Constrained optimization

Lagrange multipliers

I. Constraints with equalities

Constrained optimization problem with equalities. 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

$$\min_{x \in D} f(x) \qquad \text{subject to} \quad g_j(x) = 0 \quad j = 1, \ldots, m \tag{1}$$

Regularity. As usual, we assume that f and $g_j, \ j = 1, \ldots, m$ are smooth functions.

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 (continued)

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

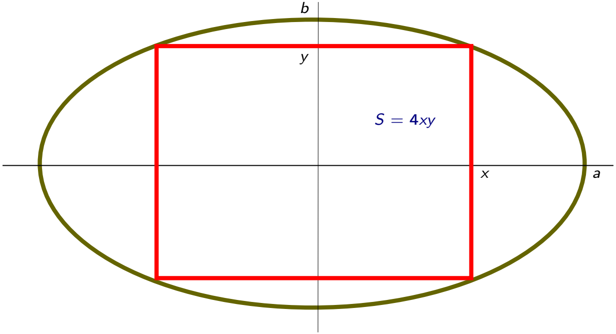
$$\nabla f(x_1, x_2) = 0 \iff \boldsymbol{x}^1 = (0,0)^T, \ \boldsymbol{x}^2 = (1,0)^T, \ \boldsymbol{x}^3 = (0,1)^T, \ \boldsymbol{x}^4 = (1/3, 1/3)^T.$$

Substituting, we obtain that $H(f)(\boldsymbol{x}^j), \ j = 1,2,3$ are indefinite while $H(f)(\boldsymbol{x}^4)$ is positive definite. Hence, $\boldsymbol{x}^4$ 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 a rectangle inscribed in the ellipse

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



Example 2 (continued)

Solution. Suppose that the upper right hand 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 $S(x) = 4xy(x)$, then $S'(x) = 0$. That 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

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

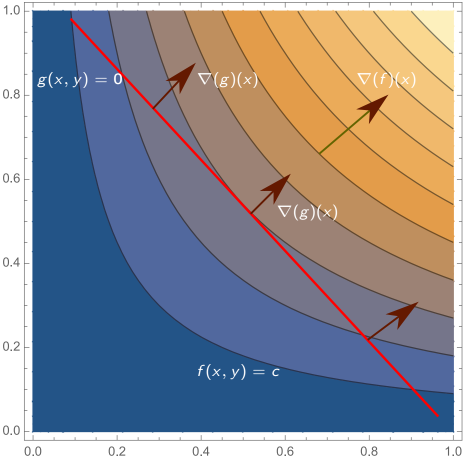
*One can check by computing the Hessian that $(x^\star, y^\star)$ is the maximum.

The substituting method

Remark. According to the examples above, a possible solution to the constrained optimization problem is to convert a constrained problem to an unconstrained problem via **substituting** some variables using the constraints. If this is possible, we end up with an unconstrained problem in fewer variables.

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

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 $\star$. Consider (1) 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 $D(g)(\boldsymbol{x}^\star)$ has rank m.

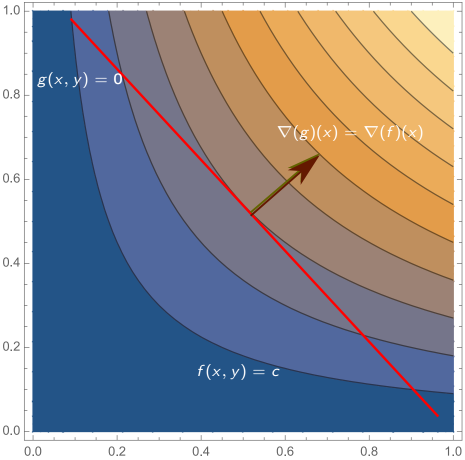
Suppose that $\boldsymbol{x}^\star$ is a (local) minimum (maximum) of f for all points $\boldsymbol{x} \in \boldsymbol{B}(\boldsymbol{x}^\star, \varepsilon)$ that also satisfy $g_i(\boldsymbol{x}) = 0,\ i = 1, ..., m$.

