CentraleSupelec
ST7 – Optimization
Part VII: Lagrange multipliers method

jean-christophe@pesquet.eu

Constrained optimization problem

Let $\mathcal{H}$ be a Hilbert space. Let $f\colon \mathcal{H} \to \,]-\infty,+\infty]$.
Let $(m,q) \in \mathbb{N}^2$. For every $i \in \{1,\ldots,m\}$, let $g_i\colon \mathcal{H} \to \mathbb{R}$ and for every $j \in \{1,\ldots,q\}$, let $h_j\colon \mathcal{H} \to \mathbb{R}$.
Let

constrains.

$$C = \{x \in \mathcal{H} \mid (\forall i \in \{1,\ldots,m\})\ \ g_i(x) = 0 \\ (\forall j \in \{1,\ldots,q\})\ \ h_j(x) \leq 0\}.$$

We want to:

$$\text{Find} \quad \widehat{x} \in \underset{x \in C}{\operatorname{Argmin}}\ f(x).$$

Remark: A vector $x \in \mathcal{H}$ is said to be feasible if $x \in \operatorname{dom} f \cap C$.

Definitions

The Lagrange function (or Lagrangian) associated with the previous problem is defined as

$$(\forall x \in \mathcal{H})(\forall \nu = (\nu^{(i)})_{1\le i\le m} \in \mathbb{R}^m)(\forall \lambda = (\lambda^{(j)})_{1\le j\le q} \in [0,+\infty[^q)$$
$$\mathcal{L}(x,\nu,\lambda) = f(x) + \sum_{i=1}^{m} \nu^{(i)} g_i(x) + \sum_{j=1}^{q} \lambda^{(j)} h_j(x).$$

The vectors ν and λ are called Lagrange multipliers.

Remark:

- $\operatorname{dom}\mathcal{L} = \operatorname{dom} f \times \mathbb{R}^m \times [0,+\infty[^q$.

Definitions

The Lagrange function (or Lagrangian) associated with the previous problem is defined as

$$(\forall x \in \mathcal{H})(\forall \nu = (\nu^{(i)})_{1 \le i \le m} \in \mathbb{R}^m)(\forall \lambda = (\lambda^{(j)})_{1 \le j \le q} \in [0, +\infty[^q)$$

$$\mathcal{L}(x, \nu, \lambda) = f(x) + \sum_{i=1}^{m} \nu^{(i)} g_i(x) + \sum_{j=1}^{q} \lambda^{(j)} h_j(x).$$

The vectors ν and λ are called Lagrange multipliers.

Remark:

- When $q = 0$ (only equality constraints), the Lagrange function simplifies to

$$(\forall x \in \mathcal{H})(\forall \nu = (\nu^{(i)})_{1 \le i \le m} \in \mathbb{R}^m) \;\; \mathcal{L}(x, \nu) = f(x) + \sum_{i=1}^{m} \nu^{(i)} g_i(x).$$

Definitions

The Lagrange function (or Lagrangian) associated with the previous problem is defined as

$$(\forall x \in \mathcal{H})(\forall \nu = (\nu^{(i)})_{1 \le i \le m} \in \mathbb{R}^m)(\forall \lambda = (\lambda^{(j)})_{1 \le j \le q} \in [0, +\infty[^q)$$

$$\mathcal{L}(x, \nu, \lambda) = f(x) + \sum_{i=1}^{m} \nu^{(i)} g_i(x) + \sum_{j=1}^{q} \lambda^{(j)} h_j(x).$$

The vectors ν and λ are called Lagrange multipliers.

Remark:

- When $m = 0$ (only inequality constraints), the Lagrange function simplifies to

$$(\forall x \in \mathcal{H})(\forall \lambda = (\lambda^{(j)})_{1 \le j \le q} \in [0, +\infty[^q)$$

$$\mathcal{L}(x, \lambda) = f(x) + \sum_{j=1}^{q} \lambda^{(j)} h_j(x).$$

Lagrange duality

Let $\overline{\mathcal{L}}$ be the primal Lagrange function defined as

$$(\forall x \in \mathcal{H}) \quad \overline{\mathcal{L}}(x) = \sup_{\nu \in \mathbb{R}^m, \lambda \in [0,+\infty[^q} \mathcal{L}(x,\nu,\lambda).$$

Then, for every $x \in C$, $\overline{\mathcal{L}}(x) = f(x)$.

Proof: For every $(x,\nu,\lambda) \in C \times \mathbb{R}^m \times [0,+\infty[^q$

$$\mathcal{L}(x,\nu,\lambda) = f(x) + \sum_{i=1}^{m} \underbrace{\nu^{(i)} g_i(x)}_{=0} + \sum_{j=1}^{q} \underbrace{\lambda^{(j)} h_j(x)}_{\leq 0} \leq f(x)$$

and $\overline{\mathcal{L}}(x) = \sup_{\nu \in \mathbb{R}^m, \lambda \in [0,+\infty[^q} \mathcal{L}(x,\nu,\lambda) = f(x)$.

Lagrange duality

Let $\overline{\mathcal{L}}$ be the primal Lagrange function defined as

$$(\forall x \in \mathcal{H}) \quad \overline{\mathcal{L}}(x) = \sup_{\nu \in \mathbb{R}^m, \lambda \in [0,+\infty[^q} \mathcal{L}(x, \nu, \lambda).$$

Then, for every $x \in C$, $\overline{\mathcal{L}}(x) = f(x)$.

Let $\underline{\mathcal{L}}$ be the dual Lagrange function defined as

$$(\forall (\nu, \lambda) \in \mathbb{R}^m \times [0, +\infty[^q) \quad \underline{\mathcal{L}}(\nu, \lambda) = \inf_{x \in \mathcal{H}} \mathcal{L}(x, \nu, \lambda)$$

Then, $-\underline{\mathcal{L}}$ is convex and s.c.i.

Lagrange duality

Let $\overline{\mathcal{L}}$ be the primal Lagrange function defined as

$$(\forall x \in \mathcal{H}) \quad \overline{\mathcal{L}}(x) = \sup_{\nu \in \mathbb{R}^m, \lambda \in [0,+\infty[^q} \mathcal{L}(x, \nu, \lambda).$$

Then, for every $x \in C$, $\overline{\mathcal{L}}(x) = f(x)$.

Let $\underline{\mathcal{L}}$ be the dual Lagrange function defined as

$$(\forall (\nu, \lambda) \in \mathbb{R}^m \times [0,+\infty[^q) \quad \underline{\mathcal{L}}(\nu, \lambda) = \inf_{x \in \mathcal{H}} \mathcal{L}(x, \nu, \lambda)$$

Then, $-\underline{\mathcal{L}}$ is convex and s.c.i.

Proof: For every $(\nu, \lambda) \in \mathbb{R}^m \times [0,+\infty[^q$,

$$-\underline{\mathcal{L}}(\nu, \lambda) = \sup_{x \in \operatorname{dom} f} (-\mathcal{L}(x, \nu, \lambda)).$$

$-\underline{\mathcal{L}}$ is thus the supremum of a set of affine functions.

Lagrange duality

For every $(x, \nu, \lambda) \in \mathcal{H} \times \mathbb{R}^m \times [0, +\infty[^q$,

$$\underline{\mathcal{L}}(\nu, \lambda) \leq \overline{\mathcal{L}}(x).$$

weak
strong.

In addition,

$$\sup_{\nu \in \mathbb{R}^m, \lambda \in [0,+\infty[^q} \underline{\mathcal{L}}(\nu, \lambda) \leq \mu = \inf_{x \in C} f(x).$$

Lagrange duality

For every $(x, \nu, \lambda) \in \mathcal{H} \times \mathbb{R}^m \times [0, +\infty[^q$,

$$\underline{\mathcal{L}}(\nu, \lambda) \leq \overline{\mathcal{L}}(x).$$

In addition,

$$\sup_{\nu \in \mathbb{R}^m, \lambda \in [0, +\infty[^q} \underline{\mathcal{L}}(\nu, \lambda) \leq \mu = \inf_{x \in C} f(x).$$

Proof: We have, for every $(x, \nu, \lambda) \in \mathcal{H} \times \mathbb{R}^m \times [0, +\infty[^q$,

$$\inf_{x'} \mathcal{L}(x', \nu, \lambda) \leq \mathcal{L}(x, \nu, \lambda) \leq \sup_{\nu', \lambda'} \mathcal{L}(x, \nu', \lambda')$$

$$\Rightarrow \qquad \underline{\mathcal{L}}(\nu, \lambda) \leq \overline{\mathcal{L}}(x).$$

We deduce that, for every $x \in C$, $\underline{\mathcal{L}}(\nu, \lambda) \leq \overline{\mathcal{L}}(x) = f(x)$, which yields the last inequality.

