Optimization

Màster de Fonaments de Ciència de Dades

PART 2. Analysis

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

Chapter 4. Constrained optimization with inequalities. Optimality conditions

Chapter 5. Convex sets and functions

Chapter 4

Constrained optimization with inequalities.

Optimality conditions

The problem

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

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

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

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

$$\min_{\boldsymbol{x} \in \mathbb{R}^n} f(\boldsymbol{x})$$

$$\text{subject to:} \quad \begin{array}{ll} g_j(\boldsymbol{x}) = 0, & i = 1, \ldots, m \\ h_j(\boldsymbol{x}) \geq 0, & j = 1, \ldots, p. \end{array} \tag{2}$$

Constructing an equality constrained problem

Remark. Problem $\mathcal{P}$ may be written as an equality constrained problem by enlarging the number of variables.

$$
\begin{aligned}
&\min_{\boldsymbol{x}\in\mathbb{R}^n} f(\boldsymbol{x}) \\
\text{subject to:}\quad & g_j(\boldsymbol{x}) = 0, \qquad\qquad i = 1,\dots,m \\
& h_j(\boldsymbol{x}) - z_j^2 = 0, \qquad j = 1,\dots,p.
\end{aligned}
\tag{3}
$$

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

Definition. The set of points $\mathcal{X} \subset D$ satisfying conditions (12) 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 $x \in \mathcal{X} \cap \boldsymbol{B}(\boldsymbol{x}^\star, \varepsilon)$.

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 $x \in \mathcal{X}$.

Definition. Let $\boldsymbol{x} \in \mathcal{X}$. A unitary vector z is called a feasible direction from x if for small enough $\delta > 0$ we have that if $|\theta| < \delta$ then

$$\{y \in \mathbb{R}^n \mid y = x + \theta z\} \subset \mathcal{X}$$

Active inequality constrains

Remark. The previous notion of local solution of $\mathcal{P}$ writes as

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

with z being a feasible direction.

Definition. We introduce the following set.

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

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

Feasible set and points and directions

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

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

where $\varepsilon_k(\theta) \to 0$ as $\theta \to 0$. Hence for θ small enough $\theta \nabla h_k(\boldsymbol{x}^\star) + \varepsilon_k(\theta) < 0$ and so $h_k(\boldsymbol{x}^\star + \theta z) < 0$, a contradiction with z 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 notation. We define the linearizing cone of $\mathcal{X}$ at $\boldsymbol{x}^\star$ as

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

Lemma. 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 $z \in \mathcal{Z}^1(\boldsymbol{x}^\star)$.

Proof. We argue by contradiction. If $z \notin \mathcal{Z}^1(\boldsymbol{x}^\star)$ then either $z^T \nabla h_k(\boldsymbol{x}^\star) < 0$ for $k \in \mathcal{I}(\boldsymbol{x}^\star)$, or $z^T \nabla g_j(\boldsymbol{x}^\star) \neq 0$. Using linear expansion of h_k, $k \in \mathcal{I}(\boldsymbol{x}^\star)$ and g_j, $j = 1, \ldots m$ these imply that either $h_k(\boldsymbol{x}^\star + \theta z) < 0$, $k \in \mathcal{I}(\boldsymbol{x}^\star)$ or $g_j(\boldsymbol{x}^\star + \theta z) \neq 0$, $j = 1, \ldots m$, for θ small enough, respectively.

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

Definition. Assume previous notation. We define the set

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

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

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

Definition. Assume previous notation. We define the generalized Lagrangian associated to $\mathcal{P}$ as the function

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

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

Necessary conditions for minimum

Theorem (Karush-Kuhn-Tucker conditions). Assume previous notation. Let $\boldsymbol{x}^\star$ be a regular local minimum for $\mathcal{P}$. Then, there exist (unique) Lagrange multiplier vectors $\boldsymbol{\lambda}^\star = (\lambda_1^\star, \dots, \lambda_m^\star)$ and $\boldsymbol{\mu}^\star = (\mu_1^\star, \dots, \mu_p^\star)$ such that

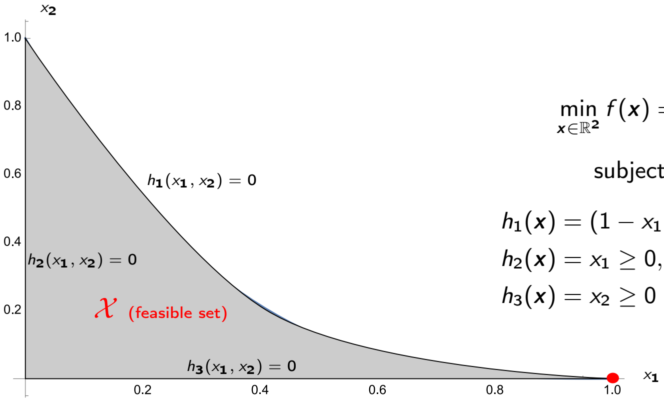
$$\nabla_x L(\boldsymbol{x}^\star, \boldsymbol{\lambda}^\star, \boldsymbol{\mu}^\star) = \nabla f(\boldsymbol{x}^\star) - \sum_{j=1}^{m} \lambda_j \nabla g_j(\boldsymbol{x}) - \sum_{j=1}^{p} \mu_j \nabla h_j(\boldsymbol{x}) = 0.$$

