§17 Penalty methods

$$(1) \quad \begin{array}{ll} \min & f(x) \\ \text{s.t.} & c_i(x) = 0 \quad i \in \mathcal{E} \\ & c_i(x) \geq 0 \quad i \in \mathcal{I} \end{array}$$

c continuous
f twice cont diff'ble

Solve (1) by solving a sequence of unconstrained problems with objective function:

$$P(\alpha, x) = f(x) + \alpha \sum_{i \in \mathcal{E}} c_i^2(x) + \alpha \sum_{i \in \mathcal{I}} \left(\min\{c_i(x), 0\}\right)^2$$

x feasible $\Rightarrow$ $P(\alpha, x) = f(x)$

x not feasible $\Rightarrow$ $P(\alpha, x) > f(x)$

the largest the α the more feasibility is prefered.

For α_k increasing sequence solve:

$$\min_{x \in \mathbb{R}^n} P(\alpha_k, x)$$

For simplicity consider equality constraints only:

$$P(\alpha, x) = f(x) + \alpha \|c(x)\|_2^2$$

More generally let

$r : \mathbb{R}^{|E|} \to \mathbb{R}$ be a function with:

- $r(y) \geq 0 \quad \forall y \in \mathbb{R}^{|E|}$
- $r(0) = 0$
- $r(y) > 0$ if $y \neq 0$
- r cont at 0

Thm Define $P(\alpha, x) = f(x) + \alpha\, r(c(x))$

Let $\{\alpha_k\}$ be monotonically strictly increasing with $\alpha_k \to \infty$.

Let $F = \{x \in \mathbb{R}^n \mid c(x) = 0\}$ be non empty and let $\{x_k\}$ be the sequence of global solutions of

$$\min_{x \in \mathbb{R}^n} P(\alpha_k, x)$$

Then the following holds:

1) $\{P(\alpha_k, x_k)\}$ is mon. increasing
2) $\{r(c(x_k))\}$ is mon. decreasing
3) $\{f(x_k)\}$ is mon. increasing
4) $\lim_{k \to \infty} c(x_k) = 0$
5) Every accumulation point of $\{x_k\}$ is a sol to (1).

$\left.\begin{array}{l}4)\\5)\end{array}\right\}$ important ones

Issues with this approach

1) need a global min of penalty function
2) Let $r(y) = \|y\|_2^2$

$$P(\alpha, x) = f(x) + \alpha \|c(x)\|_2^2$$

$$\nabla P(\alpha, x) = \nabla f(x) + 2\alpha\, c'(x)^T c(x)$$

$$\nabla^2 P(\alpha, x) = \nabla^2 f(x) + 2\alpha \sum_{i=1}^{|E|} c_i(x) \nabla^2 c_i(x) + 2\alpha\, c'(x)^T c'(x)$$

If $\alpha_k \to \infty$, $x_k \to x_*$ a min of (1) then optimality conditions for (1):

$$L(x,\lambda) = f(x) + \lambda^T c(x)$$

$$\nabla L(x_*, \lambda_*) = \nabla f(x_*) + c'(x_*)^T \lambda_* = 0$$

$$c(x_*) = 0$$

In the limit:

$$\boxed{2\alpha_k\, c(x_k) \to \lambda_*}$$

$$\nabla^2 P(\alpha_k, x_k) = \underbrace{\nabla^2 f(x_k) + \sum_{i=1}^{|E|} 2\alpha_k\, c_i(x_k) \nabla^2 c_i(x_k)}_{\xrightarrow[k\to\infty]{} \nabla^2_{xx} L(x_*,\lambda_*) \text{ constant eigenvalues}} + \underbrace{2\alpha_k c'(x)^T c'(x)}_{\text{eigenvalues that blow up}}$$

$\Rightarrow$ $\nabla^2 P(\alpha_k, x_k)$ has some eigenvalues which are constant and some others that $\to \infty$.

