# MODELOS QUADRÁTICOS

REGIÕES DE CONFIANÇA: $\min_{d} \frac{1}{2} d^t B_k d + \nabla f(x^k)^t d$

s.a. $\|d\| \leq \Delta_k$

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

s.a. $\nabla h(x^k)^t d + h(x^k) = 0$

---

COMO ESCOLHER $B_{k+1}$ ?

→ DESEJO: $B_{k+1} = \nabla^2 f(x^k)$ OU $B_k = \nabla^2 L(x^k, \lambda^k)$

→ PROBLEMA: ESTAS ESCOLHAS PODEM NÃO SER SEMIDEF. POSITIVAS (SUBPROBLEMA NÃO CONVEXO).

$\boxed{1^{\underline{a}} \text{ ESCOLHA}}$ $B_{k+1} = \nabla^2 L + \alpha I$, ONDE $\alpha > 0$ É GRANDE O SUFICIENTE PARA QUE $B_{k+1}$ SEJA DEF. POSIT.

VEJA QUE

$$0 < x^t B_{k+1} x = x^t \nabla^2 L x + \alpha \|x\|^2 , \quad \forall x \neq 0$$

$$\iff \alpha > - \frac{x^t \nabla^2 L x}{\|x\|^2} , \quad \forall x \neq 0$$

$$\iff \alpha > - \lambda_{min}(\nabla^2 L) = - \text{"MENOR AUTOVALOR"}$$

O CÁLCULO DO MENOR AUTOVALOR É CARO. PORÉM, É SUFICIENTE CALCULAR UMA COTA INFERIOR PARA $\lambda_{min}(\nabla^2 L)$.

UMA MANEIRA DE CALCULAR UMA COTA É ATRAVÉS DOS CÍRCULOS /DISCOS DE GERSHGORIN.

$$A = \nabla^2 L = (a_{ij})_{1 \le i,j \le m}.$$

DEFINIMOS O CÍRCULO DE GERSHGORIN COMO

$$D_i = \{ \lambda \in \mathbb{R} \; ; \; |\lambda - a_{ii}| \leq \underbrace{\sum_{j \neq i} |a_{ij}|}_{\text{LINHA } i \text{ DE } A} \}.$$

TEOREMA: TODO AUTOVALOR $\lambda$ DE $A$ PERTENCE A ALGUM $D_i$.

PROVA: SEJA $Av = \lambda v$, E $i$ O ÍNDICE

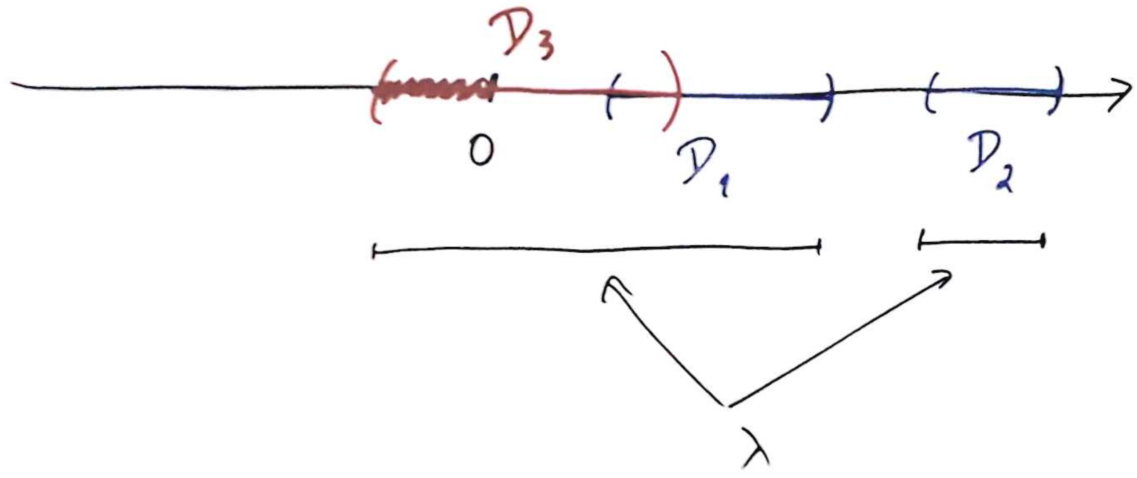
TAL QUE $|v_i| = \max_j |v_j|$. A i-ÉSIMA LINHA DE $Av = \lambda v$ É