Saddle points 鞍点

$(\widehat{x}, \widehat{\nu}, \widehat{\lambda}) \in \mathcal{H} \times \mathbb{R}^m \times [0, +\infty[^q$ is a saddle point of $\mathcal{L}$ if

$$\big(\forall (x, \nu, \lambda) \in \mathcal{H} \times \mathbb{R}^m \times [0, +\infty[^q\big) \qquad \mathcal{L}(\widehat{x}, \nu, \lambda) \leq \mathcal{L}(\widehat{x}, \widehat{\nu}, \widehat{\lambda}) \leq \mathcal{L}(x, \widehat{\nu}, \widehat{\lambda}).$$

Remark: If there exists a feasible point and $(\widehat{x}, \widehat{\nu}, \widehat{\lambda})$ is a saddle point of $\mathcal{L}$, then it follows from the right inequality that $\widehat{x} \in \operatorname{dom} f$.

Saddle points

$(\widehat{x}, \widehat{\nu}, \widehat{\lambda}) \in \mathcal{H} \times \mathbb{R}^m \times [0, +\infty[^q$ is a saddle point of $\mathcal{L}$ if

$$\big(\forall (x, \nu, \lambda) \in \mathcal{H} \times \mathbb{R}^m \times [0, +\infty[^q\big) \qquad \mathcal{L}(\widehat{x}, \nu, \lambda) \le \mathcal{L}(\widehat{x}, \widehat{\nu}, \widehat{\lambda}) \le \mathcal{L}(x, \widehat{\nu}, \widehat{\lambda}).$$

Theorem

$(\widehat{x}, \widehat{\nu}, \widehat{\lambda}) \in \mathcal{H} \times \mathbb{R}^m \times [0, +\infty[^q$ is a saddle point of $\mathcal{L}$ if and only if

$$(\forall x \in \mathcal{H}) \quad \overline{\mathcal{L}}(\widehat{x}) \le \overline{\mathcal{L}}(x)$$

$$(\forall (\nu, \lambda) \in \mathbb{R}^m \times [0, +\infty[^q) \quad \underline{\mathcal{L}}(\nu, \lambda) \le \underline{\mathcal{L}}(\widehat{\nu}, \widehat{\lambda})$$

$$\overline{\mathcal{L}}(\widehat{x}) = \underline{\mathcal{L}}(\widehat{\nu}, \widehat{\lambda}).$$

Saddle points

<u>Proof</u>: If $(\widehat{x}, \widehat{\nu}, \widehat{\lambda})$ is a saddle point of $\mathcal{L}$ then, for every $(x', \nu', \lambda') \in \mathcal{H} \times \mathbb{R}^m \times [0, +\infty[^q$,

$$
\begin{aligned}
& \mathcal{L}(\widehat{x}, \nu', \lambda') \leq \mathcal{L}(\widehat{x}, \widehat{\nu}, \widehat{\lambda}) \leq \mathcal{L}(x', \widehat{\nu}, \widehat{\lambda}) \\
\Rightarrow \quad & \sup_{\nu', \lambda'} \mathcal{L}(\widehat{x}, \nu', \lambda') \leq \inf_{x'} \mathcal{L}(x', \widehat{\nu}, \widehat{\lambda}) \\
\Leftrightarrow \quad & \overline{\mathcal{L}}(\widehat{x}) \leq \underline{\mathcal{L}}(\widehat{\nu}, \widehat{\lambda}) \\
\Rightarrow \quad & \inf_{x} \overline{\mathcal{L}}(x) \leq \overline{\mathcal{L}}(\widehat{x}) \leq \underline{\mathcal{L}}(\widehat{\nu}, \widehat{\lambda}) \leq \sup_{\nu, \lambda} \underline{\mathcal{L}}(\nu, \lambda).
\end{aligned}
$$

In addition

$$
\sup_{\nu, \lambda} \underline{\mathcal{L}}(\nu, \lambda) \leq \inf_{x} \overline{\mathcal{L}}(x).
$$

Therefore, $\inf_x \overline{\mathcal{L}}(x) = \sup_{\nu, \lambda} \underline{\mathcal{L}}(\nu, \lambda)$.
We deduce that $\overline{\mathcal{L}}(\widehat{x}) = \inf_x \overline{\mathcal{L}}(x)$, $\underline{\mathcal{L}}(\widehat{\nu}, \widehat{\lambda}) = \sup_{\nu, \lambda} \underline{\mathcal{L}}(\nu, \lambda)$, and $\overline{\mathcal{L}}(\widehat{x}) = \underline{\mathcal{L}}(\widehat{\nu}, \widehat{\lambda})$.

Saddle points

Proof: Conversely, if the last condition holds, then

$$\begin{aligned}\mathcal{L}(\widehat{x},\widehat{\nu},\widehat{\lambda}) \le \sup_{\nu,\lambda} \mathcal{L}(\widehat{x},\nu,\lambda) &= \overline{\mathcal{L}}(\widehat{x})\\ &= \underline{\mathcal{L}}(\widehat{\nu},\widehat{\lambda}) = \inf_{x} \mathcal{L}(x,\widehat{\nu},\widehat{\lambda}).\end{aligned}$$

Similarly,

$$\begin{aligned}\mathcal{L}(\widehat{x},\widehat{\nu},\widehat{\lambda}) \ge \inf_{x} \mathcal{L}(x,\widehat{\nu},\widehat{\lambda}) &= \underline{\mathcal{L}}(\widehat{\nu},\widehat{\lambda})\\ &= \overline{\mathcal{L}}(\widehat{x}) = \sup_{\nu,\lambda} \mathcal{L}(\widehat{x},\nu,\lambda).\end{aligned}$$

In conclusion, for every $(x,\nu,\lambda) \in \mathcal{H} \times \mathbb{R}^m \times [0,+\infty[^q$,

$$\mathcal{L}(\widehat{x},\nu,\lambda) \le \mathcal{L}(\widehat{x},\widehat{\nu},\widehat{\lambda}) \le \mathcal{L}(x,\widehat{\nu},\widehat{\lambda}).$$

Sufficient condition for a constrained minimum

Assume that there exists a feasible point. 充分条件.

If $(\widehat{x}, \widehat{\nu}, \widehat{\lambda}) \in \mathcal{H} \times \mathbb{R}^m \times [0, +\infty[^q$ is a saddle point of $\mathcal{L}$, then $\widehat{x}$ is a minimizer of f over C.

In addition, $\mu = f(\widehat{x}) = \underline{\mathcal{L}}(\widehat{\nu}, \widehat{\lambda})$ and the complementary slackness condition holds:

$$(\forall j \in \{1, \dots, q\}) \qquad \widehat{\lambda}_j h_j(\widehat{x}) = 0.$$

Sufficient condition for a constrained minimum

Assume that there exists a feasible point.
If $(\widehat{x}, \widehat{\nu}, \widehat{\lambda}) \in \mathcal{H} \times \mathbb{R}^m \times [0, +\infty[^q$ is a saddle point of $\mathcal{L}$,
then $\widehat{x}$ is a minimizer of f over C.
In addition, $\mu = f(\widehat{x}) = \underline{\mathcal{L}}(\widehat{\nu}, \widehat{\lambda})$ and the complementary slackness condition holds:

$$(\forall j \in \{1, \dots, q\}) \qquad \widehat{\lambda}_j h_j(\widehat{x}) = 0.$$

Proof: We know that $\widehat{x} \in \operatorname{dom} f$. We have, for every $(\nu, \lambda) \in \mathbb{R}^m \times [0, +\infty[^q$, $\mathcal{L}(\widehat{x}, \nu, \lambda) \leq \mathcal{L}(\widehat{x}, \widehat{\nu}, \widehat{\lambda})$.
For every $\nu' = (\nu'^{(i)})_{1 \leq i \leq m} \in \mathbb{R}^m$ by setting $\nu = \widehat{\nu} + \nu'$ and $\lambda = \widehat{\lambda}$,

$$\sum_{i=1}^{m} \nu'^{(i)} g_i(\widehat{x}) \leq 0$$

and, for every $\lambda' = (\lambda'^{(j)})_{1 \leq j \leq q} \in [0, +\infty[^q$, by setting $\nu = \widehat{\nu}$ and $\lambda = \widehat{\lambda} + \lambda'$,

$$\sum_{j=1}^{q} \lambda'^{(j)} h_j(\widehat{x}) \leq 0.$$

We deduce that $\begin{cases} (\forall i \in \{1, \dots, m\}) & g_i(\widehat{x}) = 0 \\ (\forall j \in \{1, \dots, q\}) & h_j(\widehat{x}) \leq 0, \end{cases}$ i.e. $\widehat{x} \in C$.