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

$$\nabla f(\boldsymbol{x}^\star) = \sum_{i=1}^{m} \lambda_i^* \nabla g_i(\boldsymbol{x}^\star). \tag{2}$$

Remark. Equation (2) has n equations and m unknowns (the multipliers λ's).

Analytic and geometric approaches meet together



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

$$\text{subject to} \quad {\color{red} x + y = 2}.$$

Proof of Theorem $\star$

(a) Without loss of generality we may assume that

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

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

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

Proof of Theorem $\star$ (cont., II)

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

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

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 prove.

(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 Theorem $\star$ (III)

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

(d2) From the Implicit Function Theorem we know that there exist functions $h_j(\hat{\boldsymbol{x}})$ defined on an open domain $\hat{D}$ containing $\hat{\boldsymbol{x}}^\star$ such that $x_j = h_j(\hat{\boldsymbol{x}})\,,\ j = 1, \ldots, m$ and

$$
\begin{aligned}
& x_j^\star = h_j(\hat{\boldsymbol{x}}^\star), \\
& g_j\left(h_1(\hat{\boldsymbol{x}}), \ldots, h_m(\hat{\boldsymbol{x}}), x_{m+1}, \ldots, x_n\right) = 0,\ j = 1, \ldots m
\end{aligned}
$$

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

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

Proof of Theorem $\star$ (IV)

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

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

Multiplying each equation by λ_j^* and adding up we have

$$\sum_{j=1}^{m} \sum_{k=1}^{m} \lambda_j^* \frac{\partial g_j\left(\boldsymbol{x}^\star\right)}{\partial x_k} \frac{\partial h_k\left(\hat{\boldsymbol{x}}^\star\right)}{\partial x_i} + \lambda_j^* \frac{\partial g_j\left(\boldsymbol{x}^\star\right)}{\partial x_i} = 0, \quad i = m+1, \ldots, n. \tag{3}$$

Proof of Theorem $\star$ (V)

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

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

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 vanish at $\boldsymbol{x} = \boldsymbol{x}^\star$.

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

Proof of Theorem $\star$ (VI)

(d4) All together this yield

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

for $i=m+1,\ldots n$. The expression in brackets 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 **Lagrange multipliers**.

Lagrange function and Lagrange multipliers

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

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

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. However, this is sub-optimal since the constraints are not taken into account.

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 (continued)

Necessary conditions. Using the Lagrange function $L(x, y, \lambda) = xy - \lambda(x^2 + y^2 - 2)$ 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.) We need to check whether $\forall \boldsymbol{z} = (z_1, z_2)^T \in \mathbb{R}^2 \setminus \{0\}$ and $\forall \boldsymbol{x} \in \boldsymbol{B}\left((x^\star, y^\star), \varepsilon\right)$ we have

$$\boldsymbol{z}^T \nabla^2 f(\boldsymbol{x}) \boldsymbol{z} > 0 \quad (\text{or } < 0).$$

But $\boldsymbol{z}^T \nabla^2 f(\boldsymbol{x}) \boldsymbol{z} = (z_1, z_2) \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} (z_1, z_2)^T = 2 z_1 z_2$ **can change sign!**

Lagrange function: Sufficient conditions

Theorem $\star\star$ (Sufficient condition). Let f and g_j, $j = 1, \ldots m$ twice continuously 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 **for every $\boldsymbol{z} \in \mathbb{R}^n$, $\boldsymbol{z} \neq 0$ satisfying**

$$\boldsymbol{z}^T \nabla g_j(\boldsymbol{x}^\star) = 0, \quad j = 1, \ldots, m,$$

it follows that

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

then f has strict local minimum at $\boldsymbol{x}^\star$ subject to $g_j = 0$, $j = 1, \ldots, m$
Similarly for the maximum (using the condition $\boldsymbol{z}^T \nabla^2_x L(\boldsymbol{x}^\star, \boldsymbol{\lambda}^\star) z < 0$).

