262B-Lecture 18

Date created: 2021.03.30
N. of Pages: 13

Sensitivity theorem:

- Assume (x_*, λ_*) satisfies FOC, SOC (sufficient)

for $\min f(x)$ s.t. $h(x) = 0$

- Assume x_* is a $\underbrace{\text{regular point}}$.

Consider a parametrized optimization:

$$\begin{cases} \min_x f(x) & \quad (1) \\ \text{s.t. } h_i(x) = v_i \quad i = 1, \dots, m \end{cases} \longrightarrow v = \begin{bmatrix} v_1 \\ \vdots \\ v_m \end{bmatrix} \text{ is a parameter}$$

There exists an $\underbrace{\text{open ball}}$ S centered at $v = 0$

and $\exists$ $x(u): S \rightarrow \mathbb{R}^n$, $\lambda(u): S \rightarrow \mathbb{R}^m$ such that

1- $(x(u), \lambda(u))$ is a pair of local min - lagrange multiplier for (1) for all $v \in S$

2- $x(u)$ and $\lambda(u)$ are continuously differentiable over S s.t. $x(0) = x_*$, $\lambda(0) = \lambda_*$,

gradient of optimal obj $= \nabla_u f(x(u)) = -\lambda(u)$ $\quad \forall v \in S$

Implications :

If $u_1, u_2, \ldots, u_m$ are small , then

$$f(x(u)) \simeq f(x_*) - \lambda_1^* u_1 - \cdots - \lambda_m^* u_m$$

That means changing each constraint a bit by u_i affects the optimal objective value by $-\lambda_i^* u_i$.

$\Rightarrow$ $\underbrace{\lambda_i^*}$

- $\to$ small : optimization is not sensitive to constraint i.
- $\to$ Large : optimization is sensitive.
- $\to$ zero : dropping the constraint has no effect on that local min.

economics: shadow price

Proof:

$$\begin{cases} \min & f(x) \\ \text{s.t} & h(x) = u \end{cases} \overset{\text{FOC}}{\Longrightarrow} \begin{cases} \nabla f(x) + \overbrace{\nabla_x (h(x) - u)}^{\nabla_x h(x)} \lambda = 0 \\ h(x) = u \end{cases}$$

Define $$g(x, \lambda, u) = \begin{bmatrix} \nabla f(x) + \nabla h(x) \lambda \\ h(x) - u \end{bmatrix}$$

$$\nabla_{(x,\lambda)} g(x_*, \lambda_*, 0) = \begin{bmatrix} \nabla^2_{xx} L(x_*, \lambda_*) & \nabla h(x_*) \\ \nabla h(x_*)^T & 0 \end{bmatrix}$$

Hw 5: SOC (sufficient) + $\underbrace{\text{regularity}}$

$\Rightarrow$ This matrix is invertible

$\Rightarrow$ $\nabla_{(x,\lambda)} g(x_*, \lambda_*, 0) =$ invertible

Implicit function theorem:

$\exists$ open ball S around $u=0$ and

$x(u): S \to \mathbb{R}^n$, $\lambda(u): S \to \mathbb{R}^m$ s.t

$$\boxed{g(x(u), \lambda(u), u) = 0 \quad \forall u \in S}$$

$\Rightarrow$ $(x(u), \lambda(u))$ satisfies FOC for $\begin{cases} \min f(x) \\ \text{s.t. } h(x) = u \end{cases}$

Also: $\nabla_x (h(x) - u) = \nabla_x h(x)$

So, $V(x(u))$ doesn't depend on u.

So, the basis matrix E doesn't depend on u either.

SOC sufficient:

$$E^T\left(\nabla^2 f(x(u)) + \sum_{i=1}^{m} \lambda_i(u) \nabla^2 h_i(x(u))\right) E \overset{?}{\succ} 0$$

This is true by assumption if $u = 0$

$\Rightarrow$ If $\underline{\underline{S}}$ is chosen to be small enough, the SOC (sufficient) is satisfied for all $u \in S$.

