Constrained Minimization

(Computational Methods for Mechatronics [140466])

Enrico Bertolazzi

DII - Dipartimento di Ingegneria Industriale – Università di Trento

AA 2014/2015

Outline

The problem (1/3)

Given the function: $\mathbb{R}^n \to \mathbb{R}$:

$$\underset{\boldsymbol{x}\in\mathbb{R}^n}{\text{minimize}} \quad \mathsf{f}(\boldsymbol{x})$$

the following regularity condition are assumed from now and forward:

Assumption (Regularity conditions)

The functione $\mathsf{f} \in \mathtt{C}^1(\mathbb{R}^n)$ *has Lipschitz continuos gradient, i.e. exists* $\gamma > 0$ *such that*

$$\left\|\nabla\mathsf{f}(\boldsymbol{x})^T - \nabla\mathsf{f}(\boldsymbol{y})^T\right\| \leq \gamma \left\|\boldsymbol{x} - \boldsymbol{y}\right\|, \qquad \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$$

The problem (2/3)

Definition (Global minimum)

Giving a function $\mathsf{f} : \mathbb{R}^n \to \mathbb{R}$ *a point* $\boldsymbol{x}^* \in \mathbb{R}^n$ *ia a global minimum if*

$$\mathsf{f}(\boldsymbol{x}^*) \leq \mathsf{f}(\boldsymbol{x}), \qquad \forall \boldsymbol{x} \in \mathbb{R}^n.$$

Definition (Local minimum)

Giving a function $\mathsf{f} : \mathbb{R}^n \to \mathbb{R}$ *a point* $\boldsymbol{x}^* \in \mathbb{R}^n$ *is a local minimum if*

$$\mathsf{f}(\boldsymbol{x}^*) \leq \mathsf{f}(\boldsymbol{x}), \qquad \forall \boldsymbol{x} \in B(\boldsymbol{x}^*; \delta).$$

Obviously a global minimum is also a local minimum. The search of a global minimum is in general a difficult task.

The problem (3/3)

Definition (Strict global minimum)

Given a function $\mathsf{f} : \mathbb{R}^n \to \mathbb{R}$ *a point* $\boldsymbol{x}^* \in \mathbb{R}^n$ *is a strict global minimum if*

$$\mathsf{f}(\boldsymbol{x}^*) < \mathsf{f}(\boldsymbol{x}), \qquad \forall \boldsymbol{x} \in \mathbb{R}^n \setminus \{\boldsymbol{x}^*\}.$$

Definition (Strict local minimum)

Given a function $\mathsf{f} : \mathbb{R}^n \to \mathbb{R}$ *a point* $\boldsymbol{x}^* \in \mathbb{R}^n$ *is a strict local minimum if*

$$\mathsf{f}(\boldsymbol{x}^*) < \mathsf{f}(\boldsymbol{x}), \qquad \forall \boldsymbol{x} \in B(\boldsymbol{x}^*; \delta) \setminus \{\boldsymbol{x}^*\}.$$

Obviously a strict global minimum is also a strict local minimum.

First order necessary conditions

Lemma (First order necessary conditions)

Given a function $\mathsf{f} : \mathbb{R}^n \to \mathbb{R}$ *that satisfy the regularity conditions, if a point* $\boldsymbol{x}^* \in \mathbb{R}^n$ *is local minimum point, then*

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

Proof.

Let $\boldsymbol{d}$ e generic direction then for δ small enough

$$\lambda^{-1}\big(\mathsf{f}(\boldsymbol{x}^* + \lambda \boldsymbol{d}) - \mathsf{f}(\boldsymbol{x}^*)\big) \geq 0, \qquad 0 < \lambda < \delta$$

and thus

$$\lim_{\lambda \to 0} \lambda^{-1}\big(\mathsf{f}(\boldsymbol{x}^* + \lambda \boldsymbol{d}) - \mathsf{f}(\boldsymbol{x}^*)\big) = \nabla \mathsf{f}(\boldsymbol{x}^*)\boldsymbol{d} \geq 0,$$

cause $\boldsymbol{d}$ is a generic direction it follows $\nabla \mathsf{f}(\boldsymbol{x}^*)^T = \mathbf{0}$. □

1. First order necessary condition do not distinguish maxima, minima or saddle point.
2. To distinguish maxima and minima we need more informations, for example second derivative of $\mathsf{f}(\boldsymbol{x})$.
3. With second order information it is possibile to build necessary and/or sufficient condition to discriminate maxima and minima.
4. In general first and second order condition are not sufficient to set both necessary and sufficient condition for a point $\boldsymbol{x}^*$ to be a maximum or minimum point.

Second order necessary conditions

Lemma (Second order necessary conditions)

Given a function $\mathsf{f} \in \mathtt{C}^2(\mathbb{R}^n)$ *if a point* $\boldsymbol{x}^* \in \mathbb{R}^n$ *is a local minimum then* $\nabla \mathsf{f}(\boldsymbol{x}^*)^T = \mathbf{0}$ *and* $\nabla^2 \mathsf{f}(\boldsymbol{x}^*)$ *is semi positive definite, i.e.*

$$\boldsymbol{d}^T \nabla^2 \mathsf{f}(\boldsymbol{x}^*) \boldsymbol{d} \geq 0, \qquad \forall \boldsymbol{d} \in \mathbb{R}^n$$

Example

This condition is necessary but not sufficient, in fact, consider $\mathsf{f}(\boldsymbol{x}) = x_1^2 - x_2^3$,

$$\nabla \mathsf{f}(\boldsymbol{x}) = \left(2x_1, -3x_2^2\right), \quad \nabla^2 \mathsf{f}(\boldsymbol{x}) = \begin{pmatrix} 2 & 0 \\ 0 & -6x_2 \end{pmatrix}$$

for the point $\boldsymbol{x}^* = \mathbf{0}$ the gradient is $\nabla \mathsf{f}(\mathbf{0}) = \mathbf{0}$ and $\nabla^2 \mathsf{f}(\mathbf{0})$ is semi positive defined, but $\mathbf{0}$ is a saddle point not a minimum point.

Proof.

Condition $\nabla \mathsf{f}(\boldsymbol{x}^*)^T = \mathbf{0}$ follows from the first order necessary conditions. Consider now a generic direction $\boldsymbol{d}$ and the finite difference:

$$\frac{\mathsf{f}(\boldsymbol{x}^* + \lambda \boldsymbol{d}) - 2\mathsf{f}(\boldsymbol{x}^*) + \mathsf{f}(\boldsymbol{x}^* - \lambda \boldsymbol{d})}{\lambda^2} \geq 0$$

using Taylor series for $\mathsf{f}(\boldsymbol{x})$

$$\mathsf{f}(\boldsymbol{x}^* \pm \lambda \boldsymbol{d}) = \mathsf{f}(\boldsymbol{x}^*) \pm \nabla \mathsf{f}(\boldsymbol{x}^*)\lambda \boldsymbol{d} + \frac{\lambda^2}{2}\boldsymbol{d}^T \nabla^2 \mathsf{f}(\boldsymbol{x}^*)\boldsymbol{d} + o(\lambda^2)$$

with the previous inequality

$$\boldsymbol{d}^T \nabla^2 \mathsf{f}(\boldsymbol{x}^*)\boldsymbol{d} + 2o(\lambda^2)/\lambda^2 \geq 0$$

so that taking limits $\lambda \to 0$ from the arbitrariety of $\boldsymbol{d}$ follows that $\nabla^2 \mathsf{f}(\boldsymbol{x}^*)$ which must be semi-positive definite. □

Second order sufficient conditions

Lemma (Second order sufficient conditions)

Given the function $\mathsf{f} \in \mathtt{C}^2(\mathbb{R}^n)$ *if a point* $\boldsymbol{x}^* \in \mathbb{R}^n$ *satisfy:*

1. $\nabla \mathsf{f}(\boldsymbol{x}^*)^T = \mathbf{0}$;
2. $\nabla^2 \mathsf{f}(\boldsymbol{x}^*)$ *is definite positive; i.e.*

$$\boldsymbol{d}^T \nabla^2 \mathsf{f}(\boldsymbol{x}^*) \boldsymbol{d} > 0, \qquad \forall \boldsymbol{d} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$$

then $\boldsymbol{x}^* \in \mathbb{R}^n$ *is a strict local minimum.*

Remark

Cause $\nabla^2 \mathsf{f}(\boldsymbol{x}^*)$ *is symmetric we have*

$$\lambda_{\min} \boldsymbol{d}^T \boldsymbol{d} \leq \boldsymbol{d}^T \nabla^2 \mathsf{f}(\boldsymbol{x}^*) \boldsymbol{d} \leq \lambda_{\max} \boldsymbol{d}^T \boldsymbol{d}$$

If $\nabla^2 \mathsf{f}(\boldsymbol{x}^*)$ *is positive definite then* $\lambda_{\min} > 0$.

Proof.

Consider a generic direction $\boldsymbol{d}$, and Taylor expansion for $\mathsf{f}(\boldsymbol{x})$

$$
\begin{aligned}
\mathsf{f}(\boldsymbol{x}^* + \boldsymbol{d}) &= \mathsf{f}(\boldsymbol{x}^*) + \nabla \mathsf{f}(\boldsymbol{x}^*)\boldsymbol{d} + \frac{1}{2}\boldsymbol{d}^T \nabla^2 \mathsf{f}(\boldsymbol{x}^*)\boldsymbol{d} + o(\|\boldsymbol{d}\|^2) \\
&\geq \mathsf{f}(\boldsymbol{x}^*) + \frac{1}{2}\lambda_{min} \|\boldsymbol{d}\|^2 + o(\|\boldsymbol{d}\|^2) \\
&\geq \mathsf{f}(\boldsymbol{x}^*) + \frac{1}{2}\lambda_{min} \|\boldsymbol{d}\|^2 \left(1 + o(\|\boldsymbol{d}\|^2)/\|\boldsymbol{d}\|^2\right)
\end{aligned}
$$

choosing $\boldsymbol{d}$ small enough

$$
\mathsf{f}(\boldsymbol{x}^* + \boldsymbol{d}) \geq \mathsf{f}(\boldsymbol{x}^*) + \frac{1}{4}\lambda_{min} \|\boldsymbol{d}\|^2 > \mathsf{f}(\boldsymbol{x}^*), \qquad \boldsymbol{d} \neq \boldsymbol{0},\ \|\boldsymbol{d}\| \leq \delta.
$$

i.e. $\boldsymbol{x}^*$ is a strict minimum. □

Constrained minimization
Problem

Let be $f \in \mathtt{C}^2(\mathbb{R}^n)$ a function and $h_k \in \mathtt{C}^2(\mathbb{R}^n)$ constraints functions with $k = 1, 2, \ldots, m$.

