262B-Lecture 19

Date created: 2021.04.01
N. of Pages: 16

Technique 2: Conversion to equality case

$$\underbrace{\begin{aligned} &\min_{x\in\mathbb{R}^n} \quad f(x) \\ &\text{s.t.} \quad h_i(x)=0 \quad i=1,\dots,m \\ &\qquad\quad g_j(x)\le 0 \quad j=1,\dots,r \end{aligned}}_{1} \quad \longleftrightarrow \quad \underbrace{\begin{aligned} &\min_{\substack{x\in\mathbb{R}^n \\ z\in\mathbb{R}^r}} \quad f(x) \\ &\text{s.t.} \quad h_i(x)=0 \quad i=1,\dots,m \\ &\qquad\quad g_j(x)+z_j^2=0 \quad j=1,\dots,r \end{aligned}}_{2}$$

x_* : Local min for 1 and regular $\Rightarrow$ (x_*, z_*) : Local min & regular for 2

where

$$z_j^* = \sqrt{-g_j(x_*)}$$

FOC for 2 : $\exists\ (\lambda_*, \mu_*)$ s.t.

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

$$\begin{bmatrix} 0 \\ \vdots \\ 0 \\ 2z_j^* \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$\Rightarrow \begin{cases} \nabla_x L(x_*, \lambda_*, \mu_*) = 0 \longrightarrow \text{stationarity} \\ \mu_j^* \underbrace{z_j^*}_{\sqrt{-g_j(x_*)}} = 0 \qquad j = 1, \dots, r \end{cases}$$

$$\mu_j^* g_j^*(x_*) = 0$$

$$\downarrow$$

Complementary slackness

SOC (necessary) for ②:

$$[\Delta x^T \ \Delta z^T] \begin{bmatrix} \nabla_{xx}^2 L(x_*, \lambda_*, \mu_*) & 0 \\ 0 & 2\,\mathrm{diag}(\mu_*) \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta z \end{bmatrix} \geq 0 \qquad ①$$

for all $[\Delta x^T \ \Delta z]^T$ s.t.

$$[\nabla h_i(x_*)^T \ \ 0] \begin{bmatrix} \Delta x \\ \Delta z \end{bmatrix} = 0 \qquad i = 1, \dots, m \qquad ②$$

$$[\nabla g_j(x_*)^T \ \ 0 \dots 0 \ \ 2z_j^* \ \ 0 \dots 0] \begin{bmatrix} \Delta x \\ \Delta z \end{bmatrix} = 0 \qquad j = 1, \dots, r \qquad ③$$

pick $\Delta x = 0$, $\Delta z = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \mu_j^* \\ 0 \\ \vdots \\ 0 \end{bmatrix} \Rightarrow$ ②, ③ ✓

$\Rightarrow$ ① $\Rightarrow (\mu_j^*)^3 \geq 0$

$\Rightarrow \quad \mu_j^* \geq 0 \qquad j = 1, \dots, r$

By eliminating Δz from ② and ③,

$\Rightarrow$ ① reduces to the SOC for the case with inequality constraints.

SOC (sufficient): Consider a feasible point x_* satisfying FOC for some pair (λ_*, μ_*) (no regularity assumption). Assume that

$$\Delta x^T \nabla^2_{xx} L(x_*, \lambda_*, \mu_*) \Delta x > 0$$

for all Δx s.t.

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

Assume strict complementary slackness:

$\mu_j^* > 0 \quad \forall j \in A(x_*)$ (μ_j^* and $g_j(x_*)$ can't be zero at the same time)

$\Longrightarrow$ x_* is a Local min.

Same as the case with only equality constraints but has the assumption of strict complementary slackness.

Sensitivity: Assume (x_*, λ_*, μ_*) satisfies SOC (sufficient) and $\underbrace{x_* \text{ is regular}}$.

Consider

$$\begin{cases} \min f(x) \\ \text{s.t. } h(x) = u \\ \quad\quad g(x) \le v \end{cases}$$

$\Rightarrow$ There exist an open ball S centered at $(u,v) = (0,0)$ and continuously differentiable functions $x(u,v)$, $\lambda(u,v)$ and $\mu(u,v)$ over S s.t.

