Subproblemas de SQP – Questões práticas

$$\min f(x) \quad s.a. \quad h(x) = 0, \quad l \leq x \leq u.$$

Subproblema:

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

$$s.a. \quad \nabla h(x^k)^t (x-x^k) - \nabla h(x^k)^t (x^k_{nor} - x^k) = 0$$

$$l \leq x \leq u, \quad \|x - x^k\| \leq \Delta_k.$$

Como escolher B_k?

- O SQP básico (= Newton sobre o sistema KKT) sugere que $B_k = \nabla^2 L(x^k, \lambda^k)$.
- Mas, como vimos, $\nabla^2 L$ pode não ser definida positiva...

Possibilidades para B_k (definida positiva)

1. $B_k = \nabla^2 L(x^k, \lambda^k) + \alpha I$, onde $\alpha \gg 1$.

$$0 < d^t B_k d = d^t \nabla^2 L\, d + \alpha \|d\|^2 \,, \quad \forall d \neq 0$$

$$\Leftrightarrow \quad \alpha > -\underbrace{\frac{d^t \nabla^2 L\, d}{\|d\|^2}}\,, \quad \forall d \neq 0$$

$$\Leftrightarrow \quad \alpha > -\lambda_{min}(\nabla^2 L)\,, \text{ onde}$$

λ_{min} denota o menor autovalor.

1.1) Computar $\lambda_{min}(\nabla^2 L)$ e tomar

$\alpha = -\lambda_{min}(\nabla^2 L) + 1$, ou ainda

$\alpha = \max\{0, -\lambda_{min}(\nabla^2 L) + 1\}$.

↳ o cálculo de $\lambda_{min}(\nabla^2 L)$ é caro (adequado p/ problemas pequenos).

1.2) Computar um limitante inferior para $\lambda_{min}(\nabla^2 L)$ de forma barata.

↳ Discos de Gerschgorin:

$A = (a_{ij})_{n \times n}$ simétrica.

$$D_i = \{\lambda \in \mathbb{R} \,;\, |\lambda - a_{ii}| \leq \sum_{j \neq i} |a_{ij}| \}$$

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & -1 & 4 \\ 3 & 4 & 5 \end{bmatrix}$$

$$\lambda_{min}(A) \geq -7.$$

TEO

$D_1 = \{\lambda \in \mathbb{R} ; |\lambda - 1| \leq |2| + |3|\} = [-4, 6]$

$D_2 = \{\lambda \in \mathbb{R} ; |\lambda - (-1)| \leq |2| + |4|\} = [-7, 5]$

$D_3 = \{\lambda \in \mathbb{R} ; |\lambda - 5| \leq |3| + |4|\} = [-2, 12]$

Teorema: Todo autovalor λ de A pertence a algum D_i.

Prova: Seja λ um autovalor de A associado ao autovetor $v \neq 0$. Considere o índice i tal que $|v_i| = \max_{j=1,\dots,n} |v_j|$. A linha i do sistema $Av = \lambda v$ é

$$\sum_{j=1}^{n} a_{ij} v_j = \lambda v_i \Leftrightarrow \lambda v_i - a_{ii} v_i = \sum_{j \neq i} a_{ij} v_j$$

$$\Rightarrow |\lambda v_i - a_{ii} v_i| = \left| \sum_{j \neq i} a_{ij} v_j \right| \leq \sum_{j \neq i} |a_{ij}| \cdot |v_j|$$

$v_i \neq 0$ pois $v \neq 0$.

$$\Rightarrow |\lambda - a_{ii}| \le \sum_{j \neq i} |a_{ij}| \underbrace{\frac{|v_j|}{|v_i|}}_{\le 1} \le \sum_{j \neq i} |a_{ij}| .$$

Ou seja, $\lambda \in D_i$. ■

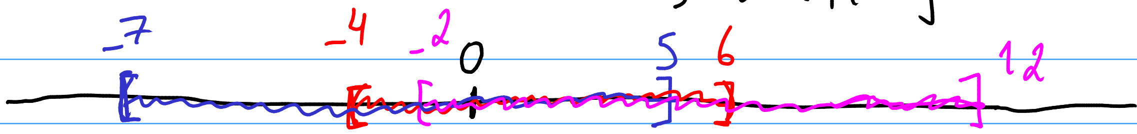
Exemplos:

1) $A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & -1 & 4 \\ 3 & 4 & 5 \end{bmatrix}$

