Otimização não linear
Otimização diferenciável: Métodos do gradiente

Ana Maria A. C. Rocha

Departamento de Produção e Sistemas

Universidade do Minho

arocha@dps.uminho.pt

Métodos numéricos de resolução

Problema sem restrições

$$\min_{x \in \mathbb{R}^n} f(x) \quad (n > 1)$$

- Métodos de procura direta (que **não** usam derivadas) ;
- Métodos do gradiente.

Métodos do gradiente

- usam informação da função e das derivadas (gradiente e/ou Hessiana);
- só podem ser usados na resolução de **problemas diferenciáveis**;
- convergem mais rapidamente do que os métodos de procura direta.

Métodos do gradiente

Métodos do gradiente

- geram uma sucessão de aproximações $\{x^{(k)}\}$ à solução:

$$\boxed{x^{(k+1)} = x^{(k)} + \alpha^k \, d^{(k)}}$$

em que $d^{(k)}$ (vetor) é a direção de procura (ou passo) e α^k (escalar) é o comprimento do passo.

- A equação iterativa para o cálculo da direção de procura é diferente para cada método.

Algoritmo geral dos métodos do gradiente

Dados: aproximação inicial $x^{(1)}$, $\varepsilon > 0\,(\approx 0)$, $k \leftarrow 1$

Enquanto $\underbrace{\left\| \nabla f(x^{(k)}) \right\|_2}_{\text{medida de estacionaridade}} > \varepsilon$ **fazer**

- calcular $d^{(k)}$ (direção de procura ou passo)
- calcular $\alpha^{(k)}$ (comprimento do passo)
- definir $x^{(k+1)} \leftarrow x^{(k)} + \alpha^{(k)} d^{(k)}$
- $k = k + 1$

Fim de enquanto

Solução: $\begin{cases} x^* \approx x^{(k+1)} \\ f^* \approx f\left(x^{(k+1)}\right) \end{cases}$

Métodos do gradiente mais usados

Em cada iteração k, para calcular a direção de procura ou passo $d^{(k)}$:

- método de Segurança de Newton

direção é $d_{SN}^{(k)}$

- método quasi-Newton

direção é $d_{QN}^{(k)}$

para calcular o comprimento do passo $\alpha^{(k)}$ – se a direção for descendente para $f(x)$ em $x^{(k)}$:

- critério de Armijo.

Método de Newton básico

O método de Newton é baseado numa aproximação local de $f(x)$ por uma função quadrática.

Seja $x^{(k)}$ uma aproximação a x^*.
Usando a expansão em série de Taylor de f, no ponto $x = x^{(k)} + d$:

$$f\left(x^{(k)} + d\right) = f(x^{(k)}) + \nabla f(x^{(k)})^T d + \frac{1}{2} d^T \nabla^2 f(x^{(k)}) d + \ldots$$

e apenas os três primeiros termos da série, obtém-se uma função quadrática em d

$$\begin{aligned} f\left(x^{(k)} + d\right) &\approx f(x^{(k)}) + \nabla f(x^{(k)})^T d + \frac{1}{2} d^T \nabla^2 f(x^{(k)}) d \\ &\approx q(d) \end{aligned}$$

O mínimo de $q(d)$ vai originar uma nova aproximação ao mínimo de $f(x)$.

... a direção Newton

Derivando em ordem a d e igualando a zero $(\nabla q\,(d) = 0)$, obtém-se

$$\nabla f(x^{(k)}) + \nabla^2 f(x^{(k)})\, d = 0$$
$$\Leftrightarrow$$

$$\nabla^2 f(x^{(k)})\, d_N^{(k)} = -\nabla f(x^{(k)}) \qquad (1)$$

A solução do **sistema (linear) Newton** (1) é o vetor $d_N^{(k)}$ – direção Newton.

- Usa-se EGPP para resolver o sistema Newton.
- A nova aproximação

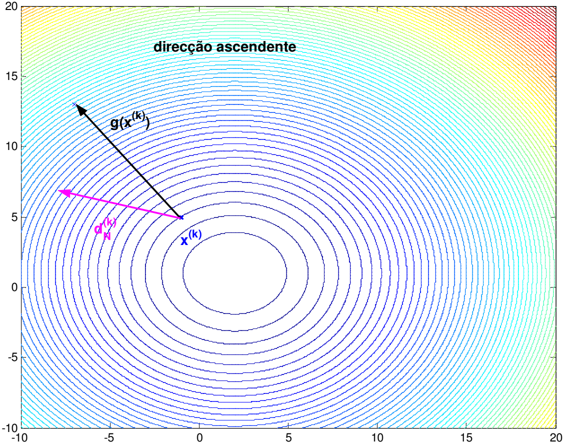
$$x^{(k+1)} \leftarrow x^{(k)} + d_N^{(k)}$$

não é necessariamente o minimizante de $f\,(x)$ e o processo deve ser repetido a partir de $x^{(k+1)}$.

1. direção Newton ascendente (!)

