Conceitos Fundamentais

Fernanda Costa

Departamento de Matemática
Universidade do Minho

Outline

Vetores e Matrizes

- Trabalhamos apenas com vetores e matrizes reais. Vetores são denotados por letras minúsculas e matrizes por letras maiúsculas.
- $\mathbb{R}^d$ denota o espaço dos vetores reais de dimensão d e $\mathbb{R}^{m\times d}$ o espaço das matrizes reais $m \times d$.
- Para $w \in \mathbb{R}^d$, w_i representa a i-ésima componente. Assumimos sempre que w é um vetor coluna:

$$w = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_d \end{bmatrix}.$$

- A transposta de w é o vetor linha

$$w^T = [\, w_1 \; w_2 \; \cdots \; w_d \,],$$

e também é frequentemente escrito com parênteses como $w = (w_1, w_2, \ldots, w_d)$.
- A notação $w \geq 0$ (resp. $w > 0$) indica que $w_i \geq 0$ (resp. $w_i > 0$) para todo $i = 1, \ldots, d$.

- Para $w, z \in \mathbb{R}^d$, o produto interno é

$$w^T z = w_1 z_1 + w_2 z_2 + \ldots + w_d z_d = \sum_{i=1}^{d} w_i z_i.$$

- Dada uma matriz $A \in \mathbb{R}^{m \times d}$, a_{ij} denota o elemento da linha i e coluna j.
- A transposta de A, denotada por A^T, é a matriz $d \times m$ obtida trocando linhas por colunas.

$$\underline{\text{Exemplo:}} \quad \begin{bmatrix} 1 & -2 & 0 \\ 0 & 2 & 1 \end{bmatrix}^T = \begin{bmatrix} 1 & 0 \\ -2 & 2 \\ 0 & 1 \end{bmatrix}.$$

- A matriz A diz-se *quadrada de ordem* d se $m = d$.

- Uma matriz quadrada A é *simétrica* se $A = A^T$.

  Exemplo: $\begin{bmatrix} 1 & a & b \\ 1 & 2 & 3 \\ 2 & c & 3 \end{bmatrix}$ é simétrica para $a = 1,\ b = 2,\ c = 3$.

- Uma matriz A quadrada $d \times d$ diz-se:
  - *definida positiva* se $s^T As > 0$ para todo $s \in \mathbb{R}^d,\ s \neq 0$;
  - *semidefinida positiva* se $s^T As \geq 0$ para todo $s \in \mathbb{R}^d$;
  - *definida negativa* se $s^T As < 0$ para todo $s \in \mathbb{R}^d,\ s \neq 0$;
  - *semidefinida negativa* se $s^T As \leq 0$ para todo $s \in \mathbb{R}^d$;
  - *indefinida* se não for nem semidefinida positiva e nem semidefinida negativa.

Notação: Dizemos que A é *SDP* se A é simétrica e definida positiva, e *SSDP* se A é simétrica e semidefinida positiva. Similarmente, *SDN* e *SSDN* indicam matrizes simétricas definida negativa e semidefinida negativa, respectivamente.

Exemplo: Verifique se a matriz $A = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}$ é SDP.

Resolução:

Reparar que $A = A^{\mathsf{T}}$, i.e. A é uma matriz simétrica.

Para qualquer $s = (s_1, s_2) \in \mathbb{R}^2$ e $s \neq 0$,

$$s^{\mathsf{T}} A s = \begin{bmatrix} s_1 & s_2 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} = \begin{bmatrix} 2s_1 - s_2, & -s_1 + s_2 \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix}$$
$$= 2s_1^2 - s_1 s_2 - s_1 s_2 + s_2^2 = s_1^2 + (s_1 - s_2)^2 \geq 0.$$

Note-se que $s_1^2 + (s_1 - s_2)^2 = 0$ se e só se $s_1 = s_2 = 0$.

Portanto, a matriz A é SDP.

