \text{contradiction!!}$$

– Assuming $F'(0) < 0$ would lead to a similar contradiction (taking $\lambda > 0$).

Thus

$$F'(0) = \nabla f(\mathbf{x}^*)^T \mathbf{s} = 0.$$

Since \mathbf{s} is an arbitrary nonzero vector ($\|\mathbf{s}\| = 1$), we must have:

$$\nabla f(\mathbf{x}^*)^T = 0.$$

Necessary and sufficient conditions for extrema

Theorem (Sufficient conditions)

Let \mathbf{x}^* be an interior point of \mathcal{C} at which f is **twice continuously differentiable**. If

$$\nabla f(\mathbf{x}^*) = \mathbf{0}, \quad \mathbf{z}^T \nabla^2 f(\mathbf{x}^*) \mathbf{z} > 0, \quad \forall \mathbf{z} \neq \mathbf{0},$$

then f has a local minimum at \mathbf{x}^* .

If

$$\nabla f(\mathbf{x}^*) = \mathbf{0}, \quad \mathbf{z}^T \nabla^2 f(\mathbf{x}^*) \mathbf{z} < 0, \quad \forall \mathbf{z} \neq \mathbf{0},$$

then f has a local maximum at \mathbf{x}^* .

Moreover, the extrema are **strict local extrema**.

Proof. Use the Taylor expansion of f around \mathbf{x}^* .

Remark. Note that the converses of the above assertions are not true. For instance, $f(x) = -x^4$, has a maximum at $x^* = 0$, is twice differentiable at 0 but $\nabla f(0) = f''(0) = 0$, which is not strictly less than 0.

Remark. The condition $\mathbf{z}^T \nabla^2 f(\mathbf{x}^*) \mathbf{z} > 0, \quad \forall \mathbf{z} \neq \mathbf{0}$ means that $\nabla^2 f(\mathbf{x}^*)$ is positive definite.

Necessary and sufficient conditions for extrema

Example. Let

$$f(x) = x^{2p}, \quad x \in \mathbb{R}, \quad p \in \mathbb{Z}_+,$$

and let \mathcal{C} be the whole real line.

- ▶ The gradient of f is

$$\nabla f(x) = f'(x) = 2px^{2p-1}.$$

Clearly $\nabla f(0) = 0$, that is $x = 0$ **satisfies the necessary condition** for a minimum or a maximum.

- ▶ The Hessian of f is

$$\nabla^2 f(x) = f''(x) = (2p-1)2px^{2p-2}.$$

For $p = 1$, $\nabla^2 f(0) = 2 > 0$, that is, the **sufficient conditions for a strict local minimum are satisfied**.

- ▶ If we take $p > 1$, then $\nabla^2 f(0) = 0$ and **the sufficient conditions for a local minimum are not satisfied**, yet f has a minimum at the origin.

By taking any neighborhood of the origin, it can be verified that all the conditions for a local minimum given in the next Theorem are satisfied for this example.

Necessary and sufficient conditions for extrema

Theorem

Let \mathbf{x}^* be an interior point of \mathcal{C} and assume that f is **twice continuously differentiable** on \mathcal{C} , then:

(a) **Necessary** conditions for a local minimum of f at \mathbf{x}^* are:

$$\nabla f(\mathbf{x}^*) = 0, \quad \mathbf{z}^T \nabla^2 f(\mathbf{x}^*) \mathbf{z} \geq 0, \quad \forall \mathbf{z} \in \mathbb{R}^n.$$

(b) **Sufficient** conditions for a local minimum are:

$$\nabla f(\mathbf{x}^*) = 0,$$

and that **for every \mathbf{x}** in some neighborhood $N_\epsilon(\mathbf{x}^*)$ and **for every $\mathbf{z} \in \mathbb{R}^n$** , we have:

$$\mathbf{z}^T \nabla^2 f(\mathbf{x}) \mathbf{z} \geq 0.$$

(c) If the sense of the inequalities is reversed, then the theorem applies to a local maximum.

Proof of the Theorem

Proof

(a) The first order condition for local minimum

$$\mathbf{x}^* \text{ minimum} \Rightarrow \nabla f(\mathbf{x}^*) = 0,$$

has already been proved (pages 8 and 9).

Turning to the second-order conditions, we have by Taylor's theorem applied to the function $F(\theta) = f(\mathbf{x}^* + \theta \mathbf{s})$, with $\|\mathbf{s}\| = 1$ fixed,

$$F(\theta) = F(0) + \nabla F(0)\theta + \frac{1}{2}\nabla^2 F(\lambda\theta)\theta^2, \quad \lambda \in (0, 1).$$

If $\nabla^2 F(0) < 0$, then, by continuity, there exists $\epsilon' > 0$ such that $\nabla^2 F(\lambda\theta) < 0$ for $\lambda \in (0, 1)$ and $|\theta| < \epsilon'$. Since $\nabla F(0) = 0$, this inequality implies that for such a θ :

$$F(\theta) < F(0),$$

which is a contradiction. Consequently

$$\nabla^2 F(0) = \mathbf{s}^T \nabla^2 f(\mathbf{x}^*) \mathbf{s} \geq 0.$$

Since this inequality holds for all unitary vector \mathbf{s} , it must hold for all vector \mathbf{z} .

Proof of the Theorem (cont.)

- (b) Assume that $\nabla f(\mathbf{x}^*) = 0$ and that $\mathbf{z}^T \nabla^2 f(\mathbf{x}^*) \mathbf{z} \geq 0$ for all $\mathbf{x} \in N_\delta(\mathbf{x}^*)$ and all $\mathbf{z} \in \mathbb{R}^n$, but that \mathbf{x}^* is not a local minimum.

Then there exists a $\mathbf{w} \in N_\delta(\mathbf{x}^*)$ such that $f(\mathbf{w}) < f(\mathbf{x}^*)$.

