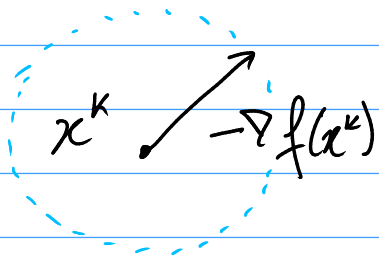


Regiões de confiança

Referência 1: Ribeiro, A. A; Karas, E. W. Otimização contínua. Cengage, 2014

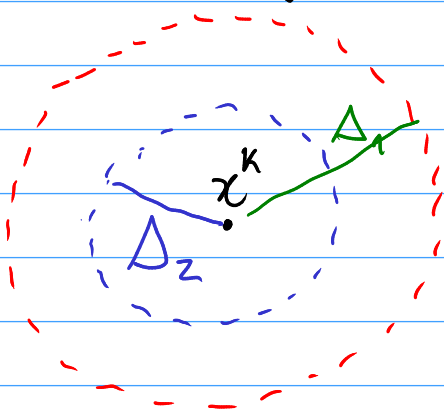
Estratégia de busca linear



$$x^{k+1} = x^k + t_k(-\nabla f(x^k)),$$
$$f(x^{k+1}) < f(x^k) \quad t_k \in (0, 1]$$

- diminuir f ao longo de uma direção de descida (local) a partir do ponto corrente.

Estratégia de regiões de confiança



$\min_x \cancel{f(x)}$ trocamos por um modelo de f .

s.a. $\|x - x^k\| \leq \Delta_1$

O modelo de f (aproxima localmente f) é fácil de minimizar.

Problema: o minimizador do modelo na bola de raio Δ pode aumentar f .!
 \hookrightarrow diminuimos o raio Δ até f diminuir.

- Buscar um ponto que minimiza um modelo simplificado de f restrito à uma vizinhança de x^k .

Que modelo simplificado?

$$\min_x f(x) \quad (\text{restrito})$$
$$\text{s.a. } x \in \mathbb{R}^n$$

Modelo usual: aproximação quadrática de f em x^k

$$f(x) = f(x^k) + \nabla f(x^k)^t (x - x^k) + \frac{1}{2} (x - x^k)^t \nabla^2 f(x^k) (x - x^k) + r(\|x - x^k\|)$$

$$\text{onde } \frac{r(\|x - x^k\|)}{\|x - x^k\|} \xrightarrow{x \rightarrow x^k} 0 \quad (\text{Taylor de ordem 2}).$$

(supomos f com derivadas de 2ª ordem contínuas)

Assim, para $x \approx x^k$,

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

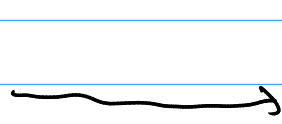
modelo m

Notação: $d = x - x^k$

$$m(d) = f(x^k) + \nabla f(x^k)^T d + \frac{1}{2} d^T \nabla^2 f(x^k) d.$$

$$\min_x f(x)$$

s.a. $x \in \mathbb{R}^n$



$$\min_d m(d)$$

$$\|d\| \leq \Delta_k.$$

\downarrow d^k solução

$$x^{k+1} = x^k + d^k.$$

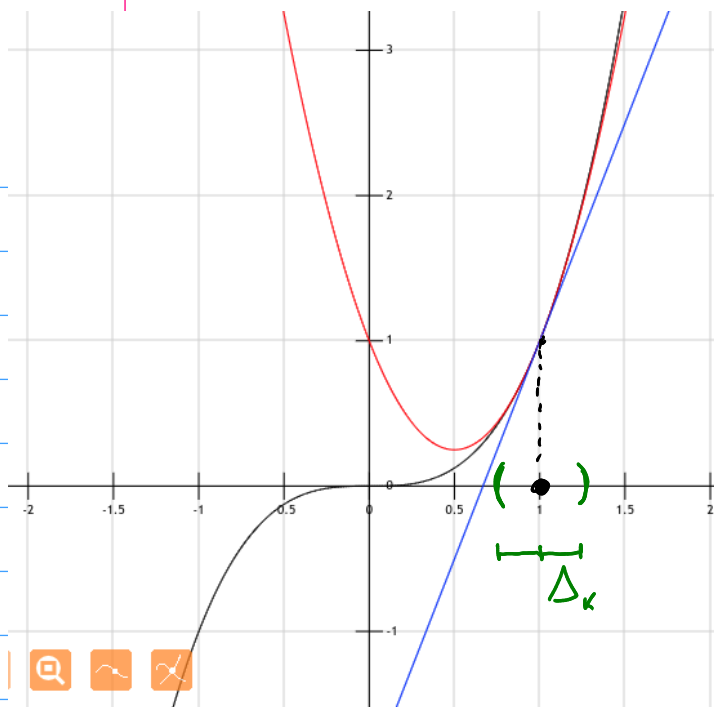
m aproxima f localmente...

