Optimization

Màster de Fonaments de Ciència de Dades

PART 2. Analysis

Chapter 3. Unconstrained and constrained optimization with equalities. Optimality conditions

Chapter 4. (Inequality) Constrained optimization. Optimality conditions

Chapter 5. Convex sets and functions

Chapter 3

Unconstrained
and
constrained optimization with equalities

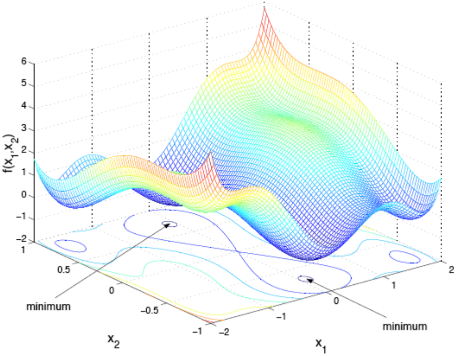
Optimality conditions

Local and global minima

Let $D \subset \mathbb{R}^n$ an open set. Let $f : D \longrightarrow \mathbb{R}$ be a real valued function. Denote by $\mathrm{B}(\boldsymbol{x}^\star, \varepsilon) := \{x \in \mathbb{R}^n \mid \|\boldsymbol{x} - \boldsymbol{x}^*\| \leq \varepsilon\}$.

1. We say that $\boldsymbol{x}^\star \in D$ is a local minimum of f if there exists $\epsilon > 0$ such that $\mathrm{B}(\boldsymbol{x}^\star, \varepsilon) \subset D$ and $f(\boldsymbol{x}^*) \leq f(\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathrm{B}(\boldsymbol{x}^\star, \varepsilon)$. If $f(\boldsymbol{x}^*) < f(\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathrm{B}(\boldsymbol{x}^\star, \varepsilon) \setminus \{\boldsymbol{x}^\star\}$ we say that $\boldsymbol{x}^\star$ is a strict local minimum of f.
2. We say that $\boldsymbol{x}^\star \in D$ is a global minimum of f if $f(\boldsymbol{x}^*) \leq f(\boldsymbol{x})$ for all $\boldsymbol{x} \in D$. If $f(\boldsymbol{x}^*) < f(\boldsymbol{x})$ for all $\boldsymbol{x} \in D \setminus \{\boldsymbol{x}^\star\}$ we say that $\boldsymbol{x}^\star \in D$ is a strict global minimum of f.
3. Similar definitions correspond to (local and global) maxima of f instead of minima of f. Finally we write extrema of f when we refer to max or min, indiscriminately.

Local and global minima



Local (4) versus global (2) minima for f above.

A first result

Theorem : Let $f : [a, b] \to \mathbb{R}$ be a continuous function. Let $c \in (a, b)$. If for some $\delta > 0$, f is decreasing on $(c - \delta, c)$ and increasing on $(c, c + \delta)$, then f has a local minimum at c.

Proof: Let x_1 and x such that $c - \delta < x_1 < x < c$. Then

$$f(x_1) \geq \lim_{x \to c^-} f(x) = f(c).$$

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

$$f(x_2) \geq \lim_{x \to c^+} f(x) = f(c).$$

Exercise. Prove that the condition of the theorem is not necessary.

Necessary conditions for minima

Theorem. Let $D \subset \mathbb{R}^n$ open and let $f : D \to \mathbb{R}$ be a $\mathcal{C}^1$ function (continuously differentiable). Assume $\boldsymbol{x}^\star \in D$ is a local minima of f. Then

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

Proof. We argue by contradiction. By hypothesis, if $\lambda > 0$ is small enough, for all $\boldsymbol{v} \in \mathbb{R}^n$, we have that

$$f(\boldsymbol{x}^*) \leq f(\boldsymbol{x}^* + \lambda \boldsymbol{v}). \tag{1}$$

Fix $\boldsymbol{v}$ (w.l.o.g. $||v|| = 1$). Define $F_{\boldsymbol{v}}(\lambda) := F(\lambda) = f(\boldsymbol{x}^* + \lambda \boldsymbol{v})$. Then, (1) writes as

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

for some $\delta > 0$. From the Mean Value Theorem, we have

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

for some $\theta \in [0, 1]$.