- Uma matriz A simétrica quadrada $d \times d$ é:
  - *definida positiva* se e só se todos os seus valores próprios são positivos;
  - *semidefinida positiva* se e só se todos os seus valores próprios são não negativos;
  - *definida negativa* se e só se todos os seus valores próprios são negativos;
  - *semidefinida negativa* se e só se todos os seus valores próprios são não positivos;
  - *indefinida* se e só se A tem pelo menos um valor próprio positivo e pelo menos um valor próprio negativo.

- A diagonal de uma matriz $A \in \mathbb{R}^{m \times d}$ é formada pelos elementos A_{ii}, com $i = 1, \ldots, \min(m, d)$.

- A matriz $A \in \mathbb{R}^{m \times d}$ é *triangular inferior* se $a_{ij} = 0$ para $i < j$.

$$\underline{\text{Exemplo:}} \begin{bmatrix} 1 & 0 & 0 \\ 4 & 2 & 0 \\ 3 & 5 & 6 \end{bmatrix}$$

- É *triangular superior* se $a_{ij} = 0$ para $i > j$.

$$\underline{\text{Exemplo:}} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix}$$

- A matriz A é *diagonal* se $a_{ij} = 0$ para $i \neq j$.

$$\underline{\text{Exemplo:}} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

- A <u>matriz identidade</u>, denotada por I, é uma matriz quadrada diagonal cujos elementos da diagonal são todos iguais a 1.

- Uma matriz quadrada $A \in \mathbb{R}^{n \times n}$ é não singular se existir uma matriz $B \in \mathbb{R}^{n \times n}$ tal que $AB = BA = I$.
    - Esta matriz B é denotada por A^{-1} e chamada de inversa de A.
    - A inversa de A^T é a transposta de A^{-1}, ou seja, $(A^T)^{-1} = (A^{-1})^T$.
- Uma matriz quadrada Q é ortogonal se $Q\,Q^T = I$, ou seja, a inversa de uma matriz ortogonal é a sua transposta, $Q^{-1} = Q^T$.

- O determinante de uma matriz quadrada $A \in \mathbb{R}^{d\times d}$, denotado por $\det(A)$ ou $|A|$, é definido por

$$\det(A) = \begin{cases} a_{11}, & d = 1, \\ \displaystyle\sum_{j=1}^{d} (-1)^{1+j} a_{1j} \det(\widetilde{A}_{1j}), & d \geq 2, \end{cases}$$

onde $\widetilde{A}_{1j}$ é a submatriz obtida eliminando a linha 1 e a coluna j.

Nota:

1. $\det\left(\left|\begin{array}{cc} a & b \\ c & d \end{array}\right|\right) = a \times d - b \times c$

2. $\det\left(\left|\begin{array}{ccc} a & b & c \\ d & e & f \\ g & h & i \end{array}\right|\right) =$

$$= a \times \det\left(\left|\begin{array}{cc} e & f \\ h & i \end{array}\right|\right) - b \times \det\left(\left|\begin{array}{cc} d & f \\ g & i \end{array}\right|\right) + c \times \det\left(\left|\begin{array}{cc} d & e \\ g & h \end{array}\right|\right)$$
$$= a \times e \times i - a \times f \times h - b \times d \times i + b \times f \times g + c \times d \times h - c \times e \times g$$

- Os menores principais de $A \in \mathbb{R}^{d \times d}$ são os determinantes $\det(A_k)$ das suas submatrizes principais: $A_k = [a_{ij}]_{i,j=1}^{k}$ para $k = 1, \ldots d$.

Exemplo: Calcular os menores principais da matriz $A = \begin{bmatrix} -2 & -1 & -1 \\ 1 & 2 & 1 \\ -1 & -1 & -2 \end{bmatrix}$.

