Optimization Methods
Lecture 10

Solmaz S. Kia
Mechanical and Aerospace Engineering Dept.
University of California Irvine
solmaz@uci.edu

Consult: pages 276-297 (section 3.1 and 3.2) from Ref[1]. The slides are only part of Lecture 10; consult your class notes for complete discussion.

Constrained optimization (review)

We consider the following standard form:

$$\begin{aligned} x^\star = & \operatorname*{argmin}_{x\in\mathbb{R}^n} f(x) \quad \text{s.t.} \\ & h_i(x) = 0, \quad i \in \{1, \cdots, m\} \\ & g_i(x) \leqslant 0, \quad i \in \{1, \cdots, r\} \end{aligned} \qquad \text{or} \qquad \begin{aligned} x^\star = & \operatorname*{argmin}_{x\in\mathbb{R}^n} f(x) \quad \text{s.t.} \\ & h(x) = 0, \\ & g(x) \leqslant 0, \end{aligned}$$

$h^i : \mathbb{R}^n \to \mathbb{R}$, $g^i : \mathbb{R}^n \to \mathbb{R}$ $\qquad$ $h : \mathbb{R}^n \to \mathbb{R}^m$, $g : \mathbb{R}^n \to \mathbb{R}^r$

- f,h,g: continuously differentiable function of x
  e.g., $f, h, g \in C^1$ continuously differentiable
  e.g., $f, h, g \in C^2$ both f and its first derivative are continuously differentiable
- the equality constraints are underdetermined. It is usually assume that $m \leqslant n$
- no restriction on r

Feasible set: set up points that satisfy the constraints

$$\Omega = \{x \in \mathbb{R}^n | h(x) = 0,\ \ g(x) \leqslant 0\}.$$

The constrained optimization can also be written as

$$x^\star = \operatorname*{argmin}_{x\in\Omega} f(x)$$

First order necessary condition for optimality (review)

$x^\star$ is a local minimizer:

$$f(x) \geqslant f(x^\star), \ \forall x \in \Omega \text{ s.t. } \|x - x^\star\| \leqslant \epsilon$$

Fitst order necessary condition analysis: consider $x \in \Omega$ that are in small neighborhood of a local minimum $x^\star$: $x = x^\star + \Delta x$

$$f(x + \Delta x) \approx f(x^\star) + \nabla f(x^\star)^\top \Delta x + \text{H.O.T} \overset{f(x) \geqslant f(x^\star)}{\Longrightarrow} \nabla f(x^\star)^\top \Delta x \geqslant 0$$

$x = x^\star + \Delta x \in \Omega$:

$$h(x{+}\Delta x) = 0 \Rightarrow h(x{+}\Delta x) \approx h(x^\star) + \nabla h(x^\star)^\top \Delta x = 0 \overset{h(x^\star)=0}{\Longrightarrow} \nabla h(x^\star)^\top \Delta x = 0$$

$$g(x{+}\Delta x) \leqslant 0 \Rightarrow g(x{+}\Delta x) \approx g(x^\star) + \nabla g(x^\star)^\top \Delta x = 0 \overset{g_i(x^\star)\leqslant 0}{\Longrightarrow} \begin{cases} \nabla g_i(x^\star)^\top \Delta \leqslant 0 & g_i(x^\star) = 0 \\ \text{none} & g_i(x^\star) < 0 \end{cases}$$

- Active inequality set at x: $\mathcal{A}(x) = \left\{ i \in \{1, \cdots, r\} \;\middle|\; g_i(x) = 0 \right\}$
- Set of first order feasible variations at x:

$$V(x) = \left\{ d \in \mathbb{R}^n \;\middle|\; \nabla h_i(x)^\top d = 0, \ \nabla g_j(x)^\top d \leqslant 0, \quad j \in \mathcal{A}(x^\star) \right\}$$

$$\boxed{\text{FONC for optimality}: \ \nabla f(x^\star)^\top \Delta x \geqslant 0, \quad \text{for} \ \ \Delta x \in V(x^\star)}$$