Write $\mathbf{w} = \mathbf{x}^* + \theta \mathbf{y}$, with $\|\mathbf{y}\| = 1$ and $\theta > 0$. By Taylor's theorem:

$$f(\mathbf{w}) = f(\mathbf{x}^*) + \theta \nabla f(\mathbf{x}^*)^T \mathbf{y} + \frac{1}{2} \theta^2 \mathbf{y}^T \nabla^2 f(\mathbf{x}^* + \lambda \theta \mathbf{y}) \mathbf{y},$$

with $\lambda \in (0, 1)$. Our assumptions lead then to

$$\mathbf{y}^T \nabla^2 f(\mathbf{x}^* + \lambda \theta \mathbf{y}) \mathbf{y} < 0,$$

contradicting the hypothesis, since $\mathbf{x}^* + \lambda \theta \mathbf{y} \in N_\delta(\mathbf{x}^*)$.

Necessary and sufficient conditions for extrema

Example

Let

$$f(x, y) = \frac{1}{2}x^2 + xy + 2y^2 - 4x - 4y - y^3.$$

Then

$$\nabla f(x, y) = (x + y - 4, x + 4y - 4 - 3y^2)^T, \quad H(x, y) = \begin{pmatrix} 1 & 1 \\ 1 & 4 - 6y \end{pmatrix}.$$

$\nabla f(x, y) = 0$ has exactly two solutions, $\mathbf{x}_1 = (4, 0)$, $\mathbf{x}_2 = (3, 1)$, and

$$H(\mathbf{x}_1) = \begin{pmatrix} 1 & 1 \\ 1 & 4 \end{pmatrix}, \quad H(\mathbf{x}_2) = \begin{pmatrix} 1 & 1 \\ 1 & -2 \end{pmatrix}.$$

$$(x \ y) \begin{pmatrix} 1 & 1 \\ 1 & 4 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = x^2 + 2xy + 4y^2 = (x+y)^2 + 3y^2 > 0 \quad \text{if } (x \ y) \neq (0, 0).$$

$$(x \ y) \begin{pmatrix} 1 & 1 \\ 1 & -2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = x^2 + 2xy - 2y^2 = (x+y)^2 - 3y^2.$$

So $H(\mathbf{x}_1)$ is positive definite and $H(\mathbf{x}_2)$ is indefinite, therefore, the only extrema is the local minimum is \mathbf{x}_1 .

Convexity

Convexity notions play an important role in nonlinear programming. Some reasons for that are:

1. Convex optimization **includes least-squares and linear programming problems**, which can be solved numerically very efficiently.
2. When the cost function f is convex, every **local maximum/minimum is also global**.
3. We will see that if $f(\mathbf{x}) = \mathbf{x}^T \mathbf{Q} \mathbf{x} + \mathbf{c}^T \mathbf{x} + d$ is such that \mathbf{Q} is **positive semidefinite**, then f is convex.
4. The (first order) **necessary condition** $\nabla f(\mathbf{x}^*) = 0$ is **also sufficient** for global optimality if f is convex.
5. The behavior of convex functions allows for very **fast algorithms** to optimize them.
6. Many optimization problems admit a **convex (re)formulation**.

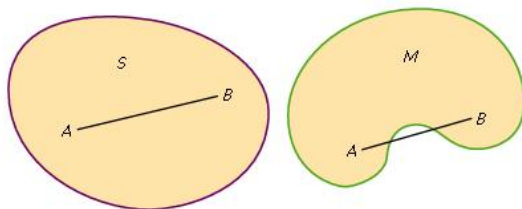
Convex sets and convex functions

- ▶ Let two points $x_1, x_2 \in \mathbb{R}$, and $0 \leq \lambda \leq 1$ be given. Then, the point

$$x = \lambda x_1 + (1 - \lambda)x_2,$$

is a **convex combination** of the two points x_1, x_2 .

- ▶ The **set** $\mathcal{C} \subset \mathbb{R}^n$ is called **convex**, if all convex combinations of any two points $x_1, x_2 \in \mathcal{C}$ are again in \mathcal{C} .

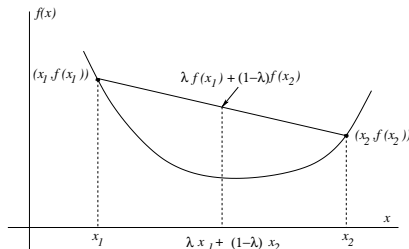


©1998 Encyclopaedia Britannica, Inc.

Convex sets and convex functions

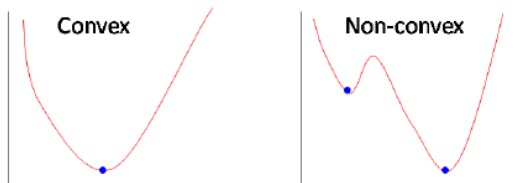
- ▶ A function $f : \mathcal{C} \rightarrow \mathbb{R}$ defined on a convex set \mathcal{C} is called **convex** if for all $x_1, x_2 \in \mathcal{C}$ and $0 \leq \lambda \leq 1$ one has:

$$f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2).$$



- ▶ For a convex function, the linear interpolation $\lambda f(x_1) + (1 - \lambda)f(x_2)$ overestimates the function value $f(\lambda x_1 + (1 - \lambda)x_2)$.

Convex sets and convex functions



- ▶ A function $f : \mathcal{C} \rightarrow \mathbb{R}$ defined on a convex set \mathcal{C} is called **strictly convex** if for all $x_1, x_2 \in \mathcal{C}$ with $x_1 \neq x_2$ and $0 < \lambda < 1$ one has.

$$f(\lambda x_1 + (1 - \lambda)x_2) < \lambda f(x_1) + (1 - \lambda)f(x_2).$$

- ▶ A function $f : \mathcal{C} \rightarrow \mathbb{R}$ defined on a convex set \mathcal{C} is called **concave** if $-f$ is convex.
- ▶ **Remark.** Recall that **the domain of a convex function must be a convex set.**
- ▶ **Remark.** Later we will give a more general definition of convex function.