- No entanto, a direção $d_N^{(k)}$ (solução do sistema Newton) pode ser ascendente para f em $x^{(k)}$, ou seja,

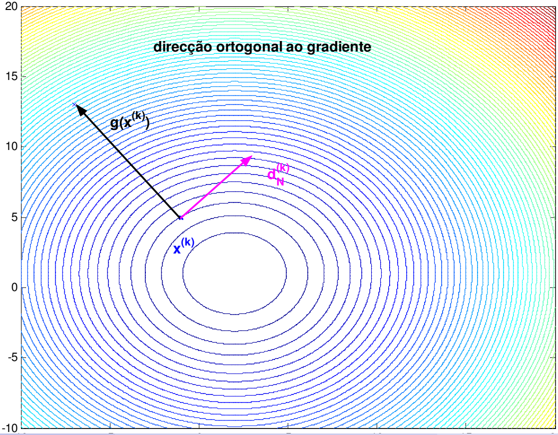
$$\nabla f(x^{(k)})^T \, d_N^{(k)} > 0$$



2. direção Newton ortogonal ao gradiente (!)

- Ou, a direção $d_N^{(k)}$ (solução do sistema Newton) pode ser ortogonal ao gradiente calculado em $x^{(k)}$, ou seja,

$$\nabla f(x^{(k)})^T \, d_N^{(k)} = 0$$

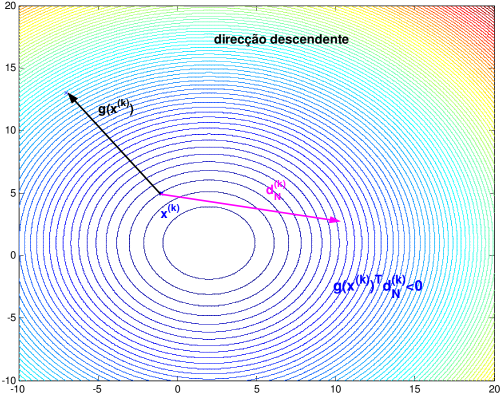


3. direção Newton descendente, mas ...

- ou, a direção $d_N^{(k)}$ pode ser descendente para f em $x^{(k)}$, i.e.

$$\nabla f(x^{(k)})^T d_N^{(k)} < 0$$

mas ser muito grande e $f(x^{(k)} + d_N^{(k)}) \geq f(x^{(k)})$ (!)



4. sistema Newton não tem solução única (!)

- ou ainda, a matriz $\nabla^2 f(x^{(k)})$ pode ser singular, o que significa que o sistema Newton não tem solução ou tem uma infinidade de soluções

$$\Updownarrow$$

$$\nexists d_N^{(k)} \text{ (não existe direção ou não é única)}$$

Limitações do Método de Newton

Para ultrapassar as limitações descritas com a direção Newton, em cada iteração k, **implementam-se** as seguintes soluções que a seguir se apresentam – dando origem ao

$$\Downarrow$$

$$\boxed{\textbf{método de Segurança de Newton}}$$

Método de Segurança de Newton

Em qualquer iteração k:

- Quando $\nabla^2 f(x^{(k)})$ é singular

$\Rightarrow$ usar $d_{SN}^{(k)} = -\nabla f\left(x^{(k)}\right)$

$\left(-\nabla f\left(x^{(k)}\right)\right.$ é a direção de descida máxima e é descendente para $\left.f\right)$

- Quando $d_N^{(k)}$ é ortogonal ao gradiente $\Leftrightarrow \nabla f(x^{(k)})^T d_N^{(k)} \approx 0$

$\Rightarrow$ usar $d_{SN}^{(k)} = -\nabla f\left(x^{(k)}\right)$

- Quando $d_N^{(k)}$ é ascendente $\Leftrightarrow \nabla f(x^{(k)})^T d_N^{(k)} > \eta$

$\Rightarrow$ usar $d_{SN}^{(k)} = -d_N^{(k)}$

- Quando $d_N^{(k)}$ é descendente

$\Rightarrow$ usar $d_{SN}^{(k)} = d_N^{(k)}$ NOTA: $d_{SN}^{(k)}$ é descendente, para todo o k.

Propriedades do método de Newton

- o método de Newton tem convergência
  - local
    (a convergência para a solução só é garantida se a aproximação inicial $x^{(1)}$ estiver na vizinhança da solução);
  - quadrática
    (verifica-se
    $$\|x^{(k+1)} - x^*\| \leq \gamma \|x^{(k)} - x^*\|^2, \gamma > 0);$$
- o método de Newton possui a propriedade da "terminação quadrática", i.e., se $f(x)$ $(x \in \mathbb{R}^n)$ for uma função quadrática e convexa o método de Newton necessita no máximo de n iterações para encontrar a solução.

Desvantagens do método de Newton

- Cálculo das segundas derivadas:
  - se a expressão de f é complicada, estas tornam-se difíceis de calcular;
  - exigem um grande esforço de cálculo quando n é grande.
- A convergência é local.

Para ultrapassar a convergência local

deve implementar-se uma técnica de globalização para garantir que o método converge para a solução, a partir de qualquer aproximação inicial.

Técnica de globalização

implementa-se

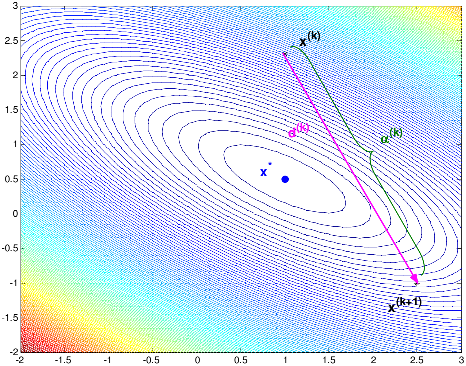
- para garantir que o método converge, qualquer que seja a aproximação inicial $x^{(1)}$ (i.e., $x^{(1)}$ pode estar fora da região de convergência do método);
- para garantir que o método converge para um ponto estacionário que é <u>minimizante</u>.

