Boa definição do SQP básico

$$\min f(x)$$
$$\text{s.a. } h(x) = 0.$$

$$QP_k: \min_{d_x} \tfrac{1}{2} d_x^t B_k d_x + \nabla f(x^k)^t d_x$$

$$\text{s.a. } \underbrace{\nabla h(x^k)^t}_{m \times n} \underbrace{d_x}_{n \times 1} + h(x^k) = 0.$$

Sendo d_x^k a solução de QP_k (B_k é simétrica e definida positiva) e $d_\lambda^{+,k}$ o multiplicador de Lagrange associado à restrição de igualdade, o passo do método é

$$x^{k+1} = x^k + d_x^k, \quad \lambda^{k+1} = \lambda^k + d_\lambda^{+,k}.$$

Como SQP básico é Newton, o que conseguimos provar é que ele está bem definido em uma vizinhança de uma solução x^* do problema original.

Hipóteses:

H1) ∇h_i, $i = 1, \ldots, m$ são contínuos

H2)

$$\nabla h(x^*) = \begin{bmatrix} | & & | \\ \nabla h_1(x^*) & \cdots & \nabla h_m(x^*) \\ | & & | \end{bmatrix}$$ tem

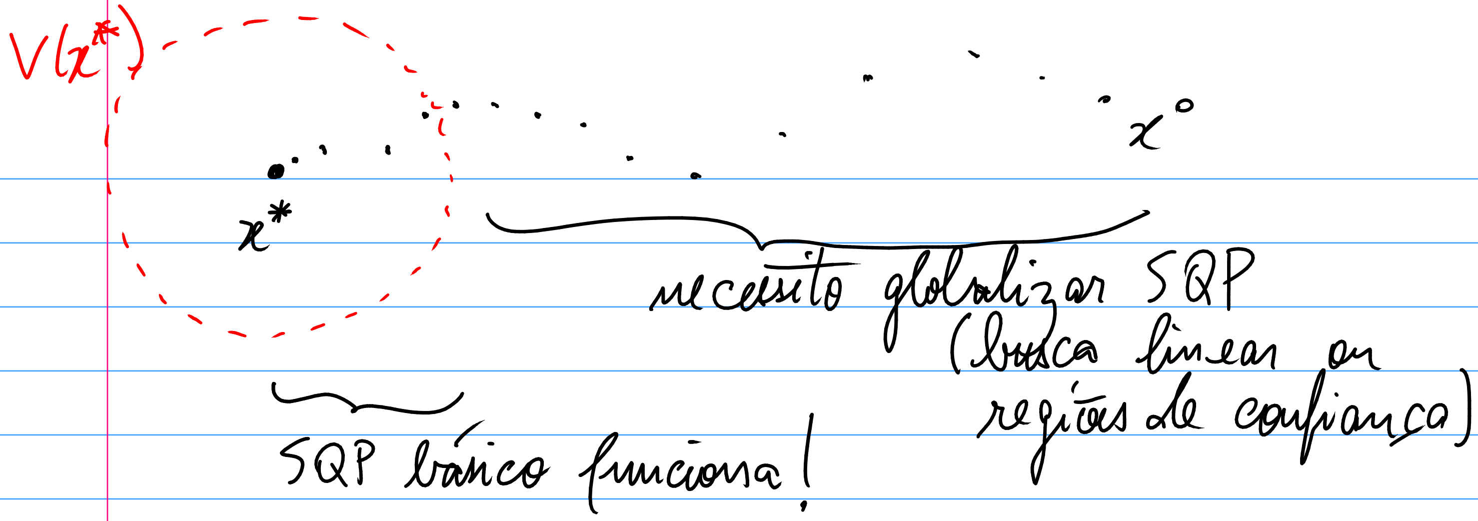
posto coluna completo. Isto é,

$\{\nabla h_1(x^*), \ldots, \nabla h_m(x^*)\}$ são L.I. $=$

x^* é regular.