Sufficient condition for a constrained minimum

Assume that there exists a feasible point.
If $(\widehat{x}, \widehat{\nu}, \widehat{\lambda}) \in \mathcal{H} \times \mathbb{R}^m \times [0, +\infty[^q$ is a saddle point of $\mathcal{L}$, then $\widehat{x}$ is a minimizer of f over C.
In addition, $\mu = f(\widehat{x}) = \underline{\mathcal{L}}(\widehat{\nu}, \widehat{\lambda})$ and the complementary slackness condition holds:

$$(\forall j \in \{1, \ldots, q\}) \qquad \widehat{\lambda}_j h_j(\widehat{x}) = 0.$$

Proof: We have shown that $(\forall i \in \{1, \ldots, m\})\ g_i(\widehat{x}) = 0$ and $(\forall j \in \{1, \ldots, q\})\ h_j(\widehat{x}) \leq 0 \Rightarrow \widehat{\lambda}_j h_j(\widehat{x}) \leq 0$.
Then, since

$$\mathcal{L}(\widehat{x}, \widehat{\nu}, 0) \leq \mathcal{L}(\widehat{x}, \widehat{\nu}, \widehat{\lambda}),$$

we have

$$\sum_{j=1}^{m} \widehat{\lambda}_j h_j(\widehat{x}) \geq 0,$$

which implies that the complementary slackness condition holds.

Sufficient condition for a constrained minimum

Assume that there exists a feasible point.
If $(\widehat{x}, \widehat{\nu}, \widehat{\lambda}) \in \mathcal{H} \times \mathbb{R}^m \times [0, +\infty[^q$ is a saddle point of $\mathcal{L}$, then $\widehat{x}$ is a minimizer of f over C.
In addition, $\mu = f(\widehat{x}) = \underline{\mathcal{L}}(\widehat{\nu}, \widehat{\lambda})$ and the complementary slackness condition holds:

$$(\forall j \in \{1, \dots, q\}) \qquad \widehat{\lambda}_j h_j(\widehat{x}) = 0.$$

Proof: As $\widehat{x} \in C$ and the complementary slackness condition holds, $\mathcal{L}(\widehat{x}, \widehat{\nu}, \widehat{\lambda}) = f(\widehat{x})$. Furthermore,

$$\begin{aligned}(\forall x \in C) \qquad & \mathcal{L}(\widehat{x}, \widehat{\nu}, \widehat{\lambda}) \leq \mathcal{L}(x, \widehat{\nu}, \widehat{\lambda}) \\ \Leftrightarrow \quad & f(\widehat{x}) \leq f(x) + \sum_{i=1}^{m} \widehat{\nu}_i g_i(x) + \sum_{j=1}^{q} \widehat{\lambda}_j h_j(x) \leq f(x).\end{aligned}$$

Finally, as a consequence of previous results, $f(\widehat{x}) = \overline{\mathcal{L}}(\widehat{x}) = \underline{\mathcal{L}}(\widehat{\nu}, \widehat{\lambda})$.

Convex case

Assume that f is a convex function, $(g_i)_{1\leq i\leq m}$ are affine functions and $(h_j)_{1\leq j\leq q}$ are convex functions. Assume that the Slater condition holds, i.e. there exists $\overline{x} \in \mathrm{int}\,(\mathrm{dom}\, f)$ such that

$$\begin{aligned} (\forall i \in \{1,\ldots,m\}) \qquad & g_i(\overline{x}) = 0 \\ (\forall j \in \{1,\ldots,q\}) \qquad & h_j(\overline{x}) < 0. \end{aligned}$$

If $\widehat{x}$ is a minimizer of f over C, then there exists $\widehat{\nu} \in \mathbb{R}^m$ and $\widehat{\lambda} \in [0,+\infty[^q$ such that $(\widehat{x},\widehat{\nu},\widehat{\lambda})$ is a saddle point of the Lagrangian.

Convex case

Assume that f is a convex function, $(g_i)_{1\le i\le m}$ are affine functions and $(h_j)_{1\le j\le q}$ are convex functions. Assume that the Slater condition holds, i.e. there exists $\overline{x} \in \operatorname{int}(\operatorname{dom} f)$ such that

$$\begin{aligned} (\forall i \in \{1,\dots,m\}) \qquad & g_i(\overline{x}) = 0 \\ (\forall j \in \{1,\dots,q\}) \qquad & h_j(\overline{x}) < 0. \end{aligned}$$

$\widehat{x}$ is a minimizer of f over C if and only if there exists $\widehat{\nu} \in \mathbb{R}^m$ and $\widehat{\lambda} \in [0,+\infty[^q$ such that $(\widehat{x},\widehat{\nu},\widehat{\lambda})$ is a saddle point of the Lagrangian.

Proof: Combine the two previous results.

Convex case

Assume that f is a convex function, $(g_i)_{1\le i\le m}$ are affine functions and $(h_j)_{1\le j\le q}$ are convex functions. Assume that the Slater condition holds, i.e. there exists $\overline{x} \in \operatorname{int}(\operatorname{dom} f)$ such that

$$\begin{aligned} (\forall i \in \{1,\ldots,m\}) \qquad & g_i(\overline{x}) = 0 \\ (\forall j \in \{1,\ldots,q\}) \qquad & h_j(\overline{x}) < 0. \end{aligned}$$

$\widehat{x}$ is a minimizer of f over C if and only if there exists $\widehat{\nu} \in \mathbb{R}^m$ and $\widehat{\lambda} \in [0,+\infty[^q$ such that $(\widehat{x},\widehat{\nu},\widehat{\lambda})$ is a saddle point of the Lagrangian.

Remark: Under the assumptions of the above theorem, if $\widehat{x}$ is a minimizer of f over C then $\mathcal{L}(\cdot,\widehat{\nu},\widehat{\lambda})$ is a convex function which is minimum at $\widehat{x}$. This optimality condition is often used to calculate $\widehat{x}$, in conjunction with the equality constraints and the complementary slackness condition.

Convex case

Only equality constraints:

Assume that f is a convex function and $(g_i)_{1\leq i\leq m}$ are affine functions. Assume that the Slater condition holds, i.e. there exists $\overline{x} \in \operatorname{int}(\operatorname{dom} f)$ such that

$$(\forall i \in \{1,\ldots,m\}) \qquad g_i(\overline{x}) = 0.$$

$\widehat{x}$ is a minimizer of f over C if and only if there exists $\widehat{\nu} \in \mathbb{R}^m$ such that $(\widehat{x},\widehat{\nu})$ is a saddle point of the Lagrangian.

Convex case

Only inequality constraints:

Assume that f is a convex function, and $(h_j)_{1\leq j\leq q}$ are convex functions. Assume that the Slater condition holds, i.e. there exists $\overline{x} \in \operatorname{dom} f$ such that

$$(\forall j \in \{1,\ldots,q\}) \qquad h_j(\overline{x}) < 0.$$

$\widehat{x}$ is a minimizer of f over C if and only if there exists $\widehat{\lambda} \in [0,+\infty[^q$ such that $(\widehat{x},\widehat{\lambda})$ is a saddle point of the Lagrangian.

Exercise 1

One wants to minimize the production cost of a factory.
The factory produces cars in quantity x_1 and trucks in quantity x_2. The production of cars and trucks require $\psi_1(x_1)$ and $\psi_2(x_2)$ machine tools, respectively. The overall number of used machine tools is equal to c. The production costs of cars and trucks are equal to $\varphi_1(x_1)$ and $\varphi_2(x_2)$, respectively.
Solve this problem by the Lagrange multiplier method, when

$$\begin{aligned}
\varphi_1(x_1) &= (x_1 - 100)^2 \\
\varphi_2(x_2) &= 2(x_2 - 50)^2 \\
\psi_1(x_1) &= x_1 \\
\psi_2(x_2) &= x_2 \\
c &= 90.
\end{aligned}$$

Exercise 2

Let f be defined as

$$(\forall x \in \mathbb{R}^N) \qquad f(x) = \frac{1}{2}x^\top Qx + c^\top x.$$

where $Q \in \mathbb{R}^{N\times N}$ is a definite positive matrix and $c \in \mathbb{R}^N$.
We are interested in finding a minimizer of f subject to the constraint:

$$Ax = b$$

where $A \in \mathbb{R}^{m\times N}$ is a matrix of rank m and $b \in \mathbb{R}^m$.