Necessary conditions for minima

Proof (continue).

If $F'(0) > 0$, then, since F' is a continuous function, we have that, if $|\lambda|$ is small enough, $F'(\theta\lambda) > 0$ for all $\theta \in [0, 1]$. Hence, we can find a $\lambda < 0$ small enough such that

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

a contradiction. Arguing similarly with $F'(0) < 0$ we conclude $F'(0) = 0$, or equivalently,

$$F'(0) = (\nabla f(\boldsymbol{x}^\star))^T \boldsymbol{v} = 0.$$

Since $\boldsymbol{v}$ is an arbitrary unitary vector, we must have:

$$\nabla f(\boldsymbol{x}^\star) = 0.$$

Remark. Similar arguments show that if $\boldsymbol{x}^\star \in D$ is a local maxima of a $\mathcal{C}^1$ map f , then

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

Sufficient conditions for minima

Theorem. Let $D \subset \mathbb{R}^n$ open and let $f : D \to \mathbb{R}$ be a $\mathcal{C}^2$ map (twice continuously differentiable). Assume

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

Then f has a strict local minimum at $\boldsymbol{x}^*$.

Exercise. Prove the Theorem above (Use the second order Taylor expansion of f around $\boldsymbol{x}^*$).

Exercise. Prove that the converse of Theorem above is not true. Consider the family of one-dimensional real valued functions $f_p(x) = x^p, \ p \geq 1$.

Necessary and sufficient conditions for minimum

Theorem A. Let $D \subset \mathbb{R}^n$ open and let $f : D \to \mathbb{R}$ be a $\mathcal{C}^2$ map (twice continuously differentiable). The following statements hold

(a) (Necessary conditions) Assume $\boldsymbol{x}^\star \in D$ is a local minima of f. Then

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

(b) (Sufficient conditions) Assume $\nabla f(\boldsymbol{x}^*) = 0$. Assume also that there exists $\varepsilon > 0$ such that for all $\boldsymbol{x} \in \boldsymbol{B}(\boldsymbol{x}^\star, \varepsilon)$ we have

$$\boldsymbol{z}^T H(f)(\boldsymbol{x})\boldsymbol{z} \geq 0, \ \forall \boldsymbol{z} \in \mathbb{R}^n.$$

Then, $\boldsymbol{x}^\star \in D$ is a local minima of f.

Necessary and sufficient conditions for minimum

Exercise. Prove the above theorem for $n = 1$. To prove (a) use similar ideas of the arguments we have used. For (b) use Taylor's Theorem to notice that for any h small enough ($|h| < \varepsilon$) we have

$$f(x^\star + h) = f(x^\star) + f'(x^\star)h + \frac{1}{2}f''(z_h)h^2,$$

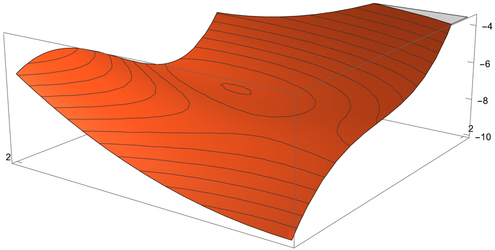
where $z_h \in (0, h)$.

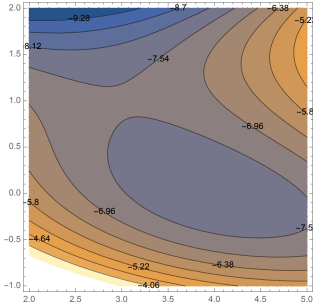
Exercise. Prove the above theorem for $n \geq 1$. Generalize previous arguments to higher dimension.