Examples of convex functions

Proposition

- a) A **linear function**, $f(\lambda \mathbf{x} + \mu \mathbf{y}) = \lambda f(\mathbf{x}) + \mu f(\mathbf{y})$, is convex.
- b) Any **vector norm**, $f(\mathbf{x}) = \|\mathbf{x}\|$, is a convex function.
- c) The **weighted sum of convex functions** $(\sum \alpha_i f_i(\mathbf{x}))$ with positive weights $(\alpha_i > 0)$ is convex.
- d) If I is an index set, $\mathcal{C} \subset \mathbb{R}^n$ is a convex set and $f_i : \mathcal{C} \rightarrow \mathbb{R}$ are convex functions for each $i \in I$, then the function:

$$\begin{aligned} h : \mathcal{C} &\longrightarrow (-\infty, \infty] \\ \mathbf{x} &\longrightarrow \sup_{i \in I} f_i(\mathbf{x}) \end{aligned}$$

is also convex (recall that the $\sup_{i \in I} f_i(\mathbf{x})$ is the least upper bound of $\{f_i(\mathbf{x}); i \in I\}$).

Proof. a) and c) are consequences of the definition of convexity.

- b) Let $\|\cdot\|$ be a norm. Then, for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and any $\lambda \in [0, 1]$ we have:

$$\|\lambda \mathbf{x} + (1 - \lambda) \mathbf{y}\| \leq \|\lambda \mathbf{x}\| + \|(1 - \lambda) \mathbf{y}\| = \lambda \|\mathbf{x}\| + (1 - \lambda) \|\mathbf{y}\|.$$

- d) For every $i \in I$ we have

$$f_i(\lambda \mathbf{x} + (1 - \lambda) \mathbf{y}) \leq \lambda f_i(\mathbf{x}) + (1 - \lambda) f_i(\mathbf{y}) \leq \lambda h(\mathbf{x}) + (1 - \lambda) h(\mathbf{y}).$$

Taking the supremum over all $i \in I$ we conclude:

$$h(\lambda \mathbf{x} + (1 - \lambda) \mathbf{y}) \leq \lambda h(\mathbf{x}) + (1 - \lambda) h(\mathbf{y}).$$

Necessary and sufficient conditions for extrema for convex functions

Theorem (Necessary condition in the convex case)

Let $f : \mathcal{C} \rightarrow \mathbb{R}$ be a convex function over the convex set $\mathcal{C} \subset \mathbb{R}^n$.

- a) A **local minimum** of f over \mathcal{C} **is also a global minimum** over \mathcal{C} .
- b) If, in addition, f is **strictly convex**, then there exists **at most one global minimum** of f .
- c) If f is convex, the set \mathcal{C} is open, and f is differentiable at \mathbf{x}^* , then $\nabla f(\mathbf{x}^*) = 0$ **is a necessary (see Theorem pg. 8) and sufficient condition** for $\mathbf{x}^* \in \mathcal{C}$ to be a **global minimum** of f over \mathcal{C} .

Proof

- a) Suppose that \mathbf{x} is a local minimum of f but not a global minimum. Then there exists some $\mathbf{y} \neq \mathbf{x}$ such that $f(\mathbf{y}) < f(\mathbf{x})$. Since f is convex:

$$f(\alpha \mathbf{x} + (1-\alpha)\mathbf{y}) \leq \alpha f(\mathbf{x}) + (1-\alpha)f(\mathbf{y}) < \alpha f(\mathbf{x}) + (1-\alpha)f(\mathbf{x}) = f(\mathbf{x}), \quad \forall \alpha \in [0, 1).$$

This contradicts the assumption that \mathbf{x} is a local minimum.

- b) Suppose that two distinct global minima \mathbf{x} and \mathbf{y} exist ($f(\mathbf{x}) = f(\mathbf{y})$). Then $(\mathbf{x} + \mathbf{y})/2 \in \mathcal{C}$, since \mathcal{C} is convex, and also:

$$f((1/2)\mathbf{x} + (1/2)\mathbf{y}) < f(\mathbf{x}), \quad f((1/2)\mathbf{x} + (1/2)\mathbf{y}) < f(\mathbf{y}),$$

and since \mathbf{x} and \mathbf{y} are global minima, we obtain a contradiction.

Proof (cont.)

- c) By the convexity of \mathcal{C} , and using the convexity of f , we have that for all $\mathbf{x} \in \mathcal{C}$ and $\alpha \in [0, 1]$:

$$\begin{aligned}\alpha \mathbf{x} + (1 - \alpha) \mathbf{x}^* &= \mathbf{x}^* + \alpha(\mathbf{x} - \mathbf{x}^*) \quad \Rightarrow \\ \Rightarrow f(\mathbf{x}^* + \alpha(\mathbf{x} - \mathbf{x}^*)) &= f(\alpha \mathbf{x} + (1 - \alpha) \mathbf{x}^*) \leq \alpha f(\mathbf{x}) + (1 - \alpha) f(\mathbf{x}^*), \\ f(\mathbf{x}^* + \alpha(\mathbf{x} - \mathbf{x}^*)) - f(\mathbf{x}^*) &\leq \alpha f(\mathbf{x}) + (1 - \alpha) f(\mathbf{x}^*) - f(\mathbf{x}^*) = \alpha(f(\mathbf{x}) - f(\mathbf{x}^*)).\end{aligned}$$

It follows that:

$$\frac{f(\mathbf{x}^* + \alpha(\mathbf{x} - \mathbf{x}^*)) - f(\mathbf{x}^*)}{\alpha} \leq f(\mathbf{x}) - f(\mathbf{x}^*).$$

Furthermore

$$\lim_{\alpha \rightarrow 0} \frac{f(\mathbf{x}^* + \alpha(\mathbf{x} - \mathbf{x}^*)) - f(\mathbf{x}^*)}{\alpha} = \nabla f(\mathbf{x}^*)^T (\mathbf{x} - \mathbf{x}^*).$$

Taking the limit as $\alpha \rightarrow 0$ in the last inequality, we get

$$\nabla f(\mathbf{x}^*)^T (\mathbf{x} - \mathbf{x}^*) \leq f(\mathbf{x}) - f(\mathbf{x}^*), \quad \forall \mathbf{x} \in \mathcal{C}.$$

If $\nabla f(\mathbf{x}^*) = 0$, we obtain

$$0 \leq f(\mathbf{x}) - f(\mathbf{x}^*) \quad \Rightarrow \quad f(\mathbf{x}) \geq f(\mathbf{x}^*) \quad \forall \mathbf{x} \in \mathcal{C},$$

so \mathbf{x}^* is a global minimum.