Problem

$$\begin{array}{lll} \text{Minimize} & f(\boldsymbol{x}) & \\ \text{With constraints:} & h_k(\boldsymbol{x}) = 0, & k = 1, 2, \ldots, m \end{array}$$

Theorem (of Lagrange multiplier)

Let $f \in \mathtt{C}^2(\mathbb{R}^n)$ *and* $\boldsymbol{h} \in \mathtt{C}^2(\mathbb{R}^n, \mathbb{R}^m)$ *a constraints map. Let* $\boldsymbol{x}^*$ *a local mininum of* $f(\boldsymbol{x})$ *which satisfy the constraints (i.e.*$\boldsymbol{h}(\boldsymbol{x}^*) = \boldsymbol{0}$*). If* $\nabla \boldsymbol{h}(\boldsymbol{x}^*)$ *has maximum rank then there exists* m *scalar* λ_k *such that*

$$\nabla f(\boldsymbol{x}^*) - \sum_{k=1}^{m} \lambda_k \nabla h_k(\boldsymbol{x}^*) = \boldsymbol{0}^T \qquad \text{(A)}$$

moreover for all $\boldsymbol{z} \in \mathbb{R}^n$ *that satisfy* $\nabla \boldsymbol{h}(\boldsymbol{x}^*)\boldsymbol{z} = \boldsymbol{0}$ *the following inequality is true*

$$\boldsymbol{z}^T \left(\nabla^2 f(\boldsymbol{x}^*) - \sum_{k=1}^{m} \lambda_k \nabla^2 h_k(\boldsymbol{x}^*) \right) \boldsymbol{z} \geq 0 \qquad \text{(B)}$$

in other words the matrix $\nabla_x^2 \big(f(\boldsymbol{x}^*) - \boldsymbol{\lambda} \cdot \boldsymbol{h}(\boldsymbol{x}^*) \big)$ *is semi-positive definite in the kernel of* $\nabla \boldsymbol{h}(\boldsymbol{x}^*)$*.*

Proof

If $\boldsymbol{x}^*$ is a local minimum then there exists $\varepsilon > 0$ such that

$$f(\boldsymbol{x}^*) \leq f(\boldsymbol{x}), \qquad \forall \boldsymbol{x} \text{ tale che: } \boldsymbol{x} \in B \text{ ed } \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}$$

where $B = \{\boldsymbol{x} \mid \|\boldsymbol{x} - \boldsymbol{x}^*\| \leq \varepsilon\}$. Consider the function sequence

$$f_k(\boldsymbol{x}) = f(\boldsymbol{x}) + k\,\|\boldsymbol{h}(\boldsymbol{x})\|^2 + \alpha\,\|\boldsymbol{x} - \boldsymbol{x}^*\|^2\,, \qquad \alpha > 0$$

and the sequence of local minimum (unconstrained) in B:

$$f_k(\boldsymbol{x}_k) = \min_{\boldsymbol{x} \in B} f_k(\boldsymbol{x})$$

theorem will be proved using the condition for unconstrained minimum and using the limit $\boldsymbol{x}_k \to \boldsymbol{x}^*$.

Proof

(2/12)

Step 1: the limit of the sequence $\boldsymbol{x}_k \to \bar{\boldsymbol{x}}$ lie on the constraint

Cause the sequence $\boldsymbol{x}_k$ is contained in the compact ball B then exist a sub-sequence converging $\boldsymbol{x}_{k_j} \to \bar{\boldsymbol{x}} \in B$. To simplify notation and proof we assume that $\boldsymbol{x}_k \to \bar{\boldsymbol{x}} \in B$. From the definition of $\boldsymbol{x}_k$

$$f_k(\boldsymbol{x}_k) \leq f_k(\boldsymbol{x}^*) = f(\boldsymbol{x}^*) + k \left\|\boldsymbol{h}(\boldsymbol{x}^*)\right\|^2 + \alpha \left\|\boldsymbol{x}^* - \boldsymbol{x}^*\right\|^2 = f(\boldsymbol{x}^*)$$

moreover

$$f_k(\boldsymbol{x}_k) = f(\boldsymbol{x}_k) + k \left\|\boldsymbol{h}(\boldsymbol{x}_k)\right\|^2 + \alpha \left\|\boldsymbol{x}_k - \boldsymbol{x}^*\right\|^2 \leq f(\boldsymbol{x}^*)$$

per cui avremo

$$k \left\|\boldsymbol{h}(\boldsymbol{x}_k)\right\|^2 + \alpha \left\|\boldsymbol{x}_k - \boldsymbol{x}^*\right\|^2 \leq f(\boldsymbol{x}^*) - \min_{\boldsymbol{x} \in B} f(\boldsymbol{x}) = C < +\infty$$

and thus

$$\lim_{k \to \infty} \left\|\boldsymbol{h}(\boldsymbol{x}_k)\right\|^2 = 0$$

and from continuity $\left\|\boldsymbol{h}(\bar{\boldsymbol{x}})\right\| = 0$

Proof

Step 2: the limit of sequence $\boldsymbol{x}_k$ is $\boldsymbol{x}^*$

Consider

$$f_k(\boldsymbol{x}_k) = f(\boldsymbol{x}_k) + k\left\|\boldsymbol{h}(\boldsymbol{x}_k)\right\|^2 + \alpha\left\|\boldsymbol{x}_k - \boldsymbol{x}^*\right\|^2 \leq f(\boldsymbol{x}^*)$$

that imply

$$\alpha\left\|\boldsymbol{x}_k - \boldsymbol{x}^*\right\|^2 \leq f(\boldsymbol{x}^*) - f(\boldsymbol{x}_k) - k\left\|\boldsymbol{h}(\boldsymbol{x}_k)\right\|^2 \leq f(\boldsymbol{x}^*) - f(\boldsymbol{x}_k)$$

taking the limit for $k \to \infty$ and using norm continuity

$$\lim_{k\to\infty} \alpha\left\|\boldsymbol{x}_k - \boldsymbol{x}^*\right\|^2 \leq \alpha\left\|\bar{\boldsymbol{x}} - \boldsymbol{x}^*\right\|^2 \leq f(\boldsymbol{x}^*) - f(\bar{\boldsymbol{x}})$$

cause $\left\|\boldsymbol{h}(\bar{\boldsymbol{x}})\right\| = 0$ and that $\boldsymbol{x}^*$ is a minimum in B that satisfy constraint it follows

$$\alpha\left\|\bar{\boldsymbol{x}} - \boldsymbol{x}^*\right\|^2 \leq f(\boldsymbol{x}^*) - f(\bar{\boldsymbol{x}}) \leq 0$$

i.e. $\bar{\boldsymbol{x}} = \boldsymbol{x}^*$.

Proof

(4/12)

Step 3: Lagrange multiplier construction

Cause $\boldsymbol{x}_k$ are unconstrained local minimum for $f_k(\boldsymbol{x})$ then

$$\nabla f_k(\boldsymbol{x}_k) = \nabla f(\boldsymbol{x}_k) + k\nabla \left\|\boldsymbol{h}(\boldsymbol{x}_k)\right\|^2 + \alpha\nabla \left\|\boldsymbol{x}_k - \boldsymbol{x}^*\right\|^2 = \boldsymbol{0}$$

remember

$$\nabla \left\|\boldsymbol{x}\right\|^2 = \nabla(\boldsymbol{x}\cdot\boldsymbol{x}) = 2\boldsymbol{x}^T,$$

$$\nabla \left\|\boldsymbol{h}(\boldsymbol{x})\right\|^2 = \nabla(\boldsymbol{h}(\boldsymbol{x})\cdot\boldsymbol{h}(\boldsymbol{x})) = 2\boldsymbol{h}(\boldsymbol{x})^T\nabla\boldsymbol{h}(\boldsymbol{x})$$

which imply (using matrix transposition)

$$\nabla f(\boldsymbol{x}_k)^T + 2k\nabla\boldsymbol{h}(\boldsymbol{x}_k)^T\boldsymbol{h}(\boldsymbol{x}_k) + 2\alpha(\boldsymbol{x}_k - \boldsymbol{x}^*) = \boldsymbol{0}$$

Proof

Step 3: Lagrange multiplier construction

Left multiply by $\nabla \boldsymbol{h}(\boldsymbol{x}_k)$

$$\nabla \boldsymbol{h}(\boldsymbol{x}_k)\nabla f(\boldsymbol{x}_k)^T + 2k\nabla \boldsymbol{h}(\boldsymbol{x}_k)\nabla \boldsymbol{h}(\boldsymbol{x}_k)^T \boldsymbol{h}(\boldsymbol{x}_k) + 2\alpha\nabla \boldsymbol{h}(\boldsymbol{x}_k)(\boldsymbol{x}_k - \boldsymbol{x}^*) = \boldsymbol{0}$$

cause $\nabla \boldsymbol{h}(\boldsymbol{x}^*) \in \mathbb{R}^{m\times n}$ is of maximum rank for large k by continuity all $\nabla \boldsymbol{h}(\boldsymbol{x}_k)$ have maximum rank, thus $\nabla \boldsymbol{h}(\boldsymbol{x}_k)\nabla \boldsymbol{h}(\boldsymbol{x}_k)^T \in \mathbb{R}^{m\times m}$ are square nonsingular and

$$2k\boldsymbol{h}(\boldsymbol{x}_k) = -\left(\nabla \boldsymbol{h}(\boldsymbol{x}_k)\nabla \boldsymbol{h}(\boldsymbol{x}_k)^T\right)^{-1}\nabla \boldsymbol{h}(\boldsymbol{x}_k)\left[\nabla f(\boldsymbol{x}_k)^T + 2\alpha(\boldsymbol{x}_k - \boldsymbol{x}^*)\right]$$

for $k \to \infty$

$$\lim_{k\to\infty} 2k\boldsymbol{h}(\boldsymbol{x}_k) = -\left(\nabla \boldsymbol{h}(\boldsymbol{x}^*)\nabla \boldsymbol{h}(\boldsymbol{x}^*)^T\right)^{-1}\nabla \boldsymbol{h}(\boldsymbol{x}^*)\nabla f(\boldsymbol{x}^*)^T$$

Proof

Step 3: Lagrange multiplier construction

Defining $\lim_{k\to\infty} 2k\boldsymbol{h}(\boldsymbol{x}_k) = \boldsymbol{\lambda}$ where

$$\boldsymbol{\lambda} = \left(\nabla\boldsymbol{h}(\boldsymbol{x}^*)\nabla\boldsymbol{h}(\boldsymbol{x}^*)^T\right)^{-1}\nabla\boldsymbol{h}(\boldsymbol{x}^*)\nabla f(\boldsymbol{x}^*)^T$$

and substituting in

$$\nabla f(\boldsymbol{x}_k)^T + 2k\nabla\boldsymbol{h}(\boldsymbol{x}_k)^T\boldsymbol{h}(\boldsymbol{x}_k) + 2\alpha(\boldsymbol{x}_k - \boldsymbol{x}^*) = \boldsymbol{0}$$

and for $k \to \infty$

$$\nabla f(\boldsymbol{x}^*)^T - \nabla\boldsymbol{h}(\boldsymbol{x}^*)^T\boldsymbol{\lambda} = \boldsymbol{0}$$

Proof

(7/12)

Passo 4: condizioni necessarie di minimo

Cause $\boldsymbol{x}_k$ are unconstrained local minimum for $f_k(\boldsymbol{x})$ then matrices

$$\nabla^2 f_k(\boldsymbol{x}_k)$$

are semi-positive definite, i.e.

