① Penalty ② Barrier ③ ALM – Augmented Lagrangian Method

CHAPTER 7

① Penalty.

$\downarrow$

quadratic penalty func $Q(x;\mu) := f(x) + \frac{1}{2\mu}\sum c_i^2(x)$

$$\boxed{\nabla Q = \nabla f + \frac{1}{2\mu}\sum 2 c_i(x)\nabla c_i(x)}$$

$\nabla_x(f(x)\vec{\alpha})_x = \nabla f \cdot \alpha^T$

① penalize infeasibility.

② we want $\mu \to 0$

问题: $\boxed{\begin{array}{l}\min f(x) \\ \text{st} \quad c_i(x)=0 \quad i\in E\end{array}}$

$$\nabla^2 Q = \nabla^2 f + \frac{1}{\mu}\sum c_i(x)\nabla^2 c_i(x) + \frac{1}{\mu}\underbrace{\sum \nabla c_i(x)\nabla c_i(x)^T}_{A(x)A^T(x)}$$

Alg:
① $\mu_0 > 0$, $\tau_0 > 0$ (→ tolerance), x_0^s
② $x_k \approx \arg\min Q(x;\mu_k)$
   ↓ How Approximately?
   $\|\nabla Q(x;\mu_k)\| < \tau_k$ → start at x_k^s, end at x_k
③ choose $\mu_{k+1} \in (0,\mu_k)$ & τ_{k+1}
   $\mu_{k+1} = \rho\mu_k$ $\quad \tau_k \to 0$

$x_{k+1}^s \longrightarrow x_k$

(4) Terminate (convergence test)

$$\|\nabla f(x_k) + \sum_i \lambda_i^k \nabla c_i(x_k)\| \le \varepsilon \qquad \lambda_i^k = \frac{c_i(x_k)}{\mu_k}$$

measure the extent of KKT !

Recall: Regular Point (LICQ)

x → regular point

⇔ satisfy LICQ ⇔ $\{\nabla c_i(x) : i\in I\} \leftarrow \{c_i(x)=0, i\in I\}$

Two version of Convergence { weak, strong }

① Strong version → $x_k = \arg\min Q(x;\mu_k)$ → exact minimizer

→ $\mu_k \to 0$

⇒ $\{x_k\}$ limit points is a soln.

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

$\bar{x}$ is a soln $\Leftrightarrow$ $f(\bar{x}) \le f(x)$ for all x s.t. $c_i(x)=0$ holds

x_k is exact minimizer $\Rightarrow$ $\underset{= f(x_k) + \frac{1}{2\mu_k}\sum c_i^2(x_k)}{Q(x_k;\mu_k)} \le Q(\bar{x};\mu_k) = f(\bar{x})$

$\Rightarrow \underline{f(x_k) + \frac{1}{2\mu_k}\sum c_i^2(x_k) \le f(\bar{x})}$

$\sum c_i^2(x_k) \le 2\mu_k (f(\bar{x}) - f(x_k))$

$x^* \to$ a limit point of $\{x_k\}$

$$\sum_{i\in E} c_i^2(x^*) = \sum_{i\in E} \lim_{k'\to\infty} c_i^2(x_{k'}) = \lim_{k'\to\infty} \sum_{i\in E} c_i^2(x_{k'}) \le \lim_{k'\to\infty} \underset{\downarrow 0}{2\mu_k} \underbrace{(f(\bar{x}) - f(x_{k'}))}_{\text{Bounded}} = 0$$

$f(x_{k'}) \to f(x^*)$

$\Rightarrow x^*$ satisfy equality

$\Rightarrow \underline{f(x^*) \le f(\bar{x})}$

$\Rightarrow \underline{x^* \text{ is a minimizer}}$

$x_k \to$ approximate minimizer!

② Weak Version $\to$ useful $\Rightarrow$ need tolerance and terminate principle

Thm: $\tau_k \to 0$, $\mu_k \to 0$, $x^* \to$ limit point of $\{x_k\}$ (holds LICQ)

$\Rightarrow x^*$ is a KKT Point $\to$ difficult Part is feasibility of x^*

and $\boxed{\lambda_i^* = \lim_{k\in K} \frac{c_i(x_k)}{\mu_k}}$

Pf: $\underline{Q(x;\mu_k)}$ $\quad \nabla_x Q(x;\mu_k) = \nabla f(x) + \sum_{i\in E} \frac{c_i(x)}{\mu_k} \nabla c_i(x)$

$\nabla_x Q(x_k;\mu_k) = \nabla f(x_k) + \sum_{i\in E} \frac{c_i(x_k)}{\mu_k} \nabla c_i(x_k)$