BAD for solving Newton step

$$\nabla^2 P(\alpha, x)\Delta = -\nabla P(\alpha, x)$$

better behaved system: (and equivalent)

$$\left[\begin{array}{c|c} \nabla^2 f(x) + \sum_{i=1}^{|E|} 2\alpha\, c(x) \nabla^2 c_i(x) & c'(x)^T \\ \hline c'(x) & -\frac{1}{\alpha} I \end{array}\right] \begin{bmatrix} \Delta \\ z \end{bmatrix} = \begin{pmatrix} -\nabla P(\alpha, x) \\ 0 \end{pmatrix}$$

Augmented Lagrangian method

$$L_A(x,\lambda;\alpha) = \underbrace{f(x)+\lambda^T c(x)}_{L(x,\lambda)} + \alpha \|c(x)\|_2^2$$

$$\nabla L_A(x,\lambda;\alpha) = \nabla f(x) + c'(x)^T(\lambda + 2\alpha c(x))$$

Thm: Let x^* be a local min for (1), let $c'(x_*)$ have rank m. Let Let λ^* be the Lagrange multiplier at solution x_* and let 2nd order suff cdt° be satisfied then $\exists\, \bar{\alpha} > 0$ s.t. x_* is a local min of $L_A(x,\lambda^*;\alpha)$ for all $\alpha \geq \bar{\alpha}$.

not very practical result, need to know λ^*!
However could do iteration:

- Let λ_k be the current approx of λ^* and $\alpha_k > 0$
- solve $\min\limits_{x \in \mathbb{R}^n} L_A(x,\lambda_k;\alpha_k)$ (call x_k its solution)
- $\lambda_{k+1} = \lambda_k + 2\alpha_k c(x_k)$

Def A penalty function $P(\alpha,x) = f(x) + \alpha \pi(c(x))$ is called exact at a local min x_* of (1) if there exists $\bar{\alpha} > 0$ s.t. for all $\alpha \geq \bar{\alpha}$, x^* is a local min of $P(\alpha,\cdot)$.

<u>theorem</u> Let x_* be a local min of (1) with $\nabla f(x_*) \neq 0$. Let $P(\alpha, x)$ be exact at x_*, then $\pi(c(x))$ is not diff'ble at x_*. (BAD for N. M.)

<u>Proof</u>: Let $\alpha_1 > \alpha_2 \geq \bar{\alpha}$. Assume for contradiction $\pi(c(x))$ is diff'ble at x_*.

x^* is a loc. min of $P(\alpha_1, \cdot)$, $P(\alpha_2, \cdot)$

$$\Rightarrow 0 = \nabla P(\alpha_1, x^*) = \nabla P(\alpha_2, x^*)$$

$$\Rightarrow \underbrace{\nabla \pi(c(x_*))^T c'(x_*)}_{=0} \underbrace{(\alpha_1 - \alpha_2)}_{>0} = 0$$

$$\Rightarrow \nabla f(x_*) = 0.$$

<u>Example</u>: $P(x, \alpha) = f(x) + \alpha \|c(x)\|_1$ no power.

§ 18. Sequential Quadratic Programming

Assume that f, c are twice cont diff'ble

$$(1) \begin{cases} \min & f(x) \\ \text{s.t.} & c(x) = 0 \end{cases} \qquad \begin{aligned} &f: \mathbb{R}^n \to \mathbb{R} \\ &c: \mathbb{R}^n \to \mathbb{R}^m \\ &\operatorname{rank}(c'(x)) = m \end{aligned}$$

Necessary optimality cdt°.

$$F(x,\lambda) = \begin{pmatrix} \nabla f(x) + c'(x)^T \lambda \\ c(x) \end{pmatrix} = 0$$

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

$$F'(x,\lambda) = \left(\begin{array}{c|c} \nabla_{xx} L(x,\lambda) & c'(x)^T \\ \hline c'(x) & 0 \end{array}\right)$$

Let the 2nd order suff cdt° be satisfied at x_* and let λ^* be the corresponding Lagrange multiplier

<u>Lemma</u>: $F'(x_*, \lambda_*)$ is invertible.

