Quadrados mínimos não linear

- quadrados mín. linear: resolver $Ax=b$
por $\min_x \frac{1}{2}\|Ax-b\|_2^2$.

- quadrados mín. não linear:
resolver o sistema não linear $R(x)=0$
por $\min_x \frac{1}{2}\|R(x)\|_2^2$

$R : \mathbb{R}^n \longrightarrow \mathbb{R}^m$,

$R(x) = \begin{bmatrix} r_1(x) \\ \vdots \\ r_m(x) \end{bmatrix}$, onde $r_i : \mathbb{R}^n \longrightarrow \mathbb{R}$,

∇r_i contínuo e

r_i é possivelmente

função não linear.

$$R(x) = 0 \iff \| R(x) \|_2^2 = \sum_{i=1}^{m} (r_i(x))^2 = 0.$$

Problema de quadrados mínimos:

$$(*) \quad \min_x \frac{1}{2}\|R(x)\|^2 = \frac{1}{2}\sum_{i=1}^{m}(r_i(x))^2$$

- Se $R(x)=0$ não tem solução, $(*)$ visa reduzir os resíduos $r_i(x)$.

Como resolver $(*)$?

Estratégias similares à Newton.

$$f(x) = \frac{1}{2} \| R(x) \|_2^2 = \frac{1}{2} \sum_{i=1}^{m} (r_i(x))^2$$

$$\nabla f(x) = \sum_{i=1}^{m} r_i(x) \nabla r_i(x) = J(x)^T R(x),$$

onde

$$J(x) = \begin{bmatrix} \nabla r_1(x)^T \\ \vdots \\ \nabla r_m(x)^T \end{bmatrix} \quad (m \times n)$$

é a jacobiana de R.

$$\nabla^2 f(x) = \sum_{i=1}^{m} \left(\nabla r_i(x) \nabla r_i(x)^T + r_i(x) \nabla^2 r_i(x) \right)$$

$$= J(x)^T J(x) + S(x),$$

onde

$$S(x) = \sum_{i=1}^{m} r_i(x) \nabla^2 r_i(x).$$

Vamos desconsiderar o termo $S(x)$.

Isso é razoável se

- estamos caminhando para uma solução x^* de $R(x)=0$, dado que neste caso $r_i(x) \approx 0$, $\forall i$.
- r_i's são "quase" lineares. Neste caso, $\nabla^2 r_i \approx 0$.

Em ambos os casos, $S(x) = \sum_i r_i(x) \nabla^2 r_i(x) \approx 0$.