$$\boldsymbol{z}^T \nabla^2 f_k(\boldsymbol{x}_k)\boldsymbol{z} \geq 0, \qquad \forall \boldsymbol{z} \in \mathbb{R}^n$$

moreover

$$\begin{aligned} \nabla^2 f_k(\boldsymbol{x}_k) &= \nabla^2 f(\boldsymbol{x}_k) + k\nabla^2 \left\| \boldsymbol{h}(\boldsymbol{x}_k) \right\|^2 + 2\alpha\nabla(\boldsymbol{x}_k - \boldsymbol{x}^*) \\ &= \nabla^2 f(\boldsymbol{x}_k)^T + k\nabla^2 \sum_{i=1}^{m} h_i(\boldsymbol{x}_k)^2 + 2\alpha \boldsymbol{I} \end{aligned}$$

Proof

Step 4: necessary condition for a minimum

substituting

$$\begin{aligned} \nabla^2 h_i(\boldsymbol{x})^2 &= \nabla(2h_i(\boldsymbol{x})\nabla h_i(\boldsymbol{x})^T) \\ &= 2\nabla h_i(\boldsymbol{x})^T \nabla h_i(\boldsymbol{x}) + 2h_i(\boldsymbol{x})\nabla^2 h_i(\boldsymbol{x}) \end{aligned}$$

in the Hessian it follows

$$\begin{aligned} \nabla^2 f_k(\boldsymbol{x}_k) &= \nabla^2 f(\boldsymbol{x}_k) + 2\alpha \boldsymbol{I} \\ &\quad + 2k \sum_{i=1}^{m} \nabla h_i(\boldsymbol{x}_k)^T \nabla h_i(\boldsymbol{x}_k) \\ &\quad + 2k \sum_{i=1}^{m} h_i(\boldsymbol{x}_k) \nabla^2 h_i(\boldsymbol{x}_k) \end{aligned}$$

Proof

Step 4: necessary condition for a minimum

(9/12)

Let $\boldsymbol{z} \in \mathbb{R}^n$ then $0 \leq \boldsymbol{z}^T \nabla^2 f_k(\boldsymbol{x}_k)\boldsymbol{z}$, i.e.

$$0 \leq \boldsymbol{z}^T \nabla^2 f(\boldsymbol{x}_k)\boldsymbol{z} + \sum_{i=1}^{m} (2k h_i(\boldsymbol{x}_k))\boldsymbol{z}^T \nabla^2 h_i(\boldsymbol{x}_k)\boldsymbol{z}$$

$$+ 2\alpha \|\boldsymbol{z}\|^2 + 2k \|\nabla \boldsymbol{h}(\boldsymbol{x}_k)\boldsymbol{z}\|^2$$

Previous inequality is true for all $\boldsymbol{z} \in \mathbb{R}^n$ and thus for all sequence $\boldsymbol{z}_k$. Consider a generic sequence $\boldsymbol{z}_k \to \boldsymbol{z}$ and take the limit for $k \to \infty$

$$0 \leq \boldsymbol{z}^T \nabla^2 f(\boldsymbol{x}^*)\boldsymbol{z} + 2\alpha \|\boldsymbol{z}\|^2 + \lim_{k\to\infty} 2k \|\nabla \boldsymbol{h}(\boldsymbol{x}_k)\boldsymbol{z}\|^2$$

$$+ \sum_{i=1}^{m} \lim_{k\to\infty} (2k h_i(\boldsymbol{x}_k)) \left[\boldsymbol{z}^T \nabla^2 h_i(\boldsymbol{x}^*)\boldsymbol{z}\right]$$

Proof

Step 4: necessary condition for a minimum

from $\lim_{k\to\infty}(2kh_i(\boldsymbol{x}_k)) = -\lambda_i$ it follows

$$0 \leq \boldsymbol{z}^T\nabla^2 f(\boldsymbol{x}^*)\boldsymbol{z} + 2\alpha\left\|\boldsymbol{z}\right\|^2 - \sum_{i=1}^{m}\lambda_i\left[\boldsymbol{z}^T\nabla^2 h_i(\boldsymbol{x}^*)\boldsymbol{z}\right]$$

$$+ \lim_{k\to\infty} 2k\left\|\nabla\boldsymbol{h}(\boldsymbol{x}_k)\boldsymbol{z}_k\right\|^2$$

if $\nabla\boldsymbol{h}(\boldsymbol{x}_k)\boldsymbol{z}_k = \boldsymbol{0}$ from $\alpha > 0$ arbitrarily small

$$0 \leq \boldsymbol{z}^T\nabla^2 f(\boldsymbol{x}^*)\boldsymbol{z} - \sum_{i=1}^{m}\lambda_i\left[\boldsymbol{z}^T\nabla^2 h_i(\boldsymbol{x}^*)\boldsymbol{z}\right]$$

which is the relation searched.

Proof

Step 4: necessary condition for a minimum

Consider $\boldsymbol{z}_k$ as the projection of $\boldsymbol{z}$ in the Kernel of $\nabla\boldsymbol{h}(\boldsymbol{x}_k)$, i.e.

$$\boldsymbol{z}_k = \boldsymbol{z} - \nabla\boldsymbol{h}(\boldsymbol{x}_k)^T \left[\nabla\boldsymbol{h}(\boldsymbol{x}_k)\nabla\boldsymbol{h}(\boldsymbol{x}_k)^T\right]^{-1} \nabla\boldsymbol{h}(\boldsymbol{x}_k)\boldsymbol{z}$$

indeed

$$\begin{aligned}\nabla\boldsymbol{h}(\boldsymbol{x}_k)\boldsymbol{z}_k &= \nabla\boldsymbol{h}(\boldsymbol{x}_k)\boldsymbol{z} \\ &\quad -\nabla\boldsymbol{h}(\boldsymbol{x}_k)\nabla\boldsymbol{h}(\boldsymbol{x}_k)^T \left[\nabla\boldsymbol{h}(\boldsymbol{x}_k)\nabla\boldsymbol{h}(\boldsymbol{x}_k)^T\right]^{-1} \nabla\boldsymbol{h}(\boldsymbol{x}_k)\boldsymbol{z} \\ &= \nabla\boldsymbol{h}(\boldsymbol{x}_k)\boldsymbol{z} - \nabla\boldsymbol{h}(\boldsymbol{x}_k)\boldsymbol{z} = \boldsymbol{0}\end{aligned}$$

It now remains to prove that $\lim_{k\to\infty} \boldsymbol{z}_k = \boldsymbol{z}$ if $\boldsymbol{z}$ is in the kernel of $\nabla\boldsymbol{h}(\boldsymbol{x}^*)$.

Proof

Step 4: necessary condition for a minimum

(12/12)

Consider the limit

$$\begin{aligned}\lim_{k\to\infty} \boldsymbol{z}_k &= \boldsymbol{z} - \lim_{k\to\infty} \nabla \boldsymbol{h}(\boldsymbol{x}_k)^T \left[\nabla \boldsymbol{h}(\boldsymbol{x}_k)\nabla \boldsymbol{h}(\boldsymbol{x}_k)^T\right]^{-1} \nabla \boldsymbol{h}(\boldsymbol{x}_k)\boldsymbol{z} \\ &= \boldsymbol{z} - \nabla \boldsymbol{h}(\boldsymbol{x}^*)^T \left[\nabla \boldsymbol{h}(\boldsymbol{x}^*)\nabla \boldsymbol{h}(\boldsymbol{x}^*)^T\right]^{-1} \nabla \boldsymbol{h}(\boldsymbol{x}^*)\boldsymbol{z}\end{aligned}$$

and, thus, if $\boldsymbol{z}$ in in the kernel of $\nabla \boldsymbol{h}(\boldsymbol{x}^*)$, i.e. $\nabla \boldsymbol{h}(\boldsymbol{x}^*)\boldsymbol{z} = \boldsymbol{0}$ it follows

$$\lim_{k\to\infty} \boldsymbol{z}_k = \boldsymbol{z}$$

and this concludes the proof.

First order necessary condition

- $f \in \mathtt{C}^1(\mathbb{R}^n)$ function to be minimized
- $\boldsymbol{h} \in \mathtt{C}^1(\mathbb{R}^n, \mathbb{R}^m)$ constraints map
- $\boldsymbol{h}(\boldsymbol{x}^*) = \boldsymbol{0}$ and $\nabla \boldsymbol{h}(\boldsymbol{x}^*)$ is of maximum rank

If $\boldsymbol{x}^*$ is a local minimum of $f(\boldsymbol{x})$ then there exists m scalars λ_k such that

$$\nabla f(\boldsymbol{x}^*) = \sum_{k=1}^{m} \lambda_k \nabla h_i(\boldsymbol{x}^*)$$

i.e. the gradient of the function is in the linear space generated by gradient of the constraints:

$$\nabla f(\boldsymbol{x}^*) \in \textsc{span}\{\nabla h_1(\boldsymbol{x}^*), \nabla h_2(\boldsymbol{x}^*), \ldots, \nabla h_m(\boldsymbol{x}^*)\}$$

Second order necessary conditions

- $f \in \mathtt{C}^2(\mathbb{R}^n)$ function to be minimized
- $\boldsymbol{h} \in \mathtt{C}^2(\mathbb{R}^n, \mathbb{R}^m)$ constraints map
- $\boldsymbol{h}(\boldsymbol{x}^*) = \boldsymbol{0}$ and $\nabla \boldsymbol{h}(\boldsymbol{x}^*)$ if of maximum rank

If $\boldsymbol{x}^*$ is a local minimum of $f(\boldsymbol{x})$ in addition to satisfy first order necessary condition for all $\boldsymbol{z} \in \mathbb{R}^n$ that satisfy $\nabla \boldsymbol{h}(\boldsymbol{x}^*)\boldsymbol{z} = \boldsymbol{0}$ the following inequality must be true

$$\boldsymbol{z}^T \left(\nabla^2 f(\boldsymbol{x}^*) - \sum_{k=1}^{m} \lambda_k \nabla^2 h_k(\boldsymbol{x}^*) \right) \boldsymbol{z} \geq 0$$

in other words the matrix $\nabla_x^2 \big(f(\boldsymbol{x}^*) - \boldsymbol{\lambda} \cdot \boldsymbol{h}(\boldsymbol{x}^*) \big)$ is semi-positive definite in the Kernel of $\nabla \boldsymbol{h}(\boldsymbol{x}^*)$.

Second order sufficient conditions

- $f \in \mathtt{C}^2(\mathbb{R}^n)$ function to be minimized
- $\boldsymbol{h} \in \mathtt{C}^2(\mathbb{R}^n, \mathbb{R}^m)$ constraints map
- $\boldsymbol{h}(\boldsymbol{x}^*) = \mathbf{0}$ and $\nabla \boldsymbol{h}(\boldsymbol{x}^*)$ if of maximum rank
- $\boldsymbol{x}^*$ satisfy first order necessary conditions

If for all $\boldsymbol{z} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ that satisfy $\nabla \boldsymbol{h}(\boldsymbol{x}^*)\boldsymbol{z} = \mathbf{0}$ satisfy also

$$\boldsymbol{z}^T \left(\nabla^2 f(\boldsymbol{x}^*) - \sum_{k=1}^{m} \lambda_k \nabla^2 h_k(\boldsymbol{x}^*) \right) \boldsymbol{z} > 0$$

Then $\boldsymbol{x}^*$ is a local minimum. In other words if the matrix $\nabla_x^2 \big(f(\boldsymbol{x}^*) - \boldsymbol{\lambda} \cdot \boldsymbol{h}(\boldsymbol{x}^*) \big)$ is positive definite in the Kernel of $\nabla \boldsymbol{h}(\boldsymbol{x}^*)$ then $\boldsymbol{x}^*$ is a local minimum.