<u>proof</u>:

$$\left(\begin{array}{c|c} \nabla_{xx} L(x_*,\lambda_*) & c'(x)^T \\ \hline c'(x) & 0 \end{array}\right) \begin{pmatrix} v \\ w \end{pmatrix} = 0$$

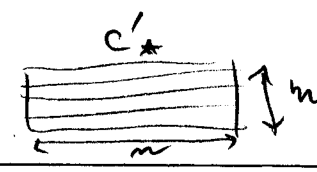
$$\Rightarrow \underbrace{v^T \nabla_{xx} L_* v}_{\geq 0} + \underbrace{v^T c_*'^T w}_{= 0 \text{ by second eq } (v \in \mathcal{N}(c'(x)))} = 0$$

but $v \in \mathcal{N}(c'(x))$ and Hessian is pos def in this subspace

$$\Rightarrow v = 0$$

$$\Rightarrow c_*'^T w = 0 \Rightarrow w = 0$$

c_*' has rank m



Idea: Apply Newton's method to solve

$$F(x,\lambda)=0$$

Newton step:

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

$$\Leftrightarrow \quad (*)\begin{cases}\nabla_{xx}L(x,\lambda)\,\Delta x+c'(x)^T\Delta\lambda=-\nabla_x L(x,\lambda)\\ c'(x)\,\Delta x \qquad\qquad\qquad\qquad =-c(x)\end{cases}$$

Lemma: Assume that $\operatorname{rank}(c'(x))=m$ and that $\nabla^2_{xx}L(x,\lambda)$ is pos def in $\mathcal{N}(c'(x_*))$ then solution $(\Delta x,\Delta\lambda)$ to Newton eq $(*)$ is equal to the sol to the (QP)

$$(2)\begin{cases}\min\ \nabla_x L(x,\lambda)^T\Delta x+\frac{1}{2}\Delta x^T\nabla_{xx}L(x,\lambda)\Delta x\\ \text{s.t.}\quad c'(x)\Delta x+c(x)=0\end{cases}$$

and $\Delta\lambda$ is the corresp Lagrange multiplier

$$(3)\begin{cases}\min\ \nabla f(x)^T\Delta x+\frac{1}{2}\Delta x^T\nabla_{xx}L(x,\lambda)\Delta x\\ \text{s.t.}\quad c'(x)\Delta x+c(x)=0\end{cases}$$

and $\lambda+\Delta\lambda$ is the corresp Lagrange mult.

To see relation between (2) and (3) gobak to (*): ($\equiv$ opt cond for (2))

$$\nabla_{xx} L(x,\lambda)\Delta x + c'(x)^T \Delta\lambda = -\nabla_x L(x,\lambda) = -\nabla f(x) - c'(x)^T \lambda$$

- New constraints are linearization of constraints
- New obj function ~ Taylor but with $\nabla_{xx} L$.
- At the solution we have: $\Delta x = 0$
  $\lambda + \Delta\lambda = \lambda^*$

We replaced full NL problem by a sequence of QP w/ linear constraints.

Globalization of iteration

Let $H \in \mathbb{R}^{n\times n}$ be a replacement of $\nabla_{xx} L(x,\lambda)$ (think Q-N methods)
s.t. H is symm and pos def on $\mathcal{N}(c'(x))$

Let Δx be the sol of:

$$(4) \quad \begin{aligned} &\min \ \nabla f(x_k)^T \Delta x + \tfrac{1}{2}\Delta x^T H_k \Delta x \\ &\text{s.t.} \ c'(x)\Delta x + c(x) = 0 \end{aligned}$$

and let λ_{k+1} ($= \lambda_k + \Delta\lambda$) be the corresp. Lagrange multiplier.
We need step to go towards feasibility and min f.