$D_1 = [-4, 6]$

$D_2 = [-7, 5]$

$D_3 = [-2, 12]$



$$\lambda_{min}(A) \in D_i \Rightarrow \lambda_{min}(A) \in D_1 \cup D_2 \cup D_3 = [-7, 12] .$$

$$\Rightarrow \lambda_{min}(A) \geq -7 .$$

2) $C = \begin{bmatrix} 5 & 1 & 2 \\ 1 & 4 & -1 \\ 2 & -1 & 7 \end{bmatrix}$

$D_1 = \{\lambda \in \mathbb{R} ; |\lambda - 5| \le 3\} = [2, 8]$

$D_2 = \{\lambda \in \mathbb{R} ; |\lambda - 4| \le 2\} = [2, 6]$

$D_3 = \{\lambda \in \mathbb{R} ; |\lambda - 7| \le 3\} = [4, 10]$

$\lambda_{min}(A) \in \bigcup D_i = [2, 10] \Rightarrow \lambda_{min}(A) \geq 2$

$\Rightarrow$ A é definida positiva.

Queremos $\alpha > -\lambda_{min}(\nabla^2 L)$.

Se σ for o extremo esquerdo do conjunto $\bigcup_{i=1}^{n} D_i$, então $\lambda_{min}(\nabla^2 L) \geq \sigma$. Assim,

$\alpha > -\sigma \geq -\lambda_{min}(\nabla^2 L)$, ou seja, o limitante obtido por Gerschgorin fornece um α válido ($\nabla^2 L + \alpha I$ definida positiva).

Podemos tomar

$$\alpha = \max\{0, -\sigma + 1\}.$$

Isso é empregado no pacote SQP WORHP (+ heurísticas para redução de α: $\nabla^2 L + \mu \alpha I$ com $\mu \in (0,1]$).

↳ adequado à problemas grandes.

② B_k = "atualização quase-Newton"

$\nabla^2 L(x,\lambda) \longrightarrow \nabla^2 f(x)$ "$x^{k+1} \approx x^k$"

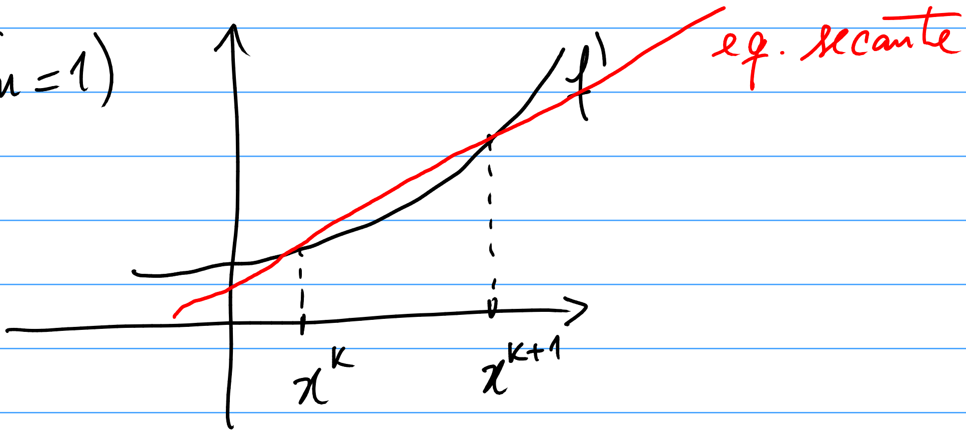
$\nabla f(x^{k+1}) \approx \nabla f(x^k) + \nabla^2 f(x^k)(x^{k+1} - x^k)$

B_{k+1} será solução do sistema

$$B_{k+1}(x^{k+1} - x^k) = \nabla f(x^{k+1}) - \nabla f(x^k)$$

$$\Rightarrow \boxed{B_{k+1} s = y} \quad \text{(equação secante)}$$

$\boxed{x \in \mathbb{R}}$ $(n=1)$



Teorema: A equação secante possui solução B.

Prova: Feito na disciplina "Otimização I" ∎

A equação possui várias soluções B !!!

- Suponha B_k simétrica e definida positiva
- $B_{k+1} = B_k + \alpha y y^t + \beta B_k s s^t B_k$

Obrigamos B_{k+1} satisfazer a eq. secante!

