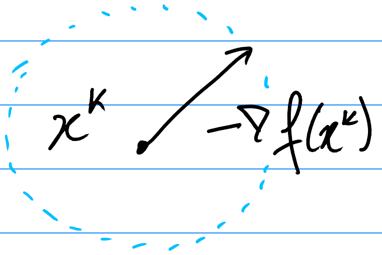
Região de confiança

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

Estratégia de busca linear

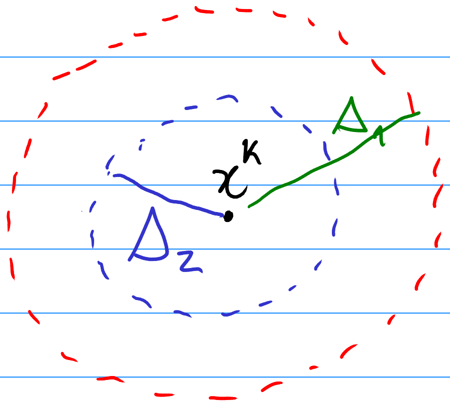


$$x^{k+1} = x^k + t_k(-\nabla f(x^k)),$$

$$f(x^{k+1}) < f(x^k) \qquad t_k \in (0,1]$$

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

Estratégia de região de confiança



$$\min_x \cancel{f(x)}$$

trocamos por um modelo de f.

$$\text{s.a. } \|x - x^k\| \le \Delta_1$$

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

Problema: o minimizador do modelo na bola de raio Δ pode aumentar f. !
↳ diminuímos 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?

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

Modelo usual: aproximação quadrática de f em x^k

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

onde $\dfrac{r(\|x - x^k\|)}{\|x - x^k\|} \xrightarrow[x \to x^k]{} 0$ (Taylor de ordem 2).

(supomos f com derivadas de 2^{a} ordem contínuas)

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

$$f(x) \approx \underbrace{f(x^k) + \nabla f(x^k)^t (x - x^k) + \frac{1}{2}(x - x^k)^t \nabla^2 f(x^k)(x - x^k)}_{\text{modelo } m}$$

Notação: $d = x - x^k$

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

$$\begin{array}{lcl} \min_x f(x) & \longrightarrow & \min_d m(d) \\ \text{s.a. } x \in \mathbb{R}^n & & \|d\| \leq \Delta_k . \end{array}$$

$$\downarrow d^k \text{ solução}$$

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

---

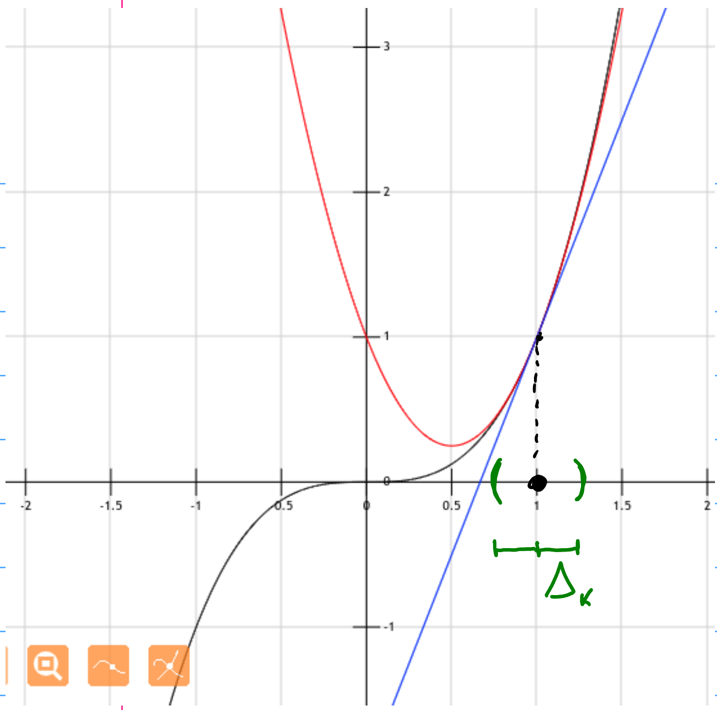
m aproxima f localmente...

↳ Δ_k deve ser controlado

↳ Se f aumentar, diminua Δ_k

↳ Se f diminuir "pouco", mantenha Δ_k

↳ 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 à x^k".