$\hookrightarrow \Delta_k$ deve ser controlado

\hookrightarrow Se f aumentar, diminua Δ_k

\hookrightarrow Se f diminuir "pouco", mantenha Δ_k

\hookrightarrow Se f diminuir "muito", aumente Δ_k



$$f(x) = x^3$$

$$x^k = 1$$

AZUL = aprox. linear

VERMELHO = aprox. quad.

"O modelo só é confiável próximo a x^k ".

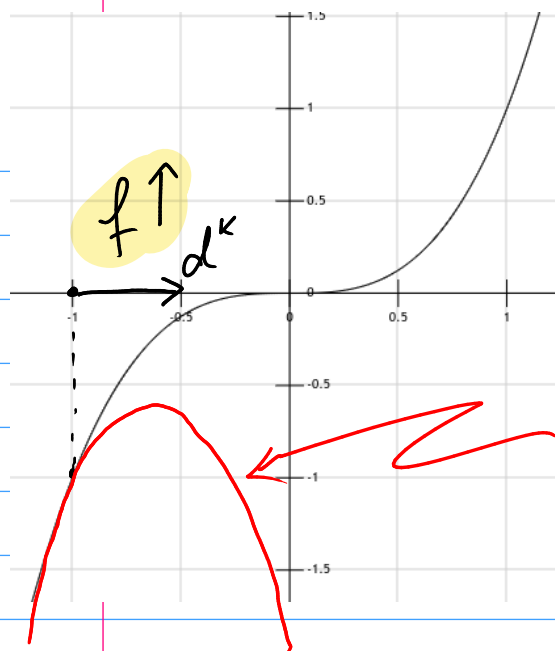
$\|x - x^k\| \leq \Delta_k$ ($\|d\| \leq \Delta_k$): Δ_k é o raio de confiança.

Problemas com uso de $\nabla^2 f(x^k)$ em $m(d)$:

- 1) alto custo computacional / armazenamento de $\nabla^2 f(x^k)$.
- 2) se $\nabla^2 f(x^k)$ não for semidefinida positiva então m não é convexa. Daí

resolver $\min m(d)$ fica complicado.
s.a. $\|d\| \leq \Delta$

Pior: d^k pode ser de subida!



$$f(x) = x^3 \quad \left\{ \begin{array}{l} \text{Em } x^k = -1, \\ m(d) = 3d - 3d^2, \\ \text{cujo minimizador} \\ \text{é } d^k = \frac{1}{2} \quad (f \uparrow) \end{array} \right.$$

aproximação quadrática não
comum.

Uma solução: Trocar $\nabla^2 f(x^k)$ por uma matriz B_k simétrica e definida positiva, e que aproxime $\nabla^2 f(x^k)$ em algum sentido.

$$m(d) = f(x^k) + \nabla f(x^k)^T d + \frac{1}{2} d^T B_k d.$$

Alternativas para B_k :

1) quase-Newton (BFGS, DFP)

↙
bons resultados
numéricos.

2) $B_k = \nabla^2 f(x^k) + \sigma_k I$, onde $\sigma_k \gg 1$
é tal que B_k seja definida positiva.
↳ σ_k : estimativa do menor autovalor de

$\nabla^2 f(x^k)$ (\uparrow custo) ou "discos de Gerschgorin"
+ heurísticas (\downarrow custo).

$$\min f(x) \longrightarrow \min m(d)$$

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

$$\downarrow$$
$$x^{k+1} = x^k + d^k$$

$f(x^{k+1})$ melhorou em relação a $f(x^k)$?

Critério de aceitação/rejeição do novo ponto

Ponto corrente: x^k

Novo ponto: $x^k + d^k$.

• redução real de f :

$$ared = f(x^k) - f(x^k + d^k)$$

• redução do modelo (redução predita)

$$pred = m(0) - m(d^k)$$

Medida de aceitação

$$\rho_k = \frac{ared}{pred}$$

Situações "boa": $ared$ é grande em relação
à $pred \longrightarrow \rho_k$ grande.

Situações "ruim": $ared$ é pequeno em relação
à $pred \longrightarrow \rho_k$ pequeno.