Example: Extrema for differentiable functions

Problem. Find the (local) extrema for the polynomial function

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





Example: Extrema for differentiable functions

Computations. We have that

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

Easily

$$\nabla f(x,y) = 0 \quad \Longleftrightarrow \quad \boldsymbol{x}_1 = (4,0)^T \text{ and } \boldsymbol{x}_2 = (3,1)^T,$$

and

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

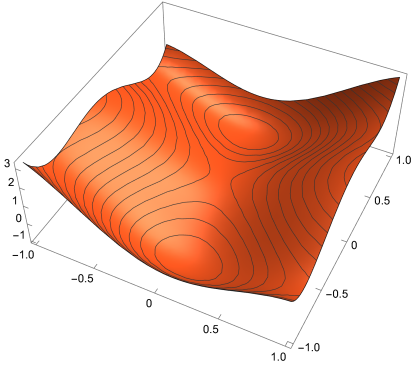
$H(\boldsymbol{x}_1)$ is positive definite and $H(\boldsymbol{x}_2)$ is indefinite.

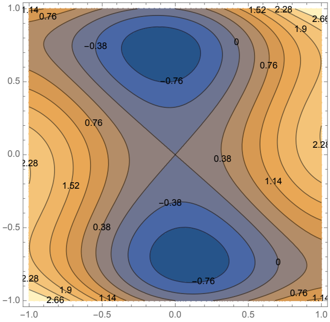
Conclusion. The only extrema is the local minimum is $\boldsymbol{x}_1 = (4,0)^T$.

Example: Extrema for differentiable functions

Problem (exercise). Find the extrema for the polynomial function

$$f(x,y) = \left(4 - \frac{21}{10}x^2 + \frac{1}{3}x^4\right)x^2 + xy + 4y^2(-1+y^2)$$





Constrained optimization with equalities

Lagrange Multipliers

The formal problem

Formal problem. Let $f : D \subset \mathbb{R}^n \to \mathbb{R}$ and $g_j : D \subset \mathbb{R}^n \to \mathbb{R}$, $j = 1, \ldots, m$ be functions defined on D. Assume $m < n$. Consider the problem

$$\min_{x \in \mathbb{R}^n} f(x) \qquad \text{subject to} \quad g_j(x) = 0 \quad j = 1, \ldots, m \quad (m < n) \quad (2)$$

Regularity. The problem can also be written for maximum instrad of minimum of f.

Regularity. As we did before, we will assume, to get a solution, certain regularity condition on f and the restriction equations g_j.

Example 1: From constrained to unconstrained problems

Problem. Find the vector $\boldsymbol{x} = (x_1, x_2, x_3) \in \mathbb{R}^3$ whose components maximize their product and such that $x_1 + x_2 + x_3 = 1$. In other words, solve

$$\max_{\boldsymbol{x}\in\mathbb{R}^3} x_1 x_2 x_3 \quad \text{subject to} \quad x_1 + x_2 + x_3 = 1.$$

Exercise. Give a geometric meaning of this problem.

Example 1

Solution. From the restriction we get $x_3 = 1 - x_1 - x_2$. Substituting, the problem becomes unrestricted and writes as

$$\max_{\boldsymbol{x}\in\mathbb{R}^2} f(x_1, x_2) := x_1 x_2 (1 - x_1 - x_2).$$

We have that

$$\nabla f(x_1, x_2) = \left(x_2\left(1 - 2x_1 - x_2\right),\, x_1\left(1 - x_1 - 2x_2\right)\right)^T$$
$$H(f)(x_1, x_2) = \begin{pmatrix} -2x_2 & 1 - 2x_1 - 2x_2 \\ 1 - 2x_1 - 2x_2 & -2x_1 \end{pmatrix}.$$

Easily

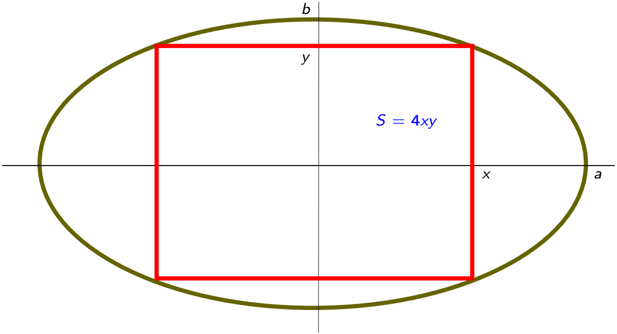
$$\nabla f(x_1, x_2) = 0 \iff \boldsymbol{x}^1 = (0,0)^T,\ \boldsymbol{x}^2 = (1,0)^T,\ \boldsymbol{x}^3 = (0,1)^T,\ \boldsymbol{x}^4 = (1/3, 1/3)^T.$$

Substitying we have that $H(f)(\boldsymbol{x}^j)$, $j = 1, 2, 3$ are indefinite while $H(f)(\boldsymbol{x}^4)$ is positive definite. So, $\boldsymbol{x}^4$ is the only maxima of f and $\boldsymbol{x} = (1/3, 1/3, 1/3) \in \mathbb{R}^3$ the only solution of our (restricted) problem.