$\Rightarrow$ $x(u)$: local min ;

Also, FOC: $\nabla_x f(x(u)) + \nabla_x h(x(u))\, \lambda(u) = 0$

$\Rightarrow$ $\nabla_u x(u)\, \nabla_x f(x(u)) + \nabla_u x(u)\, \nabla_x h(x(u))\, \lambda(u) = 0$ ①

$h(x(u)) = u \Rightarrow \nabla_u x(u)\, \nabla_x h(x(u)) = I$ ②

$\nabla_u f(x(u)) = \nabla_u x(u)\, \nabla_x f(x(u))$ ③

①, ②, ③ $\Rightarrow$ $\nabla_u f(x(u)) = -\lambda(u)$

$$\begin{cases} \min f(x) \\ \text{s.t. } h(x) = 0 \end{cases}$$

Tangent Cone: $T(x_*) = \{ \Delta x \mid \Delta x = 0$ or

$\exists \{x^{(k)}\}_{k=0}^{\infty}$ s.t. $\underbrace{h(x^{(k)}) = 0}_{\text{feasible}}$, $x^{(k)} \neq x_*$,

$\lim_{k \to \infty} x^{(k)} = x_*$, $\lim_{k \to \infty} \frac{x^{(k)} - x_*}{\|x^{(k)} - x_*\|} = \frac{\Delta x}{\|\Delta x\|} \}$

Cone of first-order feasible directions:

$V(x_*) = \{ \Delta x \mid \nabla h_i (x_*)^T \Delta x = 0 \quad i = 1, \dots, m \}$