Esquema de regiões de confiança

• Dados $x^0 \in \mathbb{R}^n$, $\Delta_0 > 0$, $\eta \in [0, \frac{1}{4})$, $k \leftarrow 0$

• Repita enquanto $\nabla f(x^k) \neq 0$ ($\|\nabla f(x^k)\| \leq \epsilon$)

\rightarrow resolver o modelo quadrático centrado em x^k :

$$\min m(d)$$

$$\text{s.a. } \|d\| \leq \Delta_k$$

 obtendo d^k

\rightarrow calcule ρ_k

→ se $\rho_k > \eta$

$$\hookrightarrow x^{k+1} = x^k + d^k$$

(redução boa,
aceitamos o ponto)

senão

$$\hookrightarrow x^{k+1} = x^k$$

(redução ruim
→ não damos o passo)

→ se $\rho_k < \frac{1}{4}$

$$\hookrightarrow \Delta_{k+1} = \frac{1}{2} \Delta_k$$

(redução ruim,
reduzimos o
raio)

senão

→ se $\rho_k > \frac{3}{4}$ e $\|d^k\| = \Delta_k$

$$\hookrightarrow \Delta_{k+1} = 2\Delta_k$$

(redução boa e o modelo
alcançou a borda da região
de confiança → aumentamos
o raio)

senão

$$\hookrightarrow \Delta_{k+1} = \Delta_k$$

(redução foi boa, mas
a borda não foi atingida
→ o raio atual é
adequado)

→ $k \leftarrow k+1$

Regiões de confiança - Convergência

<u>Problema</u>	<u>Modelo ao redor de x^k:</u>
P: $\min f(x)$	$\min_d m(d) = f(x^k) + \nabla f(x^k)^T d + \frac{1}{2} d^T B_k d$ <p>s.a. $\ d\ \leq \Delta_k$</p>

No método, damos o passo $x^{k+1} = x^k + d^k$ somente se

$0 \leq \eta < \rho_k = \frac{\text{ared}}{\text{pred}}$, onde $\text{ared} = f(x^k) - f(x^k + d^k)$
e $\text{pred} = m(0) - m(d^k)$. Como d^k é solução

aproximada do modelo, sempre $\text{pred} > 0$. Assim,

$\rho_k > 0 \Rightarrow \text{ared} > 0 \Rightarrow f(x^{k+1}) < f(x^k)$, $\forall k$,

durante o método.

Para provar convergência do método de regiões de confiança, imponho:

Hipóteses sobre o problema P:

H1) ∇f é lipschitz-contínuo, isto é, existe $L > 0$
tal que $\|\nabla f(x) - \nabla f(y)\| \leq L \|x - y\|$, $\forall x, y$.

H2) f é limitada inferiormente no conjunto de nível $N = \{x \in \mathbb{R}^n; f(x) \leq f(x^0)\}$ (x^0 pto inicial)

Hipóteses sobre as sequências geradas pelo algoritmo:

H3) d^k satisfaz, $\forall k$,

$$\rho(d^k) = m(0) - m(d^k) \geq c \|\nabla f(x^k)\| \cdot \min\{\Delta_k, \frac{\|\nabla f(x^k)\|}{\|B_k\|}\}$$
 para alguma constante $c \in (0, 1)$.

Aqui, $\|B_k\|$ é a norma matricial induzida pela norma $\|\cdot\|$ de \mathbb{R}^n .

H4) As matrizes B_k são uniformemente limitadas, isto é, existe $\beta > 0$ tal que $\|B_k\| \leq \beta, \forall k$.

Adicionalmente, lembre-se que no método temos

$$\|d^k\| \leq \Delta_k, \forall k.$$

Teorema: Suponha $H1, H2, H3$ e $H4$ válidas. Então

(i) caso $\eta = 0$, temos $\liminf_{k \rightarrow \infty} \|\nabla f(x^k)\| = 0$, isto é, existe subsequência $\{x^k\}_{k \in K}$ tal que $\lim_{k \in K} \|\nabla f(x^k)\| = 0$.

(ii) caso $\eta \in (0, 1/4)$, temos $\lim_{k \rightarrow \infty} \|\nabla f(x^k)\| = 0$ (a sequência inteira converge).

Prova: Veja seção 5.5.3 do livro de Karas e Ribeiro (5.3.3 na versão alternativa) OU

Teoremas 4.5 e 4.6 do livro de Nocedal e Wright

Observações:

- 1) O teorema garante convergência de qualquer método de regiões de confiança onde d^k é calculada para satisfazer $H3$. Exibiremos alguns deles...
- 2) Garantimos $H4$ escolhendo B_k adequadamente.