Resolução:

$\det(A_1) = |-2| = -2$;

$$\det(A_2) = \begin{vmatrix} -2 & -1 \\ 1 & 2 \end{vmatrix} = -2 \times 2 - (-1 \times 1) = -3;$$

$$\det(A_3) = \begin{vmatrix} -2 & -1 & -1 \\ 1 & 2 & 1 \\ -1 & -1 & -2 \end{vmatrix} =$$

$$= -2\det\left(\begin{vmatrix} 2 & 1 \\ -1 & -2 \end{vmatrix}\right) - (-1)\det\left(\begin{vmatrix} 1 & 1 \\ -1 & -2 \end{vmatrix}\right) + (-1)\det\left(\begin{vmatrix} 1 & 2 \\ -1 & -1 \end{vmatrix}\right)$$

$$= -2(-4+1) + (-2+1) - (-1+2) = 6 - 1 - 1 = 4.$$

- **Critério de Sylvester:** Uma matriz quadrada $A \in \mathbb{R}^{d\times d}$ é:
  - *definda positiva* se e só se todos os menores principais forem positivos;
  - *semidefinida positiva* se todos menores principais forem não negativos;
  - *definida negativa* se e só se todos os menores de ordem impar forem negativos e os todos os menores de ordem par forem positivos;
  - *semidefinida negativa* se e só se todos os menores de ordem impar forem não positivos e todos os menores de ordem par forem não negativos.

- **Função.** Seja $\text{dom } F \subseteq \mathbb{R}^d$. Uma *função real de d variáveis reais* $(w_1, w_2, \ldots, w_d)$ é uma correspondência

$$F : \text{dom } F \to \mathbb{R}$$

que a cada elemento $w = (w_1, w_2, \ldots, w_d) \in \text{dom } F$ associa um único valor real, denotado por $F(w)$ ou $F(w_1, w_2, \cdots, w_d)$, isto é,

$$\begin{aligned} F : \text{dom } F &\to \mathbb{R} \\ (w_1, \ldots, w_d) &\mapsto F(w_1, \ldots, w_d) \end{aligned}$$

- O conjunto $\text{dom } F$ é o *domínio* de F; e o *contradomínio* é o conjunto de valores que F toma em $\mathbb{R}$, isto é, $\{F(w) : w \in \text{dom } F\}$.

- **Gráfico.** Seja $F : \text{dom}\, F \subseteq \mathbb{R}^d \to \mathbb{R}$, uma função de d variáveis reais. Chama-se *gráfico de* F, denotado por $\text{graf}(F)$, o conjunto:

$$\text{graf}(F) = \{(w_1, \ldots, w_d, w_{d+1}) \in \mathbb{R}^{d+1} : w_{d+1} = F(w) \text{ e } w \in \text{dom } F\}.$$

  Neste caso, o gráfico é a superfície de equação $w_{d+1} = F(w_1, w_2, \ldots, w_d)$.

- **Superfície de nível.** Seja $F : \text{dom}\, F \subseteq \mathbb{R}^d \to \mathbb{R}$ e seja $k \in \mathbb{R}$. Chama-se *superfície de nível de* F com valor k, denotado por CN_k, ao conjunto de pontos $w = (w_1, w_2, \ldots, w_d) \in \text{dom}\, F$ tais que $F(w_1, w_2, \ldots, w_d) = k$, isto é,

$$\mathrm{CN}_k = \{w \in \text{dom}\, F : F(w) = k\}.$$

  Se $d = 2$, CN_k é chamado de **curva de nível**.

Limite e Continuidade

- **Limite.** Seja $F : \text{dom}\, F \subseteq \mathbb{R}^d \to \mathbb{R}$. Se $a = (a_1, \ldots, a_d)$ é, ou ponto interior de $\text{dom}\, F$ ou ponto fronteiro não isolado, diz-se o número real L é o limite de $F(w_1, \ldots, w_d)$ quando $w = (w_1, \ldots, w_d)$ tende para $a = (a_1, \ldots, a_d)$ se

$$\forall \varepsilon > 0, \exists \delta > 0 : (0 < \|w - a\| < \delta \text{ e } w \in \text{dom}\, F) \Rightarrow |F(w) - L| < \varepsilon,$$

e escrevemos

$$\lim_{w \to a} F(w_1, \ldots, w_d) = L$$

- **Continuidade.** A função F diz-se contínua em $a = (a_1, \ldots, a_d)$ se

$$\lim_{w \to a} F(w_1, \ldots, w_d) = F(a_1, \ldots, a_d)$$

- Exemplo: Considere a função $f(w) = \begin{cases} -w, & \text{se } w \in [-1,1],\ w \neq 0, \\ 5, & \text{para todos os outros } w \in [-10,10]. \end{cases}$

