# Ch 17: Penalty and Augmented Lagrangian Methods

## 17.1 The quadratic penalty method

Transform:

$$\min_x f(x) \quad \text{subject to:} \begin{cases} c_i(x) = 0, & i \in \mathcal{E} \\ c_i(x) \geq 0, & i \in \mathcal{I} \end{cases}$$

into:

$$Q(x;\mu) = f(x) + \frac{\mu}{2}\sum_{i\in\mathcal{E}} c_i^2(x) + \frac{\mu}{2}\sum_{i\in\mathcal{I}}\left([c_i(x)]^-\right)^2$$

with $\quad [y]^- = \max(-y, 0)$

Then make $\mu$ very large until convergence.

Problem:

* $Q(x;\mu)$ is not twice diff'ble.

For equality-constraints only:

Framework 17.1 (Quadratic Penalty Method)

Given $\mu_0 > 0$, $\{\tau_k \geq 0\}$, $\tau_k \to 0$, $x_0^s$ (starting pt)

For $k = 0, 1, 2, \ldots$

Minimize $Q(\cdot; \mu_k)$ to get $x_k$
and terminate when $\|\nabla_x Q(x; \mu_k)\| \leq \tau_k$
(unconstrained opt if smooth $Q(\cdot)$).

if final convergence satisfied
{ return $(x_k)$

Choose $\mu_{k+1} > \mu_k$

Choose $x_{k+1}^s$

end

Notes:

* $\mu_{k+1} = \begin{cases} 1.5\,\mu_k & \text{for expensive min } Q(\cdot;\mu_k) \\ 10\,\mu_k & \text{for cheap min } Q(\cdot;\mu_k) \end{cases}$

* increase $\mu_k$ if constraints not satisfied.

* For smooth $Q(\cdot;\mu_k)$, regular unconstr. opt will work.

* $\nabla_{xx}^2 Q$ is ill-cond $\Rightarrow$ Quasi-Newton + CG will not work

* Newton's method is okay for ill-conditioned Hessians but:
  - for large $\mu_k$, ill-cond $\overset{\text{can}}{\Longrightarrow}$ wrong $p_k$
  - Quadratic approx only for small regions

$\Rightarrow$ may have to set: $x^s_{k+1} = x_k$, $\mu_{k+1} = \mu_k + \delta$.

Thms 17.1, 17.2 require $\mu_k \uparrow \infty$

<u>Ill-conditioning and re-formulation</u>

Instead of solving:

$$\nabla^2_{xx} Q(x;\mu_k) = -\nabla_x Q(x;\mu_k)$$

use $\zeta = \mu A(x) p$ to get:

$$\begin{bmatrix} \nabla^2 f(x) + \sum_{i\in\mathcal{E}} \mu_k c_i(x) \nabla^2 c_i(x) & A(x)^T \\ A(x) & (-1/\mu_k) I \end{bmatrix} \begin{bmatrix} p \\ \zeta \end{bmatrix} = \begin{bmatrix} -\nabla_x Q(x;\mu_k) \\ 0 \end{bmatrix}$$

$\Rightarrow$ However, $\mu_k c_i(x)$ maybe bad approximation to $-\lambda_i^*$, the Lagrange multipliers!

# 17.2 Non-smooth penalty functions

Consider the $l_1$ penalty function:

$$\phi_1(x;\mu) = f(x) + \mu \sum_{i\in\mathcal{E}} |c_i(x)| + \mu \sum_{i\in\mathcal{I}} [c_i(x)]^-$$

which works for all $\mu > \mu^*$

where $\mu^* = \|\lambda^*\|_\infty = \max_{i\in\mathcal{E}\cup\mathcal{I}} |\lambda_i^*|$

(See thm 17.3)

A practical solution is simply:

$$\min_{p,r,s,t} \; q(p;\mu) = f(x) + \tfrac{1}{2}p^T W p + \nabla f(x)^T p + \mu \sum_{i\in\mathcal{E}} (r_i + s_i) + \mu \sum_{i\in\mathcal{I}} t_i$$

Subject to:

$$\begin{cases} \nabla c_i(x)^T p + c_i(x) = r_i - s_i, & i\in\mathcal{E} \\ \nabla c_i(x)^T p + c_i(x) \geq -t_i, & i\in\mathcal{I} \\ r, s, t \geq 0 \end{cases}$$

and plug this into framework 17.2

W is symmetric ...