Preliminares:

- $\mathbf{x}^T\mathbf{y}$ representa o produto escalar, $(\mathbf{x}, \mathbf{y})$, entre dois vetores do $\mathbb{R}^n$.
- $\mathbf{x}$ e $\mathbf{y}$ são ortogonais se e somente se $\mathbf{x}^T\mathbf{y} = 0$.
- Se A é uma matriz simétrica positiva definida então:
  - Se $\mathbf{x}$ é um vetor não-nulo então $\mathbf{x}^T A\mathbf{x} > 0$;
  - $Det(A) > 0$.
  - A tem auto-valores positivos
  - Os elementos diagonais de A são positivos;
- $(AB)^T = B^T A^T$
- $\mathbf{r} = \mathbf{b} - A\mathbf{x}$ - vetor residual. Se $\mathbf{x}$ é solução do sistema linear $A\mathbf{x} = \mathbf{b}$ então $\mathbf{r} = \mathbf{0}$ (vetor nulo).

Seja o sistema linear simétrico positivo definido $A\mathbf{x} = \mathbf{b}$ onde:

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

e consideremos a função $f : I\!R^2 \to I\!R$ definida por

$$\begin{array}{l} f : I\!R^2 \to I\!R \\ \mathbf{x} \to f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T A\,\mathbf{x} - \mathbf{b}^T\mathbf{x} + c, \qquad c \in I\!R \end{array}$$

ou seja,

$$\begin{aligned} f(\mathbf{x}) = \;& \tfrac{1}{2}[x_1\ x_2] \begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - [b_1\ b_2] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + c \\ =\;& \tfrac{1}{2}[x_1\ x_2] \begin{bmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{12}x_1 + a_{22}x_2 \end{bmatrix} - (b_1x_1 + b_2x_2) + c\,. \end{aligned}$$

Logo,

$$f(\mathbf{x}) = \frac{1}{2}\Big\{a_{11}x_1^2 + a_{22}x_2^2 + 2a_{12}x_1x_2\Big\} - \big(b_1x_1 + b_2x_2\big) + c\ .$$

Observamos que:

- A é uma matriz SPD e portanto a_{11} e a_{22} são positivos.
- Desde que $f(\mathbf{x})$ é uma função quadrática de $\mathbf{x}$, do cálculo diferencial, sabemos que necessariamente, $f(\mathbf{x})$ possue um ponto ponto critico, $\mathbf{x}^*$ (onde $f'(\mathbf{x}^*) = 0$) de modo que $f(\mathbf{x}^*)$ é um ponto de de mínimo de $f(\mathbf{x})$.

Calculando $f'(\mathbf{x})$ vem:

$$f'(\mathbf{x}) = \nabla f(\mathbf{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix} = \begin{bmatrix} \frac{1}{2}\{2a_{11}x_1 + 2a_{12}x_2\} - b_1 \\ \frac{1}{2}\{2a_{22}x_2 + 2a_{12}x_1\} - b_2 \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 - b_1 \\ a_{11}x_2 + a_{12}x_1 - b_2 \end{bmatrix} = A\mathbf{x} - \mathbf{b}\ .$$

Logo, $f'(\mathbf{x}) = 0 \Longleftrightarrow A\mathbf{x} = \mathbf{b} \Longleftrightarrow \mathbf{x}$ é solução do sistema linear $A\mathbf{x} = \mathbf{b}$.

Conclusões:

i) Se f tem ponto de mínimo então existe $\mathbf{x}^*$ tal que $f'(\mathbf{x}^*) = 0 \Longrightarrow A\mathbf{x}^* - \mathbf{b} = \mathbf{0} \Longrightarrow A\mathbf{x}^* = \mathbf{b}$.