Constrained optimization: equality constraints (review)

$$x^\star = \underset{x\in\mathbb{R}^n}{\text{argmin}}\ f(x) \quad s.t. \qquad h_i(x) = 0, \quad i \in \{1, \cdots, m\}$$

or

$$x^\star = \underset{x\in\mathbb{R}^n}{\text{argmin}}\ f(x) \quad s.t. \qquad h(x) = 0,$$

f,h,g: continuously differentiable function of x
e.g., $f, h \in C^1$ continuously differentiable
e.g., $f, h \in C^2$ both f and its first derivative are continuously differentiable

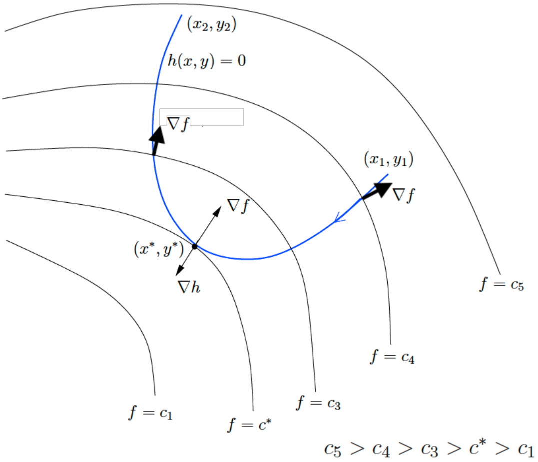
First Order Necessary Condition for Optimality: $x^\star$ is a local minimizer then

$$\nabla f(x^\star)^\top \Delta x \geqslant 0, \qquad \text{for} \ \ \Delta x \in V(x^\star)$$

- Set of first order feasible variations at x

$$V(x) = \{d \in \mathbb{R}^n \mid \nabla h_i(x)^\top d = 0\}$$

Geometric Interpretation of Lagrange Multipliers



$$\nabla f(x^\star) = -\lambda \nabla h(x^\star)$$

The methods I set forth require neither constructions nor geometric or mechanical considerations. They require only algebraic operations subject to a systematic and uniform course. -Lagrange

Lagrange Multipliers

For a given local minimizer $x^\star$ there exists scalars $\underbrace{\lambda_1, \cdots, \lambda_m}_{\text{Lagrange Multipliers}}$ such that

$$\nabla f(x^\star) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x^\star) = 0. \qquad \text{(LM-1)}$$

- $\nabla f(x^\star)$ belongs to the sub space spanned by the constraint gradients at $x^\star$:

$$\nabla f(x^\star) = -\lambda_1 \nabla h_1(x^\star) - \cdots - \lambda_m \nabla h_m(x^\star)$$

- $\nabla f(x^\star)$ is orthogonal to the subspace of first order feasible variants $V(x^\star) = \{d \in \mathbb{R}^n \mid \nabla h_i(x^\star)^\top d = 0\}$

$$\nabla f(x^\star)^\top \Delta x = (-\lambda_1 \nabla h_1(x^\star) - \cdots - \lambda_m \nabla h_m(x^\star))^\top \Delta x \Rightarrow$$
$$\nabla f(x^\star)^\top \Delta x = 0, \quad \text{for } \Delta x \in V(x^\star)$$

Thus, according to the Largrange multiplier condition (LM-1), at the local minimum $x^\star$, the first order cost variation $\nabla f(x^\star)\Delta x$ is zero for all variations Δx in $V(x^\star)$. This statement is analogous to the "zero gradient condition $\nabla f(x^\star)$ of the unconstrained optimization.