O passo de Newton para $f(x)=0$, desconsiderando $S(x)$, é

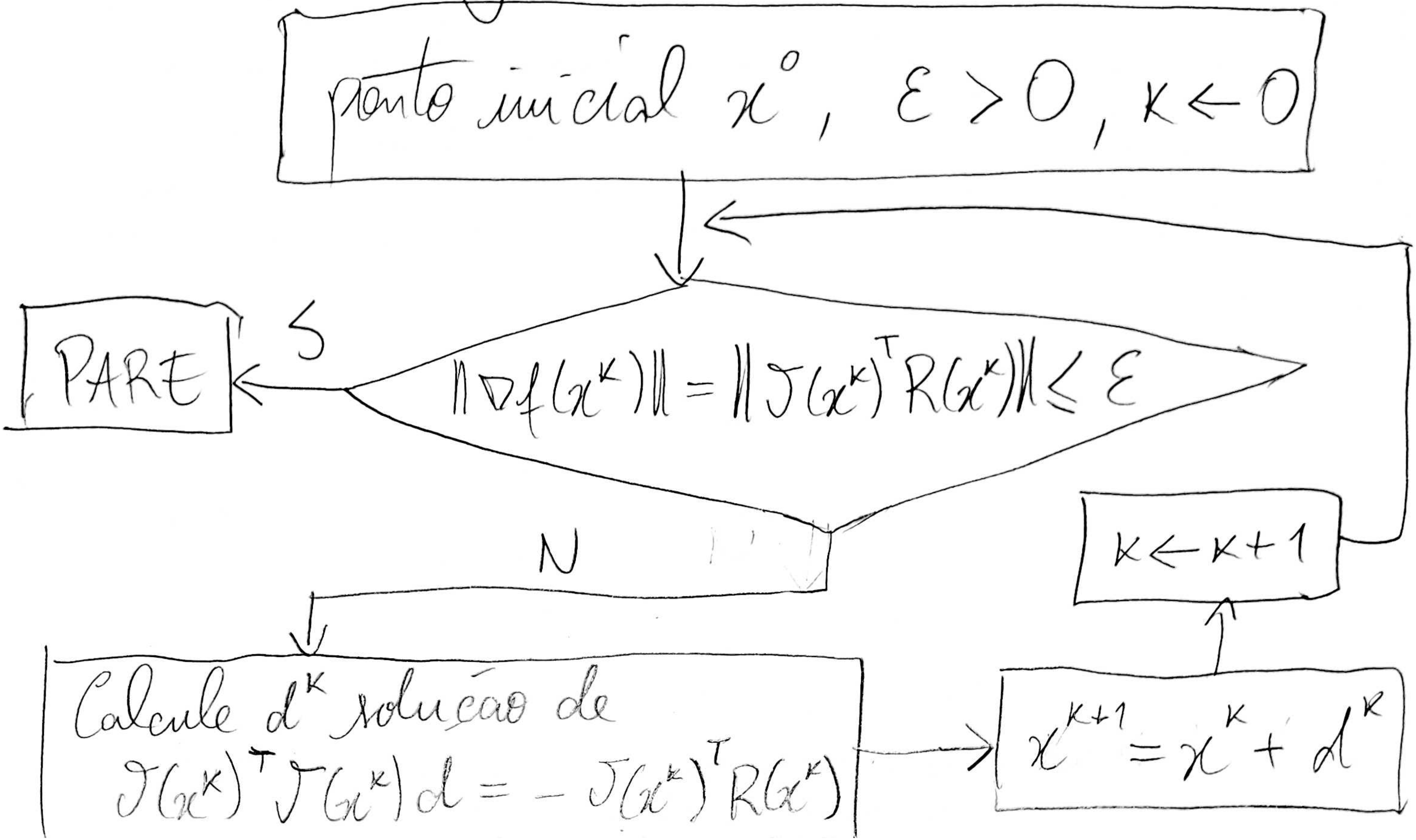
$$x^{k+1} = x^k + d^k, \quad \text{onde } d^k$$

é solução de

$$J(x^k)^T J(x^k)\, d = -J(x^k)^T R(x^k).$$

$$\left(\text{"} \nabla^2 f(x^k)\, d \approx -\nabla f(x^k) \text{"} \right).$$

Método de Gauss-Newton



Observações:

1) Descartar $S(x) = \sum_{i=1}^{m} r_i(x) \nabla^2 r_i(x)$ evita calcular m Hessianas!

2) A expressão $(J^T J)d = -J^T R$ é a equação normal do problema de quadrados mínimos linear

$$\min_{d} \frac{1}{2} \| J(x^k) d + R(x^k) \|_2^2 .$$

Logo podem ser aplicadas técnicas do caso linear (Cholesky de $J^T J$, QR ou SVD de J).

Exercício: prove esta afirmação!

3) Note que a matriz $J(x^k)^T J(x^k)$ da equação normal muda de iteração para iteração. É possível implementar o algoritmo sem precisar fazer a fatoração QR

de $J(x^k)$ do zero ($J(x^k)$ aproveita a QR de $J(x^{k-1})$ — veja livro Nocedal) 11

4) Em problemas muito grandes (e com J esparsa), podemos aplicar gradientes conjugados para calcular d^k!

5) A convergência é apenas local (pois no fundo é Newton). Logo

costuma-se empregar uma busca linear em Gauss-Newton para globalizá-lo:

$\dots \longrightarrow$ $\boxed{\text{Calcule } d^k}$ $\longrightarrow$ $\boxed{\text{Calcule } t_k \text{ satisfazendo Armijo}}$

$\dots \longleftarrow$ $\boxed{x^{k+1} = x^k + t_k d^k}$ $\longleftarrow$

Isso costuma dar certo pois d^k é uma direção de não subida para f:

$$
\begin{aligned}
\nabla f(x^k)^T d^k &= \left[J(x^k)^T R(x^k) \right]^T d^k \\
&= \left[- J(x^k)^T J(x^k) d^k \right]^T d^k \\
&= - \| J(x^k) d^k \|_2^2 \leq 0 .
\end{aligned}
$$

Quando $J(x^k)$ tem posto completo, a direção d^k é de descida, dado que

$$\nabla f(x^k)^T d^k < 0 .$$

6) A convergência vai depender de quão grande/pequeno $S(x)$ é...

↳ veja ainda Teorema 10.1 de Nocedal.

• Problema com Gauss-Newton: $J(x^*)$ não ter posto completo, e assim $J^T J \not> 0$.

↳ Outro método: <u>Levenberg-Marquardt</u>.