Região de confiança - métodos específicos

No esquema geral, devemos calcular d^k solução aproximada do modelo

$$\min_{d} \; m(d) = f(x^k) + \nabla f(x^k)^t d + \frac{1}{2} d^t B_k d$$
$$\text{s.a. } \|d\| \leq \Delta_k .$$

Para garantir convergência, d^k deve satisfazer a hipótese H3 (veja anotações sobre convergência).

O objetivo é discutir diferentes formas de fazer isso.

1ª forma: passo de Cauchy

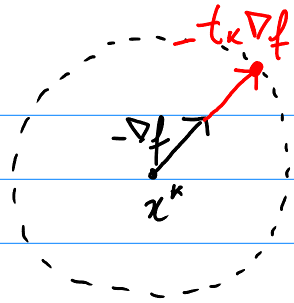
Supomos que $d^k = -t_k \nabla f(x^k)$, onde $t_k > 0$ é solução de

$$\min_{t} \; m(-t \nabla f(x^k)) \quad \text{s.a. } \|t \nabla f(x^k)\| \leq \Delta_k$$

($\|\cdot\|$ = norma Euclideana)

Resolvendo (omitindo x^k):

- Se $\nabla f^t B_k \nabla f \leq 0$ então o termo quadrático $d^t B_k d = t^2 \nabla f^t B_k \nabla f \leq 0$ não influencia na minimiza ção, e a solução estará na borda:



$$\|t_k \nabla f\| = \Delta_k \Rightarrow t_k = \frac{\Delta_k}{\|\nabla f\|}.$$

• Se $\nabla f^t B_k \nabla f > 0$ então a solução pode não estar na borda. Temos que olhar para o minimizador irrestrito de $m(d)$:

$$\frac{d}{dt} m(-t\nabla f) = -\|\nabla f\|^2 + t(\nabla f^t B_k \nabla f) = 0$$

$$\Rightarrow t^* = \frac{\|\nabla f\|^2}{\nabla f^t B_k \nabla f} > 0.$$

Caso $t^* < \frac{\Delta_k}{\|\nabla f\|}$ então $t_k = t^*$. Caso contrário, $t_k = \frac{\Delta_k}{\|\nabla f\|}$. Em resumo,

$$(*) \quad d^k = -t_k \nabla f(x^k) \quad \text{onde} \quad t_k = \begin{cases} \frac{\Delta_k}{\|\nabla f\|} & \text{se } \nabla f^t B_k \nabla f \le 0 \\ \min\left\{\frac{\Delta_k}{\|\nabla f\|}, \frac{\|\nabla f\|^2}{\nabla f^t B \nabla f}\right\} & \text{c.c.} \end{cases}$$

Com o passo de Cauchy, há convergência:

<u>Teorema:</u> O passo de Cauchy satisfaz

$$m(0) - m(d^k) \ge \frac{1}{2} \|\nabla f(x^k)\| \cdot \min\left\{\Delta_k, \frac{\|\nabla f(x^k)\|}{\|B_k\|}\right\} \quad (\text{H3 com } c = \tfrac{1}{2})$$

Prova: Como $d^k = -t_k \nabla f$, temos

$$m(0) - m(d^k) = f(x^k) - f(x^k) - \nabla f^t d^k - \frac{1}{2} d^{k^t} B_k d^k$$
$$= t_k \|\nabla f\|^2 - \frac{1}{2} t_k^2 \nabla f^t B_k \nabla f .$$

CASO 1: $\nabla f^t B_k \nabla f \leq 0$

De $(*)$, $t_k = \dfrac{\Delta_k}{\|\nabla f\|}$ e $m(0) - m(d^k) \geq t_k \|\nabla f\|^2 = \Delta_k \|\nabla f\|.$

Em particular, $m(0) - m(d^k) \geq \frac{1}{2} \|\nabla f\| \Delta_k .$ (1)

CASO 2: $\nabla f^t B_k \nabla f > 0$ e $\dfrac{\Delta_k}{\|\nabla f\|} \leq \dfrac{\|\nabla f\|^2}{\nabla f^t B_k \nabla f}$.

Neste caso, $(*)$ fornece $t_k = \dfrac{\Delta_k}{\|\nabla f\|} \leq \dfrac{\|\nabla f\|^2}{\nabla f^t B_k \nabla f}$, donde segue que $t_k^2 \nabla f^t B_k \nabla f \leq t_k \|\nabla f\|^2 = \|\nabla f\| \Delta_k$. Daí

$$m(0) - m(d^k) \geq \|\nabla f\| \Delta_k - \frac{1}{2} \|\nabla f\| \Delta_k = \frac{1}{2} \|\nabla f\| \Delta_k . \quad (2)$$

CASO 3: $\nabla f^t B_k \nabla f > 0$ e $\dfrac{\Delta_k}{\|\nabla f\|} > \dfrac{\|\nabla f\|^2}{\nabla f^t B_k \nabla f}$.