1 - $\begin{cases} x(0,0) = x_* \\ \lambda(0,0) = \lambda_* \\ \mu(0,0) = \mu_* \end{cases}$

2 - $(x(u,v), \lambda(u,v), \mu(u,v))$: local min - Lagrange multipliers

for all $(u,v) \in S$

3 - $\begin{cases} \nabla_u f(x(u,v)) = -\lambda(u,v) \\ \nabla_v f(x(u,v)) = -\mu(u,v) \end{cases}$

$$\min \quad f(x)$$

$$\text{s.t.} \quad h_i(x) = 0 \quad i = 1, \dots, m$$

$$g_j(x) \le 0 \quad j = 1, \dots, r$$

$$\underbrace{x \in X}_{\text{domain of definition}} \implies$$

you can absorb it into the objective by a penalty $p(x)$

$$= \begin{cases} 0 & x \in X \\ +\infty & x \notin X \end{cases}$$

Thm (sufficient condition for global optimality):

Consider a feasible point x_* for which there exist λ_* and μ_* s.t.

1 - $x_* \in \operatorname*{argmin}_{x \in X} L(x, \lambda_*, \mu_*)$ → global min

2 - $\lambda_* \ge 0$

3 - Complementary slackness

$\implies x_*$: global min

If $X = \mathbb{R}^n$, then

$$x_* \in \arg\min_x L(x, \lambda_*, \mu_*) \quad \text{A}$$

unconstrained opt

$$\nabla_x L(x_*, \lambda_*, \mu_*) = 0 \quad \text{B}$$

Going from B to A is not true.

So, A is much stronger than B.

So, if "stationarity condition B" in FOC is replaced with "global optimality condition A", then if there is a solution satisfying the conditions, that's a global min.

Proof: $x_* \in \arg\min_{x \in X} L(x, \lambda_*, \mu_*)$

$$\Rightarrow L(x_*, \lambda_*, \mu_*) = \min_{x \in X} \left(f(x) + \lambda_*^T h(x) + \mu_*^T g(x)\right)$$

①

$$L(x_*, \lambda_*, \mu_*) = f(x_*) + \lambda_*^T \underbrace{h(x_*)}_{=0} + \underbrace{\mu_*^T g(x_*)}_{=0}$$

Complementary

$$= f(x_*)$$

① $\Rightarrow$ $$f(x_*) \leq \min_{x \in X} \left(f(x) + \lambda_*^T h(x) + \mu_*^T g(x) \right)$$

$$\leq \min_{\substack{x \in X \\ h(x)=0 \\ g(x) \leq 0}} \left(f(x) + \underbrace{\lambda_*^T h(x)}_{\text{zero over feasible set}} + \underbrace{\mu_*^T}_{\geq 0} \underbrace{g(x)}_{\text{negative over feasible set}} \right)$$

smaller feasible set

$\Rightarrow$ higher objective

$$\leq \left(\min_{\substack{x \in X \\ h(x)=0 \\ g(x) \leq 0}} f(x) \right)$$

original problem

$\Rightarrow$ x_* gives the lowest value possible

$\Rightarrow$ global min

Special case:

Convex optimization

$\min \quad f(x) \longrightarrow$ Convex

$\text{s.t} \quad h_i(x) = 0 \quad i = 1, \dots, m \longrightarrow$ Linear

$\quad\quad g_j(x) \le 0 \quad j = 1, \dots, r \longrightarrow$ Convex

$$\Longrightarrow L(x, \lambda, \mu) = \underbrace{f(x)}_{\text{Convex}} + \underbrace{\sum_{i=1}^{m} \lambda_i h_i(x)}_{\text{Linear}} + \underbrace{\sum_{j=1}^{r} \mu_j g_j(x)}_{\substack{\text{Convex} \\ \text{if } \mu \ge 0}}$$

$$= \text{Convex over } \mu \ge 0$$

$$\underbrace{\min L(x, \lambda_*, \mu_*)}_{A} \longleftrightarrow \underbrace{\nabla_x L(x_*, \lambda_*, \mu_*) = 0}_{B}$$

$\Longrightarrow$ If x_* satisfies FOC, then x_* is a global min since $A \Leftrightarrow B$

$\Longrightarrow$ SOC is not needed

$\Rightarrow$ Convex optimization: A $\underbrace{\text{regular}}$ feasible point x_* is global / local $\underbrace{\text{if and only if}}$ it satisfies FOC

(KKT in the convex case)

Can we eliminate the regularity assumption?

Fritz-John necessary conditions:

If x_* is a local min (no regularity assumption)

then $\exists\ \lambda_* \in \mathbb{R}^m,\ \mu_* \in \mathbb{R}^r,\ \gamma_* \in \mathbb{R}$

s.t.

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

2 - $\mu_* \geq 0$

3 - Complementary slackness

4 - $(\lambda_*, \mu_*, \gamma_*) \neq (0,0,0)$

proof $\longrightarrow$ $\min F^{\kappa}(x)$ $\longrightarrow$ FOC

$\Longrightarrow$ regularity is not needed if we add γ_*.

