MATH 450/558
FALL 2015

LECTURE 8

PENALTY, BARRIER ~~AND AUGMENTED LAGRANGIAN~~ METHODS

Penalty Method

$$\text{(NEP)} \quad \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x)$$
$$c_j(x) = 0 \quad j \in E$$

Penalty function

$$P(x;\mu) := f(x) + \frac{1}{2\mu} \sum_{j \in E} [c_j(x)]^2$$

μ: penalty parameter > 0

Idea

Solve

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad P(x;\mu)$$

repeatedly for various μ, gradually decreasing μ to 0.

Ex

minimize x_2

$x_1^2 - x_2 = 4$

global minimizer $x_* = (0, -4)$

smallest value of objective $f(x_*) = -4$

Penalty function

$$P(x;\mu) := x_2 + \frac{1}{2\mu}(x_1^2 - x_2 - 4)^2$$

$$\nabla_x P(x;\mu) = \begin{bmatrix} 2x_1(x_1^2 - x_2 - 4)/\mu \\ 1 - (x_1^2 - x_2 - 4)/\mu \end{bmatrix}$$

$$\nabla^2_{xx} P(x;\mu) = \begin{bmatrix} \frac{2(x_1^2 - x_2 - 4)}{\mu} + \frac{4x_1^2(x_1^2 - x_2 - 4)}{\mu} & -\frac{2x_1}{\mu} \\ -\frac{2x_1}{\mu} & \frac{1}{\mu} \end{bmatrix}$$

x_* such that $\nabla_x P(x_*;\mu) = 0$

$$x_*(\mu) = (0, -4-\mu).$$

Furthermore,

$$\nabla^2_{xx} P(x_*(\mu);\mu) = \begin{bmatrix} 2 & 0 \\ 0 & 1/\mu \end{bmatrix}$$

is PD. Thus

(i) $x_*(\mu)$ is a local min of $P(x;\mu)$,

(ii) $\lim_{\mu \to 0^+} x_*(\mu) = (0, -4) = x_*$

Pseudocode

Given $x_s^{(0)}$ and $\mu_1 > 0$.

for $k = 1, 2, 3, \ldots$

(1) Choose a $\tau_k > 0$ $\left(\text{such that } \lim_{k \to \infty} \tau_k = 0 \right)$

(2) Find $x_s^{(k)}$ such that

$$\| \nabla_x P(x_s^{(k)}; \mu_k) \| \leq \tau_k$$

by a (modified) Newton method with a line search starting from $x^{(0)} := x_s^{(k-1)}$.

(3) Choose a $\mu_{k+1} \in (0, \mu_k)$ $\left(\text{such that } \lim_{k \to \infty} \mu_k = 0 \right)$

end

THM

Let x_* be any limit point of $\{x_s^{(k)}\}$ such that LICQ holds at x_*. There exists λ_* such that (x_*, λ_*) satisfies KKT conditions for (NEP).

PROOF

Notice that

$$\nabla_x P(x; \mu) = \nabla f(x) + \frac{1}{\mu} \sum_{j \in E} c_j(x) \nabla c_j(x)$$

Exploiting $\|\nabla_x P(x_s^{(k)}; \mu_k)\| \leqslant \tau_k$

$$\left\| \nabla f(x_s^{(k)}) + \frac{1}{\mu_k} \sum_{j \in E} c_j(x_s^{(k)}) \nabla c_j(x_s^{(k)}) \right\| \leqslant \tau_k$$

$$\Longrightarrow \left\| \sum_{j \in E} c_j(x_s^{(k)}) \nabla c_j(x_s^{(k)}) \right\| \leqslant \mu_k \left[\tau_k + \|\nabla f(x_s^{(k)})\| \right]$$

Taking the limit as $k \to \infty$,

$$\sum_{j \in E} c_j(x_*) \nabla c_j(x_*) = 0.$$

Due to LICQ $c_j(x_*) = 0$ for each $j \in E$.