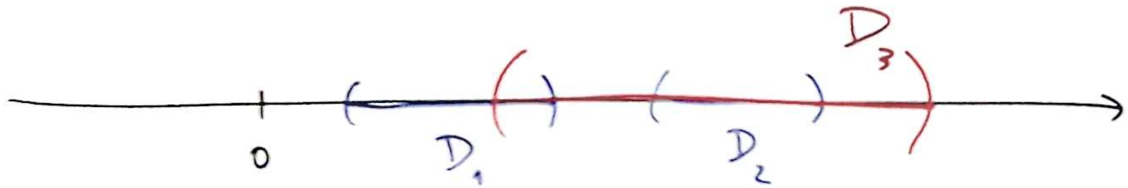
$$\sum_{j=1}^{n} a_{ij} v_j = \lambda v_i$$

$$\iff \lambda v_i - a_{ii} v_i = \sum_{j \neq i} a_{ij} v_j$$

$$\implies |\lambda - a_{ii}| \cdot |v_i| \leq \sum_{j \neq i} |a_{ij}| \, |v_j|$$

$$\underset{v_i \neq 0}{\implies} |\lambda - a_{ii}| \leq \sum_{j \neq i} |a_{ij}| \underbrace{\frac{|v_j|}{|v_i|}}_{\leq 1} \leq \sum_{j \neq i} |a_{ij}| \quad \blacksquare$$




$\Longrightarrow A$ DEF. POS.

EXEMPLO:

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

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

$$D_2 = [2, 6], \quad D_3 = [4, 10].$$

$$\Longrightarrow 0 \notin \bigcup D_i \Longrightarrow \lambda_{min}(A) \geq 2$$

ASSIM $A$ É DEF. POSIT. DIZ-SE QUE ESTA $A$ É "DIAGONALMENTE DOMINANTE". //

---

SE $\lambda_{min}(\nabla^2 L) > 0$, $\nabla^2 L$ JÁ É DEF. POSIT. E PODEMOS TOMAR $\alpha = 0$. O CASO QUE IMPORTA É QUANDO $\lambda_{min}(\nabla^2 L) \leq 0$. NA DÚVIDA, TOMAMOS O "AVANÇO MAIS NEGATIVO" DOS DISCOS $D_i$:

$$\sigma = \min_{1 \leq i \leq m} \left\{ a_{ii} - \sum_{j \neq i} |a_{ij}| \right\}.$$

VEJA QUE

$$\lambda \in D_i \iff a_{ii} - \sum_{j \neq i} |a_{ij}| \leq \lambda \leq a_{ii} + \sum_{j \neq i} |a_{ij}|$$

$$\Longrightarrow \sigma = \min \left\{ a_{ii} - \sum_{j \neq i} |a_{ij}| \right\} \leq \lambda , \quad \forall \text{ AUTOVALOR}$$

$$\Longrightarrow \sigma \leq \lambda_{min}(\nabla^2 L).$$

OBSERVE QUE $\sigma$ SERVE PARA $\alpha$ (POIS É UMA COTA INFERIOR PARA $\lambda_{min}(\nabla^2 L)$) E É FÁCIL DE CALCULAR. TOMAMOS

$$\alpha = |\sigma| + 1.$$

FINALMENTE,

$$B_{k+1} = \nabla^2 L(x^k, \lambda^k) + \rho(|\sigma|+1) I \quad , \quad \rho \in (0,1].$$

ISSO É EMPREGADO NO SQP WORHP.

- SE $\rho \approx 0$, HÁ RISCO DE $B_{k+1}$ NÃO SER DEF. POSIT. POR OUTRO LADO, $\rho = 1$, HÁ GARANTIA TEÓRICA DE $B_{k+1}$ SER DEF. POSIT,

- WORHP "FAZ UMA APOSTA" AO TOMAR $\rho < 1$, EM NOME DA ESTABILIDADE NUMÉRICA.

$\boxed{2^{\underline{a}} \text{ ESCOLHA}}$ BASEADA EM UMA APROXIMAÇÃO DE $\nabla^2 L$.

QUEREMOS A APROXIMAÇÃO $B_{k+1}$ SIMÉTRICA, DEF. POSIT, E QUE NÃO DEPENDA DO CÁLCULO DE $\nabla^2 L$.

POR SIMPLICIDADE, VAMOS CONSIDERAR UM PROBLEMA IRRESTRITO... LOGO $B_{k+1} \approx \nabla^2 f$.