1. Show that the problem has a unique solution.
2. By using the Lagrange multipliers method, find the expression of the solution.

Exercise 3

Let f be defined as

$$(\forall x = (x^{(i)})_{1 \leq i \leq N} \in [0, +\infty[^N) \qquad f(x) = \sum_{i=1}^{N} x^{(i)} \ln(x^{(i)}),$$

with $N > 1$. Find a minimizer of f on $[0, +\infty[^N$ subject to the constraints

$$\sum_{i=1}^{N} x^{(i)} = 1$$

$$\sum_{i=1}^{P} x^{(i)} = q,$$

where $P \in \{1, \ldots, N-1\}$ and $q \in]0, 1[$.

Differentiable case 可微

Assume that f, $(g_i)_{1\leq i\leq m}$, and $(h_j)_{1\leq j\leq q}$ are continuously differentiable on $\mathcal{H} = \mathbb{R}^N$.
If $\widehat{x}$ is a local minimum of f over C then the Fritz-John conditions hold, i.e. there exists a nonzero vector $(\alpha_0, \alpha_1, \ldots, \alpha_m, \alpha_{m+1}, \ldots, \alpha_{m+q}) \in [0, +\infty[\times \mathbb{R}^m \times [0, +\infty[^q$ such that

$$\alpha_0 \nabla f(\widehat{x}) + \sum_{i=1}^{m} \alpha_i \nabla g_i(\widehat{x}) + \sum_{j=1}^{q} \alpha_{m+j} \nabla h_j(\widehat{x}) = 0$$

$$(\forall j \in \{1, \ldots, q\}) \qquad \alpha_{m+j} h_j(\widehat{x}) = 0.$$

Differentiable case

Karush-Kuhn-Tucker (KKT) theorem

Assume that f, $(g_i)_{1\le i\le m}$, and $(h_j)_{1\le j\le q}$ are continuously differentiable on $\mathcal{H} = \mathbb{R}^N$.
Assume that $\widehat{x}$ is a local minimizer of f over C satisfying the following Mangasarian-Fromovitz constraint qualification conditions:

- (i) $\{\nabla g_i(\widehat{x}) \mid i \in \{1, \ldots, m\}\}$ is a family of linearly independent vectors;
- (ii) there exists $z \in \mathbb{R}^N$ such that

$$
\begin{aligned}
(\forall i \in \{1, \ldots, m\}) \qquad & \langle \nabla g_i(\widehat{x}) \mid z\rangle = 0 \\
(\forall j \in J(\widehat{x})) \qquad & \langle \nabla h_j(\widehat{x}) \mid z\rangle < 0
\end{aligned}
$$

  where $J(\widehat{x}) = \{j \in \{1, \ldots, q\} \mid h_j(\widehat{x}) = 0\}$ is the set of active inequality constraints at $\widehat{x}$.

Then, there exists $\widehat{\nu} \in \mathbb{R}^N$ and $\widehat{\lambda} \in [0, +\infty[^q$ such that $\widehat{x}$ is a critical point of $\mathcal{L}(\cdot, \widehat{\nu}, \widehat{\lambda})$ and the complementary slackness condition holds.

Differentiable case

Proof: We know that there exists a nonzero vector $(\alpha_0, \alpha_1, \ldots, \alpha_m, \alpha_{m+1}, \ldots, \alpha_{m+q}) \in [0, +\infty[\times \mathbb{R}^m \times [0, +\infty[^q$ such that

$$\alpha_0 \nabla f(\widehat{x}) + \sum_{i=1}^{m} \alpha_i \nabla g_i(\widehat{x}) + \sum_{j=1}^{q} \alpha_{m+j} \nabla h_j(\widehat{x}) = 0$$
$$(\forall j \in \{1, \ldots, q\}) \qquad \alpha_{m+j} h_j(\widehat{x}) = 0.$$

The complementary slackness condition implies that $(\forall j \notin J(\widehat{x}))\ \alpha_{m+j} = 0$ and the first equality reduces to

$$\alpha_0 \nabla f(\widehat{x}) + \sum_{i=1}^{m} \alpha_i \nabla g_i(\widehat{x}) + \sum_{j \in J(\widehat{x})} \alpha_{m+j} \nabla h_j(\widehat{x}) = 0$$

which yields

$$\alpha_0 \langle \nabla f(\widehat{x}) \mid z \rangle + \sum_{i=1}^{m} \alpha_i \langle \nabla g_i(\widehat{x}) \mid z \rangle + \sum_{j \in J(\widehat{x})} \alpha_{m+j} \langle \nabla h_j(\widehat{x}) \mid z \rangle = 0$$
$$\Leftrightarrow \quad \alpha_0 \langle \nabla f(\widehat{x}) \mid z \rangle + \sum_{j \in J(\widehat{x})} \alpha_{m+j} \langle \nabla h_j(\widehat{x}) \mid z \rangle = 0.$$

.

Differentiable case

Proof: We have proved that there exists a nonzero vector $(\alpha_0, \alpha_1, \ldots, \alpha_m, \alpha_{m+1}, \ldots, \alpha_{m+q}) \in [0, +\infty[\times \mathbb{R}^m \times [0, +\infty[^q$ such that

$$\alpha_0 \nabla f(\widehat{x}) + \sum_{i=1}^{m} \alpha_i \nabla g_i(\widehat{x}) + \sum_{j \in J(\widehat{x})} \alpha_{m+j} \nabla h_j(\widehat{x}) = 0$$

$$\alpha_0 \langle \nabla f(\widehat{x}) \mid z \rangle + \sum_{j \in J(\widehat{x})} \alpha_{m+j} \langle \nabla h_j(\widehat{x}) \mid z \rangle = 0.$$

Let us suppose that $\alpha_0 = 0$.
Since, for every $j \in J(\widehat{x})$, $\langle \nabla h_j(\widehat{x}) \mid z \rangle < 0$, in the latter equality, we would have, $\alpha_{m+j} = 0$ which, in the first equality, would lead to

$$\sum_{i=1}^{m} \alpha_i \nabla g_i(\widehat{x}) = 0.$$

Since the vectors $(\nabla g_i(\widehat{x}))_{1 \leq i \leq m}$ are linearly independent, $(\forall i \in \{1, \ldots, m\})\ \alpha_i = 0$.
In conclusion, the vector $(\alpha_i)_{1 \leq i \leq q}$ would be zero, which is impossible.
This shows that $\alpha_0 > 0$.

Differentiable case

By defining $(\forall i \in \{1,\dots,m\})\ \widehat{\nu}_i = \alpha_i/\alpha_0$, $(\forall j \in \{1,\dots,q\})$ $\widehat{\lambda}_j = \alpha_{m+j}/\alpha_0 \geq 0$, we have then

$$\nabla f(\widehat{x}) + \sum_{i=1}^{m} \widehat{\nu}_i \nabla g_i(\widehat{x}) + \sum_{j=1}^{q} \widehat{\lambda}_j \nabla h_j(\widehat{x}) = 0$$

$$(\forall j \in \{1,\dots,q\}) \qquad \widehat{\lambda}_j h_j(\widehat{x}) = 0.$$

By setting $\widehat{\nu} = (\widehat{\nu}_i)_{1\leq i\leq m} \in \mathbb{R}^m$ and $\widehat{\lambda} = (\widehat{\lambda}_j)_{1\leq j\leq q} \in [0,+\infty[^q$, the first equality also reads

$$\nabla_x \mathcal{L}(\widehat{x},\widehat{\nu},\widehat{\lambda}) = 0.$$

Differentiable case

Karush-Kuhn-Tucker (KKT) theorem

Assume that f, $(g_i)_{1\leq i\leq m}$, and $(h_j)_{1\leq j\leq q}$ are continuously differentiable on $\mathcal{H}=\mathbb{R}^N$.
Assume that $\widehat{x}$ is a local minimizer of f over C satisfying the following Mangasarian-Fromovitz constraint qualification conditions:

(*i*) $\{\nabla g_i(\widehat{x}) \mid i\in\{1,\ldots,m\}\}$ is a family of linearly independent vectors;

(*ii*) there exists $z\in\mathbb{R}^N$ such that

$$\begin{aligned}(\forall i\in\{1,\ldots,m\}) \qquad & \langle \nabla g_i(\widehat{x}) \mid z\rangle = 0\\ (\forall j\in J(\widehat{x})) \qquad & \langle \nabla h_j(\widehat{x}) \mid z\rangle < 0\end{aligned}$$

where $J(\widehat{x})=\{j\in\{1,\ldots,q\}\mid h_j(\widehat{x})=0\}$ is the set of active inequality constraints at $\widehat{x}$.