Remark

The inequality of the above page:

$$\nabla f(\mathbf{x}^*)^T (\mathbf{x} - \mathbf{x}^*) \leq f(\mathbf{x}) - f(\mathbf{x}^*), \quad \forall \mathbf{x} \in \mathcal{C},$$

can be written as:

$$f(\mathbf{x}) \geq f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)^T (\mathbf{x} - \mathbf{x}^*), \quad \forall \mathbf{x} \in \mathcal{C}.$$

In fact, we have proven the more general one:

$$\boxed{f(\mathbf{x}) \geq f(\mathbf{y}) + \nabla f(\mathbf{y})^T (\mathbf{x} - \mathbf{y}), \quad \forall \mathbf{x}, \mathbf{y} \in \mathcal{C},} \quad (1)$$

since we have not used the condition $\nabla f(\mathbf{x}^*) = 0$.

The inequality is, in fact, a consequence of the following **characterization of differentiable convex functions**.

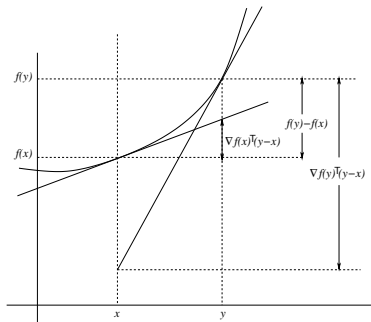
First characterization theorem of convex functions

Theorem

f is convex on \mathcal{C} if and only if for any two points $x, y \in \mathcal{C}$ one has

$$\nabla f(x)^T(y - x) \leq f(y) - f(x) \leq \nabla f(y)^T(y - x). \quad (2)$$

If the inequalities are strict whenever $x \neq y$, then f is strictly convex over \mathcal{C} .



Remarks. As it follows from the proof, the two inequalities (2) in the Theorem can be substituted by (1) ($f(x) \geq f(y) + \nabla f(y)^T(x - y)$), since one inequality is a consequence of the other.

The proof for the strictly convex case is identical to the convex case.

Proof of the characterization theorem

Proof.

Assume that f is convex. The second inequality directly follows from (1). Interchanging the roles of \mathbf{x} and \mathbf{y} in (1), one gets for all $\mathbf{x}, \mathbf{y} \in \mathcal{C}$ that

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x}), \quad \Rightarrow f(\mathbf{y}) - f(\mathbf{x}) \geq \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x}),$$

which is the first inequality in (2).

To proof the converse, suppose that (1) is true and we must proof that f is convex. We fix some $\mathbf{x}, \mathbf{y} \in \mathcal{C}$ and some $\alpha \in [0, 1]$. Let $\mathbf{z} = \alpha\mathbf{x} + (1 - \alpha)\mathbf{y}$. Using the inequality twice, we get

$$\begin{aligned} f(\mathbf{x}) &\geq f(\mathbf{z}) + \nabla f(\mathbf{z})^T(\mathbf{x} - \mathbf{z}), \\ f(\mathbf{y}) &\geq f(\mathbf{z}) + \nabla f(\mathbf{z})^T(\mathbf{y} - \mathbf{z}). \end{aligned}$$

Multiplying the first inequality by α , the second by $(1 - \alpha)$ and adding, we obtain:

$$\alpha f(\mathbf{x}) + (1 - \alpha)f(\mathbf{y}) \geq f(\mathbf{z}) + \nabla f(\mathbf{z})^T(\alpha\mathbf{x} + (1 - \alpha)\mathbf{y} - \mathbf{z}) = f(\mathbf{z}),$$

which proves that f is convex.

Applications

- ▶ Many elementary (and not so elementary) inequalities follow from the above Theorem.

Example: The well known inequality¹

$$e^x \geq 1 + x,$$

can also be proved by using the convexity of the function $f(x) = e^x$.

Since $\nabla f(x) = f'(x) = e^x$, taking $y = 0$, so $\nabla f(y) = f'(0) = 1$, and using inequality (2):

$$\nabla f(x)^T(y - x) \leq f(y) - f(x) \leq \nabla f(y)^T(y - x),$$

we get:

$$e^x(0 - x) \leq 1 - e^x \leq 0 - x \quad \Rightarrow \quad x \leq e^x - 1 \leq xe^x, \quad \forall x \in \mathbb{R},$$

or

$$e^x \geq 1 + x, \quad \text{and} \quad (1 - x)e^x \leq 1, \quad \forall x \in \mathbb{R}.$$

¹It can be easily be proved noting that $g(x) = e^x - (1 + x)$ has a minimum at $x = 0$ ▶

Exercise 3. To be delivered before 4-X-2021 as: Ex03-YourSurname.pdf

Proof, without using the above theorem, that for any $a \in \mathbb{R}$, $f(x) = e^{ax}$ is a convex function.

Characterization of convexity for twice differentiable functions

Theorem.

Let $\mathcal{C} \subset \mathbb{R}^n$ be a convex set, let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a twice continuously differentiable function over \mathcal{C} , and let Q be a real symmetric $n \times n$ matrix.

