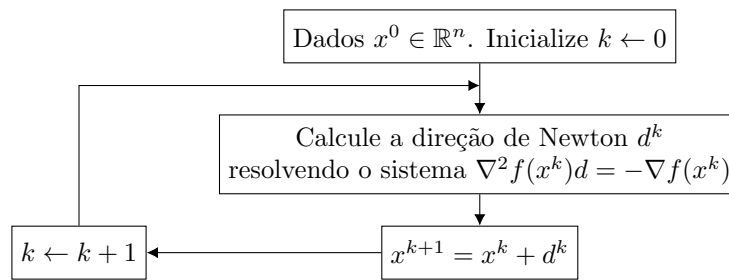


Capítulo 12

Método de Newton



Algoritmo 12.1: Método de Newton.

12.1 Convergência local

Dada uma matriz quadrada A , lembre-se que a *norma de A* é dada por

$$\|A\| = \max_{\|x\|=1} \|Ax\|$$

onde $\|Ax\|$ é norma usual do vetor Ax . Neste caso dizemos que a norma matricial $\|\cdot\|$ é induzida pela norma correspondente $\|\cdot\|$ de vetores. Nesta seção, $\|A\|$ indicará a norma matricial induzida pela norma euclidiana de vetores.

Exercício: Mostre que

1. $\|\cdot\|$ é de fato uma norma no espaço das matrizes, isto é, $\|A\| = 0 \Leftrightarrow A = 0$, $\|\lambda A\| = |\lambda|\|A\|$ e $\|A + B\| \leq \|A\| + \|B\|$;
2. $\|Ax\| \leq \|A\|\|x\|$ (A matriz, x vetor) e $\|AB\| \leq \|A\|\|B\|$ (A, B matrizes);
3. $\|I\| = 1$, onde I é matriz identidade.

Lema 2. Se F é uma matriz quadrada tal que $\|F\| < 1$, então $I - F$ é não singular e $\|(I - F)^{-1}\| \leq \frac{1}{1 - \|F\|}$.

Demonstração. Veja o Lema 2.3.3 de (Golub, Van Loan. Matrix Computations. 4 ed. The Johns Hopkins University Press, Baltimore, 2013). \square

O próximo lema diz que, dada uma matriz não singular A_* , qualquer matriz suficientemente próxima é também não singular, e a norma da sua inversa pode ser majorada pela norma da inversa de A_* . Resultados do tipo são encontrados em livros de análise matricial, veja por exemplo [4, Teorema 2.3.1]. O resultado aqui apresentado é muito similar ao Lema 5.4.1 de [2], e será útil na prova de convergência do método de Newton, dado que temos que lidar com a inversa da hessiana de f próximo à solução.

Lema 3. *Seja A_* matriz quadrada não singular. Se A é tal que $\|A - A_*\| < 1/\|A_*^{-1}\|$ então A é não singular e $\|A^{-1}\| \leq 2\|A_*^{-1}\|$.*

Demonstração. Vamos mostrar primeiro, por contraposição, que A é não singular. Suponha que A seja singular. Então existe $z \neq 0$ tal que $Az = 0$. Pelas propriedades da norma de matrizes, segue que

$$\|z\| = \|-A_*^{-1}A_*z\| = \|A_*^{-1}(A - A_*)z\| \leq \|A_*^{-1}\| \|(A - A_*)z\|.$$

Dividindo a expressão acima por $\|z\| > 0$, obtemos $\|A - A_*\| \geq 1/\|A_*^{-1}\|$, como queríamos.

Para a segunda desigualdade, note primeiro que

$$\|I - AA_*^{-1}\| = \|(A - A_*)A_*^{-1}\| \leq \|A - A_*\| \|A_*^{-1}\| < 1.$$

Aplicando o Lema 2 com $F = I - AA_*^{-1}$ obtemos

A matriz AA_*^{-1} é inversível pois A é inversível. Vamos mostrar que

$$\|(AA_*^{-1})^{-1}\| = \|[I - (I - AA_*^{-1})]^{-1}\| \leq \frac{1}{1 - \|I - AA_*^{-1}\|} \leq 2. \quad (12.1)$$

Assim,

$$\|A^{-1}\| = \|A_*^{-1}A_*A^{-1}\| = \|A_*^{-1}(AA_*^{-1})^{-1}\| \leq \|A_*^{-1}\| \|(AA_*^{-1})^{-1}\| \leq 2\|A_*^{-1}\|,$$

como queríamos demonstrar. \square

Teorema 5. *Suponha que f tenha derivadas de segunda ordem contínuas. Seja x^* tal que $\nabla f(x^*) = 0$ e suponha que $\nabla^2 f(x^*)$ seja não singular. Então existe $\varepsilon > 0$ tal que, se $\|x^0 - x^*\| \leq \varepsilon$, vale:*

1. A sequência $x^{k+1} = x^k - (\nabla^2 f(x^k))^{-1} \nabla f(x^k)$, $k \geq 0$, está bem definida (isto é, o passo de Newton está bem definido para todo $k \geq 0$);
2. $\{x^k\}$ converge a x^* com ordem superlinear;
3. Se a função $\nabla^2 f$ é Lipschitz contínua, isto é, se existe $L > 0$ tal que

$$\|\nabla^2 f(x) - \nabla^2 f(y)\| \leq L\|x - y\|, \quad \forall x, y \in \mathbb{R}^n,$$

então $\{x^k\}$ converge a x^* com ordem quadrática.