1. Procura unidimensional (line search) $\begin{cases} \text{exata} \\ \text{aproximada} \end{cases}$
2. Região de confiança (trust region)

Procura unidimensional aproximada

Dados $x^{(k)}$ e $d^{(k)}$, calcular $\alpha^{(k)}$ – comprimento do passo – que origina uma redução significativa no valor de f na nova aproximação $x^{(k)} + \alpha^{(k)} d^{(k)}$

$$f(x^{(k)} + \alpha^{(k)} d^{(k)}) \leq f(x^{(k)}) + \mu \alpha^{(k)} \nabla f(x^{(k)})^T d^{(k)}$$



Critério de Armijo

Procedimento para calcular $\alpha^{(k)}$ que origina uma redução significativa na função objetivo, isto é, verifica

- a condição de Armijo

$$f(x^{(k)} + \alpha^{(k)} d^{(k)}) \leq f(x^{(k)}) + \mu\, \alpha^{(k)}\, \nabla f(x^{(k)})^T\, d^{(k)}$$

com $0 < \mu < \frac{1}{2}$.

- Se $d^{(k)}$ for descendente para f, ou seja,

$$\nabla f(x^{(k)})^T\, d^{(k)} \, < \, 0$$

  existe um valor de

$$\alpha^{(k)} \in (0, 1]$$

  que verifica esta condição.

Algoritmo do critério de Armijo para calcular $\alpha^{(k)}$

Dados $x^{(k)}$, $d^{(k)}$, $\nabla f\left(x^{(k)}\right)$, $f\left(x^{(k)}\right)$ e μ

1. $\alpha \leftarrow 1$
2. $\bar{x} \leftarrow x^{(k)} + \alpha d^{(k)}$
3. <u>se</u> $\left(f(\bar{x}) \leq f\left(x^{(k)}\right) + \mu\,\alpha\,\nabla f\left(x^{(k)}\right)^T d^{(k)}\right)$ <u>então</u>

   fazer $\alpha^{(k)} = \alpha$

   <u>senão</u>

   $\alpha \leftarrow \alpha/2$ e voltar a 2.

Critério de paragem

Condições

- estimativa do erro relativo da aproximação

$$\frac{\left\|x^{(k+1)} - x^{(k)}\right\|_2}{\left\|x^{(k+1)}\right\|_2} \leq \varepsilon_1$$

- estimativa do erro relativo da função objetivo

$$\frac{\left|f\left(x^{(k+1)}\right) - f\left(x^{(k)}\right)\right|}{\left|f\left(x^{(k+1)}\right)\right|} \leq \varepsilon_2$$

- medida de estacionaridade

$$\left\|\nabla f\left(x^{(k+1)}\right)\right\|_2 \leq \varepsilon_3$$

ε_1, ε_2 e ε_3 são quantidades positivas e próximas de zero.

Nota: Diferentes implementações do método podem ter condições diferentes no critério de paragem.

Algoritmo geral dos métodos do gradiente

Dados: aproximação inicial $x^{(1)}$, $k \leftarrow 1$

Enquanto $\left\| \nabla f \left(x^{(k+1)} \right) \right\|_2 \leq \varepsilon$ **fazer**

- calcular $d^{(k)}$ (direção de procura ou passo)
- calcular $\alpha^{(k)}$ (comprimento do passo)
- definir $x^{(k+1)} \leftarrow x^{(k)} + \alpha^{(k)} d^{(k)}$
- $k = k + 1$

Fim de enquanto

Solução: $\begin{cases} x^* \approx x^{(k+1)} \\ f^* \approx f\left(x^{(k+1)}\right) \end{cases}$

Algoritmo para o cálculo da direção de Segurança de Newton

Dados $x^{(k)}$ e η,

Resolver o sistema linear Newton $\nabla^2 f\left(x^{(k)}\right) d_N^{(k)} = -\nabla f\left(x^{(k)}\right)$ por EGPP

<u>se</u> (o sistema linear tem solução única - $\exists d_N^{(k)}$) <u>então</u>

  <u>se</u> $\left|\nabla f\left(x^{(k)}\right)^T d_N^{(k)}\right| \leq \eta \left\|\nabla f\left(x^{(k)}\right)\right\|_2 . \left\|d_N^{(k)}\right\|_2$ com $\eta > 0\,(\approx 0)$

  <u>então</u> $d_{SN}^{(k)} \leftarrow -\nabla f\left(x^{(k)}\right)$

  <u>senão</u>

    <u>se</u> $\nabla f\left(x^{(k)}\right)^T d_N^{(k)} > \eta \left\|\nabla f\left(x^{(k)}\right)\right\|_2 . \left\|d_N^{(k)}\right\|_2$ com $\eta > 0\,(\approx 0)$