Then, there exists $\widehat{\nu}\in\mathbb{R}^N$ and $\widehat{\lambda}\in[0,+\infty[^q$ such that $\widehat{x}$ is a critical point of $\mathcal{L}(\cdot,\widehat{\nu},\widehat{\lambda})$ and the complementary slackness condition holds.

Remark: A sufficient condition for Mangasarian-Fromovitz conditions to be satisfied is that $\{\nabla g_i(\widehat{x}) \mid i\in\{1,\ldots,m\}\}\cup\{\nabla h_j(\widehat{x}) \mid j\in J(\widehat{x})\}$ is a family of linearly independent vectors.

Differentiable case

Karush-Kuhn-Tucker (KKT) theorem

Assume that f, $(g_i)_{1\le i\le m}$, and $(h_j)_{1\le j\le q}$ are continuously differentiable on $\mathcal{H}=\mathbb{R}^N$.
Assume that $\widehat{x}$ is a local minimizer of f over C satisfying the following Mangasarian-Fromovitz constraint qualification conditions :

(i) $\{\nabla g_i(\widehat{x}) \mid i\in\{1,\dots,m\}\}$ is a family of linearly independent vectors;
(ii) there exists $z\in\mathbb{R}^N$ such that

$$\begin{aligned}(\forall i\in\{1,\dots,m\}) \qquad & \langle \nabla g_i(\widehat{x}) \mid z\rangle = 0\\ (\forall j\in J(\widehat{x})) \qquad & \langle \nabla h_j(\widehat{x}) \mid z\rangle < 0.\end{aligned}$$

Then, there exists $\widehat{\nu}\in\mathbb{R}^N$ and $\widehat{\lambda}\in[0,+\infty[^q$ such that $\widehat{x}$ is a critical point of $\mathcal{L}(\cdot,\widehat{\nu},\widehat{\lambda})$ and the complementary slackness condition holds.

Remark: A sufficient condition for Mangasarian-Fromovitz conditions to be satisfied is that $\{\nabla g_i(\widehat{x}) \mid i\in\{1,\dots,m\}\}\cup\{\nabla h_j(\widehat{x}) \mid j\in J(\widehat{x})\}$ is a family of linearly independent vectors.
Proof: If this condition holds, then the matrix

$$A=\begin{bmatrix}((\nabla g_i(\widehat{x}))^\top)_{1\le i\le m}\\ ((\nabla h_j(\widehat{x}))^\top)_{j\in J(\widehat{x})}\end{bmatrix}\in\mathbb{R}^{(m+|J(\widehat{x})|)\times N}$$

has rank $m+|J(\widehat{x})|$. Let $\mathbf{1}$ be the unit vector of $\mathbb{R}^{|J(\widehat{x})|}$. Hence, the equation

$$Az=\begin{bmatrix}0\\ -\mathbf{1}\end{bmatrix}, \qquad z\in\mathbb{R}^N,$$

admits a solution. Such a solution satisfies (ii).

Differentiable case

Only equality constraints:

Assume that f and $(g_i)_{1\leq i\leq m}$ are continuously differentiable on $\mathcal{H} = \mathbb{R}^N$.
Assume that $\widehat{x}$ is a local minimizer of f over C and
$\{\nabla g_i(\widehat{x}) \mid i \in \{1, \ldots, m\}\}$ is a family of linearly independent vectors.
Then, there exists $\widehat{\nu} \in \mathbb{R}^N$ such that $\widehat{x}$ is a critical point of $\mathcal{L}(\cdot, \widehat{\nu})$.

Differentiable case

Only inequality constraints:

Assume that f and $(h_j)_{1\leq j\leq q}$ are continuously differentiable on $\mathcal{H}=\mathbb{R}^N$.
Assume that $\widehat{x}$ is a local minimizer of f over C and there exists $z\in\mathbb{R}^N$ such that

$$(\forall j\in J(\widehat{x}))\qquad \langle \nabla h_j(\widehat{x})\mid z\rangle < 0$$

where $J(\widehat{x})=\{j\in\{1,\ldots,q\}\mid h_j(\widehat{x})=0\}$ is the set of active inequality constraints at $\widehat{x}$.
Then, there exists $\widehat{\lambda}\in[0,+\infty[^q$ such that $\widehat{x}$ is a critical point of $\mathcal{L}(\cdot,\widehat{\lambda})$ and the complementary slackness condition holds.

Differentiable case

Only inequality constraints:

Assume that f and $(h_j)_{1\leq j\leq q}$ are continuously differentiable on $\mathcal{H} = \mathbb{R}^N$.
Assume that $\widehat{x}$ is a local minimizer of f over C and
there exists $z \in \mathbb{R}^N$ such that

$$(\forall j \in J(\widehat{x})) \qquad \langle \nabla h_j(\widehat{x}) \mid z \rangle < 0$$

where $J(\widehat{x}) = \{j \in \{1, \dots, q\} \mid h_j(\widehat{x}) = 0\}$ is the set of active inequality constraints at $\widehat{x}$.
Then, there exists $\widehat{\lambda} \in [0, +\infty[^q$ such that $\widehat{x}$ is a critical point of $\mathcal{L}(\cdot, \widehat{\lambda})$ and the complementary slackness condition holds.

Remark: A sufficient condition for the qualification conditions to be satisfied is that $\{\nabla h_j(\widehat{x}) \mid j \in J(\widehat{x})\}$ is a family of linearly independent vectors.

Exercise 4

By using the Lagrange multipliers method, solve the following problem

$$\underset{x=(x^{(i)})_{1\leq i\leq N}\in B}{\text{maximize}} \quad (x^{(N)})^3 - \frac{1}{2}(x^{(N)})^2$$

where B is the unit sphere, centered at 0, of $\mathbb{R}^N$.

Appendix

Preliminary result

Lemma

Let $\widehat{x} \in C \cap \operatorname{dom} f$, let $\widehat{\nu} \in \mathbb{R}^m$, and let $\widehat{\lambda} \in [0, +\infty[^q$ be such that

$$\underline{\mathcal{L}}(\widehat{\nu}, \widehat{\lambda}) = f(\widehat{x}).$$

Then $(\widehat{x}, \widehat{\nu}, \widehat{\lambda})$ is a saddle point of the Lagrange function.

Preliminary result

Lemma

Let $\widehat{x} \in C \cap \operatorname{dom} f$, let $\widehat{\nu} \in \mathbb{R}^m$, and let $\widehat{\lambda} \in [0, +\infty[^q$ be such that

$$\underline{\mathcal{L}}(\widehat{\nu}, \widehat{\lambda}) = f(\widehat{x}).$$

Then $(\widehat{x}, \widehat{\nu}, \widehat{\lambda})$ is a saddle point of the Lagrange function.

<u>Proof</u>: By looking more carefully at the proof of the theorem we provided for characterizing a saddle point, it appears that a sufficient condition for $(\widehat{x}, \widehat{\nu}, \widehat{\lambda})$ to a saddle point of $\mathcal{L}$ is

$$\underline{\mathcal{L}}(\widehat{\nu}, \widehat{\lambda}) = \overline{\mathcal{L}}(\widehat{x}).$$

In addition, we know that $(\forall x \in C)\ \overline{\mathcal{L}}(x) = f(x)$.
Therefore the sufficient condition reduces the one stated.