Resolução:
A função está definida em todos os pontos do domínio [-10,10] e é contínua em todos os pontos exceto nos pontos $w = 0$, $w = 1$ e $w = -1$. Vejamos:

Ponto $w = 0$:

$$\lim_{w\to 0^-} f(w) = \lim_{w\to 0^+} f(w) = 0 \neq f(0) = 5.$$

O limite existe mas f não é contínua em $w = 0$.

Ponto $w = -1$:

$$\lim_{w\to -1^-} f(w) = \lim_{w\to -1^-} 5 = 5, \quad \lim_{w\to -1^+} f(w) = \lim_{w\to -1^+} -w = -(-1) = 1.$$

Como os limites laterais são diferentes, não existe limite.
f não é contínua em $w = -1$.

Ponto $w = 1$:

$$\lim_{w\to 1^-} f(w) = \lim_{w\to 1^-} -w = -1, \quad \lim_{w\to 1^+} f(w) = \lim_{w\to 1^+} 5 = 5.$$

Como os limites laterais são diferentes, não existe limite.
f não é contínua em $w = 1$.

Derivadas

Seja $g : \mathbb{R} \to \mathbb{R}$ uma função real de uma variável real.

- A *primeira derivada* $g'(w)$ é definida por

$$\frac{dg}{dw} = g'(w) \stackrel{\text{def}}{=} \lim_{\epsilon \to 0} \frac{g(w+\epsilon) - g(w)}{\epsilon}.$$

- A *segunda derivada* é obtida substituindo g por g' nesta mesma fórmula; ou seja,

$$\frac{d^2 g}{dw^2} = g''(w) \stackrel{\text{def}}{=} \lim_{\epsilon \to 0} \frac{g'(w+\epsilon) - g'(w)}{\epsilon}.$$

- Suponha agora que w depende, por sua vez, de outra quantidade η (denotamos esta dependência escrevendo $w = w(\eta)$). Podemos utilizar a *regra da cadeia* para calcular a derivada de g em relação a η:

$$\frac{dg(w(\eta))}{d\eta} = \frac{dg}{dw}\,\frac{dw}{d\eta} = g'(w)\,w'(\eta).$$

Gradiente

Considere-se agora $f : \mathrm{dom}\, f \subseteq \mathbb{R}^d \to \mathbb{R}$.

- O *gradiente* de f, denotado por ∇f, é o vetor das primeiras derivadas parciais de f:

$$\nabla f(w) = \begin{bmatrix} \dfrac{\partial f}{\partial w_1} \\ \vdots \\ \dfrac{\partial f}{\partial w_d} \end{bmatrix}$$

- Cada componente do vetor gradiente quantifica a taxa de variação local da função em relação à respetiva variável, isto é, indica o declive da função ao longo de cada direção coordenada (*eixo* w_i).

- A função f é *diferenciável* em $\mathrm{dom}\, f$ se $\nabla f(w)$ existir para todo $w \in \mathrm{dom}\, f$, e *continuamente diferenciável* se $\nabla f(w)$ for uma função contínua de w.

- Para cada $\overline{w} \in \mathrm{dom}\, f$, o vetor gradiente $\nabla f(\overline{w})$ é perpendicular à curva de nível de f em $\overline{w}$ e aponta na *direção de maior crescimento da função a partir do ponto* $\overline{w}$.

Derivada direcional

- Por vezes estamos interessados na taxa de variação numa direção que não coincide com uma direção coordenada (w_i). A taxa de variação numa qualquer direção $\mathrm{p} \in \mathbb{R}^d$ é quantificada pela *derivada direcional*.