Merit functions: indicates whether progress towards both goals.
The smaller the merit function the more "progress" is made however it does not meant that f or c(x) have been reduced, only in the limit.

Example: (l_1 merit function)

$$P(x,\alpha) = f(x) + \alpha \|c(x)\|_1 = f(x) + \alpha \sum_{i=1}^{m} |c_i(x)|$$

(augmented Lagrangian can be used for same purpose)

Directional derivative: (see p 628, appendix A)

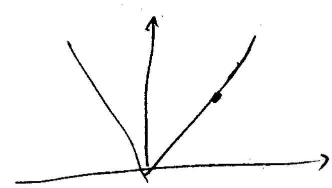
$g(x): \mathbb{R}^n \to \mathbb{R}$

$$= \lim_{\substack{h \to 0 \\ h > 0}} \frac{g(x+hp) - g(x)}{h}$$

- defined even if g is not smooth at x
- $\equiv$ derivative when g smooth at x.

example: $g: \mathbb{R} \to \mathbb{R}$, $x \to |x|$

$$g'(x; d_i) = \begin{cases} p & x > 0 \\ -p & x < 0 \\ |p| & x = 0 \end{cases}$$



$$\varphi(x) = |c_i(x)|$$

$$D(\varphi(x); p) = g'(c_i(x) \,|\, \nabla c_i(x)\, p) \quad (\sim \text{chain rule})$$

Thus:

$$P'(x, \Delta x; \alpha) = \nabla f(x)^T \Delta x + \alpha \sum_{c_i(x) > 0} \nabla c_i(x)^T \Delta x + \alpha \sum_{c_i(x) < 0} -\nabla c_i(x)^T \Delta x + \alpha \sum_{c_i(x) = 0} |\nabla c_i(x)^T \Delta x|$$

Let Δx be sol of (4) and $\alpha > \|\lambda_{k+1}\|_\infty$ where $\lambda_{k+1} = \lambda_k + \Delta\lambda$ is Lagrang mult of QP (4).

Then $P'(x_k, \Delta x; \alpha) = -\Delta x^T H_k \Delta x < 0$

if H_k is pos def

$\Rightarrow$ Δx is a descent direction for l_1 merit function

A SQP method

Given $x_0 \in \mathbb{R}^n$, $\lambda_0 \in \mathbb{R}^m$, $H_0 \in \mathbb{R}^{n\times n}$ s.p.d.
$\beta \in (0,1)$, $\sigma \in (0,1)$ (usually $\beta = \frac{1}{2}$, $\sigma = 10^{-4}$)

For $k = 0, 1, \dots$

- if (x_k, λ_k) solves (1) STOP
- solve (4), let Δx_k, λ_{k+1} be the sol and corresp. Lagrange mult.
- if $\Delta x_k = 0$ STOP
- set $\alpha > \|\lambda_{k+1}\|_\infty$
- Select smallest integer l s.t. $t_k = \beta^l$ satisfies

  $\longrightarrow P(x_{k+1}; \alpha) - P(x_k; \alpha) \leq \sigma t_k P'(x_k, \Delta x_k; \alpha)$
- Compute H_{k+1} spd.

very similar to Armijo step rule. Merit functional is only used to compute step length, not the step itself.

Choice of H_{k+1}:

- $H_{k+1} = \nabla_{xx} L(x_{k+1}, \lambda_{k+1})$

  however may not be pos def $\Rightarrow$ no guarantee to find descent direction for merit function

Take instead:

$$H_{k+1} = \nabla_{xx} L(x_{k+1}, \lambda_{k+1}) + \eta D$$, where D is pos def at "the right place"

Meaning: at solution:

$\nabla_{xx} L(x_*, \lambda_*)$ = pos def in $\mathcal{N}(C'(x))$ = subspace of dim $n-m$

$\rightsquigarrow$ there are potentially m eigenvalues where we don't know anything:

ηI not recommended. Instead take

$$D_{k+1} = C'(x_{k+1})^T C'(x_{k+1})$$

Since by FTOLA: $\mathcal{R}(C'(x_{k+1})^T) = (\mathcal{N}(C'(x_{k+1})))^\perp$

this D_{k+1} does not modify eigenvalues of $\nabla_{xx} L(x_*, \lambda_*)$ in $\mathcal{N}(C'(x_*))$ as ηI would.