Demonstração. Como $\nabla^2 f$ é contínua e não singular em x^* , temos $\|\nabla^2 f(x^*)^{-1}\| > 0$ e a existência de um $\varepsilon > 0$ tal que

$$\|x - x^*\| \leq \varepsilon \implies \|\nabla^2 f(x) - \nabla^2 f(x^*)\| \leq \frac{1}{4\|\nabla^2 f(x^*)^{-1}\|}. \quad (12.2)$$

Assim, tomando $\|x^0 - x^*\| \leq \varepsilon$, o Lema 3 garante que $x^1 = x^0 - (\nabla^2 f(x^0))^{-1} \nabla f(x^0)$ está bem definido. Por hipótese temos $\nabla f(x^*) = 0$, e logo

$$\begin{aligned} \|x^1 - x^*\| &= \|x^0 - x^* - \nabla^2 f(x^0)^{-1} \nabla f(x^0)\| \\ &= \|\nabla^2 f(x^0)^{-1} [-\nabla^2 f(x^0)(x^0 - x^*) + \nabla f(x^0) - \nabla f(x^*)]\| \\ &\leq \|\nabla^2 f(x^0)^{-1}\| \|\nabla f(x^0) - \nabla f(x^*) - \nabla^2 f(x^0)(x^0 - x^*)\|. \end{aligned} \quad (12.3)$$

Consideremos a função φ de uma variável definida por

$$\varphi(t) = \nabla f(tx^0 + (1-t)x^*), \quad t \in [0, 1].$$

Temos

$$\varphi'(t) = \nabla^2 f(tx^0 + (1-t)x^*)(x^0 - x^*).$$

Pelo Teorema do Valor Médio aplicado à φ , existe $\bar{t} \in (0, 1)$ tal que

$$\varphi'(\bar{t}) = \frac{\varphi(1) - \varphi(0)}{1 - 0},$$

o que implica

$$\nabla^2 f(\bar{t}x^0 + (1 - \bar{t})x^*)(x^0 - x^*) = \nabla f(x^0) - \nabla f(x^*).$$

Usando essa expressão em (12.3), chegamos a

$$\begin{aligned} \|x^1 - x^*\| &\leq \|\nabla^2 f(x^0)^{-1}\| \|\nabla^2 f(\bar{t}x^0 + (1 - \bar{t})x^*)(x^0 - x^*) - \nabla^2 f(x^0)(x^0 - x^*)\| \\ &\leq r_0 \|x^0 - x^*\|, \end{aligned} \quad (12.4)$$

onde

$$r_0 = \|\nabla^2 f(x^0)^{-1}\| \max_{t \in [0,1]} \|\nabla^2 f(tx^0 + (1 - t)x^*) - \nabla^2 f(x^0)\|.$$

Note que para $t \in [0, 1]$, temos

$$\|tx^0 + (1 - t)x^*\| = (1 - t)\|x^0 - x^*\| \leq \|x^0 - x^*\| \leq \varepsilon.$$

Assim, (12.2) e o Lema 3 garantem que

$$r_0 \leq 2\|\nabla^2 f(x^*)^{-1}\| \frac{1}{4\|\nabla^2 f(x^*)^{-1}\|} = \frac{1}{2}.$$

Daí, segue de (12.4) que

$$\|x^1 - x^*\| \leq r_0 \|x^0 - x^*\| \leq \varepsilon/2 \leq \varepsilon,$$

e usando novamente (12.2) e o Lema 3, concluímos que x^2 está bem definido. Repetindo indutivamente o argumento, provamos que, para todo $k \geq 0$, o iterando $x^{k+1} = x^k - (\nabla^2 f(x^k)^{-1})\nabla f(x^k)$ está bem definido (item 1) e que

$$\|x^{k+1} - x^*\| \leq r_0 \cdot r_1 \cdots r_k \|x^k - x^*\| \leq \frac{1}{2^{k+1}} \|x^k - x^*\|,$$

onde

$$r_k = \|\nabla^2 f(x^k)^{-1}\| \max_{t \in [0,1]} \|\nabla^2 f(tx^k + (1 - t)x^*) - \nabla^2 f(x^k)\| \leq 1/2.$$

Isso mostra que $\lim_k x^k = x^*$ com ordem superlinear (item 2).

