Math 273a: Optimization
Nonlinear optimization with equality constraints

Instructor: Wotao Yin
Department of Mathematics, UCLA

Fall 2015

material taken from the textbook Chong-Zak, 4th Ed.

About this lecture

- recognize a solution to the problem

$$\text{minimize } f(\boldsymbol{x})$$
$$\text{subject to } \boldsymbol{h}(\boldsymbol{x}) = \mathbf{0}.$$

  where $f : \mathbb{R}^n \to \mathbb{R}$, $\boldsymbol{h} : \mathbb{R}^n \to \mathbb{R}^m$

$$\boldsymbol{h}(\boldsymbol{x}) = \begin{bmatrix} h_1(\boldsymbol{x}) \\ \vdots \\ h_m(\boldsymbol{x}) \end{bmatrix}$$

- assume **continuous differentiability:** $f, \boldsymbol{h} \in C^1$.

- **goals of this lecture**

  1. 1st and 2nd order optimality conditions

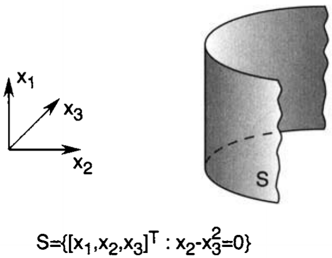
  2. solutions to certain simple/important NLP with equality constraints

- **Jacobian** of $\boldsymbol{h}$:

$$D\boldsymbol{h}(\boldsymbol{x}) = \begin{bmatrix} \nabla h_1(\boldsymbol{x})^T \\ \vdots \\ \nabla h_m(\boldsymbol{x})^T \end{bmatrix}$$

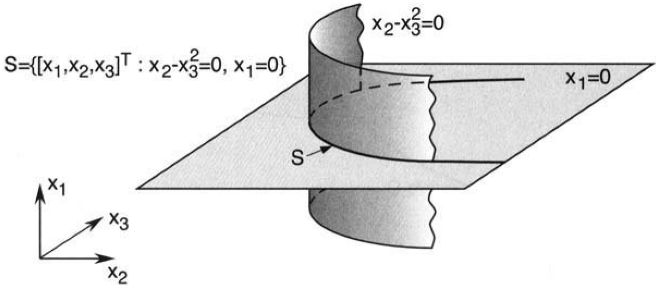
- **definition**: a point $\boldsymbol{x}^* \in \mathbb{R}^n$ satisfying $\boldsymbol{h}(\boldsymbol{x}^*) = \mathbf{0}$ is a *regular point* of the constraints if $\nabla h_1(\boldsymbol{x}^*), \ldots, \nabla h_m(\boldsymbol{x}^*)$ linearly independent, or equivalently, $\mathrm{rank}(D\boldsymbol{h}(\boldsymbol{x}^*)) = 0$.

Example 20.2



$$h_1(\boldsymbol{x}) = x_2 - x_3^2 = 0$$

Example 20.3



$$h_1(\boldsymbol{x}) = x_1 = 0$$

$$h_2(\boldsymbol{x}) = x_2 - x_3^2 = 0$$

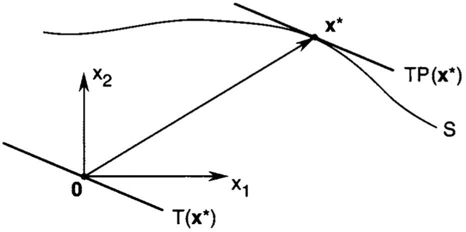
Tangent space

- **tangent space** at $\boldsymbol{x}^*$ on $S = \{x \in \mathbb{R}^n : \boldsymbol{h}(\boldsymbol{x}) = \mathbf{0}\}$ is the set

