06-720 Advanced Process Systems Engineering Spring, 2011

Homework 3 Due: 2/2/11

1. Consider the penalty function given by $\phi(x;\rho) = f(x) + \rho \sum_{j=1}^{m} \varphi(h_j(x))$, where $\varphi(\xi)$ is a smooth function of ξ where $\varphi(\xi) = 0$ if $\xi = 0$ and $\varphi(\xi) > 0$ if $\xi \neq 0$, with a finite value of ρ, and compare the KKT conditions of min $f(x)$ $s.t.$ $h(x) = 0$ with a stationary point of the penalty function. Argue why these conditions cannot yield the same solution.

2. Given a nonzero Newton step of the optimality conditions, where the KKT matrix has the correct inertia, show when this step is a descent direction for the ℓ_1 merit function $\phi_1(x;\rho) = f(x) + \rho\|h(x)\|_1$.

3. Show that the tangential step $d_t = Z^k p_Z$ and normal step $d_n = Y^k p_Y$, respectively, can be found from

$$\begin{bmatrix} W^k & A^k \\ (A^k)^T & 0 \end{bmatrix} \begin{bmatrix} d_t \\ v \end{bmatrix} = -\begin{bmatrix} \nabla f(x^k) \\ 0 \end{bmatrix}$$

and

$$\begin{bmatrix} \tilde{W} & A^k \\ (A^k)^T & 0 \end{bmatrix} \begin{bmatrix} d_n \\ v \end{bmatrix} = -\begin{bmatrix} 0 \\ h(x^k) \end{bmatrix}$$

where both KKT matrices have the correct inertia and $(Z^k)^T \tilde{W} Y^k = 0$.

4. Select solvers from the SQP, interior point and reduced gradient categories, and apply these to

$$\begin{aligned} \min \quad & x_1 + x_2 \\ s.t. \quad 1 + x_1 - (x_2)^2 + x_3 &= 0 \\ 1 - x_1 - (x_2)^2 + x_4 &= 0 \\ 0 \le x_2 \le 2,\ x_3, x_4 &\ge 0 \end{aligned}$$

Use $x^0 = [0, 0.1, 0, 0]^T$ and $x^0 = [0, 0, 0, 0]^T$ as starting points.

5. Select solvers from the SQP, interior point and reduced gradient categories, and apply these to

$$\begin{aligned} \min \quad & x_1 \\ s.t. \quad (x_1)^2 - x_2 - 1 &= 0 \\ x_1 - x_3 - 0.5 &= 0 \\ x_2, x_3 &\ge 0 \end{aligned}$$

Use $x^0 = [-2, 3, 1]^T$ as the starting point.

Solution HW 3

1) Smooth penalty

$$\phi(x;\rho) = f(x) + \rho \sum_j \phi(h_j(x))$$

Stationary pt: $\nabla \phi(x;\rho) = 0$

$$= \nabla f(x) + \rho \sum (\partial \phi / \partial h) \nabla h(x) = 0$$

Compare to: $\nabla f(x) + \nabla h(x) v = 0$

$$h(x) = 0$$

need to characterize $\partial \phi / \partial h = \phi'(\xi)$

$\phi(\xi) > 0$ and $\phi(-\xi) > 0$ for $\xi > 0$.

$\phi(0) = 0$.

$$0 < \phi(\xi) = \phi'(0)\xi + \frac{\xi^2}{2}\phi''(\bar{\xi})$$

$$0 < \phi(-\xi) = -\phi'(0)\xi + \frac{\xi^2}{2}\phi''(\hat{\xi})$$

Assume $\phi'(0) \neq 0$, then for $\xi < \frac{\phi''(\hat{\xi})}{4\phi'(0)}$

we have $\phi(\xi) < 0 \longrightarrow$ contradiction.

Hence, $\phi'(0) = 0$. Then for $h(x) = 0$,

we have:

$$\nabla \phi = \nabla f(x) + \rho \cancelto{0}{\phi'(0)} \nabla h(x) = 0$$

and the stationary point is inconsistent w/ KKT conditions.

2) $\phi_1(x;\rho) = f(x) + \rho \|h(x)\|_1$

note $\|h(x)\|_1 = \sum_i |h_i(x)|$

$$D_d \phi_1 = \lim_{\alpha \to 0} \frac{\phi_1(x^k + \alpha d_x) - \phi_1(x^k)}{\alpha}$$

$$= \nabla f(x^k)^T d_x + \rho\left(\frac{\sum h_i(x^k + \alpha d_x) - h_i(x^k)}{\alpha}\right)$$

$$= \nabla f(x^k)^T d_x - \rho \|h(x)\|_1$$

based on derivation for $\|h(x)\|_p$ and Newton step: $h(x^k) = -\nabla h(x^k)^T d_x$,