Validity of Slater condition

Assume that Slater condition holds and that $\widehat{x}$ is a minimizer of f. Let us show that there exists $\widehat{\nu} \in \mathbb{R}^m$ and $\widehat{\lambda} \in [0, +\infty[^q$ such that $f(\widehat{x}) = \underline{\mathcal{L}}(\widehat{\nu}, \widehat{\lambda})$. For every $i \in \{1, \dots, m\}$, since g_i is an affine function, there exists $v_i \in \mathcal{H}$ such that

$$(\forall x \in \mathcal{H}) \quad g_i(x) = g_i(\overline{x}) + \langle v_i \mid x - \overline{x} \rangle = \langle v_i \mid x - \overline{x} \rangle \, .$$

Assume first that the vectors $(v_i)_{1 \leq i \leq m}$ are independent.

Validity of Slater condition

Assume that Slater condition holds and that $\widehat{x}$ is a minimizer of f. Let us show that there exists $\widehat{\nu} \in \mathbb{R}^m$ and $\widehat{\lambda} \in [0, +\infty[^q$ such that $f(\widehat{x}) = \underline{\mathcal{L}}(\widehat{\nu}, \widehat{\lambda})$.
For every $i \in \{1, \dots, m\}$, since g_i is an affine function, there exists $v_i \in \mathcal{H}$ such that

$$(\forall x \in \mathcal{H}) \quad g_i(x) = g_i(\overline{x}) + \langle v_i \mid x - \overline{x} \rangle = \langle v_i \mid x - \overline{x} \rangle \,.$$

Assume first that the vectors $(v_i)_{1 \le i \le m}$ are independent.
Let

$$C_1 = \Big\{ (u^{(0)}, u^{(1)}, \dots, u^{(m)}, u^{(m+1)}, \dots, u^{(m+q)}) \in \mathbb{R}^{m+q+1} \\ \Big| \; (\exists x \in \mathcal{H}) \quad \begin{array}{l} f(x) \le u^{(0)} \\ (\forall i \in \{1, \dots, m\}) \;\; g_i(x) = u^{(i)} \\ (\forall j \in \{1, \dots, q\}) \;\; h_j(x) \le u^{(m+j)} \end{array} \Big\}$$

$$C_2 = \Big\{ (u^{(0)}, u^{(1)}, \dots, u^{(m)}, u^{(m+1)}, \dots, u^{(m+q)}) \in \mathbb{R}^{m+q+1} \\ \Big| \; \begin{array}{l} u^{(0)} < f(\widehat{x}) \\ (\forall i \in \{1, \dots, m\}) \;\; u^{(i)} = 0 \\ (\forall j \in \{1, \dots, q\}) \;\; u^{(m+j)} \le 0 \end{array} \Big\} .$$

Since f and $(h_j)_{1 \le i \le q}$ are convex and $(g_i)_{1 \le i \le m}$ are affine, C_1 is convex.
As a consequence of Slater condition, it is nonempty.
C_2 is convex and nonempty. In addition, $C_1 \cap C_2 = \varnothing$ since there does not exist $x \in C$ such that $f(x) < f(\widehat{x})$.

Validity of Slater condition

Let

$$C_1 = \Big\{\big(u^{(0)}, u^{(1)}, \ldots, u^{(m)}, u^{(m+1)}, \ldots, u^{(m+q)}\big) \in \mathbb{R}^{m+q+1} \\ \;\Big|\; (\exists x \in \mathcal{H}) \quad \begin{array}{l} f(x) \le u^{(0)} \\ (\forall i \in \{1,\ldots,m\}) \;\; g_i(x) = u^{(i)} \\ (\forall j \in \{1,\ldots,q\}) \;\; h_j(x) \le u^{(m+j)} \end{array} \Big\}$$

$$C_2 = \Big\{\big(u^{(0)}, u^{(1)}, \ldots, u^{(m)}, u^{(m+1)}, \ldots, u^{(m+q)}\big) \in \mathbb{R}^{m+q+1} \\ \;\Big|\; \begin{array}{l} u^{(0)} < f(\widehat{x}) \\ (\forall i \in \{1,\ldots,m\}) \;\; u^{(i)} = 0 \\ (\forall j \in \{1,\ldots,q\}) \;\; u^{(m+j)} \le 0 \end{array} \Big\}.$$

According to separation theorem in $\mathbb{R}^{m+q+1}$, there exists $a = (a^{(0)}, a^{(1)}, \ldots, a^{(m)}, a^{(m+1)}, \ldots, a^{(m+q)}) \in \mathbb{R}^{m+q+1} \setminus \{0\}$ such that

$$\inf_{u \in C_1} \langle a \mid u \rangle \ge \sup_{u \in C_2} \langle a \mid u \rangle = \sup_{u=(u^j)_{0 \le j \le m+q} \in C_2} \Big(a^{(0)} u^{(0)} + \sum_{j=1}^{q} a^{(j+m)} u^{(j+m)} \Big).$$

If one of the components $a^{(0)}$ or $(a^{(j+m)})_{1 \le j \le q}$ would be negative, the right-hand side term would be $+\infty$, which is prohibited.
Since these components belong to $[0, +\infty[^q$, $\sup_{u \in C_2} \langle a \mid u \rangle = a^{(0)} f(\widehat{x})$.

Validity of Slater condition

Let

$$C_1 = \Big\{ \big(u^{(0)}, u^{(1)}, \ldots, u^{(m)}, u^{(m+1)}, \ldots, u^{(m+q)}\big) \in \mathbb{R}^{m+q+1} \\ \left| \; (\exists x \in \mathcal{H}) \quad \begin{array}{l} f(x) \le u^{(0)} \\ (\forall i \in \{1, \ldots, m\}) \;\; g_i(x) = u^{(i)} \\ (\forall j \in \{1, \ldots, q\}) \;\; h_j(x) \le u^{(m+j)} \end{array} \right. \Big\}.$$

In addition,

$$(\forall x \in \operatorname{dom} f) \quad \big(f(x), g_1(x), \ldots, g_m(x), h_1(x), \ldots, h_q(x)\big) \in C_1.$$

Hence,

$$\inf_{x \in \mathcal{H}} a^{(0)} f(x) + \sum_{i=1}^{m} a^{(i)} g_i(x) + \sum_{j=1}^{q} a^{(m+i)} h_j(x)$$

$$= \inf_{x \in \operatorname{dom} f} a^{(0)} f(x) + \sum_{i=1}^{m} a^{(i)} g_i(x) + \sum_{j=1}^{q} a^{(m+j)} h_j(x) \ge \sup_{u \in C_2} \langle a \mid u \rangle = a^{(0)} f(\widehat{x}).$$

Validity of Slater condition

$$\inf_{x\in\mathcal{H}} a^{(0)}f(x) + \sum_{i=1}^{m} a^{(i)}g_i(x) + \sum_{j=1}^{q} a^{(m+i)}h_j(x)$$

$$= \inf_{x\in\operatorname{dom} f} a^{(0)}f(x) + \sum_{i=1}^{m} a^{(i)}g_i(x) + \sum_{j=1}^{q} a^{(m+j)}h_j(x) \geq \sup_{u\in C_2} \langle a \mid u\rangle = a^{(0)}f(\widehat{x}).$$

If $a^{(0)} = 0$, then

$$\inf_{x\in\operatorname{dom} f} \sum_{i=1}^{m} a^{(i)}g_i(x) + \sum_{j=1}^{q} a^{(m+j)}h_j(x) \geq 0.$$

This implies that $\sum_{j=1}^{q} a^{(m+j)}h_j(\overline{x}) \geq 0.$

Since $(\forall j\in\{1,\ldots,q\})$ $a^{(m+j)} \geq 0$ and $h_j(\overline{x}) < 0$, then $(\forall j\in\{1,\ldots,q\})$ $a^{(m+j)} = 0$.

We deduce that

$$(\forall x\in\operatorname{dom} f) \quad \sum_{i=1}^{m} a^{(i)}g_i(x) = \sum_{i=1}^{m} a^{(i)}\langle v_i \mid x-\overline{x}\rangle = \left\langle \sum_{i=1}^{m} a^{(i)}v_i \mid x-\overline{x}\right\rangle \geq 0.$$

Since $\overline{x}\in\operatorname{int}(\operatorname{dom} f)$, $\sum_{i=1}^{m} a^{(i)}v_i = 0$ and, since $(v_i)_{1\leq i\leq m}$ are independent vectors, $(\forall i\in\{1,\ldots,m\})$ $a^{(i)} = 0$, which is impossible since $a \neq 0$.