   Portanto, $\mathbf{x}^*$ é sol. do sistema linear $A\mathbf{x} = \mathbf{b}$.

ii) Se $\mathbf{x}^*$ é sol. do sistema linear $\Longrightarrow\ A\mathbf{x}^* = \mathbf{b} \Longrightarrow A\mathbf{x}^* - \mathbf{b} = \mathbf{0} \Longrightarrow f'(\mathbf{x}^*) = 0 \Longrightarrow \mathbf{x}^*$ é ponto critico de $f(\mathbf{x})$. Como A é SPD então $\mathbf{x}^*$ é ponto de mínimo de $f(\mathbf{x})$.

iii) Portanto, para encontrar a solução do sistema linear $A\mathbf{x}^* = \mathbf{b}$ basta encontrar o ponto de mínimo de $f(\mathbf{x})$.

iv) Do cálculo, sabemos que $f'(\mathbf{x}) = \nabla f(\mathbf{x})$ é um vetor que aponta na direção de maior crescimento da função $f(\mathbf{x})$ e portanto, o vetor $-\nabla f(\mathbf{x})$ aponta em uma direção de menor crescimento da função $f(\mathbf{x})$.

Isso significa que: dados $\mathbf{x}$ e $\lambda > 0$, $f\Big(\mathbf{x} - \lambda\nabla f(\mathbf{x})\Big) < f(\mathbf{x})$. Vamos utilizar essa propriedade de $f(\mathbf{x})$ para calcular seu ponto critico $\mathbf{x}^*$ (e portanto a solução do sistema linear $A\mathbf{x} = \mathbf{b}$.

Método dos Gradientes

Sem perda de generalidades, podemos considerar agora: A uma matriz SPD e

$$\begin{array}{l} f : I\!R^n \to I\!R \\ \mathbf{x} \to f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T A\,\mathbf{x} - \mathbf{b}^T\mathbf{x} + c, \qquad c \in I\!R \end{array}$$

Nesse método, a solução do sistema linear é obtida pela procura do ponto de mínimo da função $f(\mathbf{x})$ como segue.

Dada uma aproximação inicial, $\mathbf{x}^{(0)}$, para a solução do sistema linear $A\mathbf{x} = \mathbf{b}$, uma nova aproximação é calculada por meio de uma correção de $\mathbf{x}^{(0)}$ na direção $-\nabla f(\mathbf{x}^{(0)}) = \mathbf{b} - A\mathbf{x}^{(0)} = \mathbf{r}^{(0)}$ que aponta na direção em que f descresce mais rapidamente (o ponto de mínimo de f que é a solução do sistema linear).

De fato temos que $-\nabla f(\mathbf{x}^{(k)}) = \mathbf{b} - A\mathbf{x}^{(k)} = \mathbf{r}^{(k)}$. Seja $\mathbf{x}^{(1)} = \mathbf{x}^{(0)} + \lambda\mathbf{r}^{(0)}$. Sabemos que $f(\mathbf{x}^{(1)}) < f(\mathbf{x}^{(0)}), \forall\lambda > 0$. O valor de λ é calculado de modo que $f(\mathbf{x}^{(1)})$ seja minimizada em relação a λ, ou seja, quando $\frac{\partial f(\mathbf{x}^{(1)})}{\partial\lambda} = 0$.

Do cálculo, temos:

$$\frac{\partial f(\mathbf{x}^{(1)})}{\partial\lambda} = \Big(f'(\mathbf{x}^{(1)})\Big)^T\frac{\partial\mathbf{x}^{(1)}}{\partial\lambda} = \Big(f'(\mathbf{x}^{(1)})\Big)^T\mathbf{r}^{(0)} = 0. \tag{1}$$