    <u>então</u> $d_{SN}^{(k)} \leftarrow -d_N^{(k)}$

    <u>senão</u> $d_{SN}^{(k)} \leftarrow d_N^{(k)}$

<u>senão</u>

  $d_{SN}^{(k)} \leftarrow -\nabla f\left(x^{(k)}\right)$

Como evitar o cálculo das 2^{as} derivadas no método de Newton

método de Newton

$$\nabla^2 f(x^{(k)})\, d_N^{(k)} = -\nabla f(x^{(k)})$$

Para evitar o cálculo das 2^{as} derivadas para a Hessiana, $\nabla^2 f(x)$, pode usar-se uma <u>aproximação</u>:

$$B^{(k)} \approx \nabla^2 f(x^{(k)})$$

e a direção de procura seria calculada pela resolução deste sistema linear (por EGPP)

$$B^{(k)} d_{QN}^{(k)} = -\nabla f(x^{(k)})$$

A direção quasi-Newton

- Como o sistema Newton pode ser escrito na forma equivalente - embora não aconselhável na prática,

$$d_N^{(k)} = -\left(\nabla^2 f(x^{(k)})\right)^{-1} \nabla f(x^{(k)})$$

pode usar-se, em cada iteração k, uma aproximação à **inversa da Hessiana**

$$H^{(k)} \approx \left(\nabla^2 f(x^{(k)})\right)^{-1}$$

e calcular a direção pelo produto da matriz pelo vetor

$$d_{QN}^{(k)} = -H^{(k)} \nabla f(x^{(k)})$$

$d_{QN}^{(k)}$ - é a direção quasi-Newton.

Método quasi-Newton

Evita-se desta forma:

- o cálculo das segundas derivadas;
- a resolução de um sistema linear em cada iteração, substituindo-o pelo produto de uma matriz por um vetor.

As equações que descrevem o método quasi-Newton são:

$$\begin{cases} d_{QN}^{(k)} = -H^{(k)}\nabla f(x^{(k)}), \\ \\ x^{(k+1)} = x^{(k)} + \alpha^{(k)} d_{QN}^{(k)} \quad \text{para } k = 1, 2, \ldots \end{cases}$$

Propriedades da matriz H

- A matriz H deve aproximar, o melhor possível, a inversa de $\nabla^2 f\left(x^{(k)}\right)$, ou seja, deve verificar a **condição secante**:

$$\boxed{H^{(k)}y^{(k-1)} = s^{(k-1)}}$$

com

$$y^{(k-1)} = \nabla f\left(x^{(k)}\right) - \nabla f\left(x^{(k-1)}\right)$$

(variação verificada no gradiente da iteração $k-1$ para a iteração k)

$$s^{(k-1)} = x^{(k)} - x^{(k-1)} = \alpha^{(k-1)} d_{QN}^{(k-1)}$$

(variação verificada em x)

Propriedades da matriz H

- A matriz H deve preferencialmente ser

$$\begin{cases} \textbf{simétrica} & \left(\text{pois } \nabla^2 f\left(x^{(k)}\right)^{-1} \text{ também é simétrica}\right) \\ \textbf{definida positiva} & \left(\text{pois a direção } d_{QN}^{(k)} = -H^{(k)} \nabla f(x^{(k)})\right. \end{cases}$$

é **descendente** para f em $x^{(k)}$).

Para que estas propriedades se conservem ao longo do processo iterativo, a matriz inicial $H^{(1)}$, para $k = 1$, deve ser também simétrica e definida positiva. Pode usar-se, por exemplo

$$H^{(1)} = I.$$

Fórmulas de atualização que conservam a matriz H simétrica e definida positiva

Davidon, Fletcher e Powell - **DFP**

$$H^{(k)} = H^{(k-1)} - \frac{H^{(k-1)} y^{(k-1)} y^{(k-1)^T} H^{(k-1)}}{y^{(k-1)^T} H^{(k-1)} y^{(k-1)}} + \frac{s^{(k-1)} s^{(k-1)^T}}{s^{(k-1)^T} y^{(k-1)}}$$

Broyden, Fletcher, Goldfarb e Shanno - **BFGS**

$$H^{(k)} = \left(I - \frac{s^{(k-1)} y^{(k-1)^T}}{s^{(k-1)^T} y^{(k-1)}} \right) H^{(k-1)} \left(I - \frac{y^{(k-1)} s^{(k-1)^T}}{s^{(k-1)^T} y^{(k-1)}} \right) + \frac{s^{(k-1)} s^{(k-1)^T}}{s^{(k-1)^T} y^{(k-1)}}$$

Propriedades do método quasi-Newton

- o método quasi-Newton tem convergência
  - local
    (a convergência para a solução só é garantida se a aproximação inicial $x^{(1)}$ estiver na vizinhança da solução);
  - superlinear
    (verifica-se
    $$\|x^{(k+1)} - x^*\| \leq \gamma_k \|x^{(k)} - x^*\|$$
    com a sucessão $\{\gamma_k\} \to 0$ quando $k \to \infty$);
- o método quasi-Newton satisfaz a propriedade da "terminação quadrática" (ou seja, o mínimo de uma função quadrática $q(x)$, $x \in \mathbb{R}^n$, obtém-se em n, ou menos, do que n iterações).

