262B-Lecture 17

Date created: 2021.03.18
N. of Pages: 14

How to use penalty method to derive SOC ?

$$\begin{cases} \min\limits_{x \in \mathbb{R}^n} f(x) \\ \text{s.t. } h(x) = 0 \end{cases} \Rightarrow \begin{cases} \min F^k(x) \longrightarrow f(x) + \frac{k}{2}\|h(x)\|^2 + \frac{\alpha}{2}\|x - x_*\|^2 \\ \text{s.t. } x \in S \longrightarrow k : \text{large} \end{cases}$$

$\downarrow$

solution $x^{(k)}$

$\{x^{(k)}\} \longrightarrow x_*$: SOC $\longrightarrow$ $\nabla^2 F^k(x^{(k)}) \succeq 0$

Consider $\Delta x \in V(x_*)$: $\nabla h_i(x_*)^T \Delta x = 0$ $i = 1, \ldots, m$

$\Rightarrow$ $\nabla h(x_*)^T \Delta x = 0$

Note: $\nabla h(x^{(k)})^T \Delta x$ may not be zero

Let $\Delta x^{(k)}$ be projection of Δx on null space of $\nabla h(x^{(k)})^T$

$$\Rightarrow \Delta x^{(k)} = \Delta x - \nabla h(x^{(k)}) \big(\underbrace{\nabla h(x^{(k)})^T \nabla h(x^{(k)})}_{} \big)^{-1} \nabla h(x^{(k)})^T \Delta x$$

invertible if k = large
due to regularity of x_*

So, $\begin{cases} \nabla h(x^{(k)})^T \underline{\Delta x^{(k)}} = 0 \quad \circledast \\ \Delta x^{(k)} \longrightarrow \Delta x \end{cases}$

SOC for unconstrained optimization:

$$0 \leq (\Delta x^{(k)})^T \nabla^2 F^k (x^{(k)}) \Delta x^{(k)}$$

$$= (\Delta x^{(k)})^T \Big(\nabla^2 f(x^{(k)}) + k \sum_{i=1}^{m} h_i(x^{(k)}) \nabla^2 h_i(x^{(k)})$$

$$+ k \nabla h(x^{(k)}) \nabla h(x^{(k)})^T + \alpha I \Big) \Delta x^{(k)}$$

$$\overset{\circledast}{\leq} (\Delta x^{(k)})^T \Big(\nabla^2 f(x^{(k)}) + \sum_{i=1}^{m} (k h_i(x^{(k)})) \nabla^2 h_i(x^{(k)}) \Big) \Delta x^{(k)}$$

$$+ \alpha \| \Delta x^{(k)} \|^2$$

$k \to \infty$, $\alpha \to 0$: $k h_i(x^{(k)}) \longrightarrow \lambda_i^*$, $\Delta x^{(k)} \longrightarrow \Delta x$

$x^{(k)} \longrightarrow x_*$

Take the Limit $\Longrightarrow$ $$0 \leq (\Delta x)^T \Big(\nabla^2 f(x_*) + \sum_{i=1}^{m} \lambda_i^* \nabla^2 h_i(x_*) \Big) \Delta x$$

$\forall \; \Delta x \in V(x_*)$

There are infinitely many vectors in $V(x_*)$.

So, how to check the SOC inequality for all such Δx?

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

since $\nabla h_1(x_*), \dots, \nabla h_m(x_*)$ are Linearly independent

$\Rightarrow \dim V(x_*) = n - m$

- let $E_1, \dots, E_{n-m}$ be an arbitrary basis for $V(x_*)$.

- Define: $E = [E_1 \ E_2 \ \dots \ E_{n-m}]$

$\Rightarrow \Delta x \in V(x_*) \iff \Delta x = E z$ where $z \in \mathbb{R}^{n-m}$

$\Rightarrow$ SOC is equivalent to:

$$\underbrace{E^T \left(\nabla^2 f(x_*) + \sum_{i=1}^{m} \lambda_i^* \nabla^2 h_i(x_*) \right) E}_{\text{restricted hessian}} \succeq 0$$

Proofs via elimination method:

- First, let's focus on a special case:

$$\begin{cases} \min f(x) \\ \text{s.t. } Ax = b \end{cases} \Rightarrow h(x) = Ax - b$$

$A \in \mathbb{R}^{m\times n}$, regularity $\Longrightarrow$ A = full row rank

No loss of generality:

$$A = [\underbrace{B}_{m\times m} \vdots \underbrace{R}_{m\times(n-m)}] \quad \text{s.t.} \quad B = \text{invertible}$$