TAYLOR:

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

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

$B_{k+1}$ SERÁ SOLUÇÃO DO SISTEMA

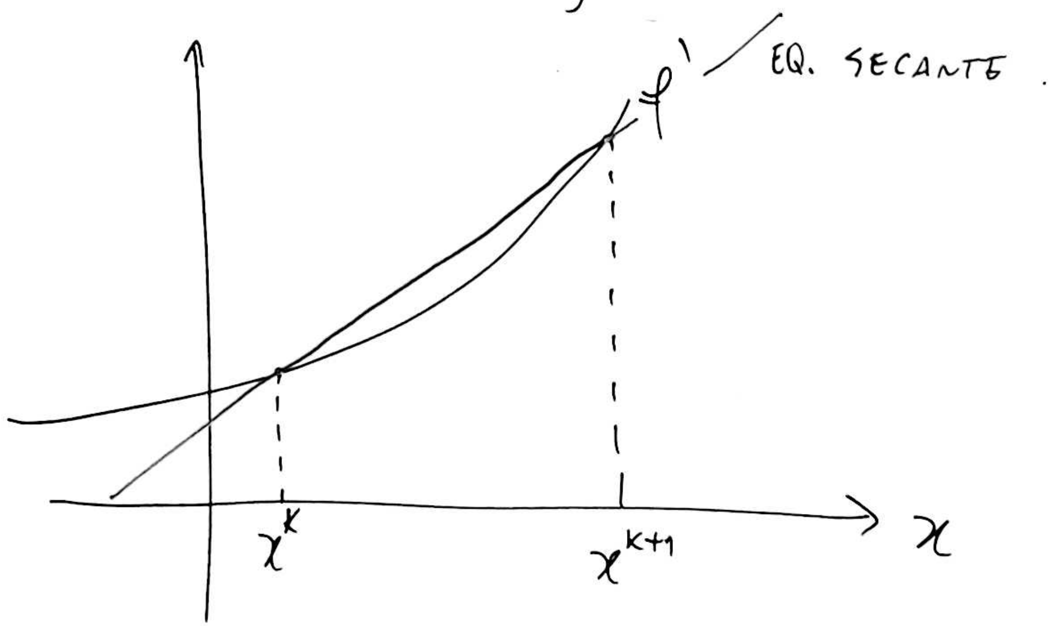
$$B_{k+1}\left(x^{k+1} - x^{k}\right) = \nabla f(x^{k+1}) - \nabla f(x^{k}),$$

OU

$$\boxed{B_{k+1}\, s = y} \qquad \text{(EQUAÇÃO SECANTE)}$$

ONDE $s = x^{k+1} - x^{k}$ E $y = \nabla f(x^{k+1}) - \nabla f(x^{k})$.

$\boxed{n=1}$



TEOREMA: A EQUAÇÃO SECANTE POSSUI SOLUÇÃO EM B.

PROVA: DEFINIMOS $g : [0,1] \longrightarrow \mathbb{R}^n$ PONDO

$$g(t) = \nabla f(x^k + t(x^{k+1} - x^k)) .$$

PELO TEO. FUNDAMENTAL PARA INTEGRAIS DE LINHA,

$$\int_0^1 \nabla g(t)\, dt = g(1) - g(0) .$$

AGORA,

$$\nabla g(t) = \nabla^2 f(x^k + t(x^{k+1} - x^k))\,(x^{k+1} - x^k)$$

ASSIM

$$\underbrace{\left[\int_0^1 \nabla^2 f(x^k + t(x^{k+1} - x^k))\, dt\right]}_{\text{MATRIZ } m \times m.} (x^{k+1} - x^k) = g(1) - g(0)$$

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

DESTA FORMA, $\int_0^1 \nabla^2 f(x^k + t(x^{k+1} - x^k))\, dt$ É UMA SOLUÇÃO DA EQUAÇÃO SECANTE.

HÁ VÁRIAS SOLUÇÕES DA EQUAÇÃO SECANTE. UMA DELAS É REALIZAR UMA "CORREÇÃO DE POSTO 2":

$$B_{k+1} = B_k + \alpha u u^t + \beta v v^t .$$

OBS.: O POSTO DE $\alpha u u^t + \beta v v^t$ É NO MÁXIMO 2.

ESCOLHEMOS $u = y$ E $v = B_k s$. DAÍ,

$$B_{k+1} = B_k + \alpha y y^t + \beta B_k s s^t B_k$$ ($B_k$ É SIMÉTRICA

QUEREMOS $y = B_{k+1} s$ (EQ. SECANTE). ASSIM,

$$y = B_{k+1} s = B_k s + \alpha y (y^t s) + \beta B_k s (s^t B_k s).$$

UMA SOLUÇÃO É

$$\alpha = \frac{1}{y^t s}, \qquad \beta = -\frac{1}{s^t B_k s}.$$

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: ALÉM DAS PROPRIEDADES TEÓRICAS (SIMETRIA / POSITIVIDADE), TEM BOM DESEMPENHO NUMÉRICO EM PROBLEMAS GERAIS.