Moreover, $\mu_j \geq 0$ and $\mu_j h_j(\boldsymbol{x}^\star) = 0$, $j = 1, \dots m$. If f, g_j and h_j are $\mathcal{C}^2$-functions then

$$y^T H_x(L)(\boldsymbol{x}^\star, \boldsymbol{\lambda}^\star, \boldsymbol{\mu}^\star)\, y \geq 0$$

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

An exemple: non-regular local minimums



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

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 exemple: non-regular local minimums

Solution. Easily we can see that the point $\boldsymbol{x}^\star = (1,0)$ is a local minimum of f under the constrains. However

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

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

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

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

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

Turning to sufficient conditions

Theorem. Assume previous notation and 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 $g_j(\boldsymbol{x}^\star) = 0,\ j = 1, \ldots, m,\ h_j(\boldsymbol{x}^\star) \geq 0,\ j = 1, \ldots, p,$

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

and

$$y^T H_x(L)(\boldsymbol{x}^\star, \boldsymbol{\lambda}^\star, \boldsymbol{\mu}^\star)\, y \geq 0$$

for all $y \in \mathbb{R}^n$ such that $(\nabla g_j(\boldsymbol{x}^\star))^T y = 0,\ j = 1, \ldots, m$ and $(\nabla h_k(\boldsymbol{x}^\star))^T y = 0,\ k \in \mathcal{I}(\boldsymbol{x}^\star)$. Assume also 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. Discuss 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.

Saddlepoints of the Lagrangian

Definition. Let $\boldsymbol{x} \in E_x \subset \mathbb{R}^n$ and $\boldsymbol{y} \in E_y \subset \mathbb{R}^m$. Let φ a (continuous) function $\varphi : E_{\boldsymbol{x}} \times E_{\boldsymbol{y}} \to \mathbb{R}$. We say that a point $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}) \in E_{\boldsymbol{x}} \times E_{\boldsymbol{y}}$ is a saddlepoint of φ if

$$\varphi\left(\hat{\boldsymbol{x}}, \boldsymbol{y}\right) \leq \varphi\left(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}\right) \leq \varphi\left(\boldsymbol{x}, \hat{\boldsymbol{y}}\right).$$

Definition. We define the problem $(\mathcal{S})$ as follows. Find a saddlepoint $\hat{\boldsymbol{x}} \in \mathbb{R}^n$, $\hat{\boldsymbol{\lambda}} \in \mathbb{R}^m$ and $\hat{\boldsymbol{\mu}} \in \mathbb{R}^p$ with $\boldsymbol{\mu} \geq 0$ for the Lagrangian. That is

$$L\left(\hat{\boldsymbol{x}}, \boldsymbol{\lambda}, \boldsymbol{\mu}\right) \leq L\left(\hat{\boldsymbol{x}}, \hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\mu}}\right) \leq L\left(\boldsymbol{x}, \hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\mu}}\right) \tag{4}$$

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

Connecting $(\mathcal{P})$ with $(\mathcal{S})$

Theorem. If $\left(\hat{\boldsymbol{x}}, \hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\mu}}\right)$ is a solution of $(\mathcal{S})$ then $\hat{x}$ is a solution of $(\mathcal{P})$.

Proof. Assume $\left(\hat{\boldsymbol{x}}, \hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\mu}}\right)$ is a solution of $(\mathcal{S})$. Then from (4) we have

$$\sum_{j=1}^{m} \left(\hat{\lambda}_j - \lambda_j\right) g_j\left(\hat{\boldsymbol{x}}\right) + \sum_{j=1}^{p} \left(\hat{\mu}_j - \mu_j\right) h_j\left(\hat{\boldsymbol{x}}\right) \leq 0 \tag{a}$$

$$f\left(\hat{\boldsymbol{x}}\right) \leq f\left(\boldsymbol{x}\right) + \sum_{j=1}^{m} \hat{\lambda}_j \left(g_j\left(\hat{\boldsymbol{x}}\right) - g_j\left(\boldsymbol{x}\right)\right) + \sum_{j=1}^{p} \hat{\mu}_j \left(h_j\left(\hat{\boldsymbol{x}}\right) - h_j\left(\boldsymbol{x}\right)\right) \tag{b}$$

After some computations we conclude that

$$g_j\left(\hat{\boldsymbol{x}}\right) = 0,\ j = 1, \ldots, m \quad \text{and} \quad \hat{\boldsymbol{\mu}}_j h_j\left(\hat{\boldsymbol{x}}\right) = 0,\ j = 1, \ldots p.$$

Hence

$$f\left(\hat{\boldsymbol{x}}\right) \leq f\left(\boldsymbol{x}\right) - \sum_{j=1}^{m} \hat{\lambda}_j g_j\left(\boldsymbol{x}\right) - \sum_{j=1}^{p} \hat{\mu}_j h_j\left(\boldsymbol{x}\right).$$

Connecting $(\mathcal{P})$ with $(\mathcal{S})$

Theorem. Suppose all functions are differentiable and suppose that $\left(\hat{\boldsymbol{x}}, \hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\mu}}\right)$ is a solution of $(\mathcal{S})$. Then $\mathcal{Z}^1(\hat{\boldsymbol{x}}) \cup \mathcal{Z}^2(\hat{\boldsymbol{x}}) \neq \emptyset$ and

$$\nabla_x L\left(\hat{\boldsymbol{x}}, \hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\mu}}\right) = 0 \quad \hat{\boldsymbol{\mu}}_j h_j(\hat{\boldsymbol{x}}) = 0, \; j = 1, \ldots p,$$

with $\hat{\boldsymbol{\mu}} \geq 0$.

These were conditions for minimum of f under general inequality constrains.