$A(x^*) = (\nabla c_1(x^*), \dots \nabla c_k(x^*))$

$= \nabla f(x_k) + A(x_k)\cdot\lambda^k$

$\Rightarrow -A(x_k)\lambda^k = \nabla f(x_k) - \nabla_x Q(x_k;\mu_k)$

$\Rightarrow \lambda^k = -[A^T A]^{-1} A^T (\nabla f(x_k) - \underset{\downarrow 0 \text{ since } \tau_k \to 0}{\underline{\nabla_x Q(x_k;\mu_k)}})$

$\Rightarrow k \to \infty$

$\lambda^* \leftarrow \lambda^k = -[A(x^*)^T A(x^*)]^{-1} A(x^*) \nabla f(x^*)$

$\lambda^k = \begin{pmatrix} \frac{c_1(x_k)}{\mu_1} \\ \vdots \\ \frac{c_i(x_k)}{\mu_i} \end{pmatrix} \to \lambda^*$

this part need $\boxed{\mu_k \to 0}$

guarantee the feasibility.

$\Rightarrow c_i(x^*) = 0 \Rightarrow \underline{x^* \text{ feasible}}$

Then

$\nabla_x Q(x_k;\mu_k) = \nabla f(x_k) + A(x_k)\cdot\lambda^k \quad k\to\infty$

$\Rightarrow 0 = \nabla f(x^*) + A(x^*)\lambda^* \Rightarrow (\lambda^*, x^*)$ KKT

矩阵求导 $\underline{\sum c_i(x_k) \nabla c_i(x_k)}$

$= ① \sum c_i(x) \nabla c_i(x_k) +$

② $\sum \nabla c_i(x_k) \nabla c_i(x_k)^T$

$\boxed{(\sum f_i(x)\cdot a_i)' = \sum \nabla f_i(x)\, a_i^T}$

$= \nabla f \cdot a^T$

$$Q(x_k;\mu_k) = f(x_k) + \frac{1}{2\mu_k}\sum_{i\in E} c_i^2(x_k)$$

$$\nabla_x Q(x_k;\mu_k) = \nabla f(x_k) + \sum_{i\in E} \frac{c_i(x_k)}{\mu_k} \nabla c_i(x_k)$$

$$= \nabla f(x_k) + A(x_k)\cdot\lambda^k$$

$$\nabla^2_{xx} Q(x_k;\mu_k) = \underline{\nabla^2 f(x_k)} + \frac{1}{\mu_k} A(x_k) A(x_k)^T + \underline{\sum \frac{c_i(x_k)}{\mu_k} \nabla^2 c_i(x_k)}$$

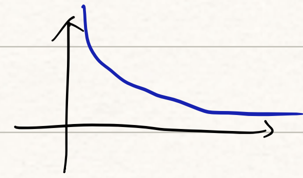
$$= \nabla^2_{xx} L(x_k, \lambda^k) + \boxed{\frac{1}{\mu_k} A(x_k) A(x_k)^T}$$

ill-condition

Barrier function Method $\Rightarrow$ for inequality

问题: $\min f(x)$

$\text{s.t.} \quad c_i(x) \geq 0 \quad i \in I$



Barrier func:

$$B(x) = \sum_{i \in I} \phi(-c_i(x)) \qquad \phi(\cdot): \mathbb{R}^+ \to \mathbb{R}^+$$

$$\phi'(y) < 0 \qquad \lim_{y \to 0^+} \phi(y) = \infty$$

$\phi = -\log(\cdot)$ $\qquad P(x;\mu) = f(x) + \mu B(x)$

$\theta(\mu) = \inf\{P(x;\mu) : x \in F^<\}$

$\boxed{\mu \downarrow \Rightarrow \text{允许 } c_i(x) \to 0}$

Alg: ① μ_0, τ_0 x_0^s

② $x_k \approx \arg\min P(x;\mu_k)$

$x_k^s \to \cdots \to x_k$ $\qquad \|\nabla P(x;\mu_k)\| \leq \tau_k$

$x_{k+1}^s \to \cdots \to x_{k+1}$

③ $\mu_{k+1} \in (0, \mu_k)$, τ_{k+1}

Convergence:
- ① Strong $\to$ Something different
- ② weak $\to$ similar to Penalty Method.

① Strong version:

i) $\min f(x)$ $\quad$ has a soln $\bar{x}$ s.t. $\boxed{F^< \cap N \neq \emptyset}$ for $\forall N$

$\text{s.t.} \quad c_i(x) \geq 0$

$\Rightarrow \underbrace{\lim_{\mu \to 0^+} \theta(\mu) = f^*}$ $\Rightarrow$ sub-problem leads to the CORRECT Obj.