Limitação do método quasi-Newton

★ Os erros de arredondamento que se cometem nos cálculos podem fazer com que $H^{(k)}$ deixe de ser definida positiva e a direção $d_{QN}^{(k)}$ deixa de ser descendente para f em $x^{(k)}$

$$\Downarrow$$

Solução: fazer $\boxed{H^{(k)} = I}$ (neste caso $H^{(k)}$ é simétrica e definida positiva)

$$\Downarrow$$

$$d_{QN}^{(k)} = -\nabla f(x^{(k)})$$

Algoritmo geral dos métodos do gradiente

Dados: aproximação inicial $x^{(1)}$, $k \leftarrow 1$

Enquanto $\left\|\nabla f\left(x^{(k+1)}\right)\right\|_2 \leq \varepsilon$ **fazer**

- calcular $d^{(k)}$ (direção de procura ou passo)
- calcular $\alpha^{(k)}$ (comprimento do passo)
- definir $x^{(k+1)} \leftarrow x^{(k)} + \alpha^{(k)} d^{(k)}$
- $k = k + 1$

Fim de enquanto

Solução: $\left\{ \begin{array}{l} x^* \approx x^{(k+1)} \\ f^* \approx f\left(x^{(k+1)}\right) \end{array} \right.$

Algoritmo para o cálculo da direção quasi-Newton

Dado $x^{(k)}$

Calcular $d_{QN}^{(k)} \leftarrow -H^{(k)} \nabla f\left(x^{(k)}\right)$, sendo $H^{(k)}$ dada por:

se $k = 1$ então
$H^{(k)} \leftarrow I$

senão

$$\begin{cases} s^{(k-1)} \leftarrow x^{(k)} - x^{(k-1)} \\ y^{(k-1)} \leftarrow \nabla f\left(x^{(k)}\right) - \nabla f\left(x^{(k-1)}\right) \\ \text{atualizar } H^{(k)} \text{ pela fórmula } \textbf{DFP} \text{ ou } \textbf{BFGS} \end{cases}$$

se $d_{QN}^{(k)}$ não é descendente $\left(\nabla f\left(x^{(k)}\right)^T d_{QN}^{(k)} \geq 0\right)$ então
fazer $d_{QN}^{(k)} \leftarrow -\nabla f\left(x^{(k)}\right)$

Exercício 2

Considere o seguinte problema de otimização sem restrições

$$f(x_1, x_2) = x_1^2 + x_2^2 - x_1 x_2.$$

Calcule o seu mínimo usando o algoritmo quasi-Newton (fórmula de atualização DFP). O processo iterativo deve ser iniciado com o ponto $(1, 0)$ e deve terminar quando o critério de paragem for verificado para $\varepsilon_3 = 0.02$.

Para calcular o comprimento do passo α, use, em cada iteração, o critério de Armijo com $\mu = 0.001$.

Resolução do Exercício 1

Determinação do vetor gradiente da função:

$$\nabla f\left(x\right) = \begin{pmatrix} 2x_1 - x_2 \\ 2x_2 - x_1 \end{pmatrix}$$

$\boxed{\mathbf{1^{a}\ iteração}}$

$$\nabla f\left(x^{(1)}\right) = \nabla f\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ -1 \end{pmatrix}$$

Cálculo da direção $d_{QN}^{(1)}$

$$d_{QN}^{(1)} = -H^{(1)}\nabla f\left(x^{(1)}\right) \Leftrightarrow d_{QN}^{(1)} = -\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 2 \\ -1 \end{pmatrix}$$

$\boxed{d_{QN}^{(1)} = \begin{pmatrix} -2 \\ 1 \end{pmatrix}}$ e é descendente $\left(\nabla f\left(x^{(k)}\right)^T d_{QN}^{(k)} \leq 0\right)$

Resolução do Exercício 1 (cont.)

Cálculo do comprimento do passo $\alpha^{(1)}$

$$\bar{x} = x^{(1)} + \alpha d_{QN}^{(1)} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \alpha \begin{pmatrix} -2 \\ 1 \end{pmatrix}$$

- Para $\alpha = 1 \Rightarrow \bar{x} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$

  $f(\bar{x}) \leq f\left(x^{(1)}\right) + \mu\alpha\nabla f\left(x^{(1)}\right)^T d_{QN}^{(1)}$

  $\underbrace{f\begin{pmatrix} -1 \\ 1 \end{pmatrix}}_{3} \leq \underbrace{f\begin{pmatrix} 1 \\ 0 \end{pmatrix}}_{=1} + \underbrace{\mu}_{=0.001}\underbrace{\alpha}_{=1}\underbrace{\nabla f\begin{pmatrix} 1 \\ 0 \end{pmatrix}^T \begin{pmatrix} -2 \\ 1 \end{pmatrix}}_{=-5}$

  Condição de Armijo: $3 \leq 1 + 0.001\,(1)\,(-5)$ (Falso) $\Rightarrow \alpha = \alpha/2$

- Para $\alpha = 0.5 \Rightarrow \bar{x} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + 0.5 \begin{pmatrix} -2 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0.5 \end{pmatrix}$ $\quad f(\bar{x}) = 0.25$

  Condição de Armijo: $0.25 \leq 1 + 0.001\,(0.5)\,(-5)$ (Verdadeiro)

$\boxed{\alpha^{(1)} = 0.5}$

Resolução do Exercício 1 (cont.)