Next letting $\lambda^{(k)}$ such that

$$\lambda_j^{(k)} := -c_j(x_s^{(k)}) / \mu_k .$$

We claim that $\lim\limits_{k \to \infty} \lambda^{(k)} = \lambda_*$ exists.

To see this, note

$$\nabla_x P(x_s^{(k)}, \mu_k) = \nabla f(x_s^{(k)}) + \underbrace{\frac{1}{\mu_k} \sum_{j \in E} c_j(x_s^{(k)}) \nabla c_j(x_s^{(k)})}_{-J_E(x_s^{(k)})^T \lambda^{(k)}}$$

$$\Longrightarrow J_E(x_s^{(k)})^T \lambda^{(k)} = \nabla f(x_s^{(k)}) - \nabla_x P(x_s^{(k)}, \mu_k)$$

$$\Longrightarrow \lambda^{(k)} = \left[J_E(x_s^{(k)}) J_E(x_s^{(k)})^T \right]^{-1} J_E(x_s^{(k)}) \left[\nabla f(x_s^{(k)}) - \nabla_x P(x_s^{(k)}, \mu_k) \right]$$

$$\Longrightarrow \lim_{k \to \infty} \lambda^{(k)} = \left[J_E(x_*) J_E(x_*)^T \right]^{-1} J_E(x_*) \nabla f(x_*)$$

Consequently,

$$\left\| \nabla f(x_s^{(k)}) - \sum_{j \in E} \lambda_j^{(k)} \nabla c_j(x_s^{(k)}) \right\| \leq \tau_k$$

implying

$$\nabla f(x_s^{(k)}) - \sum_{j \in E} \lambda_j^{(k)} \nabla c_j(x_s^{(k)}) \to 0$$

as $k \to \infty$

$$\Longrightarrow \quad \nabla f(x_*) - \sum_{j \in E} [\lambda_*]_j \nabla c_j(x_*) = 0 \quad \square$$

Calculation of Newton step

$$\nabla^2_{xx} P(x;\mu)\, p = -\nabla_x P(x;\mu)$$

where

$$\nabla_x P(x;\mu) = \nabla f(x) + \frac{1}{\mu} \sum_{j \in E} c_j(x) \nabla c_j(x)$$

$$\nabla^2_{xx} P(x;\mu) = [\nabla f(x)]' + \frac{1}{\mu} \sum_{j \in E} [\nabla c_j(x)]' c_j(x) + \frac{1}{\mu} \sum_{j \in E} \nabla c_j(x) [c_j(x)]'$$

$$= \nabla^2 f(x) + \overset{①}{\boxed{\frac{1}{\mu} \sum_{j \in E} c_j(x) \nabla^2 c_j(x)}} + \overset{②}{\boxed{\frac{1}{\mu} \underbrace{\sum_{j \in E} \nabla c_j(x) \nabla c_j(x)^T}_{J_E(x)^T J_E(x)}}}$$

① – Well-conditioned $\dfrac{-c_j(x)}{\mu} \to \lambda_*$

② – ill-conditioned

Norm $\longrightarrow \infty$ as $\mu \to 0$

Singular if $|E| < n$

To avoid ill-conditioning due to 2

$$\underset{m \times 1}{q} := \frac{1}{\mu} \underset{m \times n}{J_E(x)} \underset{n \times 1}{p}$$

Solve

$$\underset{(m+n) \times (m+n)}{\begin{bmatrix} \nabla^2 f(x) + \frac{1}{\mu} \sum_{j \in E} c_j(x) \nabla^2 c_j(x) & J_E(x)^T \\ J_E(x) & -\mu I_m \end{bmatrix}} \begin{bmatrix} p \\ q \end{bmatrix} = \begin{bmatrix} -\nabla_x P(x; \mu) \\ 0 \end{bmatrix}$$

for p and q.