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

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

Hessian of the Lagrange function. $L(x,y,\lambda) = xy - \lambda(x^2 + y^2 - 2)$.

$$\nabla_x L(x^\star, y^\star, \lambda^\star) = \begin{vmatrix} -2\lambda^\star & 1 \\ 1 & -2\lambda^\star \end{vmatrix} \qquad \nabla g(x^\star, y^\star, \lambda^\star) = (2x^\star, 2y^\star)$$

Example (continued)

Let us check the candidate $(1, 1, 1/2)$. In this case

$$\nabla_{\boldsymbol{x}} L(1, 1, 1/2) = \left| \begin{array}{cc} -1 & 1 \\ 1 & -1 \end{array} \right| \qquad \nabla g(1, 1, 1/2) = (2, 2).$$

Hence, for any $\boldsymbol{z} = (z_1, z_2)^T \neq 0$, we need to understand the sign of

$$(z_1, z_2) \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = -z_1^2 + 2z_1 z_2 - z_2^2 \quad \text{subject to} \quad 2z_1 + 2z_2 = 0.$$

This simplifies to $-4z_1^2 < 0 \Rightarrow (1, 1)$ **is a strict local maximum**.

Similarly, $(-1, -1)$ is a strict local maximum, while $(1, -1)$ and $(-1, 1)$ are strict local minima.

II. Constraints with equalities and inequalities

Let $D \subset \mathbb{R}^n$ be an open set and let

$$\begin{aligned} f &: D \to \mathbb{R}, \\ g_j &: D \to \mathbb{R}, \; j = 1, \ldots, m, \text{ and} \\ h_j &: D \to \mathbb{R}, \; j = 1, \ldots, p, \end{aligned} \tag{5}$$

with $m << n$, be $\mathcal{C}^1$-functions defined in D.

General problem. The **constrained optimization problem** $(\mathcal{P})$ is defined by

$$\min_{\boldsymbol{x} \in D} f(\boldsymbol{x})$$

$$\text{subject to:} \quad \begin{aligned} g_j(\boldsymbol{x}) &= 0, \qquad i = 1, \ldots, m \\ h_j(\boldsymbol{x}) &\geq 0, \qquad j = 1, \ldots, p. \end{aligned} \tag{6}$$

Solutions of $\mathcal{P}$. Feasible set, points, directions

Definition. The set of points $\mathcal{X} \subset D$ satisfying conditions (6) are called **feasible points** and $\mathcal{X}$ is called the **feasible set** for the constrained optimization problem.

Definition. A point $\boldsymbol{x}^\star \in \mathcal{X}$ is called a **local solution (minimum) of problem $\mathcal{P}$** if there exists ε such that $f(\boldsymbol{x}) \geq f(\boldsymbol{x}^\star)$ for all $\boldsymbol{x} \in \mathcal{X} \cap \boldsymbol{B}(\boldsymbol{x}^\star, \varepsilon)$.

Definition. A point $\boldsymbol{x}^\star \in \mathcal{X}$ is called a **global solution (minimum) of problem $\mathcal{P}$** if $f(\boldsymbol{x}) \geq f(\boldsymbol{x}^\star)$ for all $\boldsymbol{x} \in \mathcal{X}$.

Definition. Let $\boldsymbol{x} \in \mathcal{X}$. A unitary vector z is called a **feasible direction from $\boldsymbol{x}$** if there exists $\delta > 0$ such that

$$\boldsymbol{x} + \theta z \in \mathcal{X} \quad \forall \theta \in [0, \delta).$$

Active inequality constrains

Remark. If $\boldsymbol{x}^\star$ is a **local solution of $\mathcal{P}$**, then for all feasible directions z

$$f(\boldsymbol{x}^\star + \theta z) \geq f\left(\boldsymbol{x}^\star\right), \text{ for all } \; 0 \leq \theta < \delta.$$

Active inequality constrains

Remark. If $\boldsymbol{x}^\star$ is a **local solution of $\mathcal{P}$**, then for all feasible directions z

$$f(\boldsymbol{x}^\star + \theta z) \geq f(\boldsymbol{x}^\star), \text{ for all } 0 \leq \theta < \delta.$$

Definition. We introduce the following set.

$$\mathcal{I}(\boldsymbol{x}^\star) := \{j \;:\; h_j(\boldsymbol{x}^\star) = 0\}.$$

For such $j \in \mathcal{I}(\boldsymbol{x}^\star)$, we say that the inequality constraints h_j's are **saturated** or **active** at the solution $\boldsymbol{x}^\star$.