Thm: 1- $T(x_*) \subseteq V(x_*)$

2- $T(x_*) = V(x_*)$ if $\underbrace{x_* \text{ is regular}}$

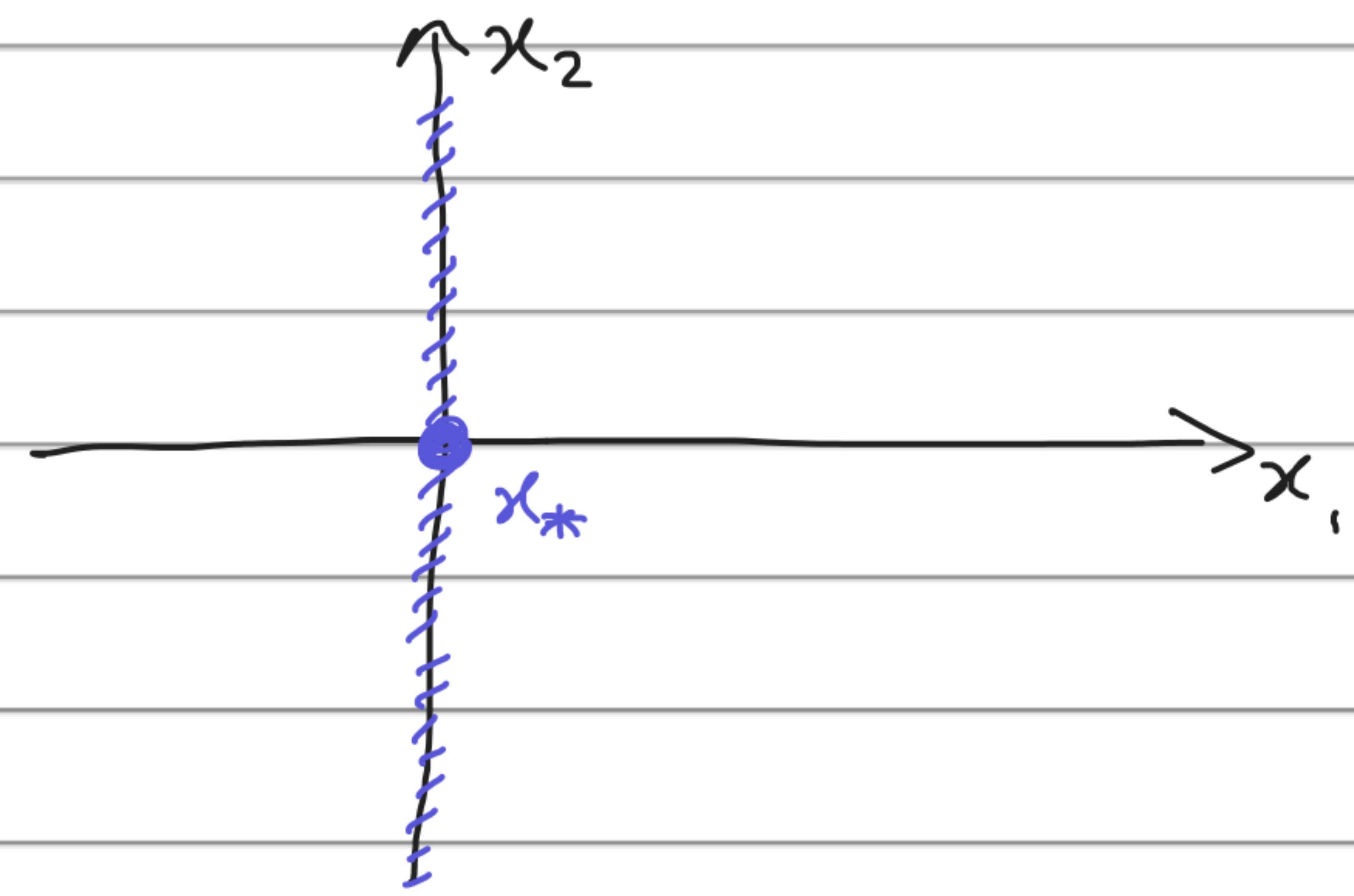
$T(x_*)$: geometric definition

$V(x_*)$: algebraic definition

Regularity allows us to characterize Tangent Cone via gradients of constraints (due to $V(x_*)$).

So, regularity bridges the gap between geometric analysis (via tangent cone) and algebraic analysis (via gradients of constraints).

Ex: $\begin{cases} \min & x_1^2 + x_2^2 \\ \text{s.t.} & x_1 = 0 \end{cases}$



$x_* = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

$\Rightarrow \nabla_x (x_1 - 0) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$: linearly independent

$\Rightarrow T(x_*) = V(x_*) = \{ \Delta x \in \mathbb{R}^2 \mid \underbrace{[1 \quad 0] \, \Delta x = 0}_{\Delta x_1 = 0} \}$

equivalent representation:

$\begin{cases} \min & x_1^2 + x_2^2 \\ \text{s.t.} & x_1^2 = 0 \end{cases}$ $\longrightarrow$ same feasible set

$\longrightarrow \nabla_x (x_1^2 - 0) \Big|_{x_*} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

$\Rightarrow$ no point is regular.

$$T(x_*) = \{ \Delta x \in \mathbb{R}^2 \mid [1 \quad 0] \Delta x = 0 \} \rightarrow \text{Line}$$

$$V(x_*) = \{ \Delta x \in \mathbb{R}^2 \mid [0 \quad 0] \Delta x = 0 \} \rightarrow \text{entire space}$$

$$\Rightarrow T(x_*) \subset V(x_*)$$

The issue is the algebraic representation of the feasible set in a way that the gradients of constraints are not informative enough.

General optimization problems:

$$\begin{cases} \min & f(x) \\ \text{s.t.} & h_i(x) = 0 \quad i = 1, \dots, m \quad \rightarrow \quad h(x) = 0 \\ & g_j(x) \le 0 \quad j = 1, \dots, r \quad \rightarrow \quad g(x) \le 0 \end{cases}$$

element-wise

Def: Given a feasible point y, define the set of active inequality constraints at y as:

$A(y) = \{ j \mid g_j(y) = 0 \ , \ j \in \{1, \dots, r\} \}$

- active or binding constraints.
- if $g_j(y) < 0$, then j^{th} constraint is called inactive or non-binding.

To write optimality conditions for a candidate solution x_*, we use the following idea:

$$\underbrace{\begin{cases} \min & f(x) \\ \text{s.t.} & h_i(x) = 0 \quad i = 1, \dots, m \\ & g_j(x) \leq 0 \quad j = 1, \dots, r \end{cases}}_{\text{general opt}} \xrightarrow[\text{treat active inequalies as equalies}]{} \underbrace{\begin{cases} \min & f(x) \\ \text{s.t.} & h_i(x) = 0 \quad i = 1, \dots, m \\ & g_j(x) \underbrace{=}_{} 0 \quad \underbrace{j \in A(x_*)}_{} \end{cases}}_{\text{optimization with only equality constraint}}$$

Intuition: FOC for problem in the right:

$$\nabla f(x_*) + \sum_{i=1}^{m} \lambda_i^* \nabla h_i(x_*) + \sum_{j \in A(x_*)} \mu_j^* \nabla g_j(x_*) = 0$$

$$\sum_{j=0}^{r} \mu_j^* \nabla g_j(x_*) = 0 \quad \text{but } \mu_j^* = 0 \text{ if } j \notin A(x_*)$$

A feasible point x_* is called regular if $\nabla h_i(x_*)$ for $i \in \{1, \dots, m\}$ and $\nabla g_j(x_*)$ for $j \in A(x_*)$ are linearly independent.

$\Rightarrow$ we only care about gradients of equality and <u>active</u> inequality constraints.

FOC or KKT (Karush - Kuhn - Tucker) :

If x_* is a local min and a regular point, then $\exists\ \lambda_* = (\lambda_1^*, \dots, \lambda_m^*)$ and $\mu_* = (\mu_1^*, \dots, \mu_r^*)$ s.t.