Example 2

Problem. Find the largest area of the rectangle inscribed in the ellipse

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



Example 2 (continuation)

Solution. Suppose that the upper righthand corner of the rectangle is at the point (x, y), then the area of the rectangle is $S = 4xy$ (as shown in the picture). Accordingly the ellipse equation implies that $y = y(x)$ in this computation (Implicit Function Theorem). In particular

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

If we write $S(x) = 4xy(x)$ we impose $S'(x) = 0$. This is,

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

Using the equation of the ellipse we get $y^2 = b^2 - y^2$ and so

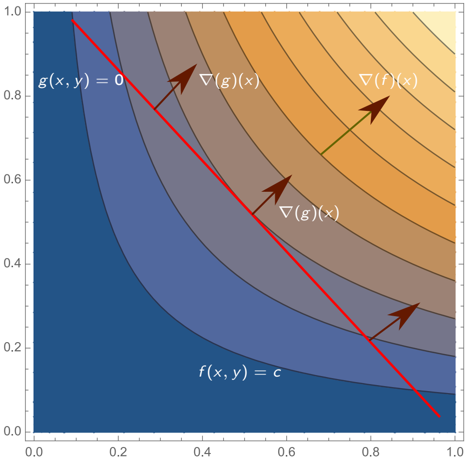
$$(x^\star, y^\star) = \left(\frac{a}{\sqrt{2}}, \frac{b}{\sqrt{2}}\right)$$

The substituting method

Remark. According to the examples above a possible solution of the problem is to convert a constrain problem to an unconstrain problem via substitute some variables using the constrains. If this is possible, we end up with an unconstrained problem with less variables.

This is, in general, quite complicated (if not impossible) since there is no an easy way to explicitly isolate the variables using the constrains, or doing this substitution implicitly using the Implicit Function Theorem applied to the constrains.

A geometric approach: A taste on Lagrange multipliers



$$\max_{(x,y)\in\mathbb{R}^2} \quad f(x,y) = xy$$

$$\text{subject to} \quad g(x,y) = 0.$$

Analytic approach: A taste on Lagrange multipliers

Theorem B. Consider (2) with f and g_j, $i = 1, ..., m$, being $\mathcal{C}^1$-functions (continuously differentiable) on $\boldsymbol{B}(\boldsymbol{x}^\star, \varepsilon)$ with $\boldsymbol{x}^\star \in D$. Assume that the Jacobian matrix $D(g)(\boldsymbol{x}^*)$ has rank m.

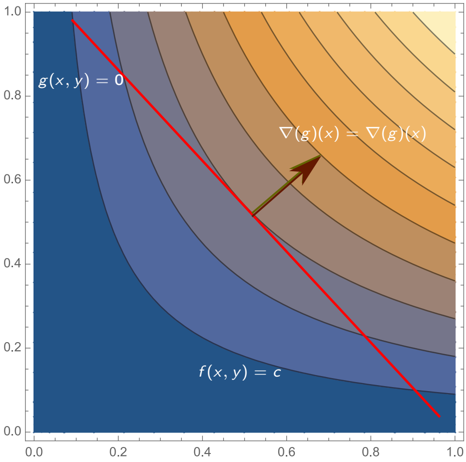
Suppose that $\boldsymbol{x}^*$ is a (local) minimum (maximum) of f for all points $\boldsymbol{x} \in \boldsymbol{B}(\boldsymbol{x}^\star, \varepsilon)$ that also satisfy $g_i(\boldsymbol{x}) = 0$, $i = 1, ..., m$.