H3) B_k é simétrica e definida positiva.

Teorema: Se valem H1, H2 e H3, então existe uma vizinhança $V(x^*)$ da solução x^* para a qual QP_k tem solução sempre que $x^k \in V(x^*)$. Ou seja SQP básico está bem definido ao redor de x^*.



Prova: Por H2, as colunas de

$$\nabla h(x^*) = \begin{bmatrix} | & & | \\ \nabla h_1(x^*) & \cdots & \nabla h_m(x^*) \\ | & & | \end{bmatrix}$$

são L.I. Da continuidade de ∇h (H1), existe uma vizinhança $V(x^*)$ de x^* tal que

$$\nabla h(x^k) = \begin{bmatrix} | & & | \\ \nabla h_1(x^k) & \cdots & \nabla h_m(x^k) \\ | & & | \end{bmatrix}$$

tem colunas L.I. para todo $x^k \in V(x^*)$.

Particionamos as colunas da matriz $m \times n$

$$\nabla h(x^k)^t = \begin{bmatrix} — \nabla h_1(x^k)^t — \\ \vdots \\ — \nabla h_m(x^k)^t — \end{bmatrix} \Bigg\} \text{linhas L.I.}$$

na forma

$$\nabla h(x^k)^t = \begin{bmatrix} C^k & N^k \end{bmatrix}$$

onde C^k $m\times m$ é inversível. Assim,

$$\nabla h(x^k)^t d_x + h(x^k) = 0$$

$$\Leftrightarrow \begin{bmatrix} C^k & N^k \end{bmatrix} \begin{bmatrix} d_x^C \\ d_x^N \end{bmatrix} + h(x^k) = 0$$

$$\Leftrightarrow C^k d_x^C + N^k d_x^N + h(x^k) = 0$$

$$\Leftrightarrow d_x^C = -(C^k)^{-1} \left[N^k d_x^N + h(x^k) \right]$$

Tomando $d_x^N = 0$, vemos que

$$(d_x^C, d_x^N) = \left(-(C^k)^{-1} h(x^k), 0 \right)$$

é solução de QP_k. Ou seja, QP_k é viável. Finalmente, como QP_k é um problema quadrático estritamente convexo ($H3 \Rightarrow H_k$ definida positiva), ele possui (única) solução. Logo, o SQP básico está bem definido. ∎

Problema: e se H2 não valer, ou seja, se $\nabla h(x^*)$ não tem posto completo?

↳ o Teorema acima não garante boa definição...

$$\nabla h(x^k)^t d_x + h(x^k) = 0$$

$$\underbrace{\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}}\begin{bmatrix} d_{x_1} \\ d_{x_2} \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \times$$

Colunas LD,
QP_k não tem solução

[Martínez, J. M.; Santos, S. A. Métodos computacionais de otimização]()

1) Uma ideia para contornar esse problema é trocar a restrição $\nabla h(x^k)^t d_x + h(x^k) = 0$ por algo que possua solução, mas que de alguma forma pareça com a restrição original.

↳ Resolvemos o problema auxiliar

$$\min_{d_x} \| \nabla h(x^k)^t d_x + h(x^k) \|^2$$

$Ay = b$ sem solução $\longleftrightarrow$ $\min_y \|Ay - b\|^2$.

Este problema auxiliar fornece uma solução d_x. Como vimos, o passo do SQP é $x^{k+1} = x^k + d_x^k \rightarrow d_x = x - x^k$. Daí, o problema auxiliar fica

$$\min_x \|\nabla h(x^k)^t (x - x^k) + h(x^k)\|^2$$

Seja x^k_{mor} uma solução desse problema. Daí trocamos a restrição original

$$\nabla h(x^k)^t (x - x^k) + h(x^k) = 0$$

por

$$\nabla h(x^k)^t (x - x^k) - \nabla h(x^k)^t (x^k_{mor} - x^k) = 0.$$

Observe que, ao resolver o problema auxiliar, $-\nabla h(x^k)^t (x^k_{mor} - x) \approx h(x^k)$. Portanto a troca da restrição faz sentido. Observe ainda que se $\nabla h(x^k)^t (x - x^k) + h(x^k) = 0$ tem solução, então

$$-\nabla h(x^k)^t(x^k_{mor} - x^k) = h(x^k),$$

e logo a troca da restrição original do QP mantém a mesma restrição.