Conditioning of a linear system

$$\underset{\substack{n\times n \\ \text{invertible} \\ \text{given}}}{A} x = \underset{\substack{n\times 1 \\ \text{given}}}{b}$$

Perturbed system

$$(A+\Delta A)(x+\Delta x) = b$$

$$\cancel{Ax} + \Delta A \cdot x + A\,\Delta x + \Delta A \cdot \Delta x = \cancel{b}$$

$$\Delta x = (A+\Delta A)^{-1}\,\Delta A\, x$$

Now ~~consider~~

$$\kappa := \limsup_{\Delta A \to 0} \frac{\|\Delta x\| / \|x\|}{\|\Delta A\| / \|A\|}$$

$$= \limsup_{\Delta A \to 0} \frac{\|(A+\Delta A)^{-1}\| \,\cancel{\|\Delta A\|}\, \cancel{\|x\|} / \cancel{\|x\|}}{\cancel{\|\Delta A\|} / \|A\|}$$

$$= \|A^{-1}\|\,\|A\|$$

$$\overset{\text{if } A \text{ is symmetric}}{=} \lambda_{max}(A)\ \lambda_{max}(A^{-1})$$

$$= \lambda_{max}(A) / \lambda_{min}(A)$$

Logarithmic Barrier Method

$$\underset{\substack{x \in \mathbb{R}^n \\ c_j(x) \geq 0 \quad j \in I}}{\text{minimize}} \quad f(x)$$

Logarithmic Barrier function

$$L(x;\mu) = f(x) - \mu \sum_{j \in I} \ln c_j(x)$$

μ ↓ Barrier parameter

Idea

Start with a feasible x. Solve

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad L(x;\mu)$$

repeatedly, letting $\mu \to 0$.

Ex

$$\underset{\substack{x \in \mathbb{R}^2 \\ x_1 + x_2 \geq 1}}{\text{minimize}} \quad x_1^2 + x_2^2$$

$x_* = (1/2, 1/2)$

optimal objective

$f_* = 1/2$

$$L(x;\mu) = x_1^2 + x_2^2 - \mu \ln(x_1 + x_2 - 1)$$

$$\nabla_x L(x;\mu) = \begin{bmatrix} 2x_1 - \mu/(x_1+x_2-1) \\ 2x_2 - \mu/(x_1+x_2-1) \end{bmatrix}$$

$$\nabla^2_{xx} L(x;\mu) = \begin{bmatrix} 2 + \dfrac{\mu}{(x_1+x_2-1)^2} & \dfrac{\mu}{(x_1+x_2-1)^2} \\ \dfrac{\mu}{(x_1+x_2-1)^2} & 2 + \dfrac{\mu}{(x_1+x_2-1)^2} \end{bmatrix}$$

An $x(\mu)$ such that $\nabla_x L(\overset{x(\mu)}{x};\mu) = 0$ is given by

$$x(\mu) = \left(\frac{1+\sqrt{1+4\mu}}{2}, \frac{1+\sqrt{1+4\mu}}{2} \right)$$

satisfies

$$\lim_{\mu \to 0^+} x(\mu) = \left(\tfrac{1}{2}, \tfrac{1}{2}\right) = x_*$$

Mixed Penalty Logarithmic Barrier Method

$$\begin{aligned} &\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x) \\ &c_j(x) = 0 \quad j \in E \\ &c_j(x) \geq 0 \quad j \in I \end{aligned}$$

equivalently

$$\underset{x,\,s}{\text{minimize}} \quad f(x)$$

$$c_j(x) = 0 \quad j \in E$$

$$c_j(x) - s_j = 0 \quad j \in I$$

slack variables

$$s_j \geq 0 \quad j \in I$$

mixed penalty logarithmic barrier function

$$M(x, s; \mu) = f(x) + \frac{1}{2\mu} \sum_{j \in E} \left[c_j(x)\right]^2 + \frac{1}{2\mu} \sum_{j \in I} \left[c_j(x) - s_j\right]^2 - \mu \sum_{j \in I} \ln s_j$$

Idea

Start with $s > 0$ and an arbitrary x. Solve

$$\underset{x,\,s}{\text{minimize}} \quad M(x, s; \mu)$$

for various μ, gradually decreasing μ to 0.