De $(*)$ vem $t_k = \dfrac{\|\nabla f\|^2}{\nabla f^t B_k \nabla f}$ e logo

$$m(0) - m(d^k) = \frac{\|\nabla f\|^4}{\nabla f^t B_k \nabla f} - \frac{1}{2} \cdot \frac{\|\nabla f\|^4}{\nabla f^t B_k \nabla f} = \frac{1}{2} \frac{\|\nabla f\|^4}{\nabla f^t B_k \nabla f} .$$

Pela desigualdade de Cauchy-Schwartz, temos $\nabla f^t(B_k \nabla f) \leq \|\nabla f\| \cdot \|B_k \nabla f\| \leq \|B_k\| \cdot \|\nabla f\|^2$. Assim,

$$m(0) - m(d^k) \geq \frac{1}{2} \frac{\|\nabla f\|^4}{\|B_k\| \cdot \|\nabla f\|^2} = \frac{1}{2} \|\nabla f\| \cdot \frac{\|\nabla f\|}{\|B_k\|} . \quad (3)$$

Finalmente, de (1), (2) e (3) chegamos ao resultado ▨

Observações:

1) O passo de Cauchy, apesar de ter baixo custo computacional, fornece um método lento próximo à solução pois a direção $d^k = -t_k \nabla f(x^k)$ é similar à do método do gradiente.

2) Note que a informação em B_k não é usada na direção (que é paralela à $-\nabla f(x^k)$). B_k é usada apenas para escalar $-\nabla f(x^k)$ (cálculo de t_k). Essa deficiência fica evidente para $B_k = \nabla^2 f(x^k)$, pois o modelo com essa B_k e "d livre" remonta ao

método de Newton, que é muito mais rápido que o método do gradiente.

3) Portanto é razoável aproveitar B_k ao máximo, idealmente no estilo Newton / Quase-Newton.

O próximo método procura fazer isso!

2ª forma: método dogleg

Neste método, a solução aproximada d^k do modelo $\min_d m(d)$ s.a. $\|d\| \leq \Delta_k$ aproveita melhor B_k.

Ele se aplica à B_k definida positiva e simétrica (opções para tal B_k em aula anterior). Quando é possível $B_k = \nabla^2 f(x^k)$, o passo do método dogleg coincide com Newton caso a direção Newtoniana satisfaça $\|d\| \leq \Delta_k$.

Dado o modelo ao redor de x^k, considere os seguintes pontos:

- x_U^k: minimizador irrestrito de m na direção $-\nabla f(x^k)$, isto é,

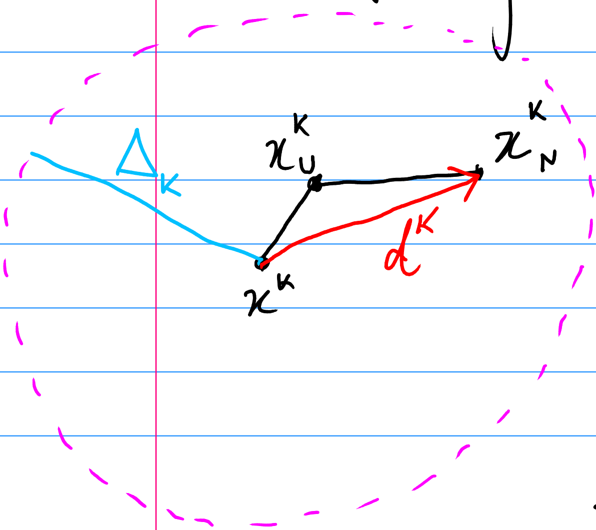
$$x_U^k = x^k - t^* \nabla f(x^k), \quad t^* = \arg\min_t \, m(-t \nabla f(x^k)).$$

- x_N^k: minimizador irrestrito de m, isto é,

$$x_N^k = x^k + d_N^*, \quad d_N^* = \arg\min_d \, m(d).$$

Supondo B_k definida positiva, esses pontos estão bem definidos pois $m(d)$ é uma quadrática estritamente convexa.