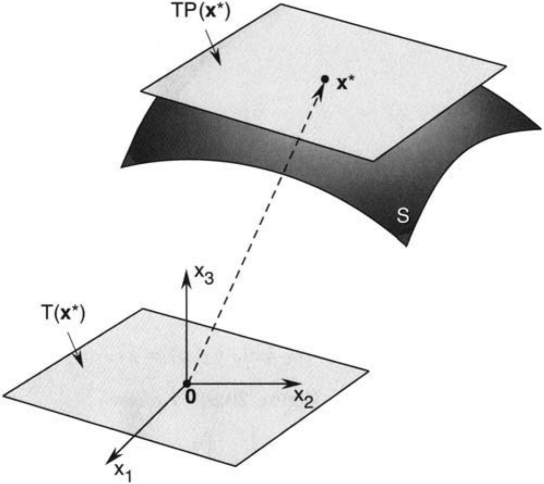
$$T(\boldsymbol{x}^*) := \{\boldsymbol{y} \in \mathbb{R}^n : D\boldsymbol{h}(\boldsymbol{x}^*)\boldsymbol{y} = \mathbf{0}\} = \mathcal{N}(D\boldsymbol{h}(\boldsymbol{x}^*)).$$

- if $\boldsymbol{x}^*$ is regular, then $\dim(T(\boldsymbol{x}^*)) = n - m$.
- **tangent plane** at x^* is the set

$$TP(\boldsymbol{x}^*) := T(\boldsymbol{x}^*) + \boldsymbol{x}^*.$$



Tangent space



Normal space

- **normal space** $N(\boldsymbol{x}^*)$ at $\boldsymbol{x}^*$ on $S = \{x \in \mathbb{R}^n : \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}\}$ is the set

$$N(\boldsymbol{x}^*) := \mathcal{R}(D\boldsymbol{h}(\boldsymbol{x}^*)^T) = \operatorname{span}\{\nabla h_1(\boldsymbol{x}^*), \ldots, \nabla h_m(\boldsymbol{x}^*)\}$$

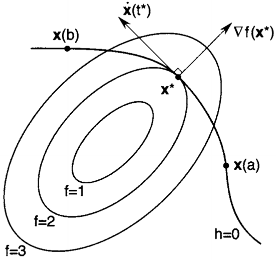
- **normal plane** is $NP(\boldsymbol{x}^*) := N(\boldsymbol{x}^*) + \boldsymbol{x}^*$
- **lemma:** $T(\boldsymbol{x}^*) = N(\boldsymbol{x}^*)^\perp$, $N(\boldsymbol{x}^*) = T(\boldsymbol{x}^*)^\perp$, $\mathbb{R}^n = N(\boldsymbol{x}^*) \oplus T(\boldsymbol{x}^*)$.

Lagrange's theorem

Theorem (Lagrange's theorem, $n = 2, m = 1$)

Let $\boldsymbol{x}^*$ be a local minimizer of $\{f(\boldsymbol{x}) : h(\boldsymbol{x}) = 0\}$. Then $\nabla f(\boldsymbol{x}^*)$ and $\nabla h(\boldsymbol{x}^*)$ are parallel, that is, if $\nabla h(\boldsymbol{x}^*) \neq \mathbf{0}$, then there exists $\lambda^* \in \mathbb{R}$ (known as the Lagrange multiplier) such that

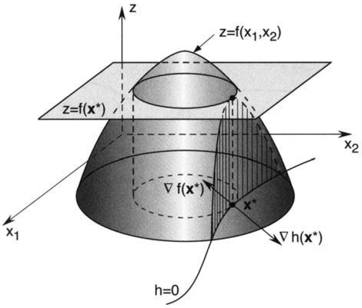
$$\nabla f(\boldsymbol{x}^*) + \lambda^* \nabla h(\boldsymbol{x}^*) = \mathbf{0}.$$



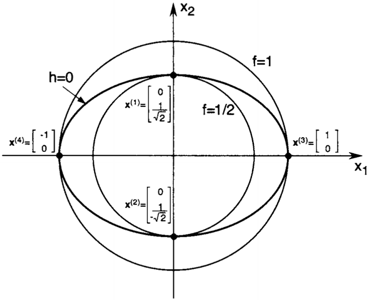
Theorem (Lagrange's theorem, 1st-order optimality condition)