Feasible set, points, directions (cont.)

Lemma. Let $\boldsymbol{x}^\star$ a local solution of $\mathcal{P}$. Suppose $k \in \mathcal{I}(\boldsymbol{x}^\star)$. Let z a feasible direction from $\boldsymbol{x}^\star$. Then $z^T \nabla h_k(\boldsymbol{x}^\star) \geq 0$.

Proof. Assume $z^T \nabla h_k(\boldsymbol{x}^\star) < 0$. We have that

$$h_k(\boldsymbol{x}^\star + \theta z) = h_k(\boldsymbol{x}^\star) + \theta \cdot z^T \nabla h_k(\boldsymbol{x}^\star) + \varepsilon_k(\theta) = \theta \cdot z^T \nabla h_k(\boldsymbol{x}^\star) + \varepsilon_k(\theta)$$

where $\varepsilon_k(\theta) \to 0$ as $\theta \to 0$. Hence for θ small enough, $\theta \cdot z^T \nabla h_k(\boldsymbol{x}^\star) + \varepsilon_k(\theta) < 0$ and so $h_k(\boldsymbol{x}^\star + \theta z) < 0$, a contradiction with z being a feasible direction.

Lemma. Let $\boldsymbol{x}^\star$ a local solution of $\mathcal{P}$. Let z a feasible direction from $\boldsymbol{x}^\star$. Then $z^T \nabla g_j(\boldsymbol{x}^\star) = 0$ for all $j = 1, \ldots m$.

The linearizing cone $\mathcal{Z}^1(\boldsymbol{x}^\star)$

Definition. Assume previous notations. We define the linearizing cone of $\mathcal{X}$ at $\boldsymbol{x}^\star$ as

$$\mathcal{Z}^1(\boldsymbol{x}^\star) := \left\{ z \;\middle|\; \begin{array}{l} z^T \nabla h_k(\boldsymbol{x}^\star) \geq 0 \text{ if } k \in \mathcal{I}(\boldsymbol{x}^\star), \text{ and} \\ \\ z^T \nabla g_j(\boldsymbol{x}^\star) = 0 \; j = 1, \ldots m \end{array} \right\}$$

Corollary to the lemmas above: If z is a feasible direction from $\boldsymbol{x}^\star \in \mathcal{X}$ (that is, $(\boldsymbol{x}^\star + \theta z) \in \mathcal{X}$ for θ small), then $z \in \mathcal{Z}^1(\boldsymbol{x}^\star)$.

The set $\mathcal{Z}^2(\boldsymbol{x}^\star)$

Definition. Assume previous notations. We define the set (decreasing cone at $\boldsymbol{x}^\star$)

$$\mathcal{Z}^2(\boldsymbol{x}^\star) := \{z \mid z^T \nabla f(\boldsymbol{x}^\star) < 0\}$$

Lemma. If $z \in \mathcal{Z}^2(\boldsymbol{x}^\star)$, then $f(\boldsymbol{x}^\star + \theta z) < f(\boldsymbol{x}^\star)$ for θ small enough.

We expect $\mathcal{Z}^1(\boldsymbol{x}^\star) \cap \mathcal{Z}^2(\boldsymbol{x}^\star) = \emptyset$ at the solution $\boldsymbol{x}^\star$

The (generalized) Lagrangian associated to $\mathcal{P}$

Definition. A **generalized Lagrangian associated to $\mathcal{P}$** is the function

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

Definition. A solution point $\boldsymbol{x}^\star$ is called **regular** if the equality constrains and the active inequality constrains at $\boldsymbol{x}^\star$ have linearly independent gradient vectors.