Validity of Slater condition

$$\inf_{x\in\mathcal{H}} a^{(0)} f(x) + \sum_{i=1}^{m} a^{(i)} g_i(x) + \sum_{j=1}^{q} a^{(m+i)} h_j(x)$$

$$= \inf_{x\in\operatorname{dom} f} a^{(0)} f(x) + \sum_{i=1}^{m} a^{(i)} g_i(x) + \sum_{j=1}^{q} a^{(m+j)} h_j(x) \geq \sup_{u\in C_2} \langle a \mid u\rangle = a^{(0)} f(\widehat{x}).$$

Hence $a^{(0)} > 0$ and, by setting

$$(\forall i \in \{1,\dots,m\}) \quad \widehat{\nu}^{(i)} = \frac{a^{(i)}}{a^{(0)}}$$

$$(\forall j \in \{1,\dots,q\}) \quad \widehat{\lambda}^{(j)} = \frac{a^{(m+j)}}{a^{(0)}} \geq 0,$$

we get $\underline{\mathcal{L}}(\widehat{\nu},\widehat{\lambda}) = \inf_{x\in\operatorname{dom} f} f(x) + \sum_{i=1}^{m} \widehat{\nu}^{(i)} g_i(x) + \sum_{j=1}^{q} \widehat{\lambda}^{(j)} h_j(x) \geq f(\widehat{x}).$

Validity of Slater condition

Hence $a^{(0)} > 0$ and, by setting

$$(\forall i \in \{1,\dots,m\}) \quad \widehat{\nu}^{(i)} = \frac{a^{(i)}}{a^{(0)}}$$
$$(\forall j \in \{1,\dots,q\}) \quad \widehat{\lambda}^{(j)} = \frac{a^{(m+j)}}{a^{(0)}} \geq 0,$$

we get $\underline{\mathcal{L}}(\widehat{\nu},\widehat{\lambda}) = \inf\limits_{x\in \operatorname{dom} f} f(x) + \sum\limits_{i=1}^{m} \widehat{\nu}^{(i)} g_i(x) + \sum\limits_{j=1}^{q} \widehat{\lambda}^{(j)} h_j(x) \geq f(\widehat{x})$.

Since $\sup_{\nu\in\mathbb{R}^m,\lambda\in[0,+\infty[^q} \underline{\mathcal{L}}(\nu,\lambda) \leq f(\widehat{x})$, $\underline{\mathcal{L}}(\widehat{\nu},\widehat{\lambda}) = f(\widehat{x})$.
If $(v_i)_{1\leq i\leq m}$ are linearly dependent, let $(v_i)_{i\in\mathbb{I}\subset\{1,\dots,m\}}$ be a maximum size subfamily of linearly independent vectors. Then the same equality holds by setting $(\forall i \in \{1,\dots,m\}\setminus\mathbb{I})\ \widehat{\nu}_i = 0$.
The final result follows from the previous lemma.

Validity of Fritz-John conditions

Let $J(\widehat{x}) = \{j \in \{1,\ldots,q\} \mid h_j(\widehat{x}) = 0\}$ be the set of active inequality constraints and let $\overline{J}(\widehat{x}) = \{1,\ldots,q\} \setminus J(\widehat{x})$.
For every $j \in \overline{J}(\widehat{x})$, $h_j(\widehat{x}) < 0$. Since functions $(h_j)_{1\le j\le q}$ are continuous, there exists $\rho \in \,]0,+\infty[$ such that $\widehat{x}$ is a global minimizer of f on $B(\widehat{x},\rho)$, the open ball centered at $\widehat{x}$ and with radius ρ, and $(\forall x \in B(\widehat{x},\rho))$ $(\forall j \in \overline{J}(\widehat{x}))$ $h_j(x) < 0$.
For every $\eta \in \,]0,+\infty[$, define

$$(\forall x \in \mathbb{R}^N) \quad f_\eta(x) = f(x) + \|x-\widehat{x}\|^2 + \frac{\eta}{2}\Big(\sum_{i=1}^m (g_i(x))^2 + \sum_{j\in J(\widehat{x})} \max\{0,h_j(x)\}^2\Big).$$

Let $\epsilon \in]0,\rho[$. Let us first show that

$$(\exists \eta_\epsilon \in \,]0,+\infty[)(\forall x \in \mathbb{R}^N) \quad \|x-\widehat{x}\| = \epsilon \Rightarrow f_{\eta_\epsilon}(x) > f(\widehat{x}). \qquad (1)$$

Validity of Fritz-John conditions

For every $\eta \in \,]0,+\infty[$, define

$$(\forall x \in \mathbb{R}^N) \quad f_\eta(x) = f(x)+\|x-\widehat{x}\|^2+\frac{\eta}{2}\Big(\sum_{i=1}^{m}(g_i(x))^2+\sum_{j\in J(\widehat{x})}\max\{0,h_j(x)\}^2\Big).$$

Let $\epsilon \in]0,\rho[$. Let us first show that

$$(\exists \eta_\epsilon \in \,]0,+\infty[)(\forall x \in \mathbb{R}^N) \quad \|x-\widehat{x}\| = \epsilon \Rightarrow f_{\eta_\epsilon}(x) > f(\widehat{x}). \qquad (1)$$

Otherwise, we could build a sequence $(\eta_n)_{n\in\mathbb{N}}$ converging to $+\infty$ and a sequence $(x_n)_{n\in\mathbb{N}}$ such that $(\forall n \in \mathbb{N})$ $\|x_n-\widehat{x}\| = \epsilon$ and $f_{\eta_n}(x_n) \le f(\widehat{x})$, i.e.

$$f(x_n)+\epsilon^2+\frac{\eta_n}{2}\Big(\sum_{i=1}^{m}(g_i(x_n))^2+\sum_{j\in J(\widehat{x})}\max\{0,h_j(x_n)\}^2\Big) \le f(\widehat{x})$$

Since $(x_n)_{n\in\mathbb{N}}$ is bounded, it has a cluster point $\widetilde{x}$. Since f, $(g_i)_{1\le i\le m}$, and $(h_j)_{1\le i\le q}$ are continuous and $\eta_n \to +\infty$, we deduce that

$$\sum_{i=1}^{m}(g_i(\widetilde{x}))^2+\sum_{j\in J(\widehat{x})}\max\{0,h_j(\widetilde{x})\}^2 = 0 \quad \Rightarrow \begin{cases}(\forall i \in \{1,\ldots,m\})\; g_i(\widetilde{x}) = 0\\ (\forall j \in J(\widehat{x}))\; h_j(\widetilde{x}) \le 0.\end{cases}$$

Since $\widetilde{x} \in B(\widehat{x},\rho) \Rightarrow (\forall j \in \overline{J}(\widehat{x}))\; h_j(\widetilde{x}) < 0$, this shows that $\widetilde{x} \in C$.

Validity of Fritz-John conditions

For every $\eta \in \,]0, +\infty[$, define

$$(\forall x \in \mathbb{R}^N) \quad f_\eta(x) = f(x) + \|x - \widehat{x}\|^2 + \frac{\eta}{2}\Big(\sum_{i=1}^{m}(g_i(x))^2 + \sum_{j \in J(\widehat{x})} \max\{0, h_j(x)\}^2\Big).$$

Let $\epsilon \in]0, \rho[$. Let us first show that

$$(\exists \eta_\epsilon \in \,]0, +\infty[)(\forall x \in \mathbb{R}^N) \quad \|x - \widehat{x}\| = \epsilon \Rightarrow f_{\eta_\epsilon}(x) > f(\widehat{x}). \tag{1}$$

Otherwise, we could build a sequence $(\eta_n)_{n\in\mathbb{N}}$ converging to $+\infty$ and a sequence $(x_n)_{n\in\mathbb{N}}$ such that $(\forall n \in \mathbb{N})$ $\|x_n - \widehat{x}\| = \epsilon$ and $f_{\eta_n}(x_n) \le f(\widehat{x})$, i.e.

$$f(x_n) + \epsilon^2 + \frac{\eta_n}{2}\Big(\sum_{i=1}^{m}(g_i(x_n))^2 + \sum_{j \in J(\widehat{x})} \max\{0, h_j(x_n)\}^2\Big) \le f(\widehat{x})$$

Since $(x_n)_{n\in\mathbb{N}}$ is bounded, it has a cluster point $\widetilde{x}$. Since f, $(g_i)_{1\le i\le m}$, and $(h_j)_{1\le i\le q}$ are continuous and $\eta_n \to +\infty$, we deduce that $\widetilde{x} \in C$ and

$$f(\widetilde{x}) \le f(\widehat{x}) - \epsilon^2,$$

which contradicts the fact that $\widehat{x}$ minimizes f over $B(\widehat{x}, \rho)$.