Under these hypotheses, there exists $\boldsymbol{\lambda}^\star = (\lambda_1^\star, \ldots, \lambda_1^\star)$ such that

$$\nabla f(\boldsymbol{x}^*) = \sum_{i=1}^{m} \lambda_i^* \nabla g_i(\boldsymbol{x}^*). \tag{3}$$

Remark. Equation (3) has n equations and m unknowns (the multipliers λ's).

Analytic and geometric approach meet together



$$\max_{(x,y)\in\mathbb{R}^2} \quad f(x,y) = xy$$

subject to $x + y = 2$.

Proof of the above theorem

(a) Without lost of generality we may assume that

$$\det \begin{pmatrix} \frac{\partial g_1(\boldsymbol{x}^*)}{\partial x_1} & \cdots & \frac{\partial g_m(\boldsymbol{x}^*)}{\partial x_1} \\ \vdots & & \vdots \\ \frac{\partial g_1(\boldsymbol{x}^*)}{\partial x_m} & \cdots & \frac{\partial g_m(\boldsymbol{x}^*)}{\partial x_m} \end{pmatrix} \neq 0.$$

(b) What we want to proof is that there exist a vector $\boldsymbol{\lambda}^\star = (\lambda_1^*, \ldots, \lambda_m^*)$ such that:

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

Proof of the above theorem (II)

(c) Because of (a) it is clear that the linear system

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

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

(d) To see that the rest of the $n - m$ components also satisfy the equality we will use the Implicit Function Theorem and the condition that $x^\star$ is a minimum of f.

Proof of the above theorem (III)

(d1) Denote by $\hat{\boldsymbol{x}} = (x_{m+1}, ..., x_n)$ the remaining variables.

(d2) From the Implicit Function Theorem we know the existence of functions $h_j(\hat{\boldsymbol{x}})$ defined on a open domain $\hat{D}$ containing $\hat{\boldsymbol{x}}^\star$ such that $x_j = h_j(\hat{\boldsymbol{x}})$, $j = 1, \dots, m$ such that

$$x_j^\star = h_j(\hat{\boldsymbol{x}}^\star), \text{ and}$$
$$g_j\left(h_1(\hat{\boldsymbol{x}}), \dots, h_m(\hat{\boldsymbol{x}}), x_{m+1}, \dots, x_n\right) = 0, \ j = 1, \dots m$$

Moreover, for every $i = m+1, \dots n$ we have

$$\sum_{k=1}^{m} \frac{\partial g_j(\boldsymbol{x}^\star)}{\partial x_k} \frac{\partial h_k(\hat{\boldsymbol{x}}^\star)}{\partial x_i} + \frac{\partial g_j(\boldsymbol{x}^\star)}{\partial x_i} = 0, \quad j = 1, \dots, m \, .$$

Proof of the above theorem (IV)

(d2) For every $i = m+1, \ldots n$ we have

$$\sum_{k=1}^{m} \frac{\partial g_j(\boldsymbol{x}^\star)}{\partial x_k} \frac{\partial h_k(\hat{\boldsymbol{x}}^\star)}{\partial x_i} + \frac{\partial g_j(\boldsymbol{x}^\star)}{\partial x_i} = 0, \quad j = 1, \ldots, m.$$

Multiplying each equation for $\lambda_j^\star$ and adding up we have

$$\sum_{j=1}^{m} \sum_{k=1}^{m} \lambda_j^\star \frac{\partial g_j(\boldsymbol{x}^\star)}{\partial x_k} \frac{\partial h_k(\hat{\boldsymbol{x}}^\star)}{\partial x_i} + \lambda_j^\star \frac{\partial g_j(\boldsymbol{x}^\star)}{\partial x_i} = 0, \quad i = m+1, \ldots, n. \quad (4)$$

Proof of the above theorem (V)

(d3) Since $x_j = h_j(\hat{\boldsymbol{x}})$, $j = 1, \dots, m$ (locally near $\hat{\boldsymbol{x}}$) we also have that

$$f(\boldsymbol{x}) = f\left(h_1(\hat{\boldsymbol{x}}), \dots, h_m(\hat{\boldsymbol{x}}), x_{m+1}, \dots, x_n\right).$$

Since by hypothesis f has a local extrema at $\boldsymbol{x} = \boldsymbol{x}^\star$ (we are now working with an unconstrained problem) the partial (implicit) derivatives must vanished at $\boldsymbol{x} = \boldsymbol{x}^\star$.