Essa equação sugere que λ seja calculado de modo que $f'(\mathbf{x}^{(1)})$ e $\mathbf{r}^{(0)}$ sejam ortogonais. Substituindo $f'(\mathbf{x}^{(1)}) = \mathbf{r}^{(1)}$ na eq. (1), vem:

$$(\mathbf{r}^{(1)})^T\mathbf{r}^{(0)} = 0 \tag{2}$$

$$\Big[\mathbf{b} - A\mathbf{x}^{(1)}\Big]^T\mathbf{r}^{(0)} = 0, \text{ desde que } \mathbf{r}^{(1)} = \mathbf{b} - A\mathbf{x}^{(1)} \tag{3}$$

$$\Big[\mathbf{b} - A(\mathbf{x}^{(0)} + \lambda\mathbf{r}^{(0)})\Big]^T\mathbf{r}^{(0)} = 0, \text{ desde que } \mathbf{x}^{(1)} = \mathbf{x}^{(0)} + \lambda\mathbf{r}^{(0)} \tag{4}$$

$$\Big[\mathbf{b} - A\mathbf{x}^{(0)}\Big]^T\mathbf{r}^{(0)} - \lambda\Big[A\mathbf{r}^{(0)}\Big]^T\mathbf{r}^{(0)} = 0 \tag{5}$$

$$(\mathbf{r}^{(0)})^T\mathbf{r}^{(0)} = \lambda\Big[A\mathbf{r}^{(0)}\Big]^T\mathbf{r}^{(0)} \tag{6}$$

$$\therefore\ \lambda = \frac{(\mathbf{r}^{(0)})^T\mathbf{r}^{(0)}}{(\mathbf{r}^{(0)})^T A\mathbf{r}^{(0)}}. \tag{7}$$

Tendo calculado, $\mathbf{x}^{(1)} = \mathbf{x}^{(0)} + \lambda\mathbf{r}^{(0)}$, o cálculo de $\mathbf{x}^{(2)}$ segue a mesma metodologia, ou seja, $\mathbf{x}^{(2)} = \mathbf{x}^{(1)} + \lambda\mathbf{r}^{(1)}$ onde o cálculo de λ conduz a $\lambda = \frac{(\mathbf{r}^{(1)})^T\mathbf{r}^{(1)}}{(\mathbf{r}^{(1)})^T A\mathbf{r}^{(1)}}$.

ALGORITMO MÉTODO DOS GRADIENTES (ou MÉTODO DE MÁXIMA DESCIDA)

Dada uma aproximação inicial $\mathbf{x}^{(0)}$ para a solução do sistema linear $A\mathbf{x} = \mathbf{b}$, o método iterativo *MÉTODOS DOS GRADIENTES* calcula aproximações $x^{(k+1)}$ como segue:

Para $k = 0, 1, \ldots$, faça

- Calcule o residuo $\mathbf{r}^{(k)}$

$$\mathbf{r}^{(k)} = \mathbf{b} - A\mathbf{x}^{(k)}$$

- Calcule λ_k

$$\lambda_k = \frac{(\mathbf{r}^{(k)})^T \mathbf{r}^{(k)}}{(\mathbf{r}^{(k)})^T A\mathbf{r}^{(k)}}$$

- Calcule $\mathbf{x}^{(k+1)}$

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \lambda_k \mathbf{r}^{(k)}$$

MÉTODO DOS GRADIENTES CONJUGADOS

Deficiência do método dos gradientes: as direções adotadas podem ser reusadas $\Longrightarrow$ convergência lenta.

Para evitar que uma mesma direção $\mathbf{r}^{(k)}$ se repita, o **Método dos Gradientes Conjugados** propõe uma alteração no Método dos Gradientes: as direções de busca são definidas por vetores $A - ortogonais$. como segue:

VETORES A-ORTOGONAIS: Dois vetores não-nulos $\mathbf{x}$ e $\mathbf{y}$ são A-ortogonais se $(\mathbf{x}^T, A\mathbf{y}) = (\mathbf{y}^T, A\mathbf{x}) = 0$.

Dado $\mathbf{x}^{(0)}$, faça $\mathbf{d}^{(0)} = \mathbf{r}^{(0)} = \mathbf{b} - A\mathbf{x}^{(0)}$. O valor de $\mathbf{x}^{(1)}$ é calculado por $\mathbf{x}^{(1)} = \mathbf{x}^{(0)} + \lambda^{(0)}\mathbf{d}^{(0)}$.