1- $\nabla f(x_*) + \sum_{i=1}^{m} \lambda_i^* \nabla h_i(x_*) + \sum_{j=1}^{r} \mu_j^* \nabla g_j(x_*) = 0$

2- $\mu_j^* \geq 0 \qquad j = 1, \dots, r$

3- $\mu_j^* g_j(x_*) = 0 \qquad j = 1, \dots, r$

Define : $L(x, \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x)$
$$+ \sum_{j=1}^{r} \mu_j g_j(x)$$

Condition $\underline{1}$: $\nabla_x L(x_*, \lambda_*, \mu_*) = 0$

$\Rightarrow$ called "stationarity"

Condition $\underline{2}$: is called dual feasibility in the convex case.

Condition $\underline{3}$: $\longleftrightarrow \mu_j^* = 0 \quad \forall j \notin A(x_*)$

By regarding μ_j^* as a sensitivity parameter, this means solution x_* is not sensitive to small changes to constraint j.

$\Rightarrow$ called "Complementary slackness"

SOC (necessary): If x_* is a local min and a regular point, for (λ_*, μ_*) satisfying FOC, we have:

$$\Delta x^T \underbrace{\left(\nabla^2 f(x_*) + \sum_{i=1}^{m} \lambda_i^* \nabla^2 h_i(x_*) + \sum_{j=1}^{r} \mu_j^* \nabla^2 g_j(x_*) \right)}_{\nabla^2_{xx} L(x_*, \lambda_*, \mu_*)} \Delta x \geq 0$$

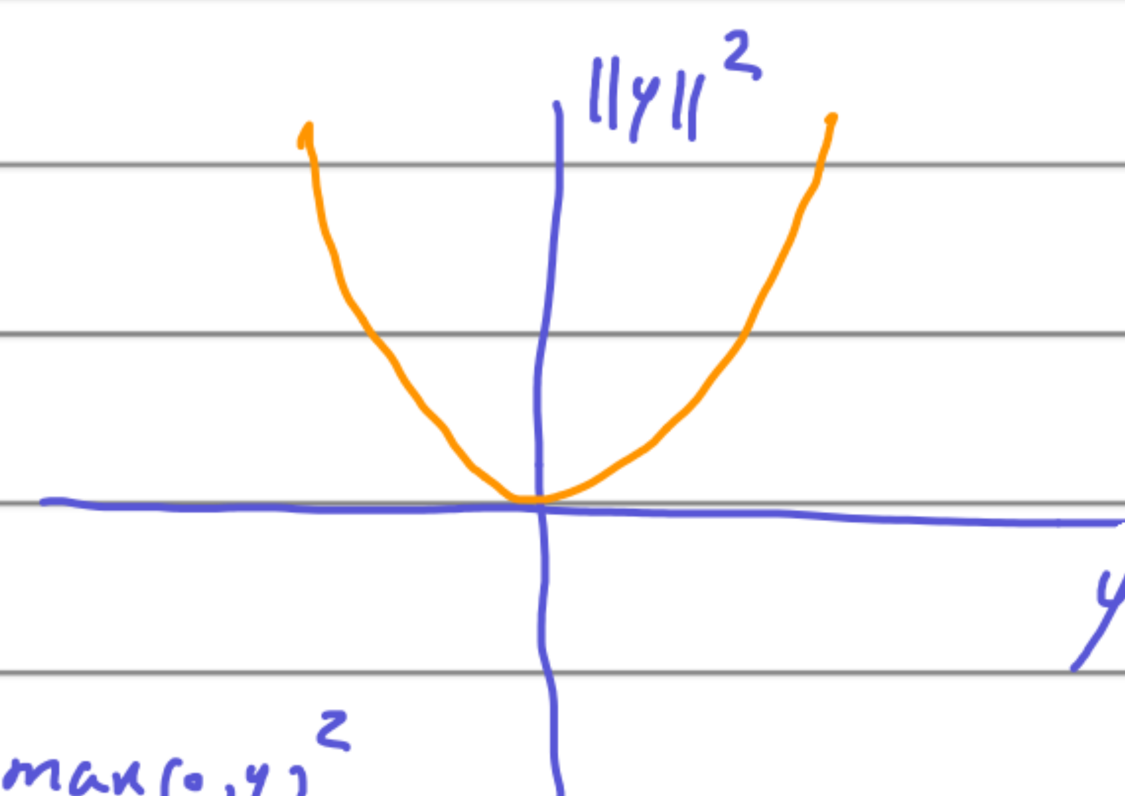
for all Δx s.t.

