Exercicio 6 - Franklin

Considere o problema

$$\begin{array}{ll} \min\limits_{x} & f(x) \\ \text{suj. a} & c(x) = 0, \end{array}$$

com $f : \mathbb{R}^n \to \mathbb{R}$ e $c : \mathbb{R}^n \to \mathbb{R}^m$, e considere aproximações x_k, y_k para a solução x^* e o multiplicador y^* associado às condições de otimalidade

$$\nabla f(x^*) + J(x^*)^T y^* = 0$$

e

$$c(x^*) = 0.$$

O problema com regularização primal-dual associado a esse problema é dado por

$$\begin{array}{ll} \min x, u & f(x) + \rho_k \frac{1}{2} \|x - x_k\|^2 + \delta_k \frac{1}{2} \|u - y_k\|^2 \\ \text{suj. a} & c(x) + \delta_k u = 0, \end{array}$$

onde $\rho_k \geq 0$ e $\delta_k > 0$ são valores escolhidos de maneira a garantir algumas condições que veremos em breve.

1. Escreva as condições de otimalidade desse problema em função de x, y e u.

A função Lagrangiana do problema é dada por:

$$L(x,u) = f(x) + \rho_k \frac{1}{2} \|x - x_k\|^2 + \delta_k \frac{1}{2} \|u - y_k\|^2 + y^T (c(x) + \delta_k u)$$

Assim, fazendo $\nabla L(x, y, u) = 0$, tem-se:

$$\nabla L(x,y,u) = \begin{bmatrix} \nabla_x L(x,y,u) \\ \nabla_y L(x,y,u) \\ \nabla_u L(x,y,u) \end{bmatrix} = \begin{bmatrix} \nabla f(x) + \rho_k (x - x_k) + J(x)^T y \\ c(x) + \delta_k u \\ \delta_k (u - y_k) + \delta_k y \end{bmatrix} = 0$$

2. Verifique que isolando e substituindo/removendo u das condições de otimalidade obtemos a condição $G_k(x, y) = 0$ onde

$$G_k(x,y) = \begin{bmatrix} \nabla f(x) + J(x)^T y + \rho_k (x - x_k) \\ c(x) - \delta_k (y - y_k) \end{bmatrix}.$$

De fato: tomando

$$\delta_k (u - y_k) + \delta_k y = 0 \qquad \Rightarrow \qquad \delta_k u - \delta_k y_k + \delta_k y = 0 \qquad \Rightarrow \qquad \delta_k u = -\delta_k (y - y_k)$$

e substituindo em

$$c(x) + \delta_k u \qquad \Rightarrow \qquad c(x) - \delta_k (y - y_k)$$

temos o resultado.

Um passo tipo Newton para $G_k(x,y)=0$ é dado pelo sistema

$$\begin{bmatrix} H_k+\rho_k I & A_k^T \\ A_k & -\delta_k I \end{bmatrix}\begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = -\begin{bmatrix} g_k + A_k^T y_k \\ c_k \end{bmatrix},$$

onde $c_k = c(x_k)$, $g_k = \nabla f(x_k)$, $A_k = J(x_k)$, e H_k é uma aproximação para a Hessiana de $f(x)+c(x)^T y$ (ou a própria Hessiana).

3. Mostre que independente da H_k e da A_k, é sempre possível escolher valores de ρ_k e δ_k tais que Δx é direção de descida para a função de mérito

$$\phi_k(x) = f(x) + y_k^T c(x) + \tfrac{1}{2}\delta_k^{-1}\|c(x)\|^2.$$

Do da segunda equação do passo tipo de Newton, temos:

$$A_k\Delta x - \delta_k I\Delta y = -c_k$$

isolando Δy

$$\Delta y = \frac{c_k}{\delta_k} + \frac{A_k \Delta x}{\delta_k}$$

Como $\dfrac{A_k\Delta x}{\delta_k} \to 0$, substituindo na primeira equação:

$$(H_k+\rho_k I)\Delta x + A_k^T\Delta y = -g_k - A_k^T y_k$$
$$(H_k+\rho_k I)\Delta x + A_k^T\frac{c_k}{\delta_k} = -g_k - A_k^T y_k$$

Note que $H_k+\rho_k I$ admite inversa pois H_k é simétrica, então basta assumir δ_k suficientemente grande. Assim temos, isolando Δx:

$$\Delta x = (H_k+\rho_k I)^{-1}(-g_k - A_k^T(y_k + \frac{c_k}{\delta_k}))$$

Para mostrar que Δx é direção de descida de $\phi_k(x)$, vamos mostrar que $\phi_k(x)\Delta x < 0$.