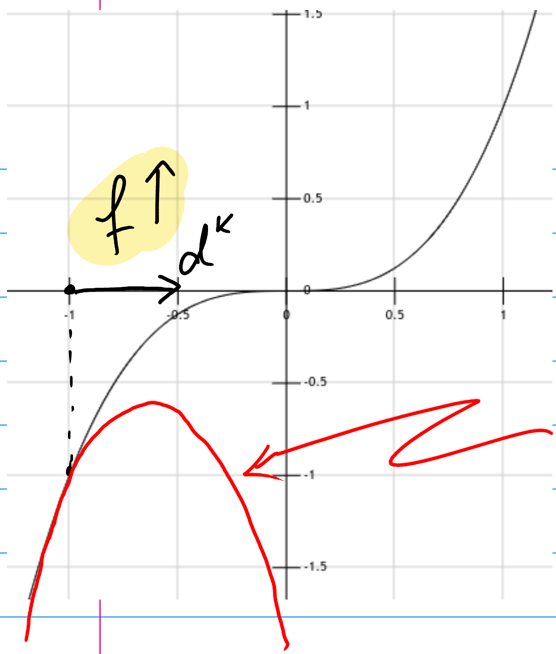
$\|x - x^k\| \leq \Delta_k$ $(\|d\| \leq \Delta_k)$: Δ_k é o raio de confiança.

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

Pior: d^k pode ser de subida!



$f(x) = x^3$

$x^k = -1$

Em $x^k = -1$,
$m(d) = 3d - 3d^2$,
cujo minimizador
é $d^k = \frac{1}{2}$ ($f\uparrow$)

aproximação quadrática não convexa.

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.

$\hookrightarrow \sigma_k$: estimativa do menor autovalor de

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

$$\min f(x) \longrightarrow \begin{array}{l} \min\ m(d) \\ \text{s.a. } \|d\| \leq \Delta \end{array}$$

$$\downarrow$$

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

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

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

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

Situação "boa": ared é grande em relação à pred $\longrightarrow$ ρ_k grande.

Situação "ruim": ared é pequeno em relação à pred $\longrightarrow$ ρ_k pequeno.

Esquema de região de confiança

- Dados $x^0 \in \mathbb{R}^n$, $\Delta_0 > 0$, $\eta \in [0, 1/4)$, $k \leftarrow 0$
- Repita enquanto $\nabla f(x^k) \neq 0$ $\left(\|\nabla f(x^k)\| \leq \varepsilon \right)$
  - $\to$ resolva o modelo quadrático centrado em x^k:

$$\min \; m(d)$$
$$\text{s.a. } \|d\| \leq \Delta_k$$

obtendo d^k

  - $\to$ calcule ρ_k

→ se $\rho_k > \eta$

↳ $x^{k+1} = x^k + d^k$

(redução boa, aceitamos o ponto)

senão

↳ $x^{k+1} = x^k$

(redução ruim ↳ não damos o passo)

→ se $\rho_k < 1/4$

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

(redução ruim, reduzimos o raio)

senão

↳ se $\rho_k > 3/4$ e $\underbrace{\|d^k\| = \Delta_k}$

↳ $\Delta_{k+1} = 2\Delta_k$

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

senão

↳ $\Delta_{k+1} = \Delta_k$

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

↳ $k \leftarrow k+1$

Região de confiança – Convergência

Problema	**Modelo ao redor de x^k:**
$P: \min f(x)$	$\min_d m(d) = f(x^k) + \nabla f(x^k)^t d + \frac{1}{2} d^t B_k d$ $s.a. \ \|d\| \leq \Delta_k$

No método, damos o passo $x^{k+1} = x^k + d^k$ somente se $0 \leq \eta < \rho_k = \dfrac{ared}{pred}$, onde $ared = f(x^k) - f(x^k + d^k)$ e $pred = m(0) - m(d^k)$. Como d^k é solução aproximada do modelo, sempre $pred > 0$. Assim,

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

Para provar convergência do método de região de confiança, impomos:

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, $\forall k$,

$$pred = m(0) - m(d^k) \geq c \, \|\nabla f(x^k)\| \cdot \min\left\{ \Delta_k , \frac{\|\nabla f(x^k)\|}{\|B_k\|} \right\}$$

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 $\liminf_{k \to \infty} \|\nabla f(x^k)\| = 0$, 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 $\lim_{k \to \infty} \|\nabla f(x^k)\| = 0$ (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 região de confiança onde d^k é calculada para satisfazer H3. Exibiremos alguns deles...

2) Garantimos H4 escolhendo B_k adequadamente.