Para obter o tamanho do passo $\lambda^{(0)}$, procedemos como no caso do Método dos Gradientes, isto é,

$$\frac{\partial\Big(f(\mathbf{x}^{(1)})\Big)}{\partial d\lambda} = f'(\mathbf{x}^{(1)})\frac{d\mathbf{x}^{(1)}}{d\lambda} = -(\mathbf{r}^{(1)})^T\mathbf{d}^{(0)} = 0 \text{ (queremos que } \lambda \text{ minize } f(\mathbf{x}^{(1)}))$$

Logo,

$$
\begin{aligned}
&(\mathbf{r}^{(1)})^T\mathbf{d}^{(0)} = 0 \Longrightarrow (\mathbf{d}^{(0)})^T\mathbf{r}^{(1)} = 0 \Longrightarrow (\mathbf{d}^{(0)})^T\Big(\mathbf{b} - A\mathbf{x}^{(1)}\Big) = 0\\
&\Longrightarrow (\mathbf{d}^{(0)})^T\Big(\mathbf{b} - A\big(\mathbf{x}^{(0)} + \lambda^{(0)}\mathbf{d}^{(0)}\big)\Big) = 0\\
&\Longrightarrow (\mathbf{d}^{(0)})^T\Big(\overbrace{\underbrace{\mathbf{b} - A\mathbf{x}^{(0)}}_{\mathbf{r}^{(0)}} - \lambda^{(0)}A\mathbf{d}^{(0)}}^{\mathbf{r}^{(1)}}\Big) = 0\\
&\Longrightarrow (\mathbf{d}^{(0)})^T\mathbf{r}^{(0)} - \lambda^{(0)}(\mathbf{d}^{(0)})^T A(\mathbf{d}^{(0)}) = 0\\
&\Longrightarrow \lambda^{(0)} = \tfrac{(\mathbf{d}^{(0)})^T\mathbf{r}^{(0)}}{(\mathbf{d}^{(0)})^T A(\mathbf{d}^{(0)})} \Longrightarrow \lambda^{(0)} = \tfrac{(\mathbf{r}^{(0)})^T\mathbf{r}^{(0)}}{(\mathbf{d}^{(0)})^T A(\mathbf{d}^{(0)})}
\end{aligned}
$$

Para que $\mathbf{r}^{(1)} = \mathbf{r}^{(1)} - \lambda^{(0)} A\mathbf{d}^{(0)}$ seja ortogonal a nova direção $\mathbf{d}^{(1)} = \mathbf{r}^{(1)} - \beta^{(1)}\mathbf{d}^{(0)}$ é necessário escolher um valor conveniente para $\beta^{(1)}$. Logo, impondo $(\mathbf{r}^{(1)})^T\mathbf{d}^{(1)} = 0$, vem

$$(\mathbf{r}^{(1)})^T\mathbf{d}^{(1)} = 0 \Longrightarrow \left(\mathbf{r}^{(0)} - \frac{(\mathbf{r}^{(0)})^T\mathbf{r}^{(0)}}{(\mathbf{d}^{(0)})^T A(\mathbf{d}^{(0)})} A\mathbf{d}^{(0)}\right)^T \left(\mathbf{r}^{(1)} + \beta^{(1)}\mathbf{d}^{(0)}\right) = 0$$

Nessa fórmula, fazendo os cálculos, simplificações e remanejamentos, pode-se mostrar que obtém-se:

$$\beta^{(1)} = \frac{(\mathbf{r}^{(1)})^T\mathbf{r}^{(1)}}{(\mathbf{r}^{(0)})^T(\mathbf{r}^{(0)})}$$

Em geral, dado $\mathbf{x}^{(0)}$, as aproximações $\mathbf{x}^{(k+1)}$ no método dos gradientes conjugados são calculádas pelo seguinte algoritmo:

- $\mathbf{d}^{(0)} = \mathbf{r}^{(0)} = \mathbf{b} - A\mathbf{x}^{(0)}$
- Para $k = 0, 1, \ldots$
  - $\lambda^{(k)} = \frac{(\mathbf{r}^{(k)})^T\mathbf{r}^{(k)}}{(\mathbf{d}^{(k)})^T A(\mathbf{d}^{(k)})}$
  - $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \lambda^{(k)}\mathbf{d}^{(k)}$
  - $\mathbf{r}^{(k+1)} = \mathbf{r}^{(k)} - \lambda^{(k)} A\mathbf{d}^{(k)}$
  - $\beta^{(k+1)} = \frac{(\mathbf{r}^{(k+1)})^T\mathbf{r}^{(k+1)}}{(\mathbf{r}^{(k)})^T\mathbf{r}^{(k)}}$
  - $\mathbf{d}^{(k+1)} = \mathbf{r}^{(k+1)} + \beta^{(k+1)}\mathbf{d}^{(k)}$

EXEMPLO: Os sistemas lineares abaixo tem sol. exata $\mathbf{x} = [1 \;\; 1 \;\; 1 \;\; 1 \;\; 1]^T$.

$$\underbrace{\overbrace{\begin{bmatrix} 4.1 & -1 & -1 & -1 & -1 \\ -1 & 4.1 & -1 & -1 & -1 \\ -1 & -1 & 4.1 & -1 & -1 \\ -1 & -1 & -1 & 4.1 & -1 \\ -1 & -1 & -1 & -1 & 4.1 \end{bmatrix}}^{b1=[0.1 \;\; 0.1 \;\; 0.1 \;\; 0.1 \;\; 0.1]}}_{A1} \quad \underbrace{\overbrace{\begin{bmatrix} 4 & -1 & -1 & -1 & -1 \\ -1 & 5 & -1 & -1 & -1 \\ -1 & -1 & 5 & -1 & -1 \\ -1 & -1 & -1 & 5 & -1 \\ -1 & -1 & -1 & -1 & 5 \end{bmatrix}}^{b2=[0 \;\; 1 \;\; 1 \;\; 1 \;\; 1]}}_{A2} \quad \underbrace{\overbrace{\begin{bmatrix} 3 & -1 & -1 & -1 & -1 \\ -1 & 4 & -1 & -1 & -1 \\ -1 & -1 & 4 & -1 & -1 \\ -1 & -1 & -1 & 5 & -1 \\ -1 & -1 & -1 & -1 & 6 \end{bmatrix}}^{b3=[-1 \;\; 0 \;\; 0 \;\; 1 \;\; 2]}}_{A3}$$

Dados utilizados: $\mathbf{x}^{(0)} = \mathbf{0}$, $TOL = 10^{-3}$ - Critério parada: $\|\mathbf{r}^{k+1}\|_2 < TOL$.

$\mathbf{x}^{(1)}$ - $\|\mathbf{r}^{(1)}\|_2 = 0.0$	$\mathbf{x}^{(27)}$ - $\|\mathbf{r}^{(27)}\|_2 = 0.8453\,10^{-3}$	$\mathbf{x}^{(105)}$ - $\|\mathbf{r}^{(105)}\|_2 = 0.9799\,10^{-3}$
0.999999	0.999418258666992	0.997215533669189
0.999999	0.999529361724854	0.997796038750797
0.999999	0.999529361724854	0.997796038750797
0.999999	0.999529361724854	0.998176249205648
0.999999	0.999529361724854	0.998444609213319

Resultados obtidos com o metodo gradiente conjugado

$\mathbf{x}^{(1)}$ - $\|\mathbf{r}^{(1)}\|_2 = 0.0$	$\mathbf{x}^{(2)}$ - $\|\mathbf{r}^{(2)}\|_2 = 0.0$	$\mathbf{x}^{(4)}$ - $\|\mathbf{r}^{(4)}\|_2 = 0.1166\,10^{-8}$
0.999999999999997	1.000000000000000	1.000000000000000
0.999999999999997	1.000000000000000	1.000000000000000
0.999999999999997	1.000000000000000	1.000000000000000
0.999999999999997	1.000000000000000	1.000000000000000
0.999999999999997	1.000000000000000	1.000000000000000