Cálculo do novo ponto

$$x^{(2)} = x^{(1)} + \alpha^{(1)} d_{QN}^{(1)} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + 0.5 \begin{pmatrix} -2 \\ 1 \end{pmatrix}$$

$$\boxed{x^{(2)} = \begin{pmatrix} 0 \\ 0.5 \end{pmatrix}} \qquad f\left(x^{(2)}\right) = 0.25$$

Testar critério de paragem

$$\nabla f\left(x^{(2)}\right) \equiv \nabla f \begin{pmatrix} 0 \\ 0.5 \end{pmatrix} = \begin{pmatrix} -0.5 \\ 1 \end{pmatrix}$$

$$\left\|\nabla f\left(x^{(2)}\right)\right\|_2 = \left\| \begin{pmatrix} -0.5 \\ 1 \end{pmatrix} \right\|_2 = 1.118 \leq 0.02 \text{ (Falso)} \Rightarrow \text{nova iteração}$$

$\boxed{\textbf{2}^{a} \textbf{ iteração}}$

- Não vamos implementar a técnica de recomeço
- Atualizar a matriz H pela fórmula de Atualização DFP

Resolução do Exercício 1 (cont.)

$$H^{(2)} = H^{(1)} - \frac{H^{(1)} y^{(1)} {y^{(1)}}^T H^{(1)}}{{y^{(1)}}^T H^{(1)} y^{(1)}} + \frac{s^{(1)} {s^{(1)}}^T}{{s^{(1)}}^T y^{(1)}}$$

$$s^{(1)} = x^{(2)} - x^{(1)} = \begin{pmatrix} 0 \\ 0.5 \end{pmatrix} - \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} -1 \\ 0.5 \end{pmatrix}$$

$$y^{(1)} = \nabla f\left(x^{(2)}\right) - \nabla f\left(x^{(1)}\right) = \begin{pmatrix} -0.5 \\ 1 \end{pmatrix} - \begin{pmatrix} 2 \\ -1 \end{pmatrix} = \begin{pmatrix} -2.5 \\ 2 \end{pmatrix}$$

Cálculo dos escalares dos denominadores

$${s^{(1)}}^T y^{(1)} = \begin{pmatrix} -1 & 0.5 \end{pmatrix} \begin{pmatrix} -2.5 \\ 2 \end{pmatrix} = 2.5 + 1 = 3.5$$

$${y^{(1)}}^T H^{(1)} y^{(1)} = \begin{pmatrix} -2.5 & 2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} -2.5 \\ 2 \end{pmatrix} = 6.25 + 4 = 10.25$$

Resolução do Exercício 1 (cont.)

Cálculo das matrizes dos numeradores

$$s^{(1)}{s^{(1)}}^T = \begin{pmatrix} -1 \\ 0.5 \end{pmatrix} \begin{pmatrix} -1 & 0.5 \end{pmatrix} = \begin{pmatrix} 1 & -0.5 \\ -0.5 & 0.25 \end{pmatrix}$$

$$H^{(1)} y^{(1)} {y^{(1)}}^T H^{(1)} = I \begin{pmatrix} -2.5 \\ 2 \end{pmatrix} \begin{pmatrix} -2.5 & 2 \end{pmatrix} I = \begin{pmatrix} 6.25 & -5 \\ -5 & 4 \end{pmatrix}$$

$$\begin{aligned} H^{(2)} &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \frac{\begin{pmatrix} 6.25 & -5 \\ -5 & 4 \end{pmatrix}}{10.25} + \frac{\begin{pmatrix} 1 & -0.5 \\ -0.5 & 0.25 \end{pmatrix}}{3.5} \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} 0.6098 & -0.4878 \\ -0.4878 & 0.3902 \end{pmatrix} + \begin{pmatrix} 0.2857 & -0.1429 \\ -0.1429 & 0.0714 \end{pmatrix} \\ &= \begin{pmatrix} 0.6759 & 0.3449 \\ 0.3449 & 0.6812 \end{pmatrix} \end{aligned}$$

Resolução do Exercício 1 (cont.)

Cálculo da direção $d_{QN}^{(2)}$

$$d_{QN}^{(2)} = -H^{(2)} \nabla f\left(x^{(2)}\right) = -\begin{pmatrix} 0.6759 & 0.3449 \\ 0.3449 & 0.6812 \end{pmatrix} \begin{pmatrix} -0.5 \\ 1 \end{pmatrix} = \begin{pmatrix} -0.0070 \\ -0.5088 \end{pmatrix}$$

Verificar se a direção é descendente

$$\nabla f\left(x^{(2)}\right)^T d_{QN}^{(2)} = \begin{pmatrix} -0.5 & 1 \end{pmatrix} \begin{pmatrix} -0.0070 \\ -0.5088 \end{pmatrix} = -0.5053 < 0$$

$\Rightarrow d_{QN}^{(2)}$ é descendente

$$\boxed{d_{QN}^{(2)} = \begin{pmatrix} -0.0070 \\ -0.5088 \end{pmatrix}}$$

Resolução do Exercício 1 (cont.)