Decompose: $x = \begin{bmatrix} x_B \\ \hline x_R \end{bmatrix}$

$$\min_{x_B, x_R} f(x_B, x_R) \quad \text{s.t.} \quad B x_B + R x_R = b$$

eliminate x_B: $x_B = B^{-1}(b - R x_R)$

$$\Longrightarrow \min_{x_R} \underbrace{f(B^{-1}(b - R x_R), x_R)}_{\text{define: } F(x_R)} \quad \text{s.t.} \quad x_R \in \mathbb{R}^{n-m}$$

Unconstrained optimization: $\min_{x_R} F(x_R)$

since (x_B^*, x_R^*) is a local min for original problem, x_R^* is a local min for new problem due to equivalence $\Longrightarrow$

FOC: $\nabla F(x_R^*) = 0$, SOC: $\nabla^2 F(x_R^*) \succeq 0$

$$0 = \nabla F(x_R^*) = \nabla_{x_R} f(B^{-1}(b - Rx_R), x_R)\Big|_{x_R = x_R^*}$$

$$= -(B^{-1}R)^T \nabla_B f(x_*) + \nabla_R f(x_*) \quad (1)$$

Define: $\lambda_* = -(B^{-1})^T \nabla_B f(x_*)$ (2)

(1) $\Rightarrow$ $\nabla_R f(x_*) + R^T \lambda_* = 0$

(2) $\Rightarrow$ $\nabla_B f(x_*) + B^T \lambda_* = 0$

$\Rightarrow$ $\nabla f(x_*) + A^T \lambda_* = 0$

$\Rightarrow$ $\nabla f(x_*) + \nabla h(x_*) \lambda_* = 0$ $\longrightarrow$ FOC

Proof for SOC is also similar.

General case: $\min f(x)$ s.t. $h(x) = 0$

since x_* is a regular point $\Rightarrow$ $\nabla h(x_*)$ = full column rank

∇h: Jacobian

No loss of generality:

$$\nabla h(x_*) = \begin{bmatrix} \boxed{m \times m}^{T} \\ \boxed{(n-m) \times m} \end{bmatrix}$$ → invertible

$$\Rightarrow \quad x = \begin{bmatrix} x_B \\ \hdashline x_R \end{bmatrix} \quad \text{and} \quad \underbrace{\nabla_B h(x_*) = \text{invertible}}_{\text{notion of regularity}}$$

Implicit function theorem:

$$\underbrace{0 = h(x) = h(x_B, x_R)}_{\text{feasible set}} \quad \text{and} \quad \nabla_B h(x_*) = \text{invertible}$$

$\Rightarrow \quad \exists\ \phi : S \to \mathbb{R}^m$ s.t.

1- ϕ: unique & twice continuously differentiable

(assumption: $h(x)$ is twice continuously differentiable)

2- S: Ball centered at x_R^*

3- $\phi(x_R^*) = x_B^*$

4 - $\phi(x_R) = x_B \quad \forall x_R \in S$

$\Rightarrow \quad h(\phi(x_R), x_R) = 0 \quad \forall x_R \in S$

(solve x_B as a function of x_R, and eliminate it)

$\nabla_R = 0$

$$\Rightarrow \nabla\phi(x_R)\underbrace{\nabla_B h(\phi(x_R), x_R)} + \underbrace{\nabla_R h(\phi(x_R), x_R)} = 0$$

(*)

$$\begin{cases} \min f(x) \\ \text{s.t. } h(x) = 0 \end{cases} \longrightarrow \min \underbrace{F(x_R)}_{f(\phi(x_R), x_R)}$$

$\Rightarrow x_R^*$ is a local min for $F(x_R)$.

FOC: $0 = \nabla F(x_R^*) = \nabla\phi(x_R^*)\nabla_B f(x_*)$

$+ \nabla_R f(x_*)$

(**)

$R^T = \nabla_R h(x_*)$, $B^T = \nabla_B h(x_*) \Rightarrow$

Same equations as if we are solving

$\min f(x)$ s.t. $\underbrace{A}_{[B \,\vdots\, R]} x = b$ (Compare to equations ①, ②)

Lagrangian function $L : \mathbb{R}^{n+m} \longrightarrow \mathbb{R}$

$$L(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x)$$

$\Rightarrow$ feasibility + FOC + SOC

$\nabla_\lambda L(x_*, \lambda_*) = 0$ $\quad$ $\nabla_x L(x_*, \lambda_*) = 0$

$\Delta x^T \nabla^2_{xx} L(x_*, \lambda_*) \Delta x \geq 0 \quad \forall \Delta x \in V(x_*)$

Note: $\left\{ \begin{array}{l} \nabla_\lambda L(x_*, \lambda_*) = 0 \\ \nabla_x L(x_*, \lambda_*) = 0 \end{array} \right. \Rightarrow (x_*, \lambda_*)$ is a stationary point for $L(x, \lambda)$. If $m \geq 1$, (x_*, λ_*) is not a local min but a saddle point.