Connection with primal-dual methods

Let us consider

$$\text{(NIP)} \quad \begin{array}{ll} \underset{x \in \mathbb{R}^n}{\text{minimize}} & f(x) \\ & c_j(x) \geq 0 \quad j \in I \end{array}$$

and associated logarithmic-barrier function

$$L(x;\mu) = f(x) - \mu \sum_{j \in I} \ln c_j(x).$$

Consider $x(\mu)$ such that $\nabla_x L(x(\mu);\mu) = 0$.

$$\nabla f(x(\mu)) - \sum_{j \in I} \underbrace{\frac{\mu}{c_j(x(\mu))}}_{\lambda_j(\mu) :=} \nabla c_j(x(\mu)) = 0.$$

Notice that $(x(\mu), \lambda(\mu))$ satisfies

$$(\text{KKT}_\mu) \quad \begin{array}{lll} \text{(i)} & \nabla_x \mathcal{L}(x(\mu), \lambda(\mu)) = 0 & \\ \text{(ii)} & c_j(x(\mu)) > 0 & j \in I \\ \text{(iii)} & \lambda_j(\mu) > 0 & j \in I \\ \text{(iv)} & c_j(x(\mu))\, \lambda_j(\mu) = \mu & \end{array}$$

Primal central-path

$$\{(x(\mu)) \mid \mu > 0,\ x(\mu) \text{ satisfies } KKT_\mu \text{ for some } \lambda(\mu)\}$$

Logarithmic-barrier method is a primal central-path following algorithm.

<u>THM</u> (Convergence)

Suppose the strictly feasible region F° for (NIP) is not empty. Suppose also that x_* is a point such that

(i) ~~KKT conditions~~ second order sufficient conditions hold at x_* for some λ_*,

(ii) LICQ holds at x_*,

(iii) strict complementarity holds at x_*.

Then the following ~~is~~ are true.

(1) There exists a ball $B(x_*, \delta)$ such that (i) $L(x;\mu)$ has a unique local minimizer $x(\mu) \in B(x_*, \delta)$; (ii) $\lim_{\mu \to 0^+} x(\mu) = x_*$.

(2) Defining $\lambda(\mu)$ by $\lambda_i(\mu) := \mu / c_i(x(\mu))$
$\lim_{\mu \to 0^+} \lambda(\mu) = \lambda_*$.

Primal-dual interior point method framework

KKT conditions for (NIP) but complementarity condition replaced by centrality condition.

(i) $\nabla f(x) - \sum_{j \in I} \lambda_j \nabla c_j(x) = 0$

(ii) $c_j(x) \geqslant 0$

(iii) $\lambda_j \geqslant 0$

(iv) $c_j(x)\, \lambda_j = \mu$

equivalently

(i') $\nabla f(x) - \sum_{j \in I} \lambda_j \nabla c_j(x) = 0$

(ii') $c_j(x) - s_j = 0$

(iii') $\lambda_j \geqslant 0, \quad s_j \geqslant 0$

(iv') $s_j \lambda_j = \mu$

Solve

(i'), (ii') and (iv')

by means of a Newton method.

Impose $\lambda_j, s_j > 0$ in a line search.

Gradually decrease $\mu \to 0^+$.

Newton's method is applied to

$$F : \mathbb{R}^{n+2m} \longrightarrow \mathbb{R}^{n+2m}$$

$$F(x,\lambda,s) := \begin{bmatrix} \nabla f(x) - \sum_{j \in I} \lambda_j \nabla c_j(x) \\ c(x) - s \\ S\Lambda e - \mu e \end{bmatrix}$$

Newton step

$$F'(x,\lambda,s) \begin{bmatrix} \Delta x \\ \Delta \lambda \\ \Delta s \end{bmatrix} = -F(x,\lambda,s)$$

$$\begin{bmatrix} \nabla^2_{xx} \mathcal{L}(x,\lambda) & -J(x)^T & 0 \\ J(x) & 0 & -I \\ 0 & S & \Lambda \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta \lambda \\ \Delta s \end{bmatrix} = -F(x,\lambda,s)$$