Cálculo do comprimento do passo $\alpha^{(2)}$

$$\bar{x} = x^{(2)} + \alpha d_{QN}^{(2)} = \begin{pmatrix} 0 \\ 0.5 \end{pmatrix} + \alpha \begin{pmatrix} -0.0070 \\ -0.5088 \end{pmatrix}$$

- Para $\alpha = 1 \Rightarrow \bar{x} = \begin{pmatrix} -0.0070 \\ -0.0088 \end{pmatrix}$

  $$f(\bar{x}) \leq f\left(x^{(2)}\right) + \mu\alpha\nabla f\left(x^{(2)}\right)^T d_{QN}^{(2)}$$

  $$\underbrace{f\begin{pmatrix} -0.0070 \\ -0.0088 \end{pmatrix}}_{=0.00006484} \leq \underbrace{f\begin{pmatrix} 0 \\ 0.5 \end{pmatrix}}_{=0.25} + \underbrace{\mu}_{=0.001}\underbrace{\alpha}_{=1}\underbrace{\nabla f\begin{pmatrix} 0 \\ 0.5 \end{pmatrix}^T \begin{pmatrix} -0.0070 \\ -0.5088 \end{pmatrix}}_{=-0.5053}$$

  Condição de Armijo: $0.00006484 \leq 0.25 + 0.001\,(1)\,(-0.5053)$ (Verdadeiro)

$$\boxed{\alpha^{(2)} = 1}$$

Resolução do Exercício 1 (cont.)

Cálculo do novo ponto

$$x^{(3)} = x^{(2)} + \alpha^{(2)} d_{QN}^{(2)} = \begin{pmatrix} 0 \\ 0.5 \end{pmatrix} + 1 \begin{pmatrix} -0.0070 \\ -0.5088 \end{pmatrix}$$

$$\boxed{x^{(3)} = \begin{pmatrix} -0.0070 \\ -0.0088 \end{pmatrix}} \qquad f\left(x^{(3)}\right) = 0.00006484$$

Testar critério de paragem

$$\nabla f\left(x^{(3)}\right) \equiv \nabla f \begin{pmatrix} -0.0070 \\ -0.0088 \end{pmatrix} = \begin{pmatrix} -0.0052 \\ -0.0106 \end{pmatrix}$$

$$\left\|\nabla f\left(x^{(3)}\right)\right\|_2 = 0.0118 \leq 0.02 \text{ (Verdadeiro)}$$

Solução:

$$\begin{cases} x^* \approx \begin{pmatrix} -0.0070 \\ -0.0088 \end{pmatrix} \\ f^* \approx 0.00006484 \end{cases}$$

Exercício 2

Dada a função $f : \mathbb{R}^2 \to \mathbb{R}$ definida por

$$f(x_1, x_2) = x_1 x_2^2 + (2 - x_1)^2$$

calcule o seu mínimo usando o algoritmo de segurança de Newton.
O processo iterativo deve ser iniciado com o ponto $(1, 1)$ e deve terminar quando $\varepsilon = 0.1$. Considere $\eta = 0.0001$.
Deve, também, implementar o algoritmo baseado no critério de Armijo para calcular o comprimento do passo α, em cada iteração (use $\mu = 0.001$).

Resolução do Exercício 2

Determinação do vector gradiente e matriz Hessiana da função:

$$\nabla f(x) = \begin{pmatrix} x_2^2 - 2(2 - x_1) \\ 2x_1x_2 \end{pmatrix} \quad \text{e} \quad \nabla^2 f(x) = \begin{pmatrix} 2 & 2x_2 \\ 2x_2 & 2x_1 \end{pmatrix}$$

1^a iteração

$$x^{(1)} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \qquad \nabla f\left(x^{(1)}\right) = \begin{pmatrix} -1 \\ 2 \end{pmatrix} \qquad \nabla^2 f\left(x^{(1)}\right) = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix}$$

Cálculo da direção $d_N^{(1)}$

$$\nabla^2 f\left(x^{(1)}\right) d_N^{(1)} = -\nabla f\left(x^{(1)}\right) \Leftrightarrow \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix} d_N^{(1)} = -\begin{pmatrix} -1 \\ 2 \end{pmatrix}$$

$$\left(\begin{array}{cc|c} 2 & 2 & 1 \\ 2 & 2 & -2 \end{array}\right) \rightarrow \left(\begin{array}{cc|c} 2 & 2 & 1 \\ 0 & 0 & -3 \end{array}\right) \Rightarrow \nabla^2 f \begin{pmatrix} 1 \\ 1 \end{pmatrix} \text{ é singular} \Rightarrow \nexists d_N^{(1)}$$

então $d_{SN}^{(1)} = -\nabla f\left(x^{(1)}\right) = -\begin{pmatrix} -1 \\ 2 \end{pmatrix} \Rightarrow d_{SN}$ é descendente $\Rightarrow \boxed{d_{SN}^{(1)} = \begin{pmatrix} 1 \\ -2 \end{pmatrix}}$

Resolução do Exercício 2 (cont.)