Resumindo, o QP_k fica

$$QP_k : \min_x \tfrac{1}{2}(x-x^k)^t B_k (x-x^k) + \nabla f(x^k)^t (x-x^k)$$

$$\text{s.a. } \nabla h(x^k)^t(x-x^k) - \nabla h(x^k)^t(x^k_{mor} - x^k) = 0 \quad (*)$$

Este QP_k é <u>sempre</u> viável pois $(*)$ vale com $x = x^k_{mor}$.

<u>Comentários</u>

1) é possível agregar limitantes em x, $l \leq x \leq u$.

$$\begin{array}{l} \min f(x) \\ \text{s.a. } h(x) = 0 \\ \quad l \leq x \leq u \end{array} \longrightarrow \begin{array}{l} QP_k : \min_x \tfrac{1}{2}(x-x^k)^t B_k (x-x^k) + \nabla f(x^k)^t (x-x^k) \\ \text{s.a. } \nabla h(x^k)^t (x - x^k) - \nabla h(x^k)^t (x^k_{mor} - x^k) = 0 \\ \qquad l \leq x \leq u. \end{array}$$

2) QP_k só aproxima bem o problema original ao redor de x^k.

↳ usamos regiões de confiança no QP.

$$QP_k: \min_x \tfrac{1}{2}(x-x^k)^t B_k (x-x^k) + \nabla f(x^k)^t (x-x^k)$$
$$\text{s.a. } \nabla h(x^k)^t (x-x^k) - \nabla h(x^k)^t (x^k_{nor} - x^k) = 0$$
$$l \leq x \leq u, \quad \|x-x^k\|_\infty \leq \Delta_k$$

(usando $\|\cdot\|_\infty$, a restrição $\|x-x^k\|_\infty \leq \Delta_k$ é equivalente à $x_i^k - \Delta_k \leq x_i \leq x_i^k + \Delta_k$, mantendo somente restrições lineares em QP_k).

Porém, veja que $x = x^k_{nor}$ pode não satisfazer $l \leq x \leq u$. Pior ainda, QP_k pode não ser mais viável com adição de novas restrições.

Para resolver esse impasse, basta computar x^k_{nor} de maneira adequada:

$$\min_x \|\nabla h(x^k)^t (x-x^k) + h(x^k)\|^2$$
$$\text{s.a. } l \leq x \leq u, \quad \|x-x^k\|_\infty \leq \Delta_k$$

Finalmente, o controle do raio de confiança Δ_k é feito como no esquema de região de confiança visto em aula.

Outra forma de definirmos QP's viáveis:

"Relaxar" as restrições $\nabla h(x^k)^t d_x + h(x^k) = 0$.

$$QP_k: \min_{d_x, \delta} \ \frac{1}{2} d_x^t B_k d_x + \nabla f(x^k)^t d_x + \frac{\eta}{2} \delta^2$$

$$\text{s.a. } \nabla h(x^k)^t d_x + (1-\delta) h(x^k) = 0$$

$$0 \leq \delta \leq 1$$

A variável adicional δ modifica o termo $h(x^k)$. Note que $\delta = 1$, $d_x = 0$ é sempre viável. O termo $\frac{\eta}{2}\delta^2$ na função objetivo visa forçar $\delta = 0$, recuperando o QP_k original. Aqui, $\eta > 0$ é um termo penalizador para δ.

↳ Essa estratégia é adotada no pacote computacional WORHP.