Proposition (Lagrange Multiplier Theorem-Necessary conditions)

Let $x^\star$ be a local minimum of f subject to $h(x) = 0$ and assume that the constraint gradients $\{\nabla h_1(x^\star), \quad , \nabla h_m(x)\}$ are linearly independent. Then there exists a unique vectors $\lambda^\star = (\lambda_1^\star, \cdots, \lambda_m^\star)$ called Lagrange multiplier vector, s.t.

$$\nabla f(x^\star) + \sum_{i=1}^{m} \lambda_i^\star \nabla h_i(x^\star) = 0.$$

If in addition f and h are twice continuously differentiable we have

$$y^\top \Big(\nabla^2 f(x^\star) + \sum_{i=1}^{m} \lambda_i^\star \nabla^2 h_i(x^\star)\Big) y \geqslant 0, \quad \forall y \in V(x^\star)$$

where $V(x^\star)$ is the space of first order feasible variations, i.e.,

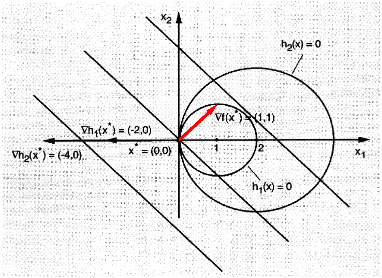
$$V(x^\star) = \{d \in \mathbb{R}^n \mid \nabla h_i(x^\star)^\top d = 0\}.$$

A Problem with no Lagrange Multipliers: regularity of optimal point

- **Regular point of a set of constraints**: A feasible vector x for which the constraint gradients $\{\nabla h_1(x), \cdots, \nabla h_m(x)\}$ are linearly independent.
- For a local minimum that is not regular, there may not exist Lagrange multipliers.

$$\text{minimize } f(x) = x_1 + x_2, \quad \text{s.t.}$$
$$h_1(x) = (x_1 - 1)^2 + x_2^2 - 1 = 0, \quad h_2(x) = (x_1 - 2)^2 + x_2^2 - 4 = 0.$$

- $x^\star$ is not regular. Therefore, this problem cannot be solved using Lagrange multiplier theorem.
- $\nabla f(x^\star)$ cannot be written as linear combination of $\nabla h_1(x^\star)$ and $\nabla h_2(x^\star)$



Necessary Conditions for Optimality

Lagrangian function $L : \mathbb{R}^{n+m} \mapsto \mathbb{R}$: $L(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x)$

Proposition (Lagrange Multiplier Theorem-Necessary conditions)

Let $x^\star$ be a local minimum of f subject to $h(x) = 0$ and assume that the constraint gradients $\{\nabla h_1(x^\star), \quad , \nabla h_m(x)\}$ are linearly independent. Then there exists a unique vectors $\lambda^\star = (\lambda_1^\star, \cdots, \lambda_m^\star)$ called Lagrange multiplier vector, s.t.

$$\nabla_x L(x^\star, \lambda^\star) = 0.$$

If in addition f and h are twice continuously differentiable we have

$$y^\top \nabla_{xx} L(x^\star, \lambda^\star)\, y \geqslant 0, \quad \forall y \in V(x^\star)$$

where $V(x^\star)$ is the space of first order feasible variations, i.e.,

$$V(x^\star) = \{d \in \mathbb{R}^n \mid \nabla h_i(x^\star)^\top d = 0\}.$$

$$h(x^\star) = 0 \Leftrightarrow \nabla_\lambda L(x^\star, \lambda^\star) = 0.$$

Proposition (Second Order Sufficiency Conditions for Optimality)

Assume that f and h are twice continuously differentiable, and let $x^\star \in \mathbb{R}^n$ and $\lambda^\star \in \mathbb{R}^m$ satisfy

$$\nabla_x L(x^\star, \lambda^\star) = 0, \quad \nabla_\lambda L(x^\star, \lambda^\star) = 0,$$
$$y^\top \nabla_{xx} L(x^\star, \lambda^\star)\, y > 0, \quad \forall\, y \neq 0 \text{ with } \nabla h(x^\star)^\top y = 0.$$