Uma solução é tomar $\alpha = \dfrac{1}{y^t s}$ e $\beta = \dfrac{-1}{s^t B_k s}$

ou seja,

$$\boxed{B_{k+1} = B_k + \frac{y y^t}{y^t s} - \frac{B_k s s^t B_k}{s^t B_k s}}$$

atualização BFGS (Broyden, Fletcher, Goldfarb, Shanno).

- BFGS é uma das mais utilizadas, e funciona bem.
- B_{k+1} como BFGS resolve a equação secante (exercício)
- B_k simétrica e definida positiva e $y^t s > 0$, então B_{k+1} é definida positiva.

↳ prova: veja Teorema abaixo

- QP_0 é construído com $B_0 = I$ (identidade)
- QP_{k+1} é " " B_{k+1} como BFGS, usando a matriz anterior B_k.
- Se em uma dada iteração $y^t s \leq 0$, então BFGS pode fornecer uma matriz B_{k+1} não definida positiva.

  ↳ uma solução $B_{k+1} = B_k$.

Teorema: Suponha que B_k seja simétrica e definida positiva. Então, se $y^t s > 0$, temos a matriz B_{k+1} da atualização BFGS satisfaz:

(i) B_{k+1} está bem definida;

(ii) B_{k+1} é simétrica;

(iii) B_{k+1} é definida positiva.

<u>Prova:</u> Como $y^t s > 0$, em particular $s \neq 0$. Assim $s^t B_k s > 0$ dado que B_k é definida positiva. Logo é possível construir B_{k+1}.

É trivial mostrar que B_{k+1} é simétrica. Isso ficará como exercício.

Vamos mostrar que B_{k+1} é definida positiva. Considere o produto $x^t B_{k+1} x$:

$$x^t B_{k+1} x = x^t B_k x + \frac{(x^t y)^2}{y^t s} - \frac{(x^t B_k s)^2}{s^t B_k s}$$

$$= \frac{(x^t y)^2}{y^t s} + \frac{(s^t B_k s)(x^t B_k x) - (x^t B_k s)^2}{s^t B_k s} \quad . (*)$$

Como B_k é simétrica e definida positiva, possui fatoração Cholesky, digamos

$$B_k = GG^t,$$

com G triangular inferior com diagonal toda positiva ($\Rightarrow G$ é inversível). Temos

- $s^t B_k s = (s^t G)(G^t s) = \|G^t s\|^2$
- $x^t B_k x = (x^t G)(G^t x) = \|G^t x\|^2$
- $x^t B_k s = (x^t G)(G^t s) = (G^t x)^t (G^t s)$.

Da desigualdade de Cauchy-Schwartz obtemos $(G^t x)^t (G^t s) \leq \|G^t x\|^2 \cdot \|G^t s\|^2$, ou seja,

$$(x^t B_k x)(s^t B_k s) - (x^t B_k s)^2 \geq 0 . \qquad (**)$$

Usando esta desigualdade em $(*)$, concluímos que $x^t B_{k+1} x \geq 0$. Sendo x arbitrário, B_{k+1} é semi-definida positiva. Agora, de $(*)$,

$$x^t B_{k+1} x = 0 \Rightarrow x^t y = 0 \text{ e } (**) = 0 .$$

Sabemos que $(**)$ só se realiza como igualdade se $G^t x$ e $G^t s$ forem colineares. Isto é, se $G^t x = \mu G^t s$ para algum $\mu \in \mathbb{R}$. Como G^t é inversível, então multiplicando à esquerda por G^{-1} obtemos $x = \mu s$. Daí $x^t y = \mu \underbrace{(y^t s)}_{>0} \Rightarrow \mu = 0 \Rightarrow x = 0$.

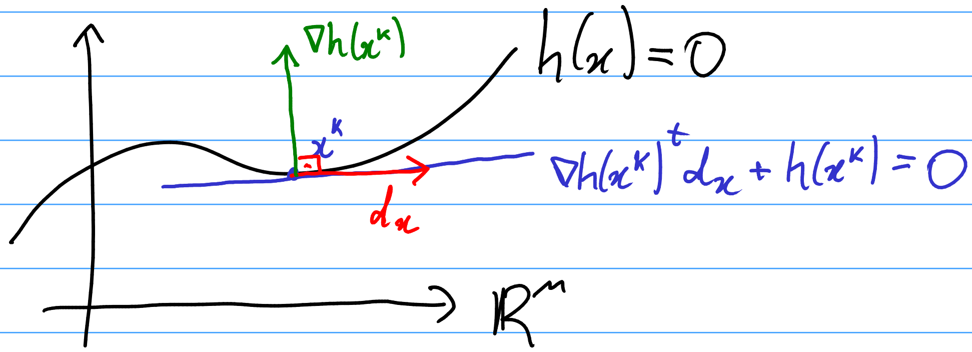
Ou seja, $x^t B_{k+1} x = 0 \Rightarrow x = 0$. Assim B_{k+1} é definida positiva. ∎