O método dogleg minimiza $m(d)$ sobre a poligonal que liga x^k, x_U^k e x_N^k, respeitando $\|d\| \leq \Delta_k$. Algumas situações:

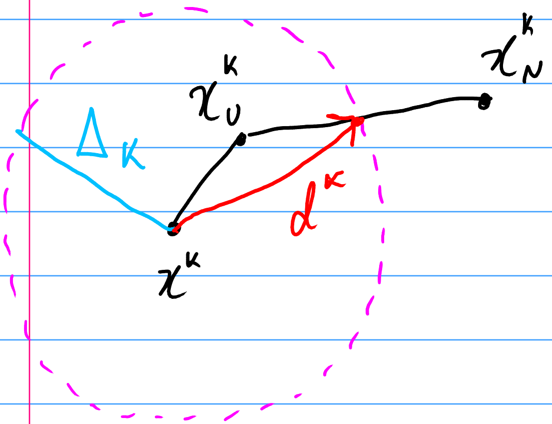


1) $\|x^k - x_N^k\| \leq \Delta_k$

A direção será $d^k = x_N^k - x^k$.

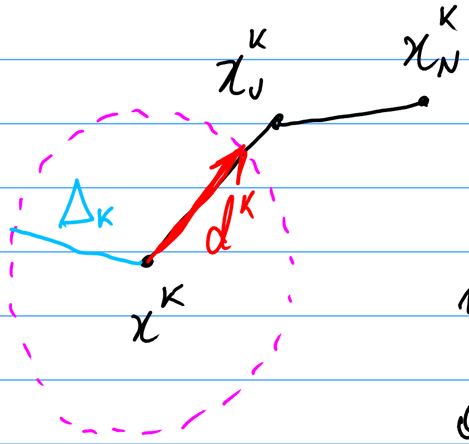
Ou seja, o ponto dogleg é x_N^k.

Se $B_k = \nabla^2 f(x^k)$, este é o passo de Newton.



2) $\|x^K - x_U^K\| \le \Delta_K < \|x^K - x_N^K\|$

O ponto dogleg $x^K + d^K$ estará na borda, intermediário entre x_U^K e x_N^K.



3) $\|x^K - x_U^K\| > \Delta_K$

O ponto dogleg $x^K + d^K$ estará na borda, e coincide com o passo de Cauchy.

¿Outra situação pode ocorrer? ¡NÃO!

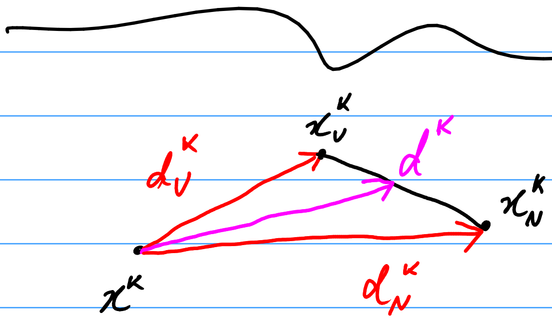
↳ "Teorema: (i) $\|x_p - x^K\|$ cresce quando o ponto x_p sobre a poligonal vai de x^K à x_N^K.

(a poligonal corta a borda no máximo 1 vez)

(ii) $m(x_p - x^K)$ é não decrescente ao longo da poligonal.

(a fim de minimizar m, devemos caminhar de x^k à x_N^k. Neste sentido, x_N^k é o melhor ponto — de fato é aquele que vem da minimização sem $\|d\| \leq \Delta_k$ com d livre)

* Veja o Lema 5.40 do livro de Karas e Ribeiro ou o Lema 4.2 de Nocedal e Wright.



(**)

Direções

$$d_U^k = x_U^k - x^k = -t^* \nabla f(x^k)$$

$$d_N^k = x_N^k - x^k \quad \text{(tipo Newton)}$$

d^k (direção dog leg)

<u>Calculando d_U^k:</u>

$$\min_t \; m(-t\nabla f(x^k))$$

$$0 = \frac{d}{dt} m(-t^*\nabla f(x^k)) = -\|\nabla f\|^2 + t^* \nabla f^t B_k \nabla f \Rightarrow t^* = \frac{\nabla f^t \nabla f}{\nabla f^t B_k \nabla f}$$

$$\Rightarrow \boxed{d_U^k = -\left(\frac{\nabla f^t \nabla f}{\nabla f^t B_k \nabla f}\right)\nabla f}$$

Calculando d_N^k: $\min\limits_{d} m(d)$

$$0 = \nabla m(d_N^k) = \nabla f + B_k d_N^k \Rightarrow \boxed{B_k d_N^k = -\nabla f(x^k)}$$