Lagrange multiplier practical usage

When you deal with a constrained minimization problem of the form:

$$\text{minimize:} \quad f(\boldsymbol{x})$$

with constraints

$$\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}$$

behove define the Lagrangia

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) - \boldsymbol{\lambda} \cdot \boldsymbol{h}(\boldsymbol{x})$$

such that the minimum/maximum point are stationary points of $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda})$

$$\nabla_x \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = \nabla_x f(\boldsymbol{x}) - \boldsymbol{\lambda}^T \nabla_x \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}$$

$$\nabla_\lambda \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}$$

Lagrange multiplier practical usage

Consider the pair $(\boldsymbol{x}, \boldsymbol{\lambda})$ that satisfy

$$\nabla_x \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = \mathbf{0} \qquad \nabla_\lambda \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = \mathbf{0}$$

and the matrix

$$\nabla_x^2 \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = \nabla_x^2 f(\boldsymbol{x}) - \sum_{k=1}^{m} \lambda_k \nabla_x^2 \boldsymbol{h}_k(\boldsymbol{x})$$

then, necessary and sufficient conditions for a local maximum/minimum are the following: (next slide)

Lagrange multiplier practical usage

- If $\boldsymbol{x}$ is a local minimum point then $\nabla_x^2 \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda})$ is semi-positive definite in the Kernel of $\nabla \boldsymbol{h}(\boldsymbol{x}^*)$, i.e.

$$\boldsymbol{z}^T \nabla_x^2 \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) \boldsymbol{z} \geq 0, \qquad \forall \boldsymbol{z} \in \mathrm{Ker}\{\nabla \boldsymbol{h}(\boldsymbol{x}^*)\}$$

  If $\boldsymbol{x}$ is a local maximum point then $\nabla_x^2 \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda})$ is semi-positive definite in the Kernel of $\nabla \boldsymbol{h}(\boldsymbol{x}^*)$, i.e

$$\boldsymbol{z}^T \nabla_x^2 \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) \boldsymbol{z} \leq 0, \qquad \forall \boldsymbol{z} \in \mathrm{Ker}\{\nabla \boldsymbol{h}(\boldsymbol{x}^*)\}$$

- If $\nabla_x^2 \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda})$ is positive definite in the Kernel of $\nabla \boldsymbol{h}(\boldsymbol{x}^*)$, i.e.

$$\boldsymbol{z}^T \nabla_x^2 \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) \boldsymbol{z} > 0, \qquad \forall \boldsymbol{z} \in \mathrm{Ker}\{\nabla \boldsymbol{h}(\boldsymbol{x}^*)\} \setminus \{\boldsymbol{0}\}$$

  then $\boldsymbol{x}$ is a local minimum point. Similarly if $\nabla_x^2 \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda})$ is positive definite in the Kernel of $\nabla \boldsymbol{h}(\boldsymbol{x}^*)$, i.e.

$$\boldsymbol{z}^T \nabla_x^2 \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) \boldsymbol{z} < 0, \qquad \forall \boldsymbol{z} \in \mathrm{Ker}\{\nabla \boldsymbol{h}(\boldsymbol{x}^*)\} \setminus \{\boldsymbol{0}\}$$

  then $\boldsymbol{x}$ is a local maximum point.

Example (1/5)

Find minimum and maximum point of the function

$$f(x,y) = e^{x^2-y^2}$$

with constraint

$$h(x,y) = x - y^2$$

build the Lagrangian

$$\mathcal{L}(x,y,\lambda) = e^{x^2-y^2} - \lambda(x - y^2)$$

the stationary points satisfy

$$\nabla_x \mathcal{L}(x,y,\lambda) = 2\,xe^{x^2-y^2} - \lambda = 0$$

$$\nabla_y \mathcal{L}(x,y,\lambda) = -2\,ye^{x^2-y^2} + 2\,\lambda\,y = 0$$

$$\nabla_\lambda \mathcal{L}(x,y,\lambda) = -x + y^2 = 0$$

Example (2/5)

the stationary points are:

x	y	λ
0	0	0
$\frac{1}{2}$	$\frac{1}{\sqrt{2}}$	$e^{-\frac{1}{4}}$
$\frac{1}{2}$	$-\frac{1}{\sqrt{2}}$	$e^{-\frac{1}{4}}$

and the gradient of the constraints

$$\nabla h(x,y) = (1, -2y)$$

while Hessian is

$$\nabla^2_{(x,y)}\mathcal{L} = \begin{pmatrix} (4x^2+2)e^{x^2-y^2} & -4\,x\,y\,e^{x^2-y^2} \\ -4\,x\,ye^{x^2-y^2} & (4y^2-2)e^{x^2-y^2} + 2\,\lambda \end{pmatrix}$$

Example

First point $x = y = \lambda = 0$:

$$\nabla h(0,0) = \big(1, 0\big)$$

$$\nabla^2_{(x,y)} \mathcal{L}(0,0,0) = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}$$

the vectors in the nel kernel of $\nabla h(0,0)$ satisfy:

$$\nabla h(0,0) \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = z_1 = 0$$

and thus are of the form $\boldsymbol{z}^T = [0, \alpha]$

$$\begin{pmatrix} 0 & \alpha \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix} \begin{pmatrix} 0 \\ \alpha \end{pmatrix} = -2\alpha^2 < 0$$

and the point is a local maximum.

Example (4/5)

Second point $x = \frac{1}{2}$, $y = \frac{1}{\sqrt{2}}$ and $\lambda = e^{-\frac{1}{4}}$

$$\nabla h\left(\frac{1}{2}, \frac{1}{\sqrt{2}}\right) = \begin{pmatrix} 1 & -\sqrt{2} \end{pmatrix}$$

$$\nabla^2_{(x,y)}\mathcal{L}\left(\frac{1}{2}, \frac{1}{\sqrt{2}}, e^{-\frac{1}{4}}\right) = e^{-1/4}\begin{pmatrix} 3 & -\sqrt{2} \\ -\sqrt{2} & 2 \end{pmatrix}$$

the vectors in the kernel of $\nabla h(0,0)$ satisfy:

$$\nabla h(0,0)\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = z_1 - \sqrt{2}\, z_2 = 0$$

and thus are of the form $\boldsymbol{z}^T = [\alpha\sqrt{2}, \alpha]$

$$e^{-1/4}\begin{pmatrix} \alpha\sqrt{2} & \alpha \end{pmatrix}\begin{pmatrix} 3 & -\sqrt{2} \\ -\sqrt{2} & 2 \end{pmatrix}\begin{pmatrix} \alpha\sqrt{2} \\ \alpha \end{pmatrix} = 4e^{-\frac{1}{2}}\alpha^2 > 0$$

and are local minimum points.

Example (5/5)

Second point $x = \frac{1}{2}$, $y = -\frac{1}{\sqrt{2}}$ and $\lambda = e^{-\frac{1}{4}}$

$$\nabla h\left(\frac{1}{2}, -\frac{1}{\sqrt{2}}\right) = \begin{pmatrix} 1 & \sqrt{2} \end{pmatrix}$$

$$\nabla^2_{(x,y)}\mathcal{L}\left(\frac{1}{2}, -\frac{1}{\sqrt{2}}, e^{-\frac{1}{4}}\right) = e^{-1/4}\begin{pmatrix} 3 & \sqrt{2} \\ \sqrt{2} & 2 \end{pmatrix}$$

the vector in the kernel of $\nabla h(0,0)$ satisfy:

$$\nabla h(0,0)\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = z_1 + \sqrt{2}\, z_2 = 0$$

and thus are of the form $\boldsymbol{z}^T = [\alpha\sqrt{2}, -\alpha]$

$$e^{-1/4}\begin{pmatrix} \alpha\sqrt{2} & -\alpha \end{pmatrix}\begin{pmatrix} 3 & \sqrt{2} \\ \sqrt{2} & 2 \end{pmatrix}\begin{pmatrix} \alpha\sqrt{2} \\ -\alpha \end{pmatrix} = 4e^{-\frac{1}{2}}\alpha^2 > 0$$

and thus is a local minimum.

Karush-Kuhn-Tucker optimality conditions (1/8)

- Add auxiliary variable ε_k for each inequality to the problem

$$\begin{aligned} \text{Minimize} \quad & f(\boldsymbol{x}) \\ \text{With constraints} \quad & h_k(\boldsymbol{x}) = 0, \qquad k = 1, 2, \ldots, m \\ & g_k(\boldsymbol{x}) \geq 0, \qquad k = 1, 2, \ldots, p \end{aligned}$$

- is thus transformed in the following minimization problem

$$\begin{aligned} \text{Minimize} \quad & \mathcal{F}(\boldsymbol{y}) = \mathcal{F}(\boldsymbol{x}, \boldsymbol{\varepsilon}) = f(\boldsymbol{x}) \\ \text{With constraints} \quad & \mathcal{H}_k(\boldsymbol{y}) = 0, \qquad k = 1, 2, \ldots, m + p \end{aligned}$$

where

$$\mathcal{F}(\boldsymbol{y}) = \mathcal{F}(\boldsymbol{x}, \boldsymbol{\varepsilon}) = f(\boldsymbol{x})$$

$$\mathcal{H}_k(\boldsymbol{y}) = \mathcal{H}_k(\boldsymbol{x}, \boldsymbol{\varepsilon}) = \begin{cases} h_k(\boldsymbol{x}) & \text{per } k \leq m \\ g_{k-m}(\boldsymbol{x}) - \frac{1}{2}\varepsilon_{k-m}^2 & \text{per } k > m \end{cases}$$

Karush-Kuhn-Tucker optimality conditions (2/8)

Given the problem

$$\text{Minimizzare} \qquad \mathcal{F}(\boldsymbol{y})$$

$$\text{Con vincoli} \qquad \mathcal{H}_k(\boldsymbol{y}) = 0, \qquad k = 1, 2, \ldots, m + p$$

characterization of maximum/minimum points are obtained using previously developed condition using Lagrange multiplier.