$$\frac{\partial f(\boldsymbol{x}^\star)}{\partial x_i} = \sum_{k=1}^{m} \frac{\partial f\left(\boldsymbol{x}^\star\right)}{\partial x_k} \frac{\partial h_k\left(\hat{\boldsymbol{x}}^\star\right)}{\partial x_i} + \frac{\partial g_j\left(\boldsymbol{x}^\star\right)}{\partial x_i} = 0, \quad i = m+1, \dots, n. \tag{5}$$

Proof of the above theorem (VI)

(d4) All together yield

$$\sum_{k=1}^{m}\left[\frac{\partial f\left(\mathbf{x}^{\star}\right)}{\partial x_k}-\sum_{j=1}^{m}\lambda_j^{\star}\frac{\partial g_j\left(\mathbf{x}^{\star}\right)}{\partial x_k}\right]\frac{\partial h_k\left(\hat{\mathbf{x}}^{\star}\right)}{\partial x_i}+\frac{\partial f\left(\mathbf{x}^{\star}\right)}{\partial x_i}-\sum_{j=1}^{m}\lambda_j^{\star}\frac{\partial g_j\left(\mathbf{x}^{\star}\right)}{\partial x_i}=0$$

for $i=m+1,\ldots n$. The expression in brakets is zero (see step (c) in this proof) and we conclude the theorem.

Lagrange function and Lagrange multipliers

Definition. Let

$$f : D \subset \mathbb{R}^n \to \mathbb{R} \quad \text{and}$$
$$g_j : D \subset \mathbb{R}^n \to \mathbb{R},\ j = 1, \ldots, m$$

be functions defined on D. Assume $m < n$. We define the Lagrange function (depending on $n + m$ variables) as

$$L(\boldsymbol{x}, \boldsymbol{\lambda}) := f(\boldsymbol{x}) - \sum_{j=1}^{m} \lambda_j g_j(\boldsymbol{x}).$$

The new components of the $\boldsymbol{\lambda}$-vector are known as Langrange multipliers.

Lagrange function and Lagrange multipliers

Corollary (necessary condition for extrema). Under the previous notation and the hypotheses of Theorem B, there exists $\boldsymbol{\lambda}^\star = (\lambda_1^\star, \ldots, \lambda_1^\star)$ such that

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

Lagrange function: Sufficient conditions

Remark. Of course once we have found the necessary condition for optimal extrema (see corollary above) we might use the sufficient conditions for unconstrained optimization problems, stated in the previous chapter (see Theorem A(b)). But this, of course it is suboptimal since Theorem A does not take into account the constrains.

Example. Find the extrema of the constrained problem given by

$$\min(\max)\ f(x,y) = xy \quad \text{subject to} \quad x^2 + y^2 = 2$$

Example (continue)

Necessary conditions. Using the Lagrange function we find four candidates $(x^\star, y^\star, \lambda^\star)$

$$\left(1, 1, \frac{1}{2}\right) \quad \left(-1, -1, \frac{1}{2}\right) \quad \left(1, -1, -\frac{1}{2}\right) \quad \left(-1, 1, -\frac{1}{2}\right)$$

(Unrestricted sufficient conditions (Theorem A(b). We need to check if $\forall \boldsymbol{z} = (z_1, z_2) \in \mathbb{R}^2$ and $\forall (x, y) \in \boldsymbol{B}\left((x^\star, y^\star), \varepsilon\right)$ we have

$$\boldsymbol{z}^T H(f)(\boldsymbol{x})\boldsymbol{z} \geq 0.$$

But $\boldsymbol{z}^T H(f)(\boldsymbol{x})\boldsymbol{z} = (z_1, z_2)^T \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} (z_1, z_2) = 2z_1 z_2.$