Validity of Fritz-John conditions

For every $\eta \in\ ]0,+\infty[$, define

$$(\forall x \in \mathbb{R}^N) \quad f_\eta(x) = f(x)+\|x-\widehat{x}\|^2+\frac{\eta}{2}\Big(\sum_{i=1}^{m}(g_i(x))^2+\sum_{j\in J(\widehat{x})}\max\{0,h_j(x)\}^2\Big).$$

Let $\epsilon \in]0,\rho[$. We have shown that

$$(\exists \eta_\epsilon \in\ ]0,+\infty[)(\forall x \in \mathbb{R}^N) \quad \|x-\widehat{x}\| = \epsilon \Rightarrow f_{\eta_\epsilon}(x) > f(\widehat{x}). \tag{1}$$

Since $\overline{B}(\widehat{x},\epsilon)$ is a compact set and f_{η_ϵ} is continuous, f_{η_ϵ} admits a minimizer $\widehat{x}_\epsilon$ on $\overline{B}(\widehat{x},\epsilon)$.
We have thus $f(\widehat{x}_\epsilon) \le f(\widehat{x})$ and we deduce from (1) that $\widehat{x}_\epsilon \in B(\widehat{x},\epsilon)$.
Since f, $(g_i)_{1\le i\le m}$ and $(h_j)_{1\le j\le q}$ are differentiable, a necessary first-order condition is

$$\nabla f_{\eta_\epsilon}(\widehat{x}_\epsilon) = 0.$$

Validity of Fritz-John conditions

For every $\eta \in\ ]0,+\infty[$, define

$$(\forall x \in \mathbb{R}^N) \quad f_\eta(x) = f(x)+\|x-\widehat{x}\|^2+\frac{\eta}{2}\Big(\sum_{i=1}^{m}(g_i(x))^2+\sum_{j\in J(\widehat{x})}\max\{0,h_j(x)\}^2\Big).$$

We have

$$\nabla f_{\eta_\epsilon}(\widehat{x}_\epsilon) = 0$$
$$\Leftrightarrow \nabla f(\widehat{x}_\epsilon) + 2(\widehat{x}_\epsilon - \widehat{x}) + \eta_\epsilon\Big(\sum_{i=1}^{m} g_i(\widehat{x}_\epsilon)\nabla g_i(\widehat{x}_\epsilon) + \sum_{j\in J(\widehat{x})}\max\{0,h_j(\widehat{x}_\epsilon)\}\nabla h_j(\widehat{x}_\epsilon)\Big) = 0.$$

Let $a_\epsilon = (1, a_\epsilon^{(1)}, \ldots, a_\epsilon^{(m)}, a_\epsilon^{(m+1)}, \ldots, a_\epsilon^{(m+q)}) \in \mathbb{R}^{m+q+1}$ with

$$(\forall i \in \{1,\ldots,m\}) \quad a_\epsilon^{(i)} = \eta_\epsilon g_i(\widehat{x}_\epsilon)$$
$$(\forall j \in \{1,\ldots,q\}) \quad a_\epsilon^{(j+m)} = \begin{cases} \eta_\epsilon \max\{0,h_j(\widehat{x}_\epsilon)\} & \text{if } j \in J(\widehat{x}) \\ 0 & \text{otherwise.} \end{cases}$$

We get

$$\nabla f(\widehat{x}_\epsilon) + 2(\widehat{x}_\epsilon - \widehat{x}) + \sum_{i=1}^{m} a_\epsilon^{(i)}\nabla g_i(\widehat{x}_\epsilon) + \sum_{j=1}^{q} a_\epsilon^{(j+m)}\nabla h_j(\widehat{x}_\epsilon) = 0.$$

Validity of Fritz-John conditions

Let $a_\epsilon = (1, a_\epsilon^{(1)}, \dots, a_\epsilon^{(m)}, a_\epsilon^{(m+1)}, \dots, a_\epsilon^{(m+q)}) \in \mathbb{R}^{m+q+1}$ with

$$(\forall i \in \{1, \dots, m\}) \quad a_\epsilon^{(i)} = \eta_\epsilon g_i(\widehat{x}_\epsilon)$$

$$(\forall j \in \{1, \dots, q\}) \quad a_\epsilon^{(j+m)} = \begin{cases} \eta_\epsilon \max\{0, h_j(\widehat{x}_\epsilon)\} & \text{if } j \in J(\widehat{x}) \\ 0 & \text{otherwise.} \end{cases}$$

We get

$$\nabla f(\widehat{x}_\epsilon) + 2(\widehat{x}_\epsilon - \widehat{x}) + \sum_{i=1}^{m} a_\epsilon^{(i)} \nabla g_i(\widehat{x}_\epsilon) + \sum_{j=1}^{q} a_\epsilon^{(j+m)} \nabla h_j(\widehat{x}_\epsilon) = 0.$$

Let $\alpha_\epsilon = a_\epsilon / \|a_\epsilon\|$.
Consider now a sequence $(\epsilon_n)_{n\in\mathbb{N}}$ of $]0, \rho[$ converging to 0.
Then $\widehat{x}_{\epsilon_n} \to \widehat{x}$, which implies that $(\widehat{x}_{\epsilon_n} - \widehat{x})/\|a_{\epsilon_n}\| \to 0$ (since $\|a_{\epsilon_n}\| \geq 1$).
In addition, $(\alpha_{\epsilon_n})_{n\in\mathbb{N}}$ being bounded, there exists a subsequence $(\alpha_{\epsilon_{n_k}})_{k\in\mathbb{N}}$ converging to some $\alpha = (\alpha^{(i)})_{0\leq i\leq m+q}$. It follows from the continuity of ∇f, $(\nabla g_i)_{1\leq i\leq m}$, and $(\nabla h_j)_{1\leq j\leq q}$ that

$$\alpha^{(0)} \nabla f(\widehat{x}) + \sum_{i=1}^{m} \alpha^{(i)} \nabla g_i(\widehat{x}) + \sum_{j=1}^{q} \alpha^{(m+j)} \nabla h_j(\widehat{x}) = 0.$$

Validity of Fritz-John conditions

Let $a_\epsilon = (1, a_\epsilon^{(1)}, \dots, a_\epsilon^{(m)}, a_\epsilon^{(m+1)}, \dots, a_\epsilon^{(m+q)}) \in \mathbb{R}^{m+q+1}$ with

$$(\forall i \in \{1, \dots, m\}) \quad a_\epsilon^{(i)} = \eta_\epsilon g_i(\widehat{x}_\epsilon)$$
$$(\forall j \in \{1, \dots, q\}) \quad a_\epsilon^{(j+m)} = \begin{cases} \eta_\epsilon \max\{0, h_j(\widehat{x}_\epsilon)\} & \text{if } j \in J(\widehat{x}) \\ 0 & \text{otherwise.} \end{cases}$$

We get

$$\nabla f(\widehat{x}_\epsilon) + 2(\widehat{x}_\epsilon - \widehat{x}) + \sum_{i=1}^{m} a_\epsilon^{(i)} \nabla g_i(\widehat{x}_\epsilon) + \sum_{j=1}^{q} a_\epsilon^{(j+m)} \nabla h_j(\widehat{x}_\epsilon) = 0.$$

Let $\alpha_\epsilon = a_\epsilon / \|a_\epsilon\|$.
Consider now a sequence $(\epsilon_n)_{n\in\mathbb{N}}$ of $]0, \rho[$ converging to 0.
Then $\widehat{x}_{\epsilon_n} \to \widehat{x}$, which implies that $(\widehat{x}_{\epsilon_n} - \widehat{x})/\|a_{\epsilon_n}\| \to 0$ (since $\|a_{\epsilon_n}\| \geq 1$).
In addition, $(\alpha_{\epsilon_n})_{n\in\mathbb{N}}$ being bounded, there exists a subsequence $(\alpha_{\epsilon_{n_k}})_{k\in\mathbb{N}}$ converging to some $\alpha = (\alpha^{(i)})_{0\leq i\leq m+q}$.
It can be finally observed that, for every $j \in \overline{J}(\widehat{x})$, $\alpha^{(j+m)} = 0$, which means that the complementarity condition is satisfied.