Using peculiar structure of the problem this condition can be rewritten without the explicit use of the slack variables (the ε_k)

This conditions are called KKT conditions (Karush-Kuhn-Tucker)

Karush-Kuhn-Tucker optimality conditions (3/8)

First order conditions:
From the Lagrangian

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\varepsilon}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\boldsymbol{x}) - \sum_{k=1}^{m} \lambda_k h_k(\boldsymbol{x}) - \sum_{k=1}^{p} \mu_k \left(g_k(\boldsymbol{x}) - \frac{1}{2}\varepsilon_k^2 \right)$$

null gradient becomes

$$\nabla_x \mathcal{L}(\boldsymbol{x}, \boldsymbol{\varepsilon}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \nabla f(\boldsymbol{x}) - \sum_{k=1}^{m} \lambda_k \nabla h_k(\boldsymbol{x}) - \sum_{k=1}^{p} \mu_k \nabla g_k(\boldsymbol{x})$$

$$\nabla_\varepsilon \mathcal{L}(\boldsymbol{x}, \boldsymbol{\varepsilon}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \begin{pmatrix} \mu_1 & & \\ & \ddots & \\ & & \mu_p \end{pmatrix} \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_p \end{pmatrix}$$

Karush-Kuhn-Tucker optimality conditions

Observing that $\frac{1}{2}\varepsilon_k^2 = g_k(\boldsymbol{x})$ condition become

$$\begin{aligned} \nabla f(\boldsymbol{x}) &= \sum_{k=1}^{m} \lambda_k \nabla h_k(\boldsymbol{x}) + \sum_{k=1}^{p} \mu_k \nabla g_k(\boldsymbol{x}) \\ 0 &= \mu_k g_k(\boldsymbol{x}) \end{aligned}$$

moreover the Hessian is

$$\nabla_x^2 \mathcal{L}(\boldsymbol{x}, \boldsymbol{\varepsilon}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \nabla^2 f(\boldsymbol{x}) - \sum_{k=1}^{m} \lambda_k \nabla^2 h_k(\boldsymbol{x}) - \sum_{k=1}^{p} \mu_k \nabla^2 g_k(\boldsymbol{x})$$

Karush-Kuhn-Tucker optimality conditions (5/8)

Evaluating Hessian respect to $\boldsymbol{x}$ and $\boldsymbol{\varepsilon}$

$$\nabla^2_{\varepsilon}\mathcal{L}(\boldsymbol{x},\boldsymbol{\varepsilon},\boldsymbol{\lambda},\boldsymbol{\mu}) = \begin{pmatrix} \mu_1 & & \\ & \ddots & \\ & & \mu_p \end{pmatrix} = \boldsymbol{M}$$

$$\nabla_x\nabla_{\varepsilon}\mathcal{L}(\boldsymbol{x},\boldsymbol{\varepsilon},\boldsymbol{\lambda},\boldsymbol{\mu}) = \boldsymbol{0}$$

and thus

$$\nabla^2_{(x,\varepsilon)}\mathcal{L}(\boldsymbol{x},\boldsymbol{\varepsilon},\boldsymbol{\lambda},\boldsymbol{\mu}) = \begin{pmatrix} \nabla^2_x\mathcal{L} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{M} \end{pmatrix}$$

Karush-Kuhn-Tucker optimality conditions (6/8)

Evaluating gradient of constraints respect to $\boldsymbol{x}$, $\boldsymbol{\varepsilon}$

$$\frac{\partial \mathcal{H}(\boldsymbol{x}, \boldsymbol{\varepsilon})}{\partial(\boldsymbol{x}, \boldsymbol{\varepsilon})} = \begin{pmatrix} \nabla \boldsymbol{h}(\boldsymbol{x}) & \boldsymbol{0} \\ \nabla \boldsymbol{g}(\boldsymbol{x}) & -\boldsymbol{E} \end{pmatrix}$$

where

$$\boldsymbol{E} = \begin{pmatrix} \varepsilon_1 & & \\ & \ddots & \\ & & \varepsilon_p \end{pmatrix}$$

The admissible direction are the vectors $(\boldsymbol{z}, \boldsymbol{w})$ such that

$$\begin{pmatrix} \nabla \boldsymbol{h}(\boldsymbol{x}) & \boldsymbol{0} \\ \nabla \boldsymbol{g}(\boldsymbol{x}) & -\boldsymbol{E} \end{pmatrix} \begin{pmatrix} \boldsymbol{z} \\ \boldsymbol{w} \end{pmatrix} = \begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix}$$

Karush-Kuhn-Tucker optimality conditions (7/8)

Necessary conditions becomes

$$\boldsymbol{z}^T \nabla_x^2 \mathcal{L} \boldsymbol{z} + \sum_{k=1}^{p} \mu_k w_k^2 \geq 0$$

for all $\boldsymbol{z}$ and $\boldsymbol{w}$ such that

$$\nabla \boldsymbol{h}(\boldsymbol{x})\boldsymbol{z} = \boldsymbol{0}$$

$$\nabla \boldsymbol{g}(\boldsymbol{x})\boldsymbol{z} = \boldsymbol{E}\boldsymbol{w}$$

Karush-Kuhn-Tucker optimality conditions (8/8)

Active constraints are the constraints for $k \in \mathcal{A}(\boldsymbol{x})$ i.e. $g_k(\boldsymbol{x}) = 0$ where $\varepsilon_k = 0$ and thus w_k can assume any values without modify $\boldsymbol{z}$. Thus using $\boldsymbol{z} = \mathbf{0}$ and choosing $(\boldsymbol{w})_i = [\delta_{ik}]$

$$\mathbf{0}^T \left(\nabla_x^2 \mathcal{L}\right) \mathbf{0} + \mu_k w_k^2 \geq 0 \qquad \mu_k \geq 0$$

$$\nabla g_k(\boldsymbol{x}) \boldsymbol{z} = 0$$

For inactive constraints i.e. $k \notin \mathcal{A}(\boldsymbol{x})$ or $g_k(\boldsymbol{x}) > 0$ the values $\varepsilon_k \neq 0$ and from first order conditions $\mu_k = 0$. Thus, w_k can assume any values without modify quadratic form, and

$$\nabla g_k(\boldsymbol{x}) \boldsymbol{z} = \varepsilon_k w_k$$

can assume any values.

Constrained minimization
Problem

Let be $f \in \mathtt{C}^2(\mathbb{R}^n)$ a function and $g_k \in \mathtt{C}^2(\mathbb{R}^n)$ $(k = 1, 2, \ldots, p)$ and $h_k \in \mathtt{C}^2(\mathbb{R}^n)$ $(k = 1, 2, \ldots, m)$ constraints.

Problem

$$
\begin{aligned}
&\text{Minimize} && f(\boldsymbol{x}) && \\
&\text{With constraints:} && g_k(\boldsymbol{x}) \geq 0, && k = 1, 2, \ldots, p \\
& && h_k(\boldsymbol{x}) = 0, && k = 1, 2, \ldots, m
\end{aligned}
$$

First order Karush-Kuhn-Tucker conditions

Theorem (F.John)

Let $f \in \mathtt{C}^1(\mathbb{R}^n)$ a function and $\boldsymbol{g} \in \mathtt{C}^1(\mathbb{R}^n, \mathbb{R}^p)$ with $\boldsymbol{h} \in \mathtt{C}^1(\mathbb{R}^n, \mathbb{R}^m)$ some constraints. Necessary condition for $\boldsymbol{x}^*$ be a local minimum is that there exists $m+p+1$ scalars (not all 0) such that the following condition are satified

$$\lambda_0 \nabla f(\boldsymbol{x}^*) - \sum_{k=1}^{p} \mu_k \nabla \boldsymbol{g}_k(\boldsymbol{x}^*) - \sum_{k=1}^{m} \lambda_k \nabla \boldsymbol{h}_k(\boldsymbol{x}^*) = \mathbf{0}^T$$

$$\begin{aligned}
&\boldsymbol{h}_k(\boldsymbol{x}^*) = 0, && k = 1, 2, \ldots, m;\\
&\boldsymbol{g}_k(\boldsymbol{x}^*) \geq 0, && k = 1, 2, \ldots, p;\\
&\mu_k \boldsymbol{g}_k(\boldsymbol{x}^*) = 0, && k = 1, 2, \ldots, p;\\
&\mu_k \geq 0, && k = 1, 2, \ldots, p;
\end{aligned}$$

Definition (Constraint qualifications)

Let be $\boldsymbol{g} \in \mathtt{C}^2(\mathbb{R}^n, \mathbb{R}^p)$ *inequality constraints and* $\boldsymbol{h} \in \mathtt{C}^2(\mathbb{R}^n, \mathbb{R}^m)$ *equality contraints. The point* $\boldsymbol{x}^*$ *is qualified if*

- $\boldsymbol{g}_k(\boldsymbol{x}^*) = 0, \qquad k \in \mathcal{A}(\boldsymbol{x}^*)$;
- $\boldsymbol{g}_k(\boldsymbol{x}^*) > 0, \qquad k \notin \mathcal{A}(\boldsymbol{x}^*)$;

moreover the vectors

$$\{\nabla \boldsymbol{g}_k(\boldsymbol{x}^*) \ : \ k \in \mathcal{A}(\boldsymbol{x}^*)\} \cup \{\nabla \boldsymbol{h}_1(\boldsymbol{x}^*), \nabla \boldsymbol{h}_2(\boldsymbol{x}^*), \ldots, \nabla \boldsymbol{h}_m(\boldsymbol{x}^*)\}$$

are linearly independent.

First order Karush-Kuhn-Tucker conditions

Theorem (First order KKT conditions)

Let $f \in \mathtt{C}^1(\mathbb{R}^n)$ a function and $\boldsymbol{g} \in \mathtt{C}^1(\mathbb{R}^n, \mathbb{R}^p)$ with $\boldsymbol{h} \in \mathtt{C}^1(\mathbb{R}^n, \mathbb{R}^m)$ constraint maps. If $\boldsymbol{x}^*$ satisfy constraint qualification then necessary condition for $\boldsymbol{x}^*$ be a local minima is that there exists $m + p$ scalars such that the following conditions are satisfied

$$
\begin{aligned}
\nabla_x \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) &= \mathbf{0}^T \\
\boldsymbol{h}_k(\boldsymbol{x}^*) &= 0, \qquad k = 1, 2, \ldots, m; \\
\mu_k^* \boldsymbol{g}_k(\boldsymbol{x}^*) &= 0, \qquad k = 1, 2, \ldots, p; \\
\mu_k^* &\geq 0, \qquad k = 1, 2, \ldots, p;
\end{aligned}
$$

where

$$
\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\boldsymbol{x}) - \sum_{k=1}^{p} \mu_k \, \boldsymbol{g}_k(\boldsymbol{x}) - \sum_{k=1}^{m} \lambda_k \, \boldsymbol{h}_k(\boldsymbol{x})
$$

Second order Karush-Kuhn-Tucker conditions

Theorem (Second order KKT conditions)

Let $f \in \mathtt{C}^1(\mathbb{R}^n)$ a function and $\boldsymbol{g} \in \mathtt{C}^1(\mathbb{R}^n, \mathbb{R}^p)$ with $\boldsymbol{h} \in \mathtt{C}^1(\mathbb{R}^n, \mathbb{R}^m)$ constraint maps. If $\boldsymbol{x}^*$ satisfy constraint qualification then necessary condition for $\boldsymbol{x}^*$ be a local minima is that there exists $m+p$ scalars that satisfy first order conditions, moreover

$$\boldsymbol{z}^T \nabla_x^2 \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \boldsymbol{z} \geq 0$$

for all $\boldsymbol{z}$ such that

$$\begin{aligned} \nabla \boldsymbol{h}_k(\boldsymbol{x}^*) \boldsymbol{z} &= 0, && k = 1, 2, \ldots, m \\ \nabla \boldsymbol{g}_k(\boldsymbol{x}^*) \boldsymbol{z} &= 0, && se\ k \in \mathcal{A}(\boldsymbol{x}^*) \end{aligned}$$

Finally $\mu_k > 0$ for all $k \in \mathcal{A}(\boldsymbol{x}^*)$.