Generalized Hessian

Definition. Let f and g_j, $j = 1, \ldots m$ twice continuous differentiable real-valued functions defined in $D \subset \mathbb{R}^n$. Let $L(\boldsymbol{x}, \boldsymbol{\lambda})$ be the associated Lagrange function. The order $p = m + 1, \ldots n$ extended hessian for L is given by the determinant

$$\mathcal{H}_p(L) := \begin{vmatrix} \frac{\partial^2 L(\boldsymbol{x}^\star, \boldsymbol{\lambda}^\star)}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 L(\boldsymbol{x}^\star, \boldsymbol{\lambda}^\star)}{\partial x_1 \partial x_p} & \frac{\partial g_1(\boldsymbol{x}^\star)}{\partial x_1} & \cdots & \frac{g_m(\boldsymbol{x}^\star)}{\partial x_1} \\ \vdots & & \vdots & \vdots & & \vdots \\ \frac{\partial^2 L(\boldsymbol{x}^\star, \boldsymbol{\lambda}^\star)}{\partial x_p \partial x_1} & \cdots & \frac{\partial^2 L(\boldsymbol{x}^\star, \boldsymbol{\lambda}^\star)}{\partial x_p \partial x_p} & \frac{\partial g_1(\boldsymbol{x}^\star)}{\partial x_p} & \cdots & \frac{\partial g_m(\boldsymbol{x}^\star)}{\partial x_p} \\ \frac{\partial g_1(\boldsymbol{x}^\star)}{\partial x_1} & \cdots & \frac{\partial g_1(\boldsymbol{x}^\star)}{\partial x_p} & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ \frac{\partial g_m(\boldsymbol{x}^\star)}{\partial x_1} & \cdots & \frac{\partial g_m(\boldsymbol{x}^\star)}{\partial x_p} & 0 & \cdots & 0 \end{vmatrix}$$

Lagrange function: Sufficient conditions

Theorem. Let f and g_j, $j = 1, \ldots m$ twice continuous differentiable real-valued functions defined in $D \subset \mathbb{R}^n$. Assume that the vector $(\boldsymbol{x}^\star, \boldsymbol{\lambda}^\star)$ with $\boldsymbol{x}^\star \in D$ and $\boldsymbol{\lambda}^\star \in \mathbb{R}^m$ satisfies $\nabla L(\boldsymbol{x}^\star, \boldsymbol{\lambda}^\star) = 0$ and

$$(-1)^m \mathcal{H}_p(L) > 0$$

for $p = m+1, \ldots, n$. Then $\boldsymbol{x}^\star$ is a strict local minimum of f subject to $g_j = 0$, $j = 1, \ldots, m$.

Theorem. Same conditions than above and

$$(-1)^p \mathcal{H}_p(L) > 0$$

for $p = m+1, \ldots, n$. Then $\boldsymbol{x}^\star$ is a strict local maximum of f subject to $g_j = 0$, $j = 1, \ldots, m$.

Example (again)

Example. Find the extrema of the constrained problem given by

$$\min(\max)\ f(x,y) = xy \quad \text{subject to} \quad x^2 + y^2 = 2$$

Four candidates $(x^\star, y^\star, \lambda^\star)$

$$\left(1, 1, \frac{1}{2}\right) \quad \left(-1, -1, \frac{1}{2}\right) \quad \left(1, -1, -\frac{1}{2}\right) \quad \left(-1, 1, -\frac{1}{2}\right)$$

Extended Hessian. $m = 1$ and $p = 2$.

$$\mathcal{H}_2(L)(x^\star, y^\star, \lambda^\star) = \begin{vmatrix} -2\lambda^\star & 1 & 2x^\star \\ 1 & -2\lambda^\star & 2y^\star \\ 2x^\star & 2y^\star & 0 \end{vmatrix}$$

Example (again)

Since

$$(-1)^2\mathcal{H}_2(L)\left(1,1,\frac{1}{2}\right)=(-1)^2\mathcal{H}_2(L)\left(-1,-1,\frac{1}{2}\right)>0$$

we have that $(1,1,\frac{1}{2})$ and $(-1,-1,\frac{1}{2})$ are strict local maximum.

Since

$$(-1)^1\mathcal{H}_2(L)\left(1,1,\frac{1}{2}\right)=(-1)^1\mathcal{H}_2(L)\left(-1,-1,\frac{1}{2}\right)>0$$

we have that $(1,1,\frac{1}{2})$ and $(-1,-1,\frac{1}{2})$ are strict local minimum.

Exercise. Geometric interpretation.