- BFGS (Quasi Newton update)

$$H_{k+1} = H_k + \frac{y_k y_k^T}{s_k^T y_k} - \frac{H_k s_k (H_k s_k)^T}{s_k^T H_k s_k}$$

where $s_k = x_{k+1} - x_k$

$$y_k = \nabla_x L(x_{k+1}, \lambda_k) - \nabla_x L(x_k, \lambda_k)$$

We saw that :

H_k sym pos def & $s_k^T \eta_k > 0$ then H_{k+1} is spd.

Ideally we would like $\eta_k = y_k$, however there is no guarantee that $s_k^T y_k > 0$.

<u>Solution</u>: damp steps:

$$\eta_k = \theta_k y_k + (1-\theta_k) H_k s_k,$$

$$\text{Here } \theta_k = \begin{cases} 1 & \text{if } s_k^T y_k \geq 0.2\, s_k^T B_k s_k \\[2ex] 0.8 \dfrac{s_k^T H_k s_k}{s_k^T H_k s_k - s_k^T y_k} & \text{if } s_k^T y_k < 0.2\, s_k^T B_k s_k \end{cases}$$

$\theta_k = 0$ gives $H_{k+1} = H_k$

$\theta_k = 1$ gives unmodified BFGS update

$\Rightarrow$ η_k interpolates between not updating H_k and the unmodified update.

θ_k chose s.t. H_{k+1} spd

<u>Maratos effect</u>: (p440-446)

$$\min f(x) = 2(x_1^2 + x_2^2 - 1) - x_1$$

$$\text{s.t. } c(x) = x_1^2 + x_2^2 - 1 = 0$$

$$x_* = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \lambda_* = -\frac{3}{2}$$

$$H_k = \nabla_{xx} L(x_*, \lambda_*) = I$$

(L1) is given by:

$$\min \frac{1}{2} \Delta x^T \Delta x + \Delta x^T \begin{pmatrix} 4x_1 - 1 \\ 4x_2 \end{pmatrix}$$

$$\text{s.t. } (2x_1 \;\; 2x_2) \Delta x + x_1^2 + x_2^2 - 1 = 0$$

Optimality conditions: (assume $x_1^2+x_2^2=1$)

$$\begin{cases} \nabla f(x) + \Delta x + c'(x)^T \lambda = 0 & (1) \\ c'(x)\Delta x = 0 \end{cases}$$

$$f(x+\Delta x) = f(x) + \nabla f(x)^T \Delta x + \frac{4}{2}\Delta x^T \Delta x$$

equality since obj func is quadratic.

$$\Delta x^T (1) \Rightarrow \quad \Delta x^T \nabla f(x) = -\Delta x^T \Delta x$$

$$\Rightarrow f(x+\Delta x) = f(x) + \|\Delta x\|^2 > f(x)$$

$$c(x+\Delta x) = \underbrace{c(x)}_{=0} + \underbrace{c'(x)^T\Delta x}_{=0} + \|\Delta x\|_2^2 > c(x)$$

$\Rightarrow P_1(x+\Delta x;\alpha) > P_1(x;\alpha)$ and we are forced to reduce stepsize; no matter how close we are to sol!

~ very different from Newton's ~~method~~ where we can expect stepsize to be eventually 1.

Problem related with l_1 merit function.
Augmented Lagrangian does not suffer from this and there 2nd order corrections to l_1 functional as well.