Second order Karush-Kuhn-Tucker conditions

Theorem (Sufficient second order KKT conditions)

Let $f \in \mathtt{C}^1(\mathbb{R}^n)$ *a function and* $\boldsymbol{g} \in \mathtt{C}^1(\mathbb{R}^n, \mathbb{R}^p)$ *with* $\boldsymbol{h} \in \mathtt{C}^1(\mathbb{R}^n, \mathbb{R}^m)$ *constraint maps. If* $\boldsymbol{x}^*$ *satisfy constraint qualification then necessary condition for* $\boldsymbol{x}^*$ *be a local minima is that there exists* $m+p$ *scalars that satisfy first order conditions, moreover*

$$\boldsymbol{z}^T \nabla_x^2 \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)\boldsymbol{z} > 0$$

for all $\boldsymbol{z} \neq \boldsymbol{0}$ *such that*

$$\begin{aligned} \nabla \boldsymbol{h}_k(\boldsymbol{x}^*)\boldsymbol{z} &= 0, & k &= 1, 2, \ldots, m \\ \nabla \boldsymbol{g}_k(\boldsymbol{x}^*)\boldsymbol{z} &= 0, & se\ k &\in \mathcal{A}(\boldsymbol{x}^*) \end{aligned}$$

Finally $\mu_k > 0$ *for all* $k \in \mathcal{A}(\boldsymbol{x}^*)$.

KKT usage example

Minimize

$$f(x, y) = x^2 - xy$$

with constraints

$$g_1(x, y) = 1 - x^2 - y^2 \geq 0$$
$$g_2(x, y) = 1 - (x - 1)^2 - y^2 \geq 0$$

KKT usage example

Solution with constraints Activation/Deactivation (1/10)

Lagrangian

$$
\begin{aligned}
\mathcal{L}(x, y, \mu_1, \mu_2) &= x^2 - xy \\
&- \mu_1(1 - x^2 - y^2) \\
&- \mu_2(1 - (x-1)^2 - y^2)
\end{aligned}
$$

gradient respect to (x, y)

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial x} &= 2x - y + 2x\mu_1 + 2(x-1)\mu_2 \\
\frac{\partial \mathcal{L}}{\partial y} &= -x + 2y(\mu_1 + \mu_2)
\end{aligned}
$$

KKT usage example

Solution with constraints Activation/Deactivation

(2/10)

Search minima in the internal part of the domain (i.e. $\mu_1 = \mu_2 = 0$). Must solve

$$0 = 2x - y$$

$$0 = -x$$

solution $x = 0$, $y = 0$. Check constraits

$$g_1(0,0) = 1 > 0$$

$$g_2(0,0) = 0 > 0$$

Then second constraints must be active, thus, solution must be discarded.

KKT usage example

Solution with constraints Activation/Deactivation

Activate first constraint only (i.e. $\mu_2 = 0$). Must solve

$$
\begin{aligned}
0 &= 2x - y + 2x\mu_1 \\
0 &= -x + 2y\mu_1 \\
1 &= x^2 + y^2
\end{aligned}
$$

found 4 solutions

x	y	μ_1
$\pm 1/2\sqrt{2-\sqrt{2}}$	$x\left(1+\sqrt{2}\right)$	$(\sqrt{2}-1)/2$
$\pm 1/2\sqrt{2+\sqrt{2}}$	$x\left(1-\sqrt{2}\right)$	$-(\sqrt{2}+1)/2$

Soltion n.3 and n.4 must be discarded cause $\mu_1 < 0$.

KKT usage example

Solution with constraints Activation/Deactivation (4/10)

Check first 2 solution for the second constraint

$$g_2(x_1, y_1) = \sqrt{2-\sqrt{2}} - 1 = -0.23\ldots < 0$$

$$g_2(x_2, y_2) = -\sqrt{2-\sqrt{2}} - 1 = -1.76\ldots < 0$$

No one satisfy constraint, solutions must be discarded.

KKT usage example

Solution with constraints Activation/Deactivation

Activate second constraint (i.e. $\mu_1 = 0$). Must solve

$$
\begin{aligned}
0 &= 2x - y + 2(x-1)\mu_2 \\
0 &= -x + 2y\mu_2 \\
1 &= (x-1)^2 + y^2
\end{aligned}
$$

found 3 solutions

x	y	μ_2
0	0	0
$(5-\sqrt{7})/4$	$(1+\sqrt{7})/4$	$\sqrt{7}/2-1$
$(5+\sqrt{7})/4$	$(1-\sqrt{7})/4$	$-\sqrt{7}/2-1$

Solution n.3 must be discarded as $\mu_2 < 0$.

KKT usage example

Solution with constraints Activation/Deactivation

(6/10)

check first 2 solution for second constraint

$$
\begin{aligned}
g_2(x_1, y_1) &= 1 > 0 \\
g_2(x_2, y_2) &= (\sqrt{7} - 3)/2 = -0.177\ldots < 0
\end{aligned}
$$

only the first satisfy constraint.

KKT usage example

Solution with constraints Activation/Deactivation

Activate both contraint. Must solve

$$
\begin{aligned}
0 &= 2x - y + 2x\mu_1 + 2(x-1)\mu_2 \\
0 &= -x + 2y(\mu_1 + \mu_2) \\
1 &= x^2 + y^2 \\
1 &= (x-1)^2 + y^2
\end{aligned}
$$

found 2 solution

x	y	μ_1	μ_2
$1/2$	$\sqrt{3}/2$	$-1/2+1/\sqrt{3}$	$1/2-1/(3\sqrt{3})$
$1/2$	$-\sqrt{3}/2$	$-1/2-1/\sqrt{3}$	$1/2+1/(3\sqrt{3})$

KKT usage example

Solution with constraints Activation/Deactivation

The candidates which satisfy first order KKT conditions are:

x	y	μ_1	μ_2
0	0	0 (*)	0
$1/2$	$\sqrt{3}/2$	$-1/2+1/\sqrt{3}$	$1/2-1/(3\sqrt{3})$
$1/2$	$-\sqrt{3}/2$	$-1/2-1/\sqrt{3}$	$1/2+1/(3\sqrt{3})$

now check second order conditions.

(*) constraint is active with null multiplier.

KKT usage example

Solution with constraints Activation/Deactivation (9/10)

gradient of the constraints and Hessian

$$\nabla \boldsymbol{g}(x,y) = \begin{pmatrix} 2x & 2y \\ 2(x-1) & 2y \end{pmatrix}$$

$$\nabla^2_{(x,y)}\mathcal{L}(x,y,\mu_1,\mu_2) = \begin{pmatrix} 2(1+\mu_1+\mu_2) & -1 \\ -1 & 2(\mu_1+\mu_2) \end{pmatrix}$$

For the first point the gradient of the active constraint:

$$\nabla g_1(0,0) = \mathbf{0}^T$$

gradient is null, thus constraint is not qualified!. Cannot apply KKT theorem.

KKT usage example

Solution with constraints Activation/Deactivation (10/10)

For the second point must solve (z_1, z_2) such that:

$$\begin{pmatrix} 1 & \sqrt{3} \\ -1 & \sqrt{3} \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

i.e. $z_1 = z_2 = 0$. Thus the point satisfy necessary conditions for a minimm but not sufficient condsitions.
For the third point must solve (z_1, z_2) such that:

$$\begin{pmatrix} 1 & -\sqrt{3} \\ -1 & -\sqrt{3} \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

i.e. $z_1 = z_2 = 0$. Thus the point satisfy necessary conditions for a minimm but not sufficient condsitions.

Least squares solution of linear equations

- Minimize

$$f(\boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{x}$$

- With constraints

$$\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}$$

Kantorovich inequality

- Minimize

$$f(\boldsymbol{x}) = (\boldsymbol{x}^T \boldsymbol{A}\boldsymbol{x})(\boldsymbol{x}^T \boldsymbol{A}^{-1}\boldsymbol{x})$$

- With constraints

$$\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{x} - 1$$

If $\boldsymbol{A}$ is symmetric ad positive definite

$$\min f(\boldsymbol{x}) = \frac{(\lambda_{\min} + \lambda_{\max})^2}{4\lambda_{\min}\lambda_{\max}}$$

Simple circuit optimization
(Chong Zak problem)

Consider the circuit in figure. Voltage generator is $20V$ while $R_2 = 10\Omega$. Resistor R_1 is unknown and must be found to minimize power loss on R_1.

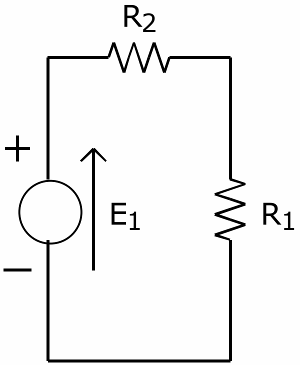
- Massimize the power loss on R_1, i.e. minimize

$$f(R_1, i) = -R_1\, i^2$$

- With constraints

$$g(R_1, i) = R_1 \geq 0$$

$$h(R_1, i) = 20 - (R_1 + 10)\, i = 0$$



Massimizzazione di un volume

Let x, y, z width height and depth of a parallelepiped. Find the dimension which maximize the volume when surface being equal to S.

- Minimize

$$f(x, y, z) = -xyz$$

- With constraints

$$\begin{aligned} h(x, y, z) &= 2(xy + yz + xz) - S &= 0 \\ g_1(x, y, z) &= x &\geq 0 \\ g_2(x, y, z) &= y &\geq 0 \\ g_3(x, y, z) &= z &\geq 0 \end{aligned}$$

links in a chain distributions

Consider a chain composed by $n+1$ links, fixed on the ceiling in $(0,0)$ and $(L,0)$. Let (x_k, y_k) the points of contacts on the links inside the chain. Compute the position of the mesh of the chain under gravity.

- Minimize the potential energy

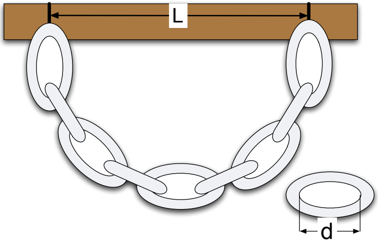
$$f(\boldsymbol{y}) = \sum_{k=1}^{n-1} y_k$$

- with constraints

$$y_0 = y_n = 0,$$

$$x_0 = 0, \qquad x_n = L,$$

$$(x_k - x_{k-1})^2 + (y_k - y_{k-1})^2 = d^2$$



SPD matrices in a subspace

Verification of KKT conditions needs the verification that a matrix $\boldsymbol{A}$ is positive definite in the kernel of another matrix $\boldsymbol{B}$.
That is, we have the problem

Problem (constrained SPD)

Verify if the matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ is positive definite in the kernel of $\boldsymbol{B} \in \mathbb{R}^{m\times n}$ $(m < n)$, namely

$$\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} > 0, \qquad \forall \boldsymbol{x} \neq \boldsymbol{0}, \quad \textit{such that} \quad \boldsymbol{B}\boldsymbol{x} = \boldsymbol{0}$$

or if the matrix $\boldsymbol{A}$ is semi-positive definite in the kernel of $\boldsymbol{B}$, namely

$$\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} \geq 0, \qquad \forall \boldsymbol{x}, \quad \textit{such that} \quad \boldsymbol{B}\boldsymbol{x} = \boldsymbol{0}$$

For the solution of the previous problem is necessary the following theorem.

