# Otimização sem restrições: Condições de otimalidade

Fernanda Costa

Departamento de Matemática
Universidade do Minho

# Outline

# Problema de otimização sem restrições

Consideramos o problema de otimização sem restrições dado por:

$$\underset{w \in \mathbb{R}^d}{\text{minimizar}} \; F(w) \tag{$P_{SR}$}$$

onde

- $w = (w_1, w_2, \cdots, w_d) \in \mathbb{R}^d$ são variáveis contínuas;
- $F : \mathbb{R}^d \to \mathbb{R}$ é não linear e *duas vezes continuamente diferenciável*.

- **Recordar.** Como $F$ é *duas vezes diferenciável em* $\mathbb{R}^d$, o vetor gradiente de $F$ e a matriz Hessiana de $F$ existem para todo $w \in \mathbb{R}^d$.

# Gradiente e Hessiana

**Recordar:** gradiente e Hessiana de $F : \mathbb{R}^d \to \mathbb{R}$ (ver slides T0).

- Vetor gradiente (1ª derivada) de $F$:

$$\nabla F(w) = \begin{bmatrix} \frac{\partial F}{\partial w_1} \\ \vdots \\ \frac{\partial F}{\partial w_d} \end{bmatrix}, \quad \text{existe } \forall w \in \mathbb{R}^d.$$

- Matriz hessiana (2ª derivada) de $F$:

$$\nabla^2 F(w) = \begin{bmatrix} \frac{\partial^2 F}{\partial w_1^2} & \cdots & \frac{\partial^2 F}{\partial w_d \partial w_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 F}{\partial w_1 \partial w_d} & \cdots & \frac{\partial^2 F}{\partial w_d^2} \end{bmatrix}_{d\times d} \quad \text{existe } \forall w \in \mathbb{R}^d.$$

**Nota.**
Como $F$ é *duas vezes continuamente diferenciável*, a matriz hessiana $\nabla^2 F$ é uma matriz simétrica, uma vez que $\frac{\partial^2 F}{\partial w_i \partial w_j} = \frac{\partial^2 F}{\partial w_j \partial w_i}$ para todo $i, j = 1 \ldots, d$

**Exemplo:**

Determinar o vetor gradiente e a matriz hessiana das seguintes funções $F : \mathbb{R}^d \to \mathbb{R}$:

- função linear: $F(w) = a^T w$, onde $a \in \mathbb{R}^d$, $a \neq 0$, é um vetor constante.

  função afim: $F(w) = a^T w + b$, onde $a \in \mathbb{R}^d$, $a \neq 0$, $b \in \mathbb{R}$ é uma constante.

  **Resolução:**
  gradiente: $\nabla F(w) = a$; Hessiana: $\nabla^2 F(w) = 0$ (matriz nula).

- função quadrática: $F(w) = \frac{1}{2} w^T Q w + a^T w + b$, onde $Q \in \mathbb{R}^{d \times d}$ é uma matriz simétrica ($Q^T = Q$).

  **Resolução:**
  gradiente: $\nabla F(w) = \left(\frac{1}{2} Q w + \frac{1}{2} Q^T w\right) + a = Q w + a$; Hessiana: $\nabla^2 F(w) = Q$.

- Determinar o vetor gradiente e a matriz hessiana da função $F : \mathbb{R}^2 \to \mathbb{R}$ dada por $F(w_1, w_2) = w_1^4 + w_1 w_2 + (1 + w_2)^2$

**Resolução:**

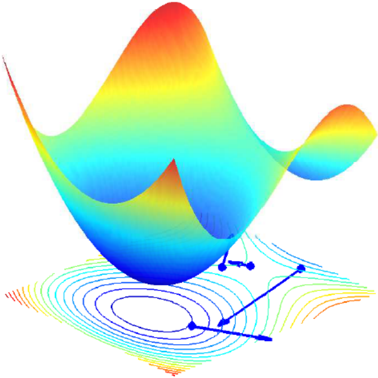
$$\nabla F(w_1, w_2) = \begin{bmatrix} \frac{\partial F(w_1, w_2)}{\partial w_1} \\ \\ \frac{\partial F(w_1, w_2)}{\partial w_2} \end{bmatrix} = \begin{bmatrix} 4w_1^3 + w_2 \\ \\ w_1 + 2(1 + w_2) \end{bmatrix}$$

$$\nabla^2 F(w_1, w_2) = \begin{bmatrix} \frac{\partial^2 F(w_1, w_2)}{\partial w_1^2} & \frac{\partial^2 F(w_1, w_2)}{\partial w_1 w_2} \\ \\ \frac{\partial^2 F(w_2, w_1)}{\partial w_2 w_1} & \frac{\partial^2 F(w_1, w_2)}{\partial w_2^2} \end{bmatrix} = \begin{bmatrix} 12w_1^2 & 1 \\ \\ 1 & 2 \end{bmatrix}$$

