Optimization Methods
Lecture 9

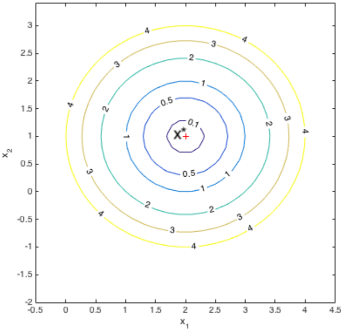
Solmaz S. Kia
Mechanical and Aerospace Engineering Dept.
University of California Irvine
solmaz@uci.edu

Consult: pages 276-297 (section 3.1 and 3.2) from Ref[1]

Constrained optimization vs. unconstrained optimization

Unconstrained optimization

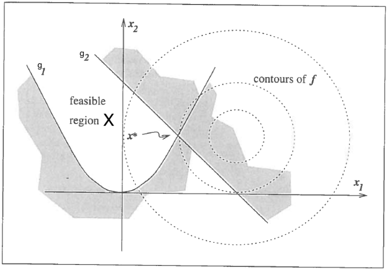
$$x^\star = \underset{x\in\mathbb{R}^2}{\text{argmin}} \underbrace{(x_1-2)^2+(x_2-1)^2}_{f(x)}$$



Constrained optimization

$$x^\star = \underset{x\in\mathbb{R}^2}{\text{argmin}} \underbrace{(x_1-2)^2+(x_2-1)^2}_{f(x)} \quad s.t.$$

$$\begin{cases} -x_1^2+x_2 \geqslant 0 \\ -x_1-x_2+2 \geqslant 0 \end{cases}$$



Constrained optimization

We consider the following standard form:

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

$$h_i(x) = 0, \quad i \in \{1, \cdots, m\} \qquad \text{or} \qquad h(x) = 0,$$

$$g_i(x) \leqslant 0, \quad i \in \{1, \cdots, r\} \qquad\qquad g(x) \leqslant 0,$$

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

- 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

$$x^\star = \underset{x \in \Omega}{\text{argmin}}\; f(x)$$

First order necessary condition for optimality

$x^\star$ 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: $\mathcal{A}(x) = \{i \in \{1, \cdots, r\} \mid g_i(x) = 0\}$
- Set of first order feasible variations at x:

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

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

Constrained optimization: equality constraints

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

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

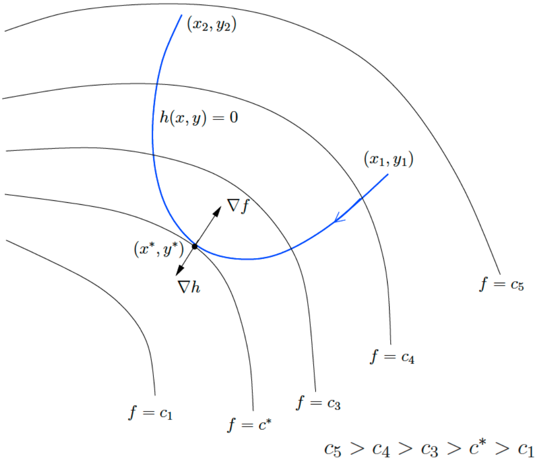
First Order Necessary Condition for Optimality: $x^\star$ is a local minimizer then

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

- 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 $\nabla h_2(x^\star)$