Theorem (Sylvester)

A symmetric matrix $\boldsymbol{A}$ *is positive definite if and only if all of the determinants of leading principal minors must be positive. In other words let* $\boldsymbol{A}$ *and* $\boldsymbol{D}_k$ *a principal minor*

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}, \qquad \boldsymbol{D}_k = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1k} \\ a_{21} & a_{22} & \dots & a_{2k} \\ \vdots & & & \vdots \\ a_{k1} & a_{k2} & \dots & a_{kk} \end{pmatrix},$$

then

$$\boldsymbol{A} \text{ è } SPD \qquad \Leftrightarrow \qquad |\boldsymbol{D}_k| > 0, \quad k = 1, 2, \dots, n$$

For semi-SPD matrix it is true

$$\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} + \varepsilon \boldsymbol{x}^T\boldsymbol{x} > 0, \qquad \forall \boldsymbol{x} \neq \boldsymbol{0}$$

and applying Sylvester theorem for $\boldsymbol{A} + \varepsilon\boldsymbol{I}$ it follows that all of the determinants of leading principal minors must be positive. One argue that if all of the determinants of leading principal minors are non-negative then the matrix $\boldsymbol{A}$ is semi-positive definite. This is false and here is a counter example for the matrix $\boldsymbol{P}$

$$\boldsymbol{P} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix} \qquad \left|\begin{pmatrix} 1 \end{pmatrix}\right| = 1, \quad \left|\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\right| = 0, \quad \left|\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix}\right| = 0$$

ma per la matrice perturbata $\boldsymbol{P} + \varepsilon\boldsymbol{I}$

$$\left|\begin{pmatrix} 1+\varepsilon \end{pmatrix}\right| = 1+\varepsilon, \quad \left|\begin{pmatrix} 1+\varepsilon & 1 \\ 1 & 1+\varepsilon \end{pmatrix}\right| = \varepsilon(2+\varepsilon),$$

$$\left|\begin{pmatrix} 1+\varepsilon & 1 & 1 \\ 1 & 1+\varepsilon & 1 \\ 1 & 1 & \varepsilon \end{pmatrix}\right| = \varepsilon(2\varepsilon + \varepsilon^2 - 2) < 0 \quad \text{se } \varepsilon < \sqrt{3} - 1$$

The matrix

$$\boldsymbol{A} = \begin{pmatrix} 3 & 2 & 1 & 1 \\ 2 & 3 & 0 & 1 \\ 1 & 0 & 3 & 2 \\ 1 & 1 & 1 & 3 \end{pmatrix}$$

is SPD, in fact

$$|(3)| = 3 > 0, \qquad \left|\begin{pmatrix} 3 & 2 \\ 2 & 3 \end{pmatrix}\right| = 5 > 0$$

$$\left|\begin{pmatrix} 3 & 2 & 1 \\ 2 & 3 & 0 \\ 1 & 0 & 3 \end{pmatrix}\right| = 12 > 0 \qquad \left|\begin{pmatrix} 3 & 2 & 1 & 1 \\ 2 & 3 & 0 & 1 \\ 1 & 0 & 3 & 2 \\ 1 & 1 & 1 & 3 \end{pmatrix}\right| = 24 > 0$$

Let $\boldsymbol{K} \in \mathbb{R}^{n \times p}$ a matrix such that

1. $\boldsymbol{B}\boldsymbol{K} = \boldsymbol{0}$
2. If $\boldsymbol{x}$ is such that $\boldsymbol{B}\boldsymbol{x} = \boldsymbol{0}$ then $\boldsymbol{x} = \boldsymbol{K}\boldsymbol{\alpha}$ for an appropriate $\boldsymbol{\alpha} \in \mathbb{R}^p$

then

$$\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} > 0, \qquad \forall \boldsymbol{x} \neq \boldsymbol{0}, \quad \text{tale che} \quad \boldsymbol{B}\boldsymbol{x} = \boldsymbol{0}$$

is equivalent to assert that matrix

$$\boldsymbol{K}^T\boldsymbol{A}\boldsymbol{K}$$

is positive definite. Analogously to check for semi-SPD.

Example (1/4)

Consider the matrices

$$\boldsymbol{A} = \begin{pmatrix} 3 & 0 & 3 & 1 \\ 0 & 3 & 0 & 0 \\ 3 & 0 & 3 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}, \qquad \boldsymbol{B} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & -1 & 1 \end{pmatrix}$$

Search the vectors $\boldsymbol{v}$ such that $\boldsymbol{B}\boldsymbol{v} = \boldsymbol{0}$:

$$\begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

and the following linear relations are obtained

$$v_1 + v_2 = 0,$$

$$v_2 - v_3 + v_4 = 0$$

Example (2/4)

Searching non trivial solution of the homogeneous linear system

$$v_1 + v_2 = 0,$$

$$v_2 - v_3 + v_4 = 0$$

and observing that $v_2 = -v_1$ we pose $v_1 = \alpha$ and thus $v_2 = -\alpha$. Substituting in the second equation

$$-\alpha - v_3 + v_4 = 0$$

set $v_3 = \beta$ obtaining $v_4 = \alpha + \beta$. Namely the vectors in the Kernel of $\boldsymbol{B}$ are of the form

$$\begin{pmatrix} \alpha \\ -\alpha \\ \beta \\ \alpha + \beta \end{pmatrix}$$

Example (3/4)

Writing previously relation a matrix-vector product

$$\begin{pmatrix} \alpha \\ -\alpha \\ \beta \\ \alpha+\beta \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$

and, thus

$$\boldsymbol{K} = \begin{pmatrix} 1 & 0 \\ -1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix}$$

Example (4/4)

Project the matrix $\boldsymbol{A}$ into the Kernel of $\boldsymbol{K}$

$$\boldsymbol{K}^T\boldsymbol{A}\boldsymbol{K} = \begin{pmatrix} 1 & -1 & 0 & 1 \\ 0 & 0 & 1 & \end{pmatrix} \begin{pmatrix} 3 & 0 & 3 & 1 \\ 0 & 3 & 0 & 0 \\ 3 & 0 & 3 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 9 & 5 \\ 5 & 4 \end{pmatrix}$$

Applying the Sylvester's criterium

$$|(9)| = 9 > 0, \qquad \left| \begin{pmatrix} 9 & 5 \\ 5 & 4 \end{pmatrix} \right| = 11 > 0,$$

namely, the matrix $\boldsymbol{A}$ is positive definite in the kernel of $\boldsymbol{B}$. Observe that the Sylvester's criterium for $\boldsymbol{A}$ is not SPD! in general.

Generale case

- How to find the matrice $\boldsymbol{K} \in \mathbb{R}^{n\times p}$ for a generic matrix $\boldsymbol{B} \in \mathbb{R}^{m\times n}$?
- A simple way to build $\boldsymbol{K}$ is by using Gauss elimination.
- Fro example after row and column elimination matrix $\boldsymbol{B}$ is in the form

$$\begin{pmatrix} \boldsymbol{I} & \boldsymbol{Q} \end{pmatrix}$$

where $\boldsymbol{I} \in \mathbb{R}^{m\times m}$ and $\boldsymbol{Q} \in \mathbb{R}^{m\times(n-m)}$. Thus, the first m components of the generic vector are given from the last components taken as free parameters.

Example

Consider the matrix

$$\boldsymbol{B} = \begin{pmatrix} 1 & 0 & 0 & 0 & 3 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & -1 & 0 & 1 & 1 & -1 \\ 1 & 0 & 0 & 0 & 3 & 1 & 0 \end{pmatrix}$$

add a row of labels and start with Gauss elimination:

$$\begin{pmatrix} v_1 & v_2 & v_3 & v_4 & v_5 & v_6 & v_7 \\ 1 & 0 & 0 & 0 & 3 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & -1 & 0 & 1 & 1 & -1 \\ 1 & 0 & 0 & 0 & 3 & 1 & 0 \end{pmatrix}$$

Example (2/5)

Delete 1 from the last row ($[4] \leftarrow [4] - [1]$)

$$\begin{pmatrix} v_1 & v_2 & v_3 & v_4 & v_5 & v_6 & v_7 \\ 1 & 0 & 0 & 0 & 3 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & -1 & 0 & 1 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

exchange second and third row ($[2] \leftrightarrow [3]$)

$$\begin{pmatrix} v_1 & v_2 & v_3 & v_4 & v_5 & v_6 & v_7 \\ 1 & 0 & 0 & 0 & 3 & 1 & 0 \\ 0 & 1 & -1 & 0 & 1 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Example (3/5)

Exchange column 3 with column 6

$$\begin{pmatrix} v_1 & v_2 & v_6 & v_4 & v_5 & v_3 & v_7 \\ 1 & 0 & 1 & 0 & 3 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & -1 & -1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Delete 1 in third column from first and second row ($[1] \leftarrow [1] - [3]$ ed $[2] \leftarrow [2] - [3]$)

$$\begin{pmatrix} v_1 & v_2 & v_6 & v_4 & v_5 & v_3 & v_7 \\ 1 & 0 & 0 & 0 & 3 & 0 & -1 \\ 0 & 1 & 0 & 0 & 1 & -1 & -2 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Example (4/5)

From last matrix get the relations

$$v_1 = 3v_5 - v_7$$
$$v_2 = v_5 - v_3 - 2v_7$$
$$v_6 = v_7$$

the free parameters are v_3, v_4, v_5, v_7. Set $v_3 = \alpha$, $v_4 = \beta$, $v_5 = \gamma$, $v_7 = \delta$ so that general solution is

$$v_1 = 3\gamma - \delta, \quad v_2 = \gamma - \alpha - 2\delta, \quad v_3 = \alpha,$$
$$v_4 = \beta, \quad v_5 = \gamma, \quad v_6 = \delta, \quad v_7 = \delta,$$

Example (5/5)

The solution

$$v_1 = 3\gamma - \delta, \quad v_2 = \gamma - \alpha - 2\delta, \quad v_3 = \alpha,$$

$$v_4 = \beta, \quad v_5 = \gamma, \quad v_6 = \delta, \quad v_7 = \delta,$$

can be written ad matrix-vector product

$$\begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ v_7 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 3 & -1 \\ -1 & 0 & 1 & -2 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \\ \gamma \\ \delta \end{pmatrix}$$

and thus, matrix $\boldsymbol{K}$ is easily determined.