Vamos mostrar a convergência quadrática (item 3). Suponha que $\nabla^2 f$ seja Lipschitz contínua com constante $L > 0$, como no enunciado. O passo de indução realizado na prova dos itens 1 e 2 fornece uma expressão análoga à (12.4), com x^{k+1} no lugar de x^1 e x^k no lugar de x^0 . Assim, usando o Lema 3, temos

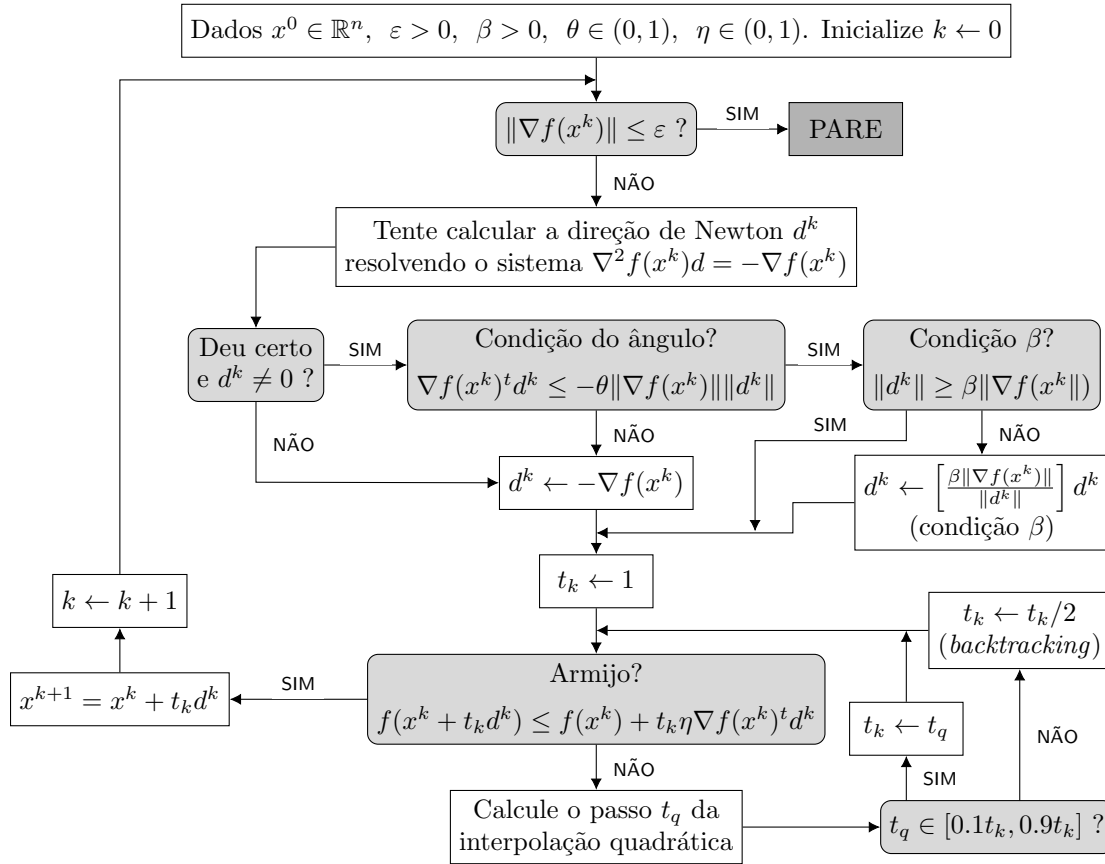
$$\begin{aligned} \|x^{k+1} - x^*\| &\leq r_k \|x^k - x^*\| \\ &= \left(\|\nabla^2 f(x^k)^{-1}\| \max_{t \in [0,1]} \|\nabla^2 f(tx^k + (1 - t)x^*) - \nabla^2 f(x^k)\| \right) \|x^k - x^*\| \\ &\leq \left(2\|\nabla^2 f(x^*)^{-1}\| \max_{t \in [0,1]} L \underbrace{\|tx^k + (1 - t)x^* - x^k\|}_{(1-t)\|x^k - x^*\|} \right) \|x^k - x^*\| \\ &\leq 2\|\nabla^2 f(x^*)^{-1}\| \|x^k - x^*\|^2 \max_{t \in [0,1]} (1 - t), \end{aligned}$$

o que implica

$$\|x^{k+1} - x^*\| \leq C \|x^k - x^*\|^2, \quad (12.5)$$

onde $C = 2\|\nabla^2 f(x^*)^{-1}\| > 0$. Ou seja, $\lim_k x^k = x^*$ com ordem quadrática. \square

12.2 Newton globalizado



Algoritmo 12.2: Método de Newton globalizado.

12.3 Cálculo da direções newtonianas usando fatoração Cholesky

Trecho código Julia

12.4 Exercícios