Lidando com restrições de desigualdade

Problema geral

$$\min_x f(x) \text{ s.a. } h(x)=0,\ g(x)\leq 0.$$

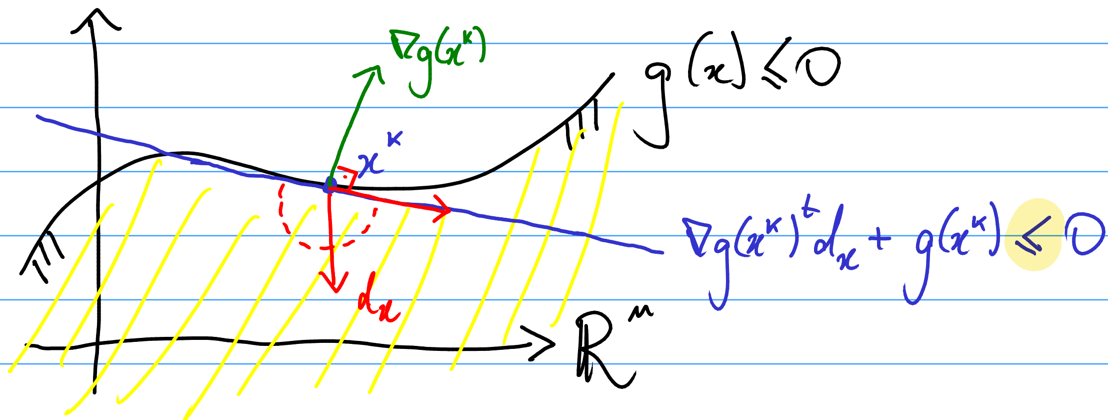
Nos subproblemas, as restrições $h(x)=0$ são linearizadas. Por simplicidade, vamos considerar o SQP básico, onde no QP a linearização é $\nabla h(x^k)^t d_x + h(x^k) = 0$.



As direções d_x são ortogonais aos gradientes $\nabla h_i(x^k)$. Assim, caminhar nessa direção a partir de x^k visa não se afastar muito de $h(x)=0$ (considere o caso em que $h(x^k)=0$, como na figura).

Considere agora as restrições de desigualdade

$g(x) \leqslant 0$:



Aqui, as direções d_x ortogonais à $\nabla g_i(x^k)$ mantém o próximo iterando próximo à viabilidade ($g(x) \approx 0$), mas também as direções d_x que entram no conjunto viável, isto é, aquelas em que $g(x^k + t d_x) < 0$, $0 < t \approx 0$. Essas são as direções que fazem um ângulo obtuso com $\nabla g_i(x^k)$ (veja figura). Assim, a linearização fica

$$\nabla g(x^k)^t d_x + g(x^k) \leq 0.$$

O subproblema "básico" resultante é

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

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

$$\nabla g(x^k)^t d_x + g(x^k) \leqslant 0$$

Não podemos aplicar o método de Newton ao sistema KKT deste QP_k, como antes, pois agora teremos uma desigualdade relativa à não-negatividade dos multiplicadores associados à $\nabla g(x^k)^t d_x + g(x^k) \leq 0$. Uma solução, que é empregada em WORHP, é usar pontos interiores em QP_k. Após adicionar folgas, QP_k é resolvido através de uma sequência de problemas com barreira logarítmica:

$$\min \ \frac{1}{2} d_x^t B_k d_x + \nabla f(x^k)^t d_x - t_k^j \sum_i \ln(w_i)$$

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

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

onde, para cada k, $t_k^j \xrightarrow[j \to \infty]{} 0$. Aqui,

k: índice da iteração SQP

j: índice da iteração de pontos interiores na resolução de QP_k.

É bom frisar que pontos interiores é aplicado para resolver 1 subproblema QP_k de SQP. Poderia-se perguntar então isto vale a pena (por que não aplicar pontos interiores direto ao problema original?)

Ocorre que há especializações de pontos interiores à problemas quadráticos, como QP_k, que são extremamente eficientes! Assim, faz sentido a estratégia descrita aqui (ela de fato é usada no pacote WORHP).