Summary of main theorems

Here a summary of fundamental theorems for the characterization of constrained minima are collected.

Definition (Ammissibile point)

A point x^* is admissible if

$$h_k(\boldsymbol{x}^*) = 0 \qquad k = 1, 2, \ldots, m$$

$$g_k(\boldsymbol{x}^*) \geq 0 \qquad k = 1, 2, \ldots, p$$

Definition (active constraints)

The following set

$$\mathcal{A}(\boldsymbol{x}^*) = \{k \mid g_k(\boldsymbol{x}^*) = 0\}$$

is named active constraints set. This set can be split in two subsets

$$\mathcal{A}^+(\boldsymbol{x}^*, \boldsymbol{\mu}^*) = \{k \mid g_k(\boldsymbol{x}^*) = 0, \quad \mu_k^* > 0\}$$

$$\mathcal{A}^0(\boldsymbol{x}^*, \boldsymbol{\mu}^*) = \{k \mid g_k(\boldsymbol{x}^*) = 0, \quad \mu_k^* = 0\}$$

$\mathcal{A}^+(\boldsymbol{x}^*, \boldsymbol{\mu}^*)$ *are the strongly active constraints e* $\mathcal{A}^0(\boldsymbol{x}^*, \boldsymbol{\mu}^*)$ *are the weakly active constraints.*

Obviously

$$\mathcal{A}^0(\boldsymbol{x}^*, \boldsymbol{\mu}^*) \bigcap \mathcal{A}^+(\boldsymbol{x}^*, \boldsymbol{\mu}^*) = \emptyset \quad \text{and} \quad \mathcal{A}^0(\boldsymbol{x}^*, \boldsymbol{\mu}^*) \bigcup \mathcal{A}^+(\boldsymbol{x}^*, \boldsymbol{\mu}^*) = \mathcal{A}(\boldsymbol{x}^*)$$

In the study of optimality condition the constraints and its gradients cannot be arbitrary. They must satisfy additional analytic/geometric properties. This properties are named constraints qualification. The easiest qualification (but also compelling) is linear independence (LI)

Definition (Constraints qualification LI)

Given the inequality constraints $\boldsymbol{g}(\boldsymbol{x})$ and equality constraints $\boldsymbol{h}(\boldsymbol{x})$, we will say than an admissible point $\boldsymbol{x}^*$ is qualified if the vectors

$$\{\nabla g_k(\boldsymbol{x}^*) \;:\; k \in \mathcal{A}(\boldsymbol{x}^*)\} \cup \{\nabla h_1(\boldsymbol{x}^*), \nabla h_2(\boldsymbol{x}^*), \ldots, \nabla h_m(\boldsymbol{x}^*)\}$$

are linearly independent.

Mangasarian-Fromovitz qualification

This qualification is less stringent of the previous

Definition (Constraints qualification MF)

Given the inequality constraints $\boldsymbol{g}(\boldsymbol{x})$ and equality constraints $\boldsymbol{h}(\boldsymbol{x})$, we will say than an admissible point $\boldsymbol{x}^*$ is qualified if does not exists a linear combination

$$\sum_{k\in\mathcal{A}(\boldsymbol{x}^*)}^{m} \alpha_k \nabla g_k(\boldsymbol{x}^*) + \sum_{k=1}^{m} \beta_k \nabla h_k(\boldsymbol{x}^*) = \boldsymbol{0}$$

with $\alpha_k \geq 0$ for $k \in \mathcal{A}(\boldsymbol{x}^*)$ and α_k and β_k not all zero. That is, there is no non trivial linear combination for the null vector with $\alpha_k \geq 0$ for $k \in \mathcal{A}(\boldsymbol{x}^*)$.

Garth P. McCormick qualification

Definition (Constraints qualification (1 ordine))

Given an admissible point $\boldsymbol{x}^*$ *the constraints are* *first order qualified* *if for all direction* $\boldsymbol{d}$ *that satisfy*

$$\nabla h_k(\boldsymbol{x}^*)\boldsymbol{d} = 0, \qquad k \in \{1,2,\ldots,m\},$$

$$\nabla g_k(\boldsymbol{x}^*)\boldsymbol{d} \geq 0, \qquad k \in \mathcal{A}(\boldsymbol{x}^*),$$

exists a curve $\boldsymbol{x} \in \mathtt{C}^1(\mathbb{R},\mathbb{R}^n)$ *and an* $\varepsilon > 0$ *such that for* $0 < t < \varepsilon$.

$$\boldsymbol{x}(0) = \boldsymbol{x}^*, \qquad h_k(\boldsymbol{x}(t)) = 0, \qquad k \in \{1,2,\ldots,m\},$$

$$\boldsymbol{x}'(0) = \boldsymbol{d}, \qquad g_k(\boldsymbol{x}(t)) \geq 0, \qquad k \in \{1,2,\ldots,p\}.$$

Garth P. McCormick qualification

Definition (Constraints qualification (2 ordine))

Given an admissible point $\boldsymbol{x}^*$ the constraints are first order qualified if for all direction $\boldsymbol{d}$ that satisfy

$$\nabla h_k(\boldsymbol{x}^*)\boldsymbol{d} = 0, \qquad k \in \{1, 2, \ldots, m\},$$

$$\nabla g_k(\boldsymbol{x}^*)\boldsymbol{d} = 0, \qquad k \in \mathcal{A}(\boldsymbol{x}^*),$$

exists a curve $\boldsymbol{x} \in \mathtt{C}^2(\mathbb{R}, \mathbb{R}^n)$ and an $\varepsilon > 0$ such that for $0 < t < \varepsilon$.

$$\boldsymbol{x}(0) = \boldsymbol{x}^*, \qquad h_k(\boldsymbol{x}(t)) = 0, \qquad k \in \{1, 2, \ldots, m\},$$

$$\boldsymbol{x}'(0) = \boldsymbol{d}, \qquad g_k(\boldsymbol{x}(t)) = 0, \qquad k \in \mathcal{A}(\boldsymbol{x}^*).$$

Theorem (First order KKT condition)

Let $f \in \mathtt{C}^1(\mathbb{R}^n)$ and $\boldsymbol{g} \in \mathtt{C}^1(\mathbb{R}^n, \mathbb{R}^p)$ with $\boldsymbol{h} \in \mathtt{C}^1(\mathbb{R}^n, \mathbb{R}^m)$ inequality and equality constraints. If $\boldsymbol{x}^*$ satisfy constraints qualification then necessary condition for local minimum is that there exists $m + p$ scalars such that

$$
\begin{aligned}
\nabla_x \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) &= \mathbf{0}^T \\
\mu_k^* g_k(\boldsymbol{x}^*) &= 0, \qquad k = 1, 2, \ldots, p; \\
\mu_k^* &\geq 0, \qquad k = 1, 2, \ldots, p;
\end{aligned}
$$

where

$$
\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\boldsymbol{x}) - \sum_{k=1}^{p} \mu_k \, g_k(\boldsymbol{x}) - \sum_{k=1}^{m} \lambda_k \, h_k(\boldsymbol{x})
$$

Theorem (Second order necessary KKT conditions)

Let $f \in \mathtt{C}^2(\mathbb{R}^n)$ *and the constraints* $\boldsymbol{g} \in \mathtt{C}^2(\mathbb{R}^n, \mathbb{R}^p)$ *and* $\boldsymbol{h} \in \mathtt{C}^2(\mathbb{R}^n, \mathbb{R}^m)$. *If* $\boldsymbol{x}^*$ *satisfy constraints qualification, then necessary condition for* $\boldsymbol{x}^*$ *be a local minimum id that there exists* $m+p$ *scalars that satisfy first order conditions and*

$$\boldsymbol{d}^T \nabla_x^2 \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)\boldsymbol{d} \geq 0$$

for all $\boldsymbol{d}$ *such that*

$$\nabla h_k(\boldsymbol{x}^*)\boldsymbol{d} = 0, \qquad k = 1, 2, \ldots, m$$

$$\nabla g_k(\boldsymbol{x}^*)\boldsymbol{d} = 0, \qquad se\ k \in \mathcal{A}(\boldsymbol{x}^*)$$

A more tighten condition:

$$\nabla g_k(\boldsymbol{x}^*)\boldsymbol{d} = 0, \qquad \text{se } k \in \mathcal{A}^+(\boldsymbol{x}^*)$$

$$\nabla g_k(\boldsymbol{x}^*)\boldsymbol{d} \geq 0, \qquad \text{se } k \in \mathcal{A}^0(\boldsymbol{x}^*)$$

Riassunto teoremi fondamentali

Theorem (Second order sufficient conditions by G.P.McCormick)

Let $f \in \mathtt{C}^2(\mathbb{R}^n)$ *and the constraints* $\boldsymbol{g} \in \mathtt{C}^2(\mathbb{R}^n, \mathbb{R}^p)$ *and* $\boldsymbol{h} \in \mathtt{C}^2(\mathbb{R}^n, \mathbb{R}^m)$. *A sufficient condition for* $\boldsymbol{x}^*$ *be a local minimum id that there exists* $m + p$ *scalars that satisfy first order conditions and*

$$
\begin{aligned}
h_j(\boldsymbol{x}^*) &= 0, \qquad & j &= 1, 2, \ldots, m \\
g_k(\boldsymbol{x}^*) &\geq 0, \qquad & k &= 1, 2, \ldots, p \\
\mu_k g_k(\boldsymbol{x}^*) &= 0, \qquad & k &= 1, 2, \ldots, p \\
\mu_k &\geq 0, \qquad & k &= 1, 2, \ldots, p \\
\nabla_x \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) &= 0
\end{aligned}
$$

(continue...)

Summary of fundamental theorem

Theorem (Second order sufficient conditions by G.P.McCormick)

(...continue)
moreover for all $\boldsymbol{d} \neq \mathbf{0}$ *such that*

$$\nabla h_k(\boldsymbol{x}^*)\boldsymbol{d} = 0, \qquad k = 1, 2, \ldots, m$$

$$\nabla g_k(\boldsymbol{x}^*)\boldsymbol{d} = 0, \qquad se\ \mu_k > 0$$

and

$$\boldsymbol{d}^T \nabla_x^2 \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)\boldsymbol{d} > 0$$

notice that constraint qualification is not necessary for sufficient condition

References

William Karush, *Minima of functions of several variables with inequalities as side constraints*.
Master's thesis, University of Chicago, Illiois, 1939.

Fritz John, *Extremum problems with inequalities as subsidiary conditions*.
In Sudies and Essays, Courant Anniversary Volume, pp. 187-–204. Interscience, New York, 1948.

Harold William Kuhn, Albert William Tucker, *Nonlinear programming*.
Second Berkeley Symposium on Mathematical Statistics and Probability, 1950, pp. 481–492, Berkeley and Los Angeles, 1951. University of California Press.

Garth P. McCormick, *Second Order Conditions for Constrained Minima*.
SIAM J. Appl. Math, Vol. **15**, n.2. 1967.

Olvi L. Mangasarian, S. Fromovitz, *The Fritz John necessary optimality conditions in the presence of equality and inequality constraints*.
Journal of Mathematical Analysis and Applications Vol. **17**, n.1, 1967.