$$\Rightarrow \begin{cases} \nabla_x L(x_*, \lambda_*) = 0 \\ \nabla_\lambda L(x_*, \lambda_*) = 0 \end{cases} \Rightarrow \quad n+m \text{ equations}$$

and $n+m$ unknowns

SOC sufficient : Consider a point x_* and

assume $\exists\ \lambda_*$ s.t.

$$\begin{cases} \nabla_x L(x_*, \lambda_*) = 0 \\ \nabla_\lambda L(x_*, \lambda_*) = 0 \\ \Delta x^T \nabla^2_{xx} L(x_*, \lambda_*) \Delta x > 0 \quad \forall\ \Delta x : \ \Delta x \neq 0,\ \Delta x \in V(x_*) \end{cases}$$

Then, $\exists\ \gamma > 0$, $\varepsilon > 0$ s.t.

$$f(x) \geq \underbrace{f(x_*) + \frac{\gamma}{2} \| x - x_* \|^2}_{\text{quadratic lower bound in } x} \quad \forall x : \ h(x) = 0,\ \| x - x_* \| < \varepsilon$$

---

1 - SOC implies x_* is a strict (isolated)

Local min $\Rightarrow$ Local strong Convexity

2 - It doesn't require x_* to be a regular point.

Proof needs a powerful lemma.

Lemma: Let P and Q be $n \times n$ symmetric matrices s.t.

1 - $Q \succeq 0$

2 - P is positive definite on the null space of Q, i.e. $x^T P x > 0 \quad \forall x : Qx = 0$

$\Rightarrow \exists \bar{c}$ s.t. $P + cQ \succ 0 \quad \forall c > \bar{c}$

Proof of Lemma: By contradiction, assume

$\exists \{c_k\}_{k=1}^{\infty}$ s.t. $c_1 < c_2 < \dots, \quad c_k \to \infty$

and $P + c_k Q \not\succ 0$

$\Rightarrow \exists x^{(k)} : \| x^{(k)} \| = 1$ and $(x^{(k)})^T (P + c_k Q) x^{(k)} \leq 0$

Since $x^{(k)}$ is bounded, $\{x^{(k)}\}$ has a convergent subsequence K.

$$\Rightarrow \{x^{(k)}\}_{k \in K} \longrightarrow \bar{x}$$

Recall : $(x^{(k)})^T (P + c_k Q) x^{(k)} \leq 0 \quad \forall k \in K$

Take the Limit

$$\underbrace{\bar{x}^T P \bar{x}}_{\text{Constant}} + \underbrace{\limsup_{\substack{k \to \infty \\ k \in K}} c_k}_{\text{goes to } +\infty} \times \underbrace{\bar{x}^T Q \bar{x}}_{\text{Constant is nonnegative since } Q \succeq 0} \leq 0$$

$$\Rightarrow \bar{x}^T Q \bar{x} = 0 \Rightarrow \bar{x}^T P \bar{x} \leq 0$$

This is a contradiction.

---

proof of SOC (sufficient) :

augmented Lagrangian

$$L_c(x, \lambda) = f(x) + \underbrace{\lambda^T h(x)}_{\sum_{i=1}^{m} \lambda_i h_i(x)} + \frac{c}{2} \|h(x)\|^2$$

$c = 0 \Rightarrow L_0(x, \lambda) = L(x, \lambda)$

observe: augmented Lagrangian is the same as Lagrangian for a new problem:

$$\min \; f(x) + \underbrace{\frac{c}{2} \| h(x) \|^2}_{\text{penalty}}$$

$$\text{s.t.} \quad h(x) = 0$$

penalty $= 0$ over the feasible set

$\Rightarrow$ It won't change local minima of the original problem.

$$\nabla_x L_c(x_*, \lambda_*) = \nabla_x L(x_*, \lambda_*) + c \nabla h(x_*) \cancelto{0}{h(x_*)}$$

$$= \nabla_x L(x_*, \lambda_*) = 0 \longrightarrow \text{by assumption}$$

$$\nabla^2_{xx} L_c(x_*, \lambda_*) = \nabla^2_{xx} L(x_*, \lambda_*)$$

$$+ c \nabla h(x_*) \nabla h(x_*)^T$$

$$+ c \sum_{i=1}^{m} \cancelto{0}{h_i(x_*)} \nabla^2 h_i(x_*)$$

$$= \underbrace{\nabla^2_{xx} L(x_*, \lambda_*)}_{P} + c \underbrace{\nabla h(x_*) \nabla h(x_*)^T}_{Q}$$

$Q \succeq 0$ and by SOC: $\Delta x^T P \Delta x > 0$ if $Q \Delta x = 0, \Delta x \neq 0$