  A *derivada direcional de f* em $\overline{w}$ na direção de $p \in \mathbb{R}^d$, denotada por $D_p f(\overline{w})$, é dada por

$$\begin{aligned} D_p f(\overline{w}) &= \lim_{\epsilon \to 0} \frac{f(\overline{w} + \epsilon p) - f(\overline{w})}{\epsilon} \\ &= \nabla f(\overline{w})^T p \quad \textbf{(projeção do gradiente na direção } p\textbf{)} \\ &= \|\nabla f(\overline{w})\| \|p\| \cos\theta. \end{aligned}$$

  onde θ é o ângulo formado por $\nabla f(\overline{w})$ e p.

- Quando p é um vetor unitário alinhado com uma das coordenadas cartesianas w_i, este produto escalar dá derivada parcial correspondente $\dfrac{\partial f}{\partial w_i}$.

  Exemplo: Projeção do vetor gradiente numa direção unitária arbitrária p de uma função de duas variáveis:

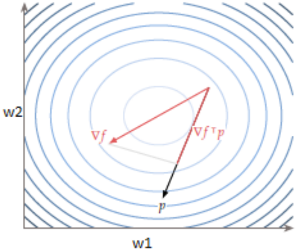


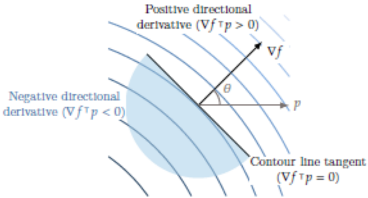
Figura: Projeção do gradiente numa direção unitária arbitrária p.

Como

$$D_{\mathrm{p}}f(\overline{w}) = \nabla f(\overline{w})^T \mathrm{p} = \|\nabla f(\overline{w})\| \, \|\mathrm{p}\| \cos\theta,$$

admite um:

- valor máximo quando p tem a mesma direção e sentido de ∇f, ou seja, $\theta = 0°$;
- valor mínimo quando p tem a mesma direção, mas sentido contrário ao de ∇f, ou seja, $\theta = 180°$;
- valor zero quando p é perpendicular a ∇f, ou seja, $\theta = \pm 90°$
- valor positivo quando $\theta \in (-90°, 90°)$, indicando uma direção de aumento para f (ver figura);
- valor negativo quando $\theta \in (90°, 180°]$, indicando uma direção de descida para f.



Hessiana

- A *matriz hessiana de f*, denotada por $\nabla^2 f$, é a matriz das segundas derivada parciais de f:

$$\nabla^2 f(w) = \begin{bmatrix} \frac{\partial^2 f}{\partial w_1^2} & \cdots & \frac{\partial^2 f}{\partial w_1 \partial w_d} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial w_d \partial w_1} & \cdots & \frac{\partial^2 f}{\partial w_d^2} \end{bmatrix}_{d\times d}$$

- A função f é *duas vezes diferenciável* em $\textbf{dom} \, f$ se $\nabla^2 f(w)$ existir para todo $w \in \textbf{dom} \, f$, e *duas vezes continuamente diferenciável* se $\nabla^2 f(w)$ for contínua em $\textbf{dom} \, f$.
- Notar que quando f é *duas vezes continuamente diferenciável*, a hessiana é uma matriz simétrica, uma vez que

$$\frac{\partial^2 f}{\partial w_i \partial w_j} = \frac{\partial^2 f}{\partial w_j \partial w_i} \quad \text{para todo } i, j = 1 \ldots, d$$

- Exemplo: Determine o gradiente e a matriz Hessiana da função $F(w_1, w_2) = 10(w_2 - w_1^2)^2 + (1 - w_1)^2$ no ponto $\overline{w} = (0, 1)$.

Resolução:

Para determinar o gradiente, temos de calcular as primeiras derivadas parciais:

$$\frac{\partial F(w_1, w_2)}{\partial w_1} = 20(w_2 - w_1^2)(-2w_1) - 2(1 - w_1),$$

$$\frac{\partial F(w_1, w_2)}{\partial w_2} = 20(w_2 - w_1^2).