# Propriedades do vetor gradiente

O vetor gradiente de uma função tem propriedades interessantes. Destacam-se algumas:

- O vetor gradiente é uma direção de crescimento da função.
- O vetor gradiente é a direção de crescimento mais rápido da função.
- O vetor gradiente é perpendicular à curva de nível da função e aponta na direção de maior crescimento da função.



# Direção de descida

## Definição (Direção de descida)

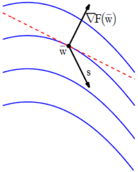
*Considere uma função* $F : \mathbb{R}^d \to \mathbb{R}$ *e um ponto* $\overline{w} \in \mathbb{R}^d$. *Uma direção* $s \in \mathbb{R}^d \setminus \{0\}$ *é uma direção de descida para* $F$ *a partir de* $\overline{w}$, *se existe* $\overline{\eta} > 0$ *tal que*

$$F(\overline{w} + \eta s) < F(\overline{w})$$

*para todo* $\eta \in (0, \overline{\eta})$.

## Teorema 1

*Se* $\nabla F(\overline{w})^T s < 0$, *então* $s$ *é uma direção de descida para* $F$ *a partir de* $\overline{w}$.




($\nabla F(\overline{w})^T s < 0 \Rightarrow$ o declive de $F$ em $\overline{w}$ na direção de $s$ é negativo)

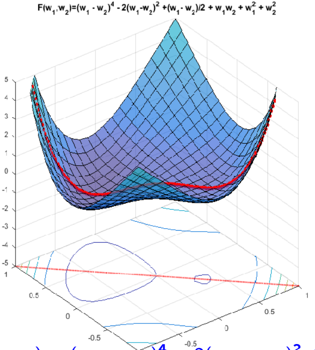
A reta que passa no ponto $\overline{w} \in \mathbb{R}^d$ e tem a direção do vetor $s$ é definida por:

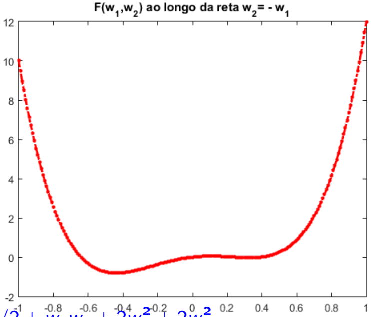
$$w = w(\eta) = \overline{w} + \eta\, s, \ \forall \eta \in \mathbb{R}$$

onde $\eta$ é uma variável escalar que representa o deslocamento ao longo da reta que passa por $\overline{w}$ e tem direção $s$.

A restrição de $F$ aos pontos desta reta é uma função unidimensional definida por:

$$F(\eta) := F(w(\eta)) = F(\overline{w} + \eta\, s)$$




$F(w_1, w_2) = (w_1 - w_2)^4 - 2(w_1 - w_2)^2 + (w_1 - w_2)/2 + w_1 w_2 + 2w_1^2 + 2w_2^2$

$$\underset{w \in \mathbb{R}^d}{\text{minimizar}} \; F(w)$$

Usa a restrição de $F(w)$ ao longo de uma reta ...
Recordar condições suficientes para um mínimo local de uma função de 1 variável $F(\eta) := F(w(\eta)) = F(\overline{w} + \eta\, s)$:

$$\frac{dF}{d\eta} = 0, \quad \text{e} \quad \frac{d^2F}{d\eta^2} > 0$$

Assim, usando a regra da cadeia, temos:

- 1ª derivada de $F(\eta) := F(\overline{w} + \eta\, s)$ é:

  $$\frac{dF}{d\eta} = s^T \nabla F(\overline{w} + \eta\, s) \quad \Rightarrow \text{declive de } F(\eta) \text{ em } \overline{w} \text{ na direção } s \; (\eta = 0).$$