$$\begin{cases} \nabla h_i(x_*)^T \Delta x = 0 & i = 1, \dots, m \\ \nabla g_j(x_*)^T \Delta x = 0 & j \in A(x_*) \end{cases}$$

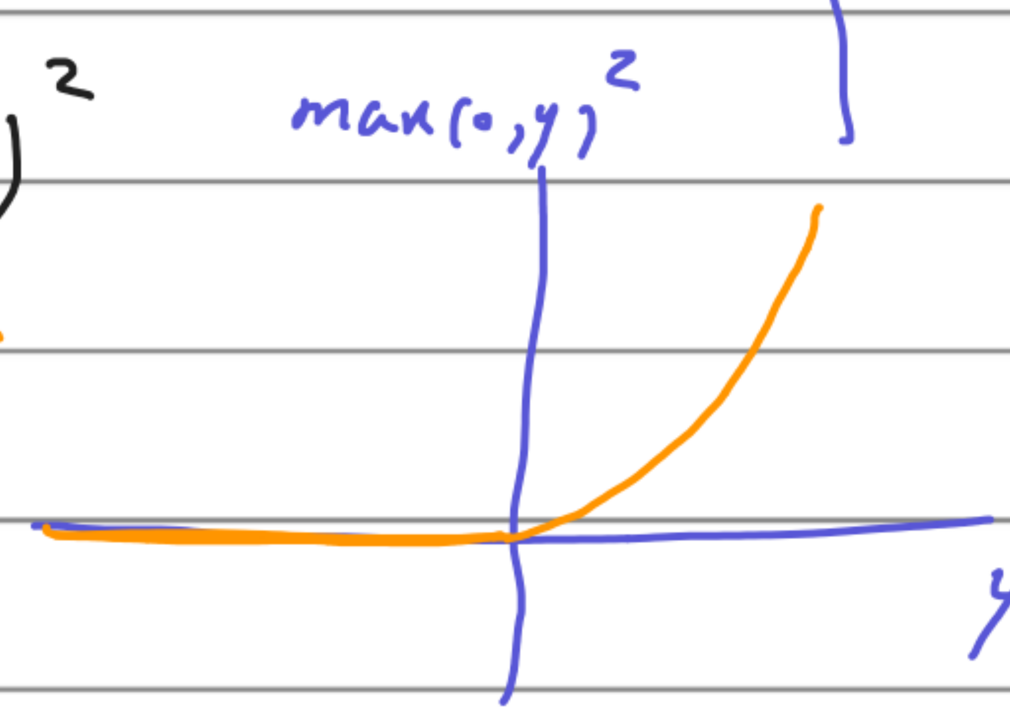
Two proof techniques:

1- penalty:

- $h_i(x) = 0 \longrightarrow \| h_i(x) \|^2$



- $g_j(x) \leq 0 \longrightarrow \underbrace{\max(0, g_j(x))^2}_{\text{differentiable}}$



$$F^k(x) = f(x) + \frac{k}{2}\|h(x)\|^2 + \frac{k}{2}\sum_{j=1}^{r}\left(\max(0, g_j(x))\right)^2 + \frac{\alpha}{2}\|x - x_*\|^2$$

As before:

$$\begin{cases} \min & f(x) \\ \text{s.t.} & h(x) = 0 \\ & g(x) \le 0 \end{cases} \longrightarrow \begin{array}{ll} \min & F^k(x) \\ \text{s.t.} & x \in S \end{array} \longrightarrow \text{solution } x^{(k)} \text{ and } k \text{: large}$$

$$\longrightarrow \min \quad F^k(x)$$

$$\downarrow$$

$$\nabla F^k(x^{(k)}) = 0$$

$$\nabla\left(\max(0, g_j(x))\right)^2 = \underbrace{2\max(0, g_j(x))\,\nabla g_j(x)}_{\text{second derivative doesn't exist}}$$

$$\Rightarrow \lambda_i^* = \lim_{k\to\infty} k\, h_i(x^{(k)}) \qquad i = 1, \dots, m$$

$$\mu_j^* = \lim_{k\to\infty} k\, \underbrace{\max(0, g_j(x^{(k)}))} \qquad j = 1, \dots, r$$

1- $\max(0, g_j(x^{(k)})) \geq 0 \implies \mu_j^* \geq 0$

2- if $g_j(x_*) < 0$ then $g_j(x^{(k)}) < 0$ for large k

$\implies \max(0, g_j(x^{(k)})) = 0 \implies \mu_j^* = 0$