Let $\boldsymbol{x}^*$ *be a local minimizer of* $\{f(\boldsymbol{x}) : \boldsymbol{h}(\boldsymbol{x}) = 0\}$*, where* $\boldsymbol{h} : \mathbb{R}^n \to \mathbb{R}^m, m < n$*. Assume that* $\boldsymbol{x}^*$ *is regular. Then, there exists* $\boldsymbol{\lambda}^* \in \mathbb{R}^m$ *(the Lagrange multipliers) such that*

$$\nabla f(\boldsymbol{x}^*) + D\boldsymbol{h}(\boldsymbol{x}^*)^T \boldsymbol{\lambda}^* = \mathbf{0}.$$

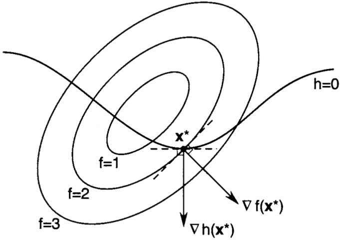


Example 20.7

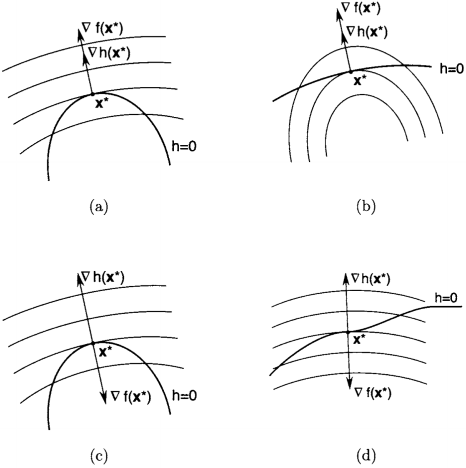


$$\text{minimize } f(\boldsymbol{x}) = x_1^2 + x_2^2$$
$$\text{subject to } h(\boldsymbol{x}) = x_1^2 + 2x_2^2 - 1 = 0.$$

An example where the Lagrange condition is not satisfied



Different cases where the Lagrange condition satisfied



(a) maximizer; (b),(c) minimizer; (d) neither. Courtesy of Seeley'70

Second-order conditions

- **Lagrangian**:

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) := f(\boldsymbol{x}) + \boldsymbol{\lambda}^T \boldsymbol{h}(\boldsymbol{x}) = f(\boldsymbol{x}) + \lambda_1 h_1(\boldsymbol{x}) + \cdots + \lambda_m h_m(\boldsymbol{x})$$

- **Hessians**: $\boldsymbol{F}$ of f and $\boldsymbol{H}_k$ of h_k
- **Hessian** of $\mathcal{L}$: $\boldsymbol{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = \boldsymbol{F}(\boldsymbol{x}) + \lambda_1 \boldsymbol{H}_1(\boldsymbol{x}) + \cdots + \lambda_m \boldsymbol{H}_m(\boldsymbol{x})$

Theorem (2nd-order necessary condition)

Let $\boldsymbol{x}^*$ be a local minimizer of $\{f(\boldsymbol{x}) : \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}\}$, where $\boldsymbol{h} : \mathbb{R}^n \to \mathbb{R}^m, m < n$. Suppose that $f, \boldsymbol{h} \in C^2$. Assume that $\boldsymbol{x}^*$ is regular. Then, there exists $\boldsymbol{\lambda}^*$ such that

1. $\nabla f(\boldsymbol{x}^*) + D\boldsymbol{h}(\boldsymbol{x}^*)^T \boldsymbol{\lambda}^* = \boldsymbol{0}$;
2. *for all $\boldsymbol{y} \in T(\boldsymbol{x}^*)$, we have $\boldsymbol{y}^T \boldsymbol{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) \boldsymbol{y} \geq 0$.*

The role of $\boldsymbol{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ is similar to the objective Hessian $\boldsymbol{F}(\boldsymbol{x}^*)$ in the unconstrained case.