Cálculo do comprimento do passo $\alpha^{(1)}$

$$\bar{x} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \alpha \begin{pmatrix} 1 \\ -2 \end{pmatrix}$$

- Para $\alpha = 1 \Rightarrow \bar{x} = \begin{pmatrix} 2 \\ -1 \end{pmatrix}$

$$f(\bar{x}) \equiv \underbrace{f\begin{pmatrix} 2 \\ -1 \end{pmatrix}}_{=2} \leq \underbrace{f\begin{pmatrix} 1 \\ 1 \end{pmatrix}}_{=2} + \underbrace{\mu}_{=0.001} \underbrace{\alpha}_{=1} \underbrace{\nabla f \begin{pmatrix} 1 \\ 1 \end{pmatrix}^T d_{SN}^{(1)}}_{\begin{pmatrix} -1 & 2 \end{pmatrix}\begin{pmatrix} 1 \\ -2 \end{pmatrix} = -5 < 0}$$

Condição de Armijo: $2 \leq 2 + 0.001\,(1)\,(-5)$ (Falso) $\Rightarrow \alpha = \alpha/2$

- Para $\alpha = 0.5 \Rightarrow \bar{x} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + 0.5 \begin{pmatrix} 1 \\ -2 \end{pmatrix} = \begin{pmatrix} 1.5 \\ 0 \end{pmatrix}$, $f(\bar{x}) = 0.25$

Condição de Armijo: $0.25 \leq 2 + 0.001\,(0.5)\,(-5)$ (Verdadeiro)

$\boxed{\alpha^{(1)} = 0.5}$

Resolução do Exercício 2 (cont.)

Cálculo do novo ponto

$$x^{(2)} = x^{(1)} + \alpha^{(1)} d_{SN}^{(1)} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + 0.5 \begin{pmatrix} 1 \\ -2 \end{pmatrix}$$

$$\boxed{x^{(2)} = \begin{pmatrix} 1.5 \\ 0 \end{pmatrix}} \qquad f\left(x^{(2)}\right) = 0.25$$

Testar critério de paragem

$$\nabla f\left(x^{(2)}\right) = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$$

$$\left\|\nabla f\left(x^{(2)}\right)\right\|_2 = \left\|\begin{pmatrix} -1 \\ 0 \end{pmatrix}\right\|_2 = 1 \leq 0.1 \text{ (Falso)} \Rightarrow \text{nova iteração}$$

$\boxed{\mathbf{2^a \text{ iteração}}}$

$$\nabla^2 f\left(x^{(2)}\right) = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}$$

Resolução do Exercício 2 (cont.)

Cálculo da direção $d_N^{(2)}$

$$\nabla^2 f\left(x^{(2)}\right) d_N^{(2)} = -\nabla f\left(x^{(2)}\right) \Leftrightarrow \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix} d_N^{(2)} = -\begin{pmatrix} -1 \\ 0 \end{pmatrix}$$

$$\left(\begin{array}{cc|c} 2 & 0 & 1 \\ 0 & 3 & 0 \end{array}\right) \Rightarrow \nabla^2 f\left(x^{(2)}\right) \text{ é não singular} \Rightarrow d_N^{(2)} = \begin{pmatrix} 0.5 \\ 0 \end{pmatrix}$$

Verificar tipo de direção:

Se $\underbrace{\left|\nabla f\left(x^{(2)}\right)^T d_N^{(2)}\right|}_{=0.5} \leq \eta \Leftrightarrow 0.5 \leq 0.0001$ (F) - não ortogonal ao gradiente

Se $\underbrace{\left|\nabla f\left(x^{(2)}\right)^T d_N^{(2)}\right|}_{=0.5} > \eta \Leftrightarrow 0.5 \leq 0.0001$ (F) - direção de descida

$d_N^{(2)}$ é descendente $\Rightarrow \boxed{d_{SN}^{(2)} = \begin{pmatrix} 0.5 \\ 0 \end{pmatrix}}$

Resolução do Exercício 2 (cont.)

Cálculo do comprimento do passo $\alpha^{(2)}$

$$\bar{x} = \begin{pmatrix} 1.5 \\ 0 \end{pmatrix} + \alpha \begin{pmatrix} 0.5 \\ 0 \end{pmatrix}$$

- Para $\alpha = 1 \Rightarrow \bar{x} = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$

  $$f(\bar{x}) \equiv \underbrace{f\begin{pmatrix} 2 \\ 0 \end{pmatrix}}_{=0} \leq \underbrace{f\begin{pmatrix} 1.5 \\ 0 \end{pmatrix}}_{=0.25} + \underbrace{\mu}_{=0.001} \underbrace{\alpha}_{=1} \underbrace{\nabla f\left(x^{(2)}\right)^T d^{(2)}}_{-0.5}$$

  Condição de Armijo: $0 \leq 0.25 + 0.001\,(1)\,(-0.5)$ (Verdadeiro)

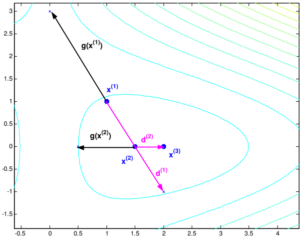
$$\boxed{\alpha^{(2)} = 1}$$

Resolução do Exercício 2 (cont.)

Cálculo do novo ponto

$$x^{(3)} = x^{(2)} + \alpha^{(2)} d_{SN}^{(2)} = \begin{pmatrix} 1.5 \\ 0 \end{pmatrix} + 1 \begin{pmatrix} 0.5 \\ 0 \end{pmatrix}$$

$$\boxed{x^{(3)} = \begin{pmatrix} 2 \\ 0 \end{pmatrix}} \qquad f\left(x^{(3)}\right) = 0$$



Testar critério de paragem

$$\nabla f\left(x^{(3)}\right) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\left\|\nabla f\left(x^{(3)}\right)\right\|_2 = \left\| \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right\|_2 = 0 \leq 0.1 \text{ (V)}$$

Solução: $\left\{ \begin{array}{l} x^* = (2,0)^T \\ f^* = 0 \end{array} \right.$