Using $\begin{bmatrix} W^k & A^k \\ A^{kT} & 0 \end{bmatrix}\begin{bmatrix} d_x \\ \bar{v} \end{bmatrix} = -\begin{bmatrix} \nabla f^k \\ h(x^k) \end{bmatrix}$

where KKT matrix has correct inertia, we have: $d_x = Y p_Y + Z p_Z$ where

$$Y p_Y = -Y(A^T Y)^{-1} h(x^k)$$

$$Z p_Z = -Z\left(H^{-1}(Z^T \nabla f^k + w^k)\right)$$

$$\bar{v} = -(Y^T A)^{-1}\left[Y^T \nabla f^k + Y^T W (Z p_Z + Y p_Y)\right]$$

where $H = Z^T W Z$, $w^k = Z^T W Y p_Y$

substitute for d_x to get:

$$\nabla f(x^k)^T d_x - \rho \|h(x)\|_1 = \nabla f(x^k)^T Y p_Y + \nabla f(x^k)^T Z p_Z - \rho \|h(x)\|_1$$

$$= -\nabla f(x^k)^T Y (A^T Y)^{-1} h(x^k) - \nabla f(x^k)^T Z H^{-1} Z \nabla f(x^k)$$

$$- \nabla f(x^k)^T Z H^{-1} w^k - \rho \|h(x^k)\|$$

for a descent direction, it is sufficient to choose ρ so that $\nabla f^T d_x - \rho \|h^k\| < 0$

__but__ for a strongly descent direction we would like $D_d \phi_1 \leq -\varepsilon(\|Z^T \nabla f\|^2 + \|h\|_1)$

- from inertia props, H^{-1} is pd & bdded so $-\nabla f^k (Z^T H^{-1} Z) \nabla f^k \leq -\varepsilon \|Z^T \nabla f\|^2$
- also from $\nabla f^{k^T} (Y (A^T Y)^{-1} h^k + Z H^{-1} w^k)$

$$= \nabla f^{k^T} (Y + Z H^{-1} Z^T W Y)(A^T Y)^{-1} h^k$$

$$\leq \| \nabla f^{k^T} (Y + Z H^{-1} Z^T W Y)(A^T Y)^{-1} \| \|h^k\|$$

$$\leq \gamma \|h^k\|$$

Choosing $\rho \geq \gamma + \varepsilon$ leads to

$$D_d \phi_1 \leq -\varepsilon (\|Z^T \nabla f^k\|^2 + \|h^k\|_1)$$

3) <u>Tangential step</u>: show $d_t = Zp_Z$

i) $W d_t + A v = -\nabla f^k$
ii) $A^T d_t = 0$

let $d_t = Zp_Z + Yp_Y$. from ii) we have

$$A^T Z p_Z + A Y p_Y$$
$$= A Y p_Y = 0 \Rightarrow p_Y = 0$$

from i) $\begin{bmatrix} Y^T \\ Z^T \end{bmatrix} (W Z p_Z + A v) = -\begin{bmatrix} Y^T \\ Z^T \end{bmatrix} \nabla f^k$

leaving: $Z^T W Z p_Z = -Z^T \nabla f^k$
$Y^T W Z p_Z + Y^T A v = -Y^T \nabla f^k$

with first equation providing p_Z and v discarded.

<u>Normal step</u>: show $d_n = Y p_Y$

i) $\bar{W} d_n + A v = 0$
ii) $A^T d_n = -h(x^k)$ let $d_n = Y p_Y + Z p_Z$

from i)

$$\begin{bmatrix} Y^T \\ Z^T \end{bmatrix} (\bar{W} d_n + A v) = 0, \quad Z^T \bar{W} Y = 0$$

$$Y^T \bar{W}(Y p_Y + Z p_Z) + (Y^T A) v = 0$$
$$Z^T \bar{W}(Y p_Y + Z p_Z) = 0$$

$$Z^T \bar{W} Z p_Z = 0 \Rightarrow p_Z = 0$$

so $d_n = Y p_Y$

4. Min $x_1 + x_2$

$$1 + x_1 - (x_2)^2 + x_3 = 0$$
$$1 - x_1 - (x_2)^2 + x_4 = 0$$
$$0 \le x_2 \le 2, \; x_3, x_4 \ge 0$$

$x^0 = 0$ compare IPOPT - solves
CONOPT } infeasible
SNOPT }

- linearization at zero, makes this problem inconsistent and the KKT system is singular

- barrier approach needs to move away from origin → leads to convergence with $x^* = [-3, 1, 2, 6]^T$.

- choosing $x^0 = [0, 0.1, 0, 0]$ as starting point allows all methods to converge

5. Min x_1

s.t. $x_1^2 - x_2 - 1 = 0$
$x_1 - x_3 - 0.5 = 0$
$x_2, x_3 \ge 0$

This problem causes IPOPT to fail at min infeasibility.

Active set solvers (CONOPT, SNOPT) project into the active constraints + continue until convergence at $x^* = [1, 0, \frac{1}{2}]^T$.