Remark. This definition generalizes the previous technical condition of the Jacobian matrix $D(g)(\boldsymbol{x}^\star)$ having rank m.

The necessary condition

Theorem (Karush-Kuhn-Tucker (KKT) conditions). Let $\boldsymbol{x}^\star$ be a regular local minimum for $\mathcal{P}$. Then there exist unique vectors $\boldsymbol{\lambda}^\star = (\lambda_1^\star, \ldots, \lambda_m^\star)$ and $\boldsymbol{\mu}^\star = \left(\mu_1^\star, \ldots, \mu_p^\star\right)$ (called **Lagrange multipliers**) such that

$$\begin{aligned}
\nabla_{\boldsymbol{x}} L\left(\boldsymbol{x}^\star, \boldsymbol{\lambda}^\star, \boldsymbol{\mu}^\star\right) = \nabla f\left(\boldsymbol{x}^\star\right) - \sum_{j=1}^{m} \lambda_j \nabla g_j(\boldsymbol{x}) - \sum_{j=1}^{p} \mu_j \nabla h_j(\boldsymbol{x}) = 0,\\
\mu_j h_j\left(\boldsymbol{x}^\star\right) = 0,\ j = 1, \ldots m,\\
\mu_j \geq 0.
\end{aligned}$$

The necessary condition

Theorem (Karush-Kuhn-Tucker (KKT) conditions). Let $\boldsymbol{x}^\star$ be a regular local minimum for $\mathcal{P}$. Then there exist unique vectors $\boldsymbol{\lambda}^\star = (\lambda_1^\star, \ldots, \lambda_m^\star)$ and $\boldsymbol{\mu}^\star = \left(\mu_1^\star, \ldots, \mu_p^\star\right)$ (called **Lagrange multipliers**) such that

$$\nabla_{\boldsymbol{x}} L\left(\boldsymbol{x}^\star, \boldsymbol{\lambda}^\star, \boldsymbol{\mu}^\star\right) = \nabla f\left(\boldsymbol{x}^\star\right) - \sum_{j=1}^{m} \lambda_j \nabla g_j(\boldsymbol{x}) - \sum_{j=1}^{p} \mu_j \nabla h_j(\boldsymbol{x}) = 0,$$

$$\mu_j h_j\left(\boldsymbol{x}^\star\right) = 0,\ j = 1, \ldots m,$$

$$\mu_j \geq 0.$$

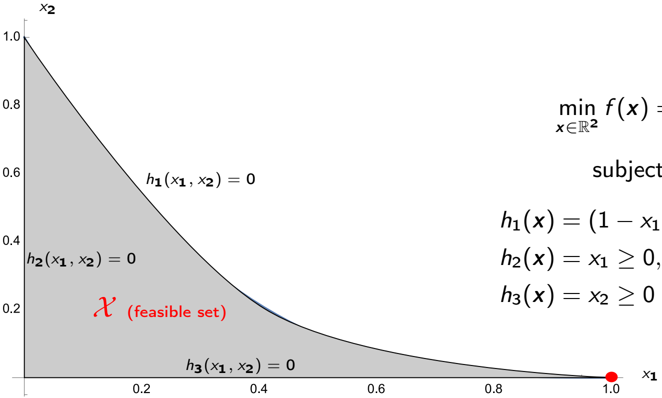
If f, g_j and h_j are $\mathcal{C}^2$-functions, then

$$\boldsymbol{z}^T \nabla_x^2 L\left(\boldsymbol{x}^\star, \boldsymbol{\lambda}^\star, \boldsymbol{\mu}^\star\right) \boldsymbol{z} \geq 0$$

for all $\boldsymbol{z} \in \mathbb{R}^n$ such that

$$\boldsymbol{z}^T \nabla g_j\left(\boldsymbol{x}^\star\right) = 0,\ j = 1, \ldots, m \quad \text{and} \quad \boldsymbol{z}^T \nabla h_k\left(\boldsymbol{x}^\star\right) = 0,\ k \in \mathcal{I}\left(\boldsymbol{x}^\star\right).$$