Note: By re-scaling, we can always assume $\gamma_* \in \{0, 1\}$

- If regularity is satisfied $\longrightarrow$ $\gamma_* = 1$
- Using this theorem, we can obtain conditions replacing "regularity".
- Any condition that makes FOC hold is called Constraint qualification.

First constraint qualification :

Linear / Concave constraints

Thm: If $h_i(x)$ are linear and $g_j(x)$ are concave, then FOC is satisfied for every local min x_*.

(this replaces regularity).

proof: Based on Fritz-John conditions $\Rightarrow \gamma_* = 1$

Second constraint qualification:

Linear independence / interior point

Thm: A local min x_* satisfies FOC if

i) $\nabla h_1(x_*), \ldots, \nabla h_m(x_*)$ are linearly independent.

ii) $\exists\ \Delta x$ s.t.

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

Fritz-John: $\exists\, (\gamma_*, \lambda_*, \mu_*)$

and by contradiction assume $\gamma_* = 0$

$$\Rightarrow \quad 0 \times \nabla f(x_*) + \sum_{i=1}^{m} \lambda_i^* \nabla h_i(x_*)$$

$$+ \sum_{j=1}^{r} \mu_j^* \nabla g_j(x_*) = 0$$

Also, $\mu_* \neq 0$ since if $\mu_* = 0$, due to Linear independence of $\nabla h_i(x_*)$'s, $\lambda_* = 0$

$$\Rightarrow \sum_{i=1}^{m} \lambda_i^* \underbrace{\nabla h_i(x_*)^T \Delta x}_{=0} + \sum_{j=1}^{r} \underbrace{\mu_j^*}_{\substack{\geq 0 \\ \text{Fritz-John}}} \underbrace{\nabla g_j(x_*)^T \Delta x}_{< 0} = 0$$

and since $\mu_* \neq 0$, this is contradiction

$$\Longrightarrow \gamma_* = 1$$

Convex Case:

Thm: Assume $h_i(x)$ are Linear and $g_j(x)$ are Convex. Then a local min x_* satisfies KKT if $\exists\ y \in \mathbb{R}^n$

s.t. 1- y = feasible

2- $g_j(y) < 0 \quad \forall j \in A(x_*)$

proof: $\Delta x = y - x_* \implies$

$$0 = h_i(y) \underset{\text{Linearity}}{=} \overset{0}{\cancel{h_i(x_*)}} + \nabla h_i(x_*)^T \underbrace{(y - x_*)}_{\Delta x}$$

$$\implies \nabla h_i(x_*)^T \Delta x = 0$$

Also,

$$\underset{\text{by assumption}}{0 >} g_j(y) \underset{\text{Convexity}}{\geq} \underbrace{g_j(x_*)}_{=0 \text{ if } j \in A(x_*)} + \nabla g_j(x_*)^T \underbrace{(y - x_*)}_{\Delta x}$$

$$\implies \nabla g_j(x_*)^T \Delta x < 0 \quad \forall j \in A(x_*)$$

$\min \quad f(x)$

s.t. $\quad h_i(x) = 0 \qquad i = 1, \dots, m$

$\qquad g_j(x) \le 0 \qquad j = 1, \dots, r$ $\Longrightarrow$ it's expected that $n-m$ constraints are binding.

So, the set $A(x_*)$ has $\binom{r}{n-m}$ possibilities

To simplify this:

$g_j(y) < 0 \quad \forall \, j \in A(x_*)$

$\Big\downarrow$ stronger

$g_j(y) < 0 \quad j \in \{1, \dots, r\}$ $\Longrightarrow$ slater

slater: $\exists \underbrace{y}_{\text{in the relative interior of feasible set}} : h(y) = 0 \;, \; g(y) < 0$

Remember $\min \underbrace{f(x)}_{\text{non-convex}}$ s.t. $x \in \underbrace{X}_{\text{convex}}$

FOC (geometric) : $\nabla f(x_*)^T (x - x_*) \geq 0 \quad \forall x \in X$

$\updownarrow$ proved

$\nabla f(x_*)^T \Delta x \geq 0 \quad \forall \Delta x \in T(x_*)$

This is true even if X is non convex

FOC (geometric) : If x_* is a local min (without any assumptions), then

$$\underbrace{\nabla f(x_*)^T \Delta x}_{} \geq 0 \qquad \forall \Delta x \in T(x_*)$$

(there is no descent direction to take)

So, FOC (geometric) : no assumption

FOC (algebric) : Constraint qualification

why ? gap ?