---

<u>TEOREMA</u>: SUPONHA QUE $B_k$ SEJA SIMÉTRICA E DEFINIDA POSITIVA. ENTÃO, SE $\underline{y^t s > 0}$,

- $B_{k+1}$ ESTÁ BEM DEFINIDA;
- $B_{k+1}$ É SIMÉTRICA;
- $B_{k+1}$ É DEF. POSITIVA.

PROVA: SE $B_{k+1}$ É BEM DEFINIDA, É SIMÉTRICA POIS É SOMA DE MATRIZES SIMÉTRICAS. SUPONHA QUE $B_k$ ESTEJA BEM DEFINIDA, E $y^t s > 0$ (NA ITERAÇÃO $k-1$). MULTIPLICANDO $B_k$ POR $s$ EM AMBOS OS LADOS,

$$s^t B_k s = \cancel{s^t B_{k-1} s^t} + s^t y - \cancel{s^t B_{k-1} s} = s^t y > 0.$$

ASSIM A CONTA PARA A MATRIZ $B_{k+1}$ É POSSÍVEL, E A ITERAÇÃO BFGS É BEM DEFINIDA.

AGORA, CONSIDERE O PRODUTO $x^t B_{k+1} x$:

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

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

COMO $B_k$ É DEFINIDA POSITIVA, $B_k$ POSSUI DECOMPOSIÇÃO DE CHOLESKY

$$B_k = G G^t.$$

ASSIM,

$$s^t B_k s = (s^t G)(G^t s) = \|\underline{G^t s}\|^2,$$

$$x^t B_k x = (x^t G)(G^t x) = \|\underline{G^t x}\|^2 \quad \text{E}$$

$$x^t B_k s = (x^t G)(G^t s) = (\underline{G^t x})^t (\underline{G^t s})$$

DE CAUCHY-SCHWARTZ OBTEMOS

$$(G^t x)^t (G^t s) \leq \|G^t x\|^2 \cdot \|G^t s\|^2 .$$

OU SEJA,

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

ASSIM,

$$x^t B_{k+1} x \geq 0 \qquad (B_{k+1} \text{ É SEMI-DEF. POSITIVA}).$$

AGORA, $\underline{x^t B_{k+1} x = 0} \Longrightarrow x^t y = 0$ E $(*) = 0$.

SABEMOS QUE $(*)$ SÓ SE REALIZA COMO IGUALDADE SE $G^t x$ E $G^t s$ FOREM COLINEARES. ISTO É, $G^t x = \mu G^t s$. COMO $G^t$ É INVERSÍVEL (POIS É O FATOR DE CHOLESKY DE $B_k$), ENTÃO $x = \mu s$.

DAÍ, $\underbrace{x^t y}_{=0} = \mu \underbrace{s^t y}_{>0 \text{ (HIP)}} \implies \mu = 0 \implies x = 0.$

OU SEJA,

$$x^t B_{k+1} x > 0 \,, \quad \forall x \neq 0 .$$

($B_{k+1}$ É DEF. POSITIVA). ∎

---

<u>OBS.</u>: O TEOREMA DIZ QUE $B_k$ DEF. POSIT./SIMÉTRICA $\implies B_{k+1}$ DEF. POSIT. /SIMÉTRICA (SE $y^t s > 0$).

INICIAMOS ENTÃO $B_0 = I$ (IDENTIDADE).

CASO ACONTEÇA DE $y^t s \not> 0$, A ATUALIZAÇÃO BFGS PODE NÃO SER BEM DEFINIDA OU NÃO SER DEF. POSITIVA. SOLUÇÃO: GAMBIARRA:

TROCAR $y$ POR

$$\hat{y} := \theta y + (1-\theta) B_k s$$

PARA UM $\theta \in (0,1]$ ADEQUADO.

- $\theta = 1 \Rightarrow$ BFGS USUAL
- $\theta \approx 0 \Longrightarrow \hat{y}^t s = \theta (y^t s) + (1-\theta) \underbrace{(s^t B_k s)}_{>0}$ TENDE A FICAR POSITIVO.

PROBLEMA: SE $\theta < 1$, A EQUAÇÃO SECANTE PODE FALHAR $(B_{k+1} s \neq \hat{y})$.