- a) If $\nabla^2 f(\mathbf{x})$ is positive semidefinite for all $\mathbf{x} \in \mathcal{C}$, then f is convex over \mathcal{C} .
- b) If $\nabla^2 f(\mathbf{x})$ is positive definite for all $\mathbf{x} \in \mathcal{C}$, then f is strictly convex over \mathcal{C} .
- c) If $\mathcal{C} = \mathbb{R}^n$ and f is convex, then $\nabla^2 f(\mathbf{x})$ is positive semidefinite for all $\mathbf{x} \in \mathcal{C}$.
- d) The quadratic function $f(\mathbf{x}) = \mathbf{x}^T Q \mathbf{x}$, where Q is a symmetric matrix, is convex if and only if Q is positive semidefinite. Furthermore, f is strictly convex if and only if Q is positive definite.

Proof.

- a) According to Taylor's formula, for all $\mathbf{x}, \mathbf{y} \in \mathcal{C}$

$$f(\mathbf{y}) = f(\mathbf{x}) + \nabla f(\mathbf{x})^T (\mathbf{y} - \mathbf{x}) + \frac{1}{2} (\mathbf{y} - \mathbf{x})^T \nabla^2 f(\mathbf{x} + \alpha(\mathbf{y} - \mathbf{x})) (\mathbf{y} - \mathbf{x}),$$

for some $\alpha \in [0, 1]$. Therefore, using the positive semi-definiteness of $\nabla^2 f(\mathbf{x})$, $(\mathbf{z}^T \nabla^2 f(\mathbf{x}) \mathbf{z} \geq 0, \forall \mathbf{z})$, we obtain

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T (\mathbf{y} - \mathbf{x}), \quad \forall \mathbf{x}, \mathbf{y} \in \mathcal{C},$$

from which we can conclude that f is convex.

Proof of the Theorem (cont.)

- b) Similar to the proof of part a).
- c) Suppose that $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is convex and that $\mathbf{x} \in \mathcal{C}$. For some small $\alpha > 0$ and any $\mathbf{y} \in \mathbb{R}^n$, we have that $\mathbf{x} + \alpha\mathbf{y} \in \mathcal{C}$. From Taylor's formula:

$$f(\mathbf{x} + \alpha\mathbf{y}) = f(\mathbf{x}) + \alpha \nabla f(\mathbf{x})^T \mathbf{y} + \frac{\alpha^2}{2} \mathbf{y}^T \nabla^2 f(\mathbf{x}) \mathbf{y} + o(\|\alpha\mathbf{y}\|^2).$$

Since f is convex, we know that for any \mathbf{a} and \mathbf{b} :

$$f(\mathbf{a}) \geq f(\mathbf{b}) + \nabla f(\mathbf{b})^T (\mathbf{a} - \mathbf{b}),$$

so, taking $\mathbf{a} = \mathbf{x} + \alpha\mathbf{y}$ and $\mathbf{b} = \mathbf{x}$, we get

$$f(\mathbf{x} + \alpha\mathbf{y}) \geq f(\mathbf{x}) + \alpha \nabla f(\mathbf{x})^T \mathbf{y}.$$

Therefore, we have that for any $\mathbf{y} \in \mathbb{R}^n$

$$\frac{\alpha^2}{2} \mathbf{y}^T \nabla^2 f(\mathbf{x}) \mathbf{y} + o(\|\alpha\mathbf{y}\|^2) \geq 0.$$

Dividing by $\alpha^2/2$ and taking $\alpha \rightarrow 0$, we get

$$\mathbf{y}^T \nabla^2 f(\mathbf{x}) \mathbf{y} \geq 0, \quad \forall \mathbf{y} \in \mathbb{R}^n.$$

Proof of the Theorem (cont.)

- d) If $f(\mathbf{x}) = \mathbf{x}^T Q \mathbf{x}$ then $\nabla^2 f(\mathbf{x}) = 2Q$. Hence, from a) and c) it follows that f is convex if and only if Q is positive semidefinite.

For the strict convexity, suppose that f is strictly convex, then, according to c), Q is positive semidefinite and it remains to show that Q is positive definite.

It can be shown that Q is positive definite if and only if all its eigenvalues are positive.

Assume that zero is an eigenvalue, then there exists some $\mathbf{x} \neq 0$ such that $Q\mathbf{x} = 0$. It follows that

$$0 = f(0) = f\left(\frac{1}{2}\mathbf{x} + \frac{1}{2}(-\mathbf{x})\right) = \frac{1}{2}f(\mathbf{x}) + \frac{1}{2}f(-\mathbf{x}) = 0,$$

which contradicts the strict convexity of f .

Optimization with equality constraints. Lagrange multiplier theory

- ▶ 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 only to the equality constraints

$$g_i(\mathbf{x}) = 0, \quad i = 1, \dots, m, \quad m < n, \quad (3)$$

where each of the g_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 (3).

- ▶ The first and most **intuitive method** of solution of such a problem involves the **elimination of m variables from the problem by using equations (3)**.
- ▶ The conditions for such an elimination are stated by the **Implicit Function Theorem**, that assumes **differentiability** of the functions g_i and that the $n \times m$ **Jacobian** matrix $(\partial g_i / \partial x_j)$ has **rank m** .
- ▶ **The actual solution of the unconstraint equations for m variables in terms of the remaining $n - m$ can often be a difficult, if not impossible, task.**

Optimization with equality constraints

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

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

Solution Suppose that the upper righthand corner of the rectangle is at the point (x, y) , then the area of the rectangle is $S = 4xy$. We have:

$$\frac{2x}{a^2} dx + \frac{2y}{b^2} dy = 0 \quad \Rightarrow \quad \frac{2x}{a^2} + \frac{2y}{b^2} \frac{dy}{dx} = 0 \quad \Rightarrow \quad \frac{dy}{dx} = -\frac{b^2 x}{a^2 y},$$

so

$$\frac{dS}{dx} = 4y + 4x \frac{dy}{dx} = 4y - \frac{4b^2 x^2}{a^2 y}, \quad \frac{dS}{dx} = 0 \quad \Rightarrow \quad y^2 = \frac{b^2 x^2}{a^2}.$$

Since, according to the equation of the ellipse

$$y^2 = b^2 - \frac{b^2 x^2}{a^2},$$

we get

$$y^2 = b^2 - y^2 \quad \Rightarrow \quad y = \frac{b}{\sqrt{2}} \quad \text{and} \quad x = \frac{a}{\sqrt{2}} \quad \Rightarrow \quad S_{\max} = 2ab.$$

Lagrange multipliers

Another method for finding the minimum, also based on the idea of **transforming a constrained problem into an unconstrained one**, was proposed by Lagrange.