Pf: $\theta(\mu) = \inf\{ f(x) + \mu \phi(-c_i(x)) : x \in F^< \}$

① $\geq \inf\{ f(x) : x \in F^\leq \}$

$= f^*$

② $\theta(\mu) \leq f(\hat{x}) + \mu \phi(-c_i(\hat{x}))$ $\quad \underline{\hat{x} \in F^<} \sim x^*$

$\leq f(x^*) + \varepsilon + \mu \sum \phi(-c_i(\hat{x}))$

#

ii) $x_\mu = \text{argmin}\{ P(x; \mu) : x \in F^< \} \in F^<$ $\Rightarrow$ $\boxed{\text{exact minimizer}}$

$\Rightarrow$ ① $\mu B(x_\mu) \Rightarrow 0$ ② $x_\mu \rightarrow$ opt soln x^*

$P(x_\mu; \mu) = f(x_\mu) + \mu B(x_\mu)$

$f(\bar{x}) \leq f(x_\mu) \leq f(x_\mu) + \mu B(x_\mu) = \theta(\mu) \to f(\bar{x})$

$\Rightarrow$ ① $\mu B(x_\mu) \to 0$

② $f(x_\mu) \to f(x^*) = f(x)$ $\quad \} \Rightarrow$ opt

& $c(x_\mu) \leq 0 \Rightarrow c(x^*) \leq 0$

② Weak Version

$\{x_k\}$ $\quad t_k \to 0.$ $\quad x_k \to \bar{x}$ which LICQ holds

$\Rightarrow \bar{x}$ KKT with $\lambda_i = \lim_{k \in K} \mu_k \phi'(-c_i(x_k))$

Pf: $P(x_k; \mu_k) = f(x_k) + \mu_k \sum_{i \in E} \phi(-c_i(x_k))$

$\nabla_x P(x_k; \mu_k) = \nabla f(x_k) \mp \mu_k \sum_{i \in E} \phi'(-c_i(x_k)) \nabla c_i(x_k)$

$= \nabla f(x_k) + \sum_{i \in E} \underbrace{-\mu_k \phi'(-c_i(x_k))} \nabla c_i(x_k)$

$x_{k_i} \to x^*$ $\quad x^* \to$ LICQ $\quad E \begin{cases} A \Rightarrow c_i(x^*) = 0 \\ E \setminus A \Rightarrow c_i(x^*) < 0 \Rightarrow \mu_k \phi' \to 0 \end{cases}$

$\nabla_x P = \nabla f(x_k) + \sum_{i \in A} \boxed{} \nabla c_i(x^*) + \sum_{i \in E \setminus A}$

$= \nabla f(x_k) + A(x_k) \cdot \lambda^k + \cdots$

$\Rightarrow A(x_k) \lambda^k = -\nabla f(x_k) + \cdots$

$\Rightarrow \quad \lambda^k = -(A^T A)^{-1} A^T \nabla f(x_k) + \cdots \quad k \to \infty$

$\lambda^* = -(A^{*T} A^*)^{-1} A^{*T} \nabla f(x^*) \Rightarrow$ 极限存在.

Augmented Lagrangian.

$$\begin{cases} \min_x f(x) \\ \text{s.t. } c_i(x)=0 \quad i\in E \end{cases} \quad (1) \qquad (\bar{x}, \bar{\lambda})$$

$$L(x;\lambda) = f(x) + \sum \lambda_i c_i(x)$$

$$P(x;\mu) = f(x) + \frac{1}{2\mu}\sum c_i^2(x)$$

$$L_A(x;\lambda,\mu) = f(x) + \sum_{i\in E}\lambda_i c_i(x) + \frac{1}{2\mu}\sum_{i\in E} c_i^2(x)$$

Our aim is to : want x^* to be a KKT Point of $(1) \to (\bar{x}, \bar{\lambda})$

$\Leftrightarrow$ $\bar{x} \to$ regular Point. $c(\bar{x})=0$

$\boxed{\nabla f(\bar{x}) + \sum \bar{\lambda}_i c_i(\bar{x}) = 0} \to$ desired condition!

Consider $\nabla P_x(x;\mu) = \nabla f(x) + \sum \frac{c_i(x)}{\mu}\nabla c_i(x)$

In order to guarantee this, we must want $\mu \to 0$!

Consider $\nabla L_A(x;\lambda,\mu) = \nabla f(x) + \sum \frac{c_i(x)}{\mu}\nabla c_i(x) + \sum \lambda_i \nabla c_i(x)$

$$= \nabla f(x) + \sum\left(\frac{c_i(x)}{\mu} + \lambda_i\right)\nabla c_i(x)$$

Then we want $\boxed{\begin{array}{l}\mu \text{ to be positive} \\ \lambda_i \to \bar{\lambda}_i\end{array}} \to$ More well-condition.