An example: non-regular local minima



$$\min_{\boldsymbol{x}\in\mathbb{R}^2} f(\boldsymbol{x}) = -x_1$$

subject to

$$h_1(\boldsymbol{x}) = (1 - x_1)^3 - x_2 \geq 0$$
$$h_2(\boldsymbol{x}) = x_1 \geq 0,$$
$$h_3(\boldsymbol{x}) = x_2 \geq 0$$

An example: non-regular local minima

Solution. It is easy to check that the point $\boldsymbol{x}^\star = (1,0)$ is a local minimum of f under the constrains. However

$$\nabla h_1(\boldsymbol{x}) = \left(-3(1-x_1)^2, -1\right), \quad \nabla h_2(\boldsymbol{x}) = (1,0), \quad \nabla h_3(\boldsymbol{x}) = (0,1),$$

and so, observe that $\nabla h_1(\boldsymbol{x}^\star) = (0,-1)$ and $\nabla h_2(\boldsymbol{x}^\star) = (0,1)$ are not linearly independent. Moreover, $\mu_2 = 0$ and

$$\nabla f(\boldsymbol{x}^\star) = (1,0) \neq \mu_1(0,-1) + \mu_3(0,1),$$

and so $\boldsymbol{x}^\star$ does not satisfy the necessary conditions.

Exercise. Prove that $\mathcal{Z}^1(\boldsymbol{x}^\star) \cap \mathcal{Z}^2(\boldsymbol{x}^\star) \neq \emptyset$. Indeed, this is the condition that characterizes non-regular candidates.

Turning to sufficient conditions

Theorem. Assume that all functions are of class $\mathcal{C}^2$. Assume that $\boldsymbol{x}^\star \in \mathbb{R}^n$, $\boldsymbol{\lambda}^\star \in \mathbb{R}^m$ and $\boldsymbol{\mu}^\star \in \mathbb{R}^p$ satisfy $g_j(\boldsymbol{x}^\star) = 0,\ j = 1, \dots, m,\ h_j(\boldsymbol{x}^\star) \geq 0,\ j = 1, \dots, p,$

$$\nabla_x L(\boldsymbol{x}^\star, \boldsymbol{\lambda}^\star, \boldsymbol{\mu}^\star) = 0, \quad \mu_j \geq 0, \quad \mu_j h_j(\boldsymbol{x}^\star) = 0,\ j = 1, \dots m,$$

and

$$\boldsymbol{z}^T \nabla^2_{\boldsymbol{x}}(L)(\boldsymbol{x}^\star, \boldsymbol{\lambda}^\star, \boldsymbol{\mu}^\star)\, \boldsymbol{z} \geq 0$$

for all $\boldsymbol{z} \in \mathbb{R}^n$ such that $\boldsymbol{z}^T \nabla g_j(\boldsymbol{x}^\star) = 0,\ j = 1, \dots, m$ and $\boldsymbol{z}^T \nabla h_k(\boldsymbol{x}^\star) = 0,\ k \in \mathcal{I}(\boldsymbol{x}^\star)$. Further assume that $\mu_k^\star > 0$ for all $k \in \mathcal{I}(\boldsymbol{x}^\star)$.

Then, $\boldsymbol{x}^\star$ is a strict local minimum of f subject to the constrains given by $\mathcal{P}$.

An interesting example

Exercise. Solve the following optimization problem in terms of the parameter $\beta > 0$.

$$\min_{\boldsymbol{x} \in \mathbb{R}^2} f(\boldsymbol{x}) = (x_1 - 1)^2 + x_2^2$$

subject to

$$h(x_1, x_2) = -x_1 + \beta x_2^2 \geq 0$$

Interpret the solutions geometrically in terms of the level curves and the restriction.