By Lemma : $P + cQ \succ 0$ if c = Large

$$\Rightarrow \begin{cases} \nabla_x L_c(x_*, \lambda_*) = 0 \\ \nabla^2_{xx} L_c(x_*, \lambda_*) \succ 0 \end{cases} \quad (*)$$

$\Rightarrow$ x_* is a strict local min for

$$\min_{x \in \mathbb{R}^n} L_c(x, \lambda_*)$$

more precisely : (due to $(*)$)

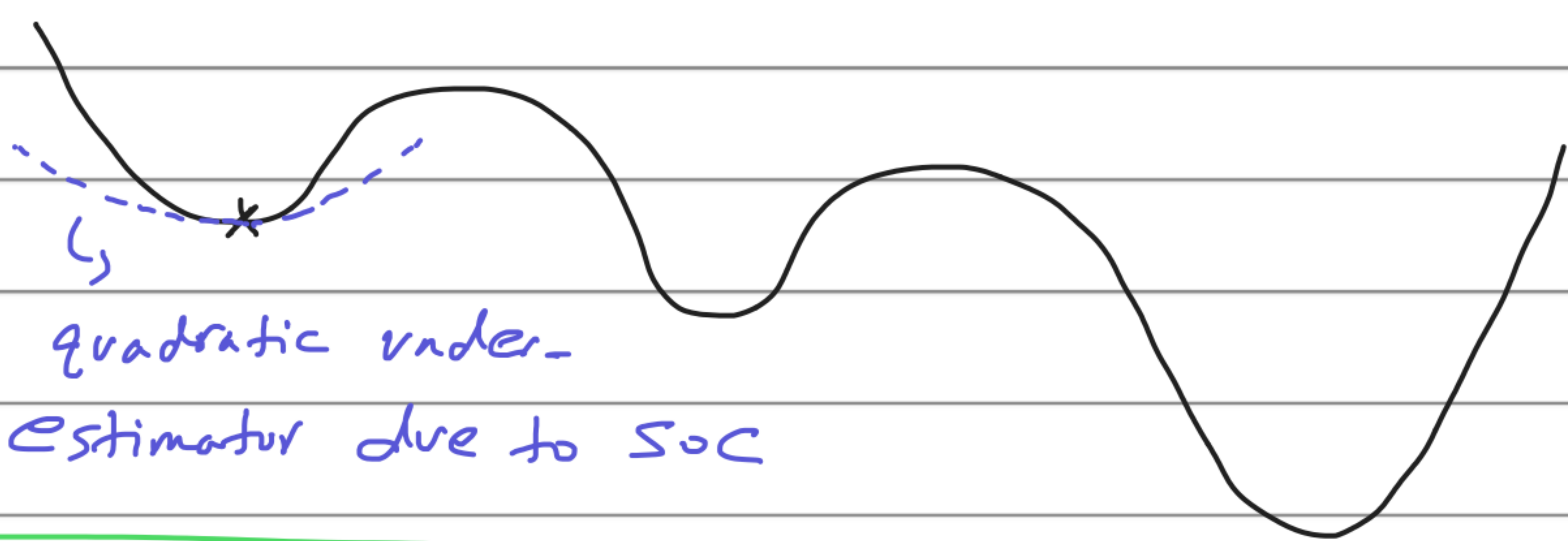
$\exists\ \gamma > 0,\ \varepsilon > 0$ s.t.

$$L_c(x, \lambda_*) \geq L_c(x_*, \lambda_*) + \frac{\gamma}{2} \|x - x_*\|^2$$

$\forall x : \|x - x_*\| < \varepsilon$

$(*)$

Unconstrained opt



If $h(x) = 0 \Rightarrow L_c(x, \lambda_*) = f(x) + \lambda_*^T \cancel{h(x)}^{0}$

$+ \frac{c}{2} \| \cancel{h(x)}^{0} \|^2 = f(x)$

$\circledast \Rightarrow f(x) \geq f(x_*) + \frac{\gamma}{2} \| x - x_* \|^2$

$\forall x : \| x - x_* \| < \varepsilon,$
$h(x) = 0$

$$\left\{ \begin{array}{ll} \min & f(x) \\ \text{s.t.} & h(x) = 0 \end{array} \right. \Longrightarrow \min \ f(x) + \lambda_*^T h(x) + \frac{c}{2} \| h(x) \|^2$$

Constrained — unconstrained

If c is large, then the penalty $\| h(x) \|^2$ is exact

and we correctly solve the original problem.

Challenge : Don't know λ_* in advance.

what if we approximate λ_* ?

$\Rightarrow$ Algorithm: augmented Lagrangian method
or method of multipliers