Then $x^\star$ is a strict local minimum of f subject to $h(x) = 0$. In fact, there exists scalars $\gamma > 0$ and $\epsilon > 0$ such that

$$f(x) \geqslant f(x^\star) + \frac{\gamma}{2}\|x - x^\star\|, \quad \forall x \text{ with } h(x) = 0 \text{ and } \|x - x^\star\| < \epsilon.$$

Second order necessary and sufficient conditions

$$\text{SONC: } y^\top \nabla_x^2 L(x^\star, \lambda^\star) y \geqslant 0, \quad \forall y \in V(x^\star)$$
$$\text{SOSC: } y^\top \nabla_x^2 L(x^\star, \lambda^\star) y > 0, \quad \forall y \in V(x^\star)$$

where $V(x^\star) = \left\{ y \in \mathbb{R}^n \middle| \nabla_x h(x)^\top y = 0 \right\}$ Because $(\nabla_x h_1(x), \cdots, \nabla_x h_m(x))$ are linearly independent, then

$$\text{rank}(\nabla_x h(x)^\top) = \text{rank}\left(\begin{bmatrix} \nabla_x h_1(x) & \cdots & \nabla_x h_m(x) \end{bmatrix} \right) = m < n.$$

Therefore,
$V(x^\star) = \left\{ y \in \mathbb{R}^n \middle| \nabla_x h(x)^\top y = 0 \right\} = \left\{ y \in \mathbb{R}^n \middle| N z = 0, \ \forall z \in \mathbb{R}^{n-m} \right\}$ where columns of N span the null-space of $\nabla_x h(x)^\top$.

$$\text{SONC: } z^\top \underbrace{N^\top \nabla_x^2 L(x^\star, \lambda^\star) N}_{M} z \geqslant 0, \quad \forall z \in \mathbb{R}^{n-m} \qquad \Leftrightarrow \qquad \text{SONC: } M \geqslant 0$$
$$\text{SOSC: } z^\top \underbrace{N^\top \nabla_x^2 L(x^\star, \lambda^\star) N}_{M} z > 0, \quad \forall z \in \mathbb{R}^{n-m} \qquad \Leftrightarrow \qquad \text{SOSC: } M > 0$$

Numerical example: maximum volume cardboard box with given area.

Problem: Construct a cardboard box of maximum volume, given a fixed area $c \in \mathbb{R}_{>0}$ of cardboard.
Solution: Denote the dimensions of the box by x, y and z. Then the problem becomes

$$\begin{aligned} \text{minimize} \quad & -xyz \qquad \text{(equivalent to maximize } xyz\text{)} \\ \text{subject to} \quad & (xy + yz + xz) = \frac{c}{2} \end{aligned}$$

$$- - - - - - - - - - -$$

$$L(x, y, z, \lambda) = -xyz + \lambda(xy + yz + xz - \frac{c}{2})$$

$$\text{FONC: } \begin{cases} \nabla_x L(x, y, z, \lambda) = 0 \Rightarrow \; -yz + \lambda(y + z) = 0 \\ \nabla_y L(x, y, z, \lambda) = 0 \Rightarrow \; -xz + \lambda(x + z) = 0 \\ \nabla_z L(x, y, z, \lambda) = 0 \Rightarrow \; -xy + \lambda(x + y) = 0 \\ \nabla_\lambda L(x, y, z, \lambda) = 0 \Rightarrow xy + yz + xz - \frac{c}{2} = 0 \end{cases}$$

$x^\star = y^\star = z^\star = \sqrt{\frac{c}{6}}$ and $\lambda^\star = \frac{1}{2}\sqrt{\frac{c}{6}}$ is the unique solution.

Numerical example: maximum volume cardboard box with given area.

$$\nabla^2_{x,y,z} L(x^\star, y^\star, z^\star, \lambda^\star) = \begin{bmatrix} 0 & -z^\star+\lambda^\star & -y^\star+\lambda^\star \\ -z^\star+\lambda^\star & 0 & -x^\star+\lambda^\star \\ -y^\star+\lambda^\star & -x^\star+\lambda^\star & 0 \end{bmatrix} = (-\tfrac{1}{2}\sqrt{\tfrac{c}{6}}) \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

Note that $\nabla^2_{x,y,z} L(x^\star, y^\star, z^\star, \lambda^\star)$ is indefinite because $\text{eig}\left(\begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}\right) = \{-1, -1, 2\}$.