Algorithm of Augmented Lagrangian Method

① μ_0, t_0, x_0^s, λ_0

② $x^k \approx \arg\min L_A(x;\lambda_k,\mu_k)$

i.e. start from x_k^s, find x_k s.t $\|\nabla_x L_A(x;\lambda_k,\mu_k)\| \le t_k$.

③ final convergence test.

④ set $\lambda_i^{k+1} = \lambda_i^k + \frac{c_i(x_k)}{\mu_k}$

⑤ Update μ_0, τ_0, x_0^s, λ_0

Strong Version Convergence. (threshold $\bar{\mu}$)

Thm. $\bar{x} \to$ local sln of (1) and $\bar{x}$ is a regular point with $\bar{\lambda}$

and $(\bar{x}, \bar{\lambda})$ satisfy second order sufficient cond

then exist $\bar{\mu} > 0$, s.t. $\mu \in (0, \bar{\mu}) \Rightarrow \bar{x} = \arg\min_x L_A(x, \bar{\lambda}, \mu)$.

$\boxed{\text{we don't know!}}$

Pf: ① $\nabla L(\bar{x}; \bar{\lambda}) = 0$

$\Rightarrow \nabla f(\bar{x}) + \sum \bar{\lambda}_i \nabla c_i(\bar{x}) = 0$

② $\nabla_x L_A(\bar{x}; \mu, \bar{\lambda})$

$= \nabla f(\bar{x}) + \sum \bar{\lambda}_i \nabla c_i(\bar{x}) + \sum \frac{c_i(\bar{x})}{\mu} \nabla c_i(\bar{x}) = 0$

③ $\nabla_{xx}^2 L_A(\bar{x}; \mu, \bar{\lambda})$

$= \nabla^2 f(x) + \sum \bar{\lambda}_i \nabla^2 c_i(x) + \cancel{\sum \frac{c_i(\bar{x})}{\mu} \nabla^2 c_i(\bar{x})} + \frac{1}{\mu} \sum \nabla c_i(x) \nabla c_i(x)^T$

$= \underbrace{\nabla_{xx}^2 L(\bar{x}; \bar{\lambda})}_{\text{partial PD}} + \frac{1}{\mu} \sum \underbrace{\nabla c_i(\bar{x}) \nabla c_i(\bar{x})^T}_{\text{PD semi}}$

if $\exists \bar{\mu} > 0$, s.t. $\nabla_{xx}^2 L_A(\bar{x}; \bar{\lambda}, \mu) > 0$ for $0 < \mu < \bar{\mu}$ ✓

If not! $\forall \bar{\mu}$ $\nabla_{xx}^2 L_A(\bar{x}; \bar{\lambda}, \mu)$ is not PD

$\Rightarrow$ choose $\bar{\mu}_k = \frac{1}{k} \Rightarrow \mu_k \to 0$. s.t

$d_k^T \nabla_{xx}^2 L_A(\bar{x}; \bar{\lambda}, \mu_k) d_k \le 0$

$\Rightarrow d^T \left(\nabla_{xx}^2 L(\bar{x}, \bar{\lambda}) + \frac{1}{\mu_k} \sum \bigcirc \bigcirc \right) d^T \le 0$

$\Rightarrow d^T \nabla_{xx}^2 L d + \frac{1}{\mu_k} \left(\sum \nabla c_i(\bar{x})^T d\right)^2 \le 0$

$\Rightarrow d \in C \Rightarrow d^T \nabla_{xx}^2 L d \le 0$ 矛盾!

Weak Version

$(\bar{x}, \bar{\lambda}) \Rightarrow$ regular point satisfy second order sufficient cond.

$\Downarrow$

$\bar{\mu} \to$ threshold.

① for $\|\lambda^k - \bar{\lambda}\| \le \frac{\delta}{\mu_k}$ $\to$ sub prob $\min_x L_A(x, \lambda^k, \mu_k)$

s.t. $\|x - \bar{x}\| < \varepsilon$

$\Rightarrow \|x_k - \bar{x}\| \le M \mu_k \|\lambda_k - \bar{\lambda}\|$

② $\|\lambda^{k+1} - \bar{\lambda}\| \leq \mu\mu_k \|\lambda^k \cdot \bar{\lambda}\|$

③ $\|\lambda^{k+1} - \bar{\lambda}\| \leq \mu\mu_k \|\lambda^k \cdot \lambda\| \Rightarrow \{\lambda^k\} \to$ Q- linearly

$\{x^k\} \to$ R- linearly