Para $\delta_k > 0$ e $\rho_k \geq 0$, derivando $\phi_k(x)$ temos:

$$\nabla\phi_k(x) = \nabla f(x) + J(x)^T y + J(x)^T\frac{c_k}{\delta_k} = g_k + A_k^T(y_k + \frac{c_k}{\delta_k})$$

logo:

$$\begin{aligned} &\nabla\phi_k(x)\Delta x = \\ &= [g_k + A_k^T(y_k+\frac{c_k}{\delta_k})]^T(H_k+\rho_k I)^{-1}[-g_k - A_k^T(y_k+\frac{c_k}{\delta_k})] = \\ &= -[g_k + A_k^T(y_k+\frac{c_k}{\delta_k})]^T(H_k+\rho_k I)^{-1}[g_k + A_k^T(y_k+\frac{c_k}{\delta_k})] = \\ &= -\|g_k + A_k^T(y_k+\frac{c_k}{\delta_k})\|^2(H_k+\rho_k I)^{-1} \leq 0 \end{aligned}$$

Escolhendo $g_k + A_k^T(y_k + \dfrac{c_k}{\delta_k}) \neq 0$ temos que $\phi_k(x)\Delta x < 0$.

4. Uma estratégia para resolver este problema é escrevê-lo como um problema de quadrados mínimos! Embora não seja óbvio a priori, com a ajuda do exercício 4, escreva o cálculo do passo de Newton como um problema de quadrados mínimos com regularização. Para simplificar, assuma que $\rho_k = 0$ e H_k é definida positiva. Algumas dicas: escreva $\Delta y = \overline{\Delta y} + \gamma$ onde γ é escolhido de maneira a zerar uma certa parte do vetor do lado direito. Com o sistema resultante, identifique-o com um problema de quadrados mínimos regularizado na variável $\overline{\Delta y}$.

Considere $\Delta y = \bar{\Delta} y + \gamma$. Assim o sistema

$$\begin{bmatrix} H_k & A_k^T \\ A_k & -\delta_k I \end{bmatrix} \begin{bmatrix} \Delta x \\ \bar{\Delta} y + \gamma \end{bmatrix} = - \begin{bmatrix} g_k + A_k^T y_k \\ c_k \end{bmatrix},$$

pode ser escrito como:

$$\begin{cases} H_k \Delta x + A_k^T (\bar{\Delta} y + \gamma) = -g_k - A_k^T y_k \\ A_k \Delta x - \delta_k (\bar{\Delta} y + \gamma) = -c_k \end{cases}$$

ou,

$$\begin{cases} H_k \Delta x + A_k^T \bar{\Delta} y + A_k^T \gamma = -g_k - A_k^T y_k \\ A_k \Delta x - \delta_k \bar{\Delta} y - \delta_k \gamma = -c_k \end{cases}$$

Assim,

$$\begin{cases} H_k \Delta x + A_k^T \bar{\Delta} y = -[g_k + A_k^T (y_k + \gamma)] \\ A_k \Delta x - \delta_k \bar{\Delta} y = -[c_k - \delta_k \gamma] \end{cases}$$

Fazendo $c_k - \delta_k \gamma = 0$ ou seja $\gamma = c_k / \delta_k$

$$\begin{cases} H_k \Delta x + A_k^T \bar{\Delta} y = -[g_k + A_k^T (y_k + c_k/\delta_k)] \\ A_k \Delta x - \delta_k \bar{\Delta} y = -[c_k - \delta_k c_k/\delta_k] = 0 \end{cases}$$

Assim, podemos escrever:

$$\begin{bmatrix} H_k & A_k^T \\ A_k & -\delta_k I \end{bmatrix} \begin{bmatrix} \Delta x \\ \bar{\Delta} y \end{bmatrix} = - \begin{bmatrix} g_k + A_k^T (y_k + c_k/\delta_k) \\ 0 \end{bmatrix},$$

Que corresponde ao problema de mínimos quadrados regularizado na variável $\bar{\Delta} y$:

$$\begin{array}{ll} \min\limits_{\Delta x, \bar{\Delta} y} & \frac{1}{2} \|\Delta x\|_{H_k}^2 + \frac{1}{2} \|\bar{\Delta} y\|_{\delta_k I}^2 \\ \text{suj. a} & A_k \Delta x - \delta_k \bar{\Delta} y = 0, \end{array}$$