- 2ª derivada de $F(\eta) := F(\overline{w} + \eta\, s)$ é:

  $$\frac{d^2F}{d\eta^2} = s^T \nabla^2 F(\overline{w} + \eta\, s)s \Rightarrow \text{curvatura de } F(\eta) \text{ em } \overline{w} \text{ na direção } s \; (\eta = 0).$$

# Condição necessária de primeira ordem

$$\underset{w \in \mathbb{R}^d}{\text{minimizar}} \; F(w)$$

Num minimizante local $w^*$ de $F$

- o declive de $F(w)$ ao longo da direção $s$ é zero

$$\Rightarrow s^T \nabla F(w^*) = 0, \quad \forall s \in \mathbb{R}^d;$$

- a curvatura de $F(w)$ ao longo da direção $s$ é não negativa

$$\Rightarrow s^T \nabla^2 F(w^*) s \geq 0 \quad \forall s \in \mathbb{R}^d.$$

**Teorema 2 (Condição necessária de 1ª ordem para minimizante)**

*Se $w^*$ é um minimizante local de $F$ então $\nabla F(w^*) = 0$.*

**Definição (Ponto estacionário)**

*Um ponto $w^*$ que satisfaça a condição $\nabla F(w^*) = 0$ é designado por ponto estacionário de $F$.*

# Condição necessária de segunda ordem

**Teorema 3 (Condição necessária de 2ª ordem para minimizante)**

*Se $w^*$ é um minimizante local de $F$ então $\nabla F(w^*) = 0$ e $\nabla^2 F(w^*)$ é semi-definida positiva.*

**Nota.**

- As condições necessários otimalidade ajudam a identificar os pontos candidatos a minimizante. Caso não as verifiquem, não é ponto minimizante.

**Recordar:** Seja $A \in \mathbb{R}^{d \times d}$ uma matriz simétrica.

- $A$ é semi-definida positiva se e só se $s^T As \geq 0$ para todo $s \in \mathbb{R}^d,\ s \neq 0$; OU
- $A$ é semi-definida positiva se e só se os valores próprios de $A$ são não negativos; OU
- $A$ é semi-definida positiva se e só se os determinantes das submatrizes principais de $A$ são não negativos.

# Condição suficientes de segunda ordem

**Teorema 4 (Condições suficientes de 2ª ordem para minimizante)**

*Se $\nabla F(w^*) = 0$ e $\nabla^2 F(w^*)$ é definida positiva então $w^*$ é um minimizante local de $F$.*

**Teorema 5 (Ponto sela)**

*Seja $\overline{w}$ um ponto estacionário de $F$. Se $\nabla^2 F(\overline{w})$ é indefinida, então $\overline{w}$ é ponto sela de $F$.*

**Recordar:** Seja $A \in \mathbb{R}^{d \times d}$ uma matriz simétrica.

- $A$ é definida positiva se e só se $s^T As > 0$ para todo $s \in \mathbb{R}^d$, $s \neq 0$; OU
- $A$ é definida positiva se e só se os valores próprios de $A$ são positivos; OU
- $A$ é definida positiva se e só se os determinantes das submatrizes principais de $A$ são positivos.

- $A$ é indefinida se e só se não for nem semi-definida positiva nem semi-definida negativa.

### Teorema 6

*Se $F$ é convexa, então qualquer minimizante local $w^*$ é um minimizante global de $F$. Mais ainda, se $F$ é diferenciável, então qualquer ponto estacionário é minimizante global de $F$.*

Analogamente, as condições necessárias e suficientes de segunda ordem para um maximizante são:

- $w^*$ maximizante local de $F \Rightarrow \nabla F(w^*) = 0$ e $\nabla^2 F(w^*)$ semi-definida negativa.
- $\nabla F(w^*) = 0$ e $\nabla^2 F(w^*)$ definida negativa $\Rightarrow w^*$ maximizante local de $F$.

**Recordar:** Seja $A \in \mathbb{R}^{d\times d}$ uma matriz simétrica.

- $A$ é semi-definida negativa se e só se $s^T As \leq 0$ para todo $s \in \mathbb{R}^d,\ s \neq 0$; ou
- $A$ é semi-definida negativa se e só se os valores próprios de $A$ são não positivos; ou
- $A$ é semi-definida negativa se e só se os determinantes das submatrizes principais de $A$ de ordem ímpar forem não positivos, e os de ordem par forem não negativos.