Recall that you do not necessarily need $\nabla^2_{x,y,z} L(x^\star, y^\star, z^\star, \lambda^\star)$ to be positive semi-definite for $(x^\star, y^\star, z^\star, \lambda^\star)$ to be a minimizer.

SONC: $p^\top \nabla^2_{x,y,z} L(x^\star, y^\star, z^\star, \lambda^\star) p \geqslant 0 \;\forall p \in V(x^\star, y^\star, z^\star)$

$$V(x^\star, y^\star, z^\star) = \left\{ p \in \mathbb{R}^3 \middle| \begin{bmatrix} y^\star+z^\star & x^\star+z^\star & z^\star+y^\star \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix} = 0 \right\} = \left\{ p \in \mathbb{R}^3 \middle| p_1+p_2+p_3 = 0 \right\}$$

$$p^\top \nabla^2_{x,y,z} L(x^\star, y^\star, z^\star, \lambda^\star) p = (-\tfrac{1}{2}\sqrt{\tfrac{c}{6}}) \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix}^\top \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix} =$$

$$(-\tfrac{1}{2}\sqrt{\tfrac{c}{6}})(p_2+p_3)p_1 + (p_1+p_3)p_2 + (p_1+p_2)p_3 = -(-\tfrac{1}{2}\sqrt{\tfrac{c}{6}})(p_1^2+p_2^2+p_3^2) > 0$$

Because $p \in V(x^\star, y^\star, z^\star)$, we used $p_2+p_3 = -p_1$, $p_1+p_3 = -p_2$ and $p_1+p_2 = -p_3$.

Since the Second Order Sufficiency Condition is satisfied: $(x^\star = \sqrt{\frac{c}{6}}, y^\star = \sqrt{\frac{c}{6}}, z^\star = \sqrt{\frac{c}{6}})$ is the unique global minimizer (note that the cost function is bounded over the feasible set, therefore we have a global minimizer).

Numerical example: maximum volume cardboard box with given area.

- Alternative way to check for second order optimality condition:

$$\boxed{\text{SONC: } p^\top \nabla^2_{x,y,z} L(x^\star, y^\star, z^\star, \lambda^\star) p \geqslant 0 \; \forall p \in V(x^\star, y^\star, z^\star)}$$

$$V(x^\star, y^\star, z^\star) = \left\{ p \in \mathbb{R}^3 \middle| \begin{bmatrix} y^\star + z^\star & x^\star + z^\star & z^\star + y^\star \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix} = 0 \right\} =$$

$\left\{ p \in \mathbb{R}^3 \middle| p = N\,z,\ \forall z \in \mathbb{R}^2 \right\}$ where N is a matrix whose columns span the null-space of $\nabla h(x^\star, y^\star, z^\star)$: $N = \begin{bmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{bmatrix}$.

$$p^\top \nabla^2_{x,y,z} L(x^\star, y^\star, z^\star, \lambda^\star) p = z \underbrace{\begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \left(-\frac{1}{2}\sqrt{\frac{c}{6}}\right) \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{bmatrix}}_{M = (-\frac{1}{2}\sqrt{\frac{c}{6}}) \begin{bmatrix} -2 & 1 \\ 1 & -2 \end{bmatrix}} z.$$

Since matrix M is positive definite, we conclude that in fact the Second Order Sufficiency Condition is satisfied: $(x^\star = \sqrt{\frac{c}{6}}, y^\star = \sqrt{\frac{c}{6}}, z^\star = \sqrt{\frac{c}{6}})$ is the unique global minimizer (note that the cost function is bounded over the feasible set, therefore we have a global minimizer).