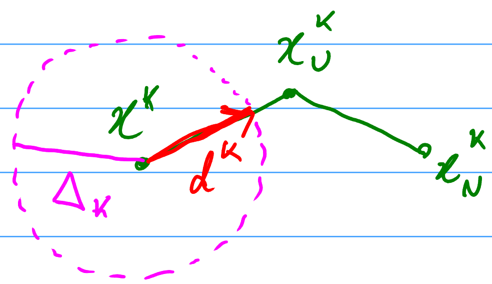
Algoritmo dog leg (para cálculo de d^k)

Entrada: x^k, $\Delta_k > 0$

→ calcule $d_U^k = -\left(\frac{\nabla f(x^k)^t \nabla f(x^k)}{\nabla f(x^k)^t B_k \nabla f(x^k)}\right)\nabla f(x^k)$

→ Se $\|d_U^k\| > \Delta_k$

→ $d^k = -\frac{\Delta_k}{\|\nabla f(x^k)\|}\nabla f(x^k)$

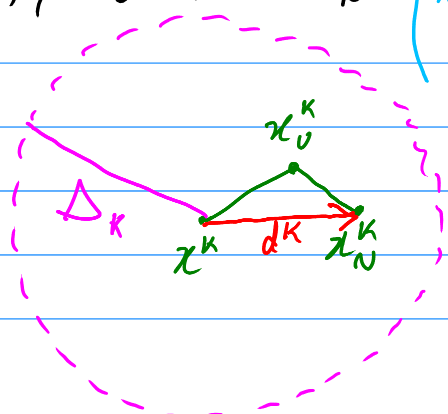


senão

→ resolva $B_k d_N = -\nabla f(x^k)$, obtendo d_N^k (por exemplo usando Cholesky)

→ Se $\|d_N^k\| \le \Delta_k$

$d^k = d_N^k$

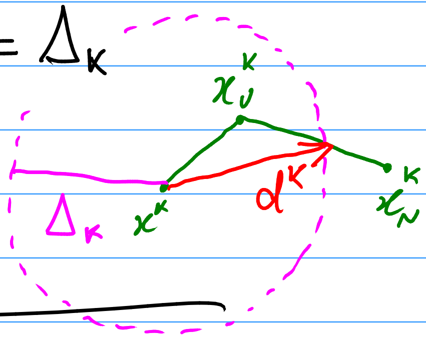


senão

→ calcule $\alpha_k \in [0,1]$ tal que

$$\|d_N^k + \alpha_k(d_N^k - d_U^k)\| = \Delta_k$$

→ $d^k = d_U^k + \alpha_k(d_N^k - d_U^k)$



No último caso ($\|x_U^k - x^k\| \leq \Delta_k < \|x_N^k - x^k\|$), a direção dogleg é combinação convexa das direções d_U^k e d_N^k (veja a figura (**)):

$$d^k = d_U^k + \alpha_k(d_N^k - d_U^k) = (1-\alpha_k)\, d_U^k + \alpha_k d_N^k \quad , \quad \alpha_k \in [0,1]$$

__Exercício:__ mostre que é possível calcular α_k como raiz positiva de uma equação do 2° grau.

Vamos mostrar agora que o método de região de confiança com passo dogleg converge, mostrando que H3 é satisfeita.

__Teorema:__ O passo dogleg satisfaz

$$m(0) - m(d^k) \geq \frac{1}{2} \|\nabla f(x^k)\| \cdot \min\left[\Delta_k, \frac{\|\nabla f(x^k)\|}{\|B_k\|}\right].$$

Prova: Seja d_C^k a direção de Cauchy. Como mostramos, $m(0) - m(d_C^k) \geq \frac{1}{2} \|\nabla f\| \cdot \min\left\{\Delta_k, \frac{\|\nabla f\|}{\|B_k\|}\right\}$. Agora, como $m(x - x^k)$ não cresce ao longo da poligonal saindo de x^k à x_N^k (comentário anterior), temos

$$m(d^k) = m(x_{dogleg} - x^k) \leq m(x_{Cauchy} - x^k) = m(d_C^k).$$

Logo $m(0) - m(d^k) \geq m(0) - m(d_C^k)$, donde segue o resultado. ∎

3ª forma: gradientes conjugados de Steinhaug

Gradientes conjugados (GC) é aplicado à minimização irrestrita de quadráticas (disciplina "Otimização I"). Aqui, GC é adaptado para lidar com a restrição $\|d\| \leq \Delta_k$. Este método é adequada à problemas grandes. Veja o livro de Karas e Ribeiro para mais detalhes.