- $A$ é definida negativa se e só se $s^T As < 0$ para todo $s \in \mathbb{R}^d,\ s \neq 0$; ou
- $A$ é definida negativa se e só se os valores próprios de $A$ são negativos; ou
- $A$ é definida negativa se e só se os determinantes das submatrizes principais de $A$ de ordem ímpar forem negativos, e os de ordem par forem positivos.

Classificação dos pontos estacionários:

- $\nabla^2 F(w^*)$ definida positiva $\Rightarrow w^*$ **minimizante local**;
- $\nabla^2 F(w^*)$ definida negativa $\Rightarrow w^*$ **maximizante local**;
- $\nabla^2 F(w^*)$ indefinida $\Rightarrow w^*$ **ponto sela**;

- $\nabla^2 F(w^*)$ semi-definida positiva $\Rightarrow w^*$ ou **minimizante local** ou **ponto sela**;
- $\nabla^2 F(w^*)$ semi-definida negativa $\Rightarrow w^*$ ou **maximizante local** ou **ponto sela**.

**Exemplo:**
Considere a função $F : \mathbb{R}^2 \to \mathbb{R}$ dada por $F(w) = (w_1^2 - w_2)(2w_1^2 - w_2)$.
Verificar que $\overline{w} = (0,0)^T$ é um ponto estacionário mas não é minimizante de $F$.

**Resolução:**
Temos $F(w_1, w_2) = 2w_1^4 - 3w_1^2 w_2 + w_2^2$.
$\nabla F(w_1, w_1) = \begin{bmatrix} 8w_1^3 - 6w_1 w_2 \\ -3w_1^2 + 2w_2 \end{bmatrix}$. Como $\nabla F(0,0) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, $\overline{w} = (0,0)^T$ é um ponto estacionário.

$\nabla^2 F(w_1, w_2) = \begin{bmatrix} 24w_1^2 - 6w_2 & -6w_1 \\ -6w_1 & 2 \end{bmatrix}$. Como $\nabla^2 F(0,0) = \begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix}$

é semi-definida positiva, uma vez que os valores próprios são: $0$ e $2$. Logo $\overline{w} = (0,0)^T$ ou é minimizante local ou é ponto sela.

Notar que, a função é negativa para $w_1^2 < w_2 < 2w_1^2$. Por exemplo, para $w_2 = \frac{3}{2}w_1^2$ tem-se $F(w_1, \frac{3}{2}w_1^2) = -\frac{1}{4}w_1^4 < 0$, para $w_1 \neq 0$.

Donde se conclui que, $F(w_1, \frac{3}{2}w_1^2) < F(0,0) = 0$, para $w_1 \neq 0$.

Pelo que $\overline{w} = (0,0)$ satisfaz as condições de otimalidade de segunda ordem, mas não é um minimizante local.

**Exemplo:**
Classifique os pontos estacionários da função $F : \mathbb{R}^2 \to \mathbb{R}$ dada por:

$$F(w_1, w_2) = 2w_1^3 - 3w_1^2 - 6w_1 w_2(w_1 - w_2 - 1)$$

**Resolução:**
Temos $F(w_1, w_2) = 2w_1^3 - 3w_1^2 - 6w_1^2 w_2 + 6w_1 w_2^2 + 6w_1 w_2$.

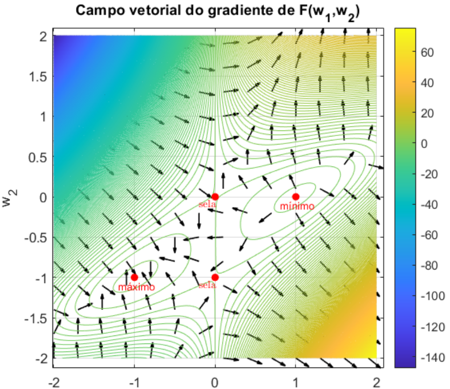
- **gradiente:** $\nabla F(w_1, w_2) = \begin{bmatrix} 6w_1^2 - 6w_1 - 12w_1 w_2 + 6w_2^2 + 6w_2 \\ -6w_1^2 + 12w_1 w_2 + 6w_1 \end{bmatrix}$
- **hessiana:**
  $\nabla^2 F(w_1, w_2) = \begin{bmatrix} 12w_1 - 12w_2 - 6 & -12w_1 + 12w_2 + 6 \\ -12w_1 + 12w_2 + 6 & 12w_1 \end{bmatrix}$
- **pontos estacionários:** Temos de resolver os sistema $\nabla F(w_1, w_2) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$:

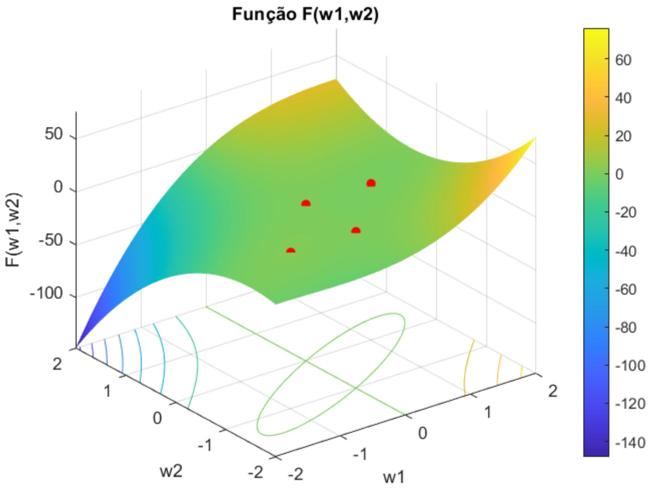
  $$\begin{cases} 6w_1^2 - 6w_1 - 12w_1 w_2 + 6w_2^2 + 6w_2 = 0 \\ -6w_1^2 + 12w_1 w_2 + 6w_1 = 0 \end{cases} \Leftrightarrow \begin{cases} 6w_2^2 + 6w_2 = 0, \\ -6w_1^2 + 12w_1 w_2 + 6w_1 = 0. \end{cases}$$

  Obtendo-se 4 pontos estacionários:
  $\overline{w}_1 = (0, 0)$, $\overline{w}_2 = (0, -1)$, $\overline{w}_3 = (1, 0)$, $\overline{w}_4 = (-1, -1)$.

**Nota:** usar a função eig(H) do matLab.

- $\nabla^2 F(0,0) = \begin{bmatrix} -6 & 6 \\ 6 & 0 \end{bmatrix}$ é indefinida, pois os valores próprios são: -9.7082, 3.7082. Logo $(0,0)$ é ponto sela.
- $\nabla^2 F(1,0) = \begin{bmatrix} 12 & -6 \\ -6 & 12 \end{bmatrix}$ é definida positiva, pois os valores próprios são positivos: 6, 18. Logo $(1,0)$ é um minimizante.
- $\nabla^2 F(0,-1) = \begin{bmatrix} 6 & 18 \\ 18 & 0 \end{bmatrix}$ é indefinida, pois os valores próprios são: -15.2483, 21.2483. Logo $(0,-1)$ é ponto sela.
- $\nabla^2 F(-1,-1) = \begin{bmatrix} -6 & 6 \\ 6 & -12 \end{bmatrix}$ é definida negativa, pois os valores próprios são negativos: -15.7082, -2.2918. Logo $(-1,-1)$ é um maximizante.






- A função $F$ não tem mínimo global, uma vez que $F$ é ilimitada inferiormente. Notar que, $\lim\limits_{w_1 \to -\infty} F(w_1, 0) = -\infty$.
- A função $F$ não tem máximo global, uma vez que $F$ é ilimitada superiormente. Notar que, $\lim\limits_{w_1 \to +\infty} F(w_1, -1) = +\infty$.

# Algumas notas finais sobre as condições de otimalidade

**Limitações das Condições de Otimalidade**

Em geral, as condições de primeira e segunda ordem apenas garantem optimalidade local. É praticamente impossível afirmar algo sobre a optimalidade global, especialmente em problemas não convexos.

Porque é que as condições de otimalidade são importantes?

- Fornecem garantias de que um ponto candidato $w^*$ é um minimizante local.
- Indicam quando um ponto não é ótimo: não satisfaz as condições necessárias.
- Fornecem uma condição de paragem para algoritmos, por exemplo

$$\|\nabla F(w^{(k)})\| \leq \epsilon$$

  onde $\epsilon > 0$ é uma pequena tolerância.
- Orientam o desenvolvimento de métodos numéricos, por exemplo,

$$\underset{w \in \mathbb{R}^d}{\text{minimizar}} \; F(w) \quad \text{``}\Leftrightarrow\text{''} \quad \nabla F(w) = 0$$

  ... sistema de equações não lineares $\cdots$ utiliza o método de Newton.