Before introducing this method, we present the following result:

Theorem

Let f and g_i , $i = 1, \dots, m$, be real-valued functions on $C \subset \mathbb{R}^n$ ($m < n$) and continuously differentiable on a neighborhood $N_\epsilon(\mathbf{x}^) \subset C$. Suppose that \mathbf{x}^* is a local minimum (or maximum) of f for all points $\mathbf{x} \in N_\epsilon(\mathbf{x}^*)$ that also satisfy:*

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

Assume also that the Jacobian matrix $(\partial g_i / \partial x_j)$ at \mathbf{x}^ has rank m . Under these hypotheses, there exist real numbers λ_i^* such that*

$$\nabla f(\mathbf{x}^*) = \sum_{i=1}^m \lambda_i^* \nabla g_i(\mathbf{x}^*).$$

Before the proof. Example

Consider the problem

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

subject to the constraint

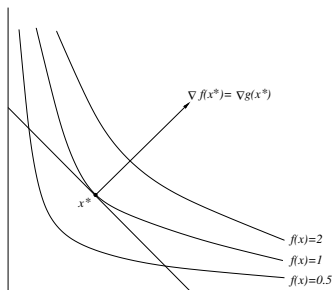
$$g(x, y) = x + y - 2 = 0.$$

The solution of this system problem is:

$$x^* = y^* = 1.$$

In this case:

$$\nabla f(x^*) = \left(\begin{array}{c} y \\ x \end{array} \right)_{(x,y)=(1,1)} = \left(\begin{array}{c} 1 \\ 1 \end{array} \right) = \nabla g(x^*) = \left(\begin{array}{c} 1 \\ 1 \end{array} \right)_{(x,y)=(1,1)}.$$



Proof of the Theorem

Proof. By suitable rearrangement and relabeling of rows, we can always assume that the $m \times m$ matrix formed by taking the first m rows of the Jacobian $(\partial g_i(\mathbf{x}^*)/\partial x_j)$, is nonsingular since it has rank m .

What we want to proof is that there exist $\lambda_1^*, \dots, \lambda_m^*$ such that:

$$\nabla f(\mathbf{x}^*) = \lambda_1^* \nabla g_1(\mathbf{x}^*) + \lambda_2^* \nabla g_2(\mathbf{x}^*) + \dots + \lambda_m^* \nabla g_m(\mathbf{x}^*),$$

that can also be written as:

$$\begin{pmatrix} \frac{\partial f(\mathbf{x}^*)}{\partial x_1} \\ \vdots \\ \frac{\partial f(\mathbf{x}^*)}{\partial x_m} \\ \vdots \\ \frac{\partial f(\mathbf{x}^*)}{\partial x_n} \end{pmatrix} = \begin{pmatrix} \frac{\partial g_1(\mathbf{x}^*)}{\partial x_1} & \dots & \frac{\partial g_m(\mathbf{x}^*)}{\partial x_1} \\ \vdots & & \vdots \\ \frac{\partial g_1(\mathbf{x}^*)}{\partial x_m} & \dots & \frac{\partial g_m(\mathbf{x}^*)}{\partial x_m} \\ \vdots & & \vdots \\ \frac{\partial g_1(\mathbf{x}^*)}{\partial x_n} & \dots & \frac{\partial g_m(\mathbf{x}^*)}{\partial x_n} \end{pmatrix} \begin{pmatrix} \lambda_1^* \\ \vdots \\ \lambda_m^* \end{pmatrix}.$$

We will first proof that there exist $\lambda_1^*, \dots, \lambda_m^*$ such that:

$$\begin{pmatrix} \frac{\partial f(\mathbf{x}^*)}{\partial x_1} \\ \vdots \\ \frac{\partial f(\mathbf{x}^*)}{\partial x_m} \end{pmatrix} = \begin{pmatrix} \frac{\partial g_1(\mathbf{x}^*)}{\partial x_1} & \dots & \frac{\partial g_m(\mathbf{x}^*)}{\partial x_1} \\ \vdots & & \vdots \\ \frac{\partial g_1(\mathbf{x}^*)}{\partial x_m} & \dots & \frac{\partial g_m(\mathbf{x}^*)}{\partial x_m} \end{pmatrix} \begin{pmatrix} \lambda_1^* \\ \vdots \\ \lambda_m^* \end{pmatrix}.$$

Proof of the Theorem (cont. 1)

Since the matrix of the above linear system is non-singular, the set of linear equations

$$\sum_{i=1}^m \frac{\partial g_i(\mathbf{x}^*)}{\partial x_j} \lambda_i = \frac{\partial f(\mathbf{x}^*)}{\partial x_j}, \quad j = 1, \dots, m, \quad (4)$$

has a unique solution: λ_i^* , $i = 1, \dots, m$. In this way we have seen that **the first m components of the gradients verify the equality** that we want to proof.