Theorem (2nd-order sufficient condition)

Suppose that $f, \boldsymbol{h} \in C^2$ *and there exists feasible* $\boldsymbol{x}^* \in \mathbb{R}^n$, $\boldsymbol{\lambda}^* \in \mathbb{R}^m$ *such that*

1. $\nabla f(\boldsymbol{x}^*) + D\boldsymbol{h}(\boldsymbol{x}^*)^T \boldsymbol{\lambda}^* = \boldsymbol{0}$;
2. *for all* $\boldsymbol{y} \in T(\boldsymbol{x}^*), \boldsymbol{y} \neq \boldsymbol{0}$, *we have* $\boldsymbol{y}^T \boldsymbol{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)\boldsymbol{y} > 0$.

Then, $\boldsymbol{x}^*$ *is a strict local minimizer of* $\{f(\boldsymbol{x}) : \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}\}$.

Exercise (Example 20.9)

- solve

$$\text{maximize} \ \frac{\boldsymbol{x}^T \boldsymbol{Q} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{P} \boldsymbol{x}}$$

  where

$$\boldsymbol{P} = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}, \quad \boldsymbol{Q} = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}$$

- reformulate as

$$\text{maximize} \ f(\boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{Q} \boldsymbol{x}$$
$$\text{subject to} \ h(\boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{P} \boldsymbol{x} - 1 = 0$$

- the Lagrange condition gives $\boldsymbol{x}^* = \pm[\frac{1}{\sqrt{2}}, 0]^T$
- the second-order sufficient condition is satisfied by $\boldsymbol{x}^*$
- any $t\boldsymbol{x}^*$, $t \neq 0$, is the solution to the original problem

Quadratic minimization with linear constraints

- suppose $\boldsymbol{Q}$ invertible and $\text{rank}(\boldsymbol{A}) = m$. consider

$$\begin{aligned} &\text{minimize } \frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} \\ &\text{subject to } \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \end{aligned}$$

- Lagrangian: $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{\lambda}^T(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x})$
- first-order condition: $\nabla_x \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \boldsymbol{Q}\boldsymbol{x}^* - \boldsymbol{A}^T\boldsymbol{\lambda}^* = \boldsymbol{0}$
- hence, $\boldsymbol{x}^* = \boldsymbol{Q}^{-1}\boldsymbol{A}^T\boldsymbol{\lambda}^*$
- since $\boldsymbol{A}\boldsymbol{x}^* = \boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^T\boldsymbol{\lambda}^* = \boldsymbol{b}$, we get $\boldsymbol{\lambda}^* = (\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^T)^{-1}\boldsymbol{b}$
- $\boldsymbol{x}^* = \boldsymbol{Q}^{-1}\boldsymbol{A}^T\boldsymbol{\lambda}^* = \boldsymbol{Q}^{-1}\boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^T)^{-1}\boldsymbol{b}$ is a candidate sol
- if Lagrangian Hessian $\boldsymbol{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \boldsymbol{Q} \succ \boldsymbol{0}$, then $\boldsymbol{x}^*$ is a strict minimizer

Minimal norm solution to linear equations

- consider $\boldsymbol{Q}$ invertible and

$$\begin{aligned} \text{minimize } & \frac{1}{2}\|\boldsymbol{x}\|^2 \\ \text{subject to } & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \end{aligned}$$

- assume $\text{rank}(\boldsymbol{A}) = m$. then, the solution is

$$\boldsymbol{x}^* = \boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{A}^T)^{-1}\boldsymbol{b}.$$

- the solution is the projection of $\mathbf{0}$ to the affine set $\{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}\}$

Summary

- the equality constraints in an NLP define the “constraint surface”
- at a “stationary point” of an NLP, the objective gradient is perpendicular to the tangent of the constraint surface
- the 1st-order conditions equalize ∇f with a linear combination of all ∇h_i
- like before, 1st-order conditions give candidate solutions, 2nd-order conditions help determine extremal status