Let us see that **the remaining $n - m$ components also fulfil the same equality.**

Let $\hat{\mathbf{x}} = (x_{m+1}, \dots, x_n)$, then, applying the Implicit Function Theorem to the equations $g_i(\mathbf{x}^*) = 0$, it follows that there exist real functions

$$h_j(\hat{\mathbf{x}}) = h_j(x_{m+1}, \dots, x_n), \quad j = 1, \dots, m,$$

defined in an open set $\hat{D} \subset \mathbb{R}^{n-m}$ containing \mathbf{x}^* such that

$$x_j^* = h_j(\hat{\mathbf{x}}^*) = h_j(x_{m+1}^*, \dots, x_n^*), \quad j = 1, \dots, m, \quad (5)$$

$$f(\mathbf{x}^*) = f(h_1(\hat{\mathbf{x}}^*), \dots, h_m(\hat{\mathbf{x}}^*), x_{m+1}^*, \dots, x_n^*). \quad (6)$$

Using the same Theorem, we have also that for $j = m+1, \dots, n$

$$\sum_{k=1}^m \frac{\partial g_i(\mathbf{x}^*)}{\partial x_k} \frac{\partial h_k(\hat{\mathbf{x}}^*)}{\partial x_j} + \frac{\partial g_i(\mathbf{x}^*)}{\partial x_j} = 0, \quad i = 1, \dots, m. \quad (7)$$

Proof of the Theorem (cont. 2)

If \mathbf{x}^* is a minima of f its first partial derivatives with respect to x_{m+1}, \dots, x_n must vanish at \mathbf{x}^* . Thus

$$\frac{\partial f(\mathbf{x}^*)}{\partial x_j} = \sum_{k=1}^m \frac{\partial f(\mathbf{x}^*)}{\partial x_k} \frac{\partial h_k(\hat{\mathbf{x}}^*)}{\partial x_j} + \frac{\partial f(\mathbf{x}^*)}{\partial x_j} = 0, \quad j = m+1, \dots, n. \quad (8)$$

Multiplying each of the equations in (7) by λ_i^* and adding up, we get:

$$\sum_{i=1}^m \left(\sum_{k=1}^m \lambda_i^* \frac{\partial g_i(\mathbf{x}^*)}{\partial x_k} \frac{\partial h_k(\hat{\mathbf{x}}^*)}{\partial x_j} + \lambda_i^* \frac{\partial g_i(\mathbf{x}^*)}{\partial x_j} \right) = 0, \quad j = m+1, \dots, n.$$

Subtracting this equality from (8) we get:

$$\sum_{k=1}^m \left[\frac{\partial f(\mathbf{x}^*)}{\partial x_k} - \sum_{i=1}^m \lambda_i^* \frac{\partial g_i(\mathbf{x}^*)}{\partial x_j} \right] \frac{\partial h_k(\hat{\mathbf{x}}^*)}{\partial x_j} + \frac{\partial f(\mathbf{x}^*)}{\partial x_j} - \sum_{i=1}^m \lambda_i^* \frac{\partial g_i(\mathbf{x}^*)}{\partial x_j} = 0,$$

for $j = m+1, \dots, n$. Since, due to (4), the expression in the brackets is zero, we get the desired result

$$\frac{\partial f(\mathbf{x}^*)}{\partial x_j} - \sum_{i=1}^m \lambda_i^* \frac{\partial g_i(\mathbf{x}^*)}{\partial x_j} = 0, \quad j = m+1, \dots, n.$$

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

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

Theorem (Necessary conditions)

Suppose that f and g_i , $i = 1, \dots, m$, are real-valued functions that satisfy the hypotheses of the preceding Theorem, this is:

- ▶ $f : \mathcal{C} \subset \mathbb{R}^n \longrightarrow \mathbb{R}$, and $g_i : \mathcal{C} \subset \mathbb{R}^n \longrightarrow \mathbb{R}$, $i = 1, \dots, m$.
- ▶ They are all continuously differentiable on a neighborhood $N_\epsilon(\mathbf{x}^*) \subset \mathcal{C} \subset \mathbb{R}^n$.
- ▶ The Jacobian matrix $(\partial g_i(\mathbf{x}^*)/\partial x_j)$ has rank m .
- ▶ \mathbf{x}^* is a local minimum (or maximum) of f in $N_\epsilon(\mathbf{x}^*)$.
- ▶ If $\mathbf{x} \in N_\epsilon(\mathbf{x}^*)$, then

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

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

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

Proof. Follows directly from the definition of L and the preceding Theorem.

Lagrange's method

Theorem (Sufficient conditions).

Let f, g_1, \dots, g_m be twice continuously differentiable real-valued functions in \mathbb{R}^n .

If there exist vectors $\mathbf{x}^* \in \mathbb{R}^n$, $\boldsymbol{\lambda}^* \in \mathbb{R}^m$ such that

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

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

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

it follows that

$$\mathbf{z}^T \nabla_{\mathbf{x}}^2 L(\mathbf{x}^*, \boldsymbol{\lambda}^*) \mathbf{z} > 0,$$

then, f has a strict local minimum at \mathbf{x}^* subject to $g_i(\mathbf{x}) = 0$, $i = 1, \dots, m$,
(similar for a maximum if $\mathbf{z}^T \nabla_{\mathbf{x}}^2 L(\mathbf{x}^*, \boldsymbol{\lambda}^*) \mathbf{z} < 0$).

Proof of the Theorem

Proof.

Assume that \mathbf{x}^* is not a strict local minimum. Then there exist a neighborhood $N_\delta(\mathbf{x}^*)$ and a sequence $\{\mathbf{z}^k\}_{k \in \mathbb{Z}}$, $\mathbf{z}^k \in N_\delta(\mathbf{x}^*)$, $\mathbf{z}^k \neq \mathbf{x}^*$, converging to \mathbf{x}^* such that for every \mathbf{z}^k in the sequence

$$g_i(\mathbf{z}^k) = 0, \quad i = 1, \dots, m, \quad f(\mathbf{x}^*) \geq f(\mathbf{z}^k). \quad (9)$$

Let $\mathbf{z}^k = \mathbf{x}^* + \theta^k \mathbf{y}^k$, where $\theta^k > 0$ and $\|\mathbf{y}^k\| = 1$. The sequence $\{(\theta^k, \mathbf{y}^k)\}_{k \in \mathbb{Z}}$ has a subsequence that converges to $(0, \bar{\mathbf{y}})$, where $\|\bar{\mathbf{y}}\| = 1$. By the Mean Value Theorem, for each k in this subsequence

$$g_i(\mathbf{z}^k) - g_i(\mathbf{x}^*) = \theta^k (\mathbf{y}^k)^T \nabla g_i(\mathbf{x}^* + \eta_i^k \theta^k \mathbf{y}^k) = 0, \quad i = 1, \dots, m. \quad (10)$$

with $0 < \eta_i^k < 1$ and

$$f(\mathbf{z}^k) - f(\mathbf{x}^*) = \theta^k (\mathbf{y}^k)^T \nabla f(\mathbf{x}^* + \xi^k \theta^k \mathbf{y}^k) \leq 0, \quad (11)$$

with $0 < \xi_i^k < 1$. Dividing (10) and (11) by θ^k and taking limits as $k \rightarrow \infty$, we get

$$\begin{aligned} \bar{\mathbf{y}}^T \nabla g_i(\mathbf{x}^*) &= 0, \quad i = 1, \dots, m \\ \bar{\mathbf{y}}^T \nabla f(\mathbf{x}^*) &\leq 0. \end{aligned}$$

Proof of the Theorem (cont)

From Taylor's theorem we have

$$\begin{aligned} L(\mathbf{z}^k, \boldsymbol{\lambda}^*) &= L(\mathbf{x}^*, \boldsymbol{\lambda}^*) + \theta^k (\mathbf{y}^k)^T \nabla_{\mathbf{x}} L(\mathbf{x}^*, \boldsymbol{\lambda}^*) \\ &\quad + \frac{1}{2} (\theta^k)^2 (\mathbf{y}^k)^T \nabla_{\mathbf{x}}^2 L(\mathbf{x}^* + \eta^k \theta^k \mathbf{y}^k, \boldsymbol{\lambda}^*) \mathbf{y}^k, \end{aligned} \quad (12)$$

with $0 < \eta^k < 1$. Dividing this equality by $(\theta^k)^2/2$, using the definition of L , the hypothesis $\nabla L(\mathbf{x}^*, \boldsymbol{\lambda}^*) = 0$ and the conditions (9), we get (exercise)

$$(\mathbf{y}^k)^T \nabla_{\mathbf{x}}^2 L(\mathbf{x}^* + \eta^k \theta^k \mathbf{y}^k, \boldsymbol{\lambda}^*) \mathbf{y}^k \leq 0.$$

Letting $k \rightarrow \infty$, we obtain $\bar{\mathbf{y}}$ verifying

$$\bar{\mathbf{y}}^T \nabla_{\mathbf{x}}^2 L(\mathbf{x}^*, \boldsymbol{\lambda}^*) \bar{\mathbf{y}} \leq 0.$$

This completes the proof, since $\bar{\mathbf{y}} \neq 0$ and satisfies $\bar{\mathbf{y}}^T \nabla g_i(\mathbf{x}^*) = 0$ for $i = 1, \dots, m$.

Example

Consider again the problem

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

subject to the constraint

$$g(x, y) = x + y - 2 = 0.$$

The Lagrangian is

$$L(\mathbf{x}, \lambda) = xy - \lambda(x + y - 2).$$

Setting $\nabla L(\mathbf{x}, \lambda) = 0$, we get:

$$\begin{aligned}\frac{\partial L(\mathbf{x}, \lambda)}{\partial x} &= y - \lambda = 0, \\ \frac{\partial L(\mathbf{x}, \lambda)}{\partial y} &= x - \lambda = 0, \\ \frac{\partial L(\mathbf{x}, \lambda)}{\partial \lambda} &= -x - y + 2 = 0.\end{aligned}$$

The solution of this system of equations is

$$x^* = y^* = \lambda^* = 1.$$

According to the Theorem on necessary conditions, the point $(\mathbf{x}^*, \lambda^*) = (1, 1, 1)$ satisfies the necessary conditions for a maximum.

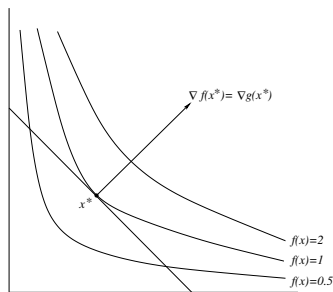
Example (cont.)

The linear dependence between ∇f and ∇g at the maxima, is clearly illustrated in the figure. In fact, in this case they coincide, since

$$\nabla f(\mathbf{x}^*) = \begin{pmatrix} y \\ x \end{pmatrix}_{(x,y)=(1,1)} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

and

$$\nabla g(\mathbf{x}^*) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}_{(x,y)=(1,1)} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$



Example (cont.)

Turning to the sufficient conditions, we compute $\nabla_x^2 L(\mathbf{x}^*, \lambda^*)$:

$$\frac{\partial^2 L(\mathbf{x}^*, \lambda^*)}{\partial x \partial x} = 0, \quad \frac{\partial^2 L(\mathbf{x}^*, \lambda^*)}{\partial x \partial y} = 1, \quad \frac{\partial^2 L(\mathbf{x}^*, \lambda^*)}{\partial y \partial y} = 0.$$

Hence

$$\mathbf{z}^T \nabla_x^2 L(\mathbf{x}^*, \lambda^*) \mathbf{z} = (z_1, z_2) \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = 2z_1 z_2,$$

According to the last Theorem, we must determine the sign of $2z_1 z_2$ for all $\mathbf{z} \neq 0$ such that $\mathbf{z}^T \nabla g(\mathbf{x}^*) = 0$.

Since

$$\frac{\partial g(\mathbf{x}^*)}{\partial x} = \frac{\partial g(\mathbf{x}^*)}{\partial y} = 1,$$

the last condition $\mathbf{z}^T \nabla g(\mathbf{x}^*) = 0$ is equivalent to $z_1 + z_2 = 0$, from which we get

$$\mathbf{z}^T \nabla_x^2 L(\mathbf{x}^*, \lambda^*) \mathbf{z} = -2z_1^2 < 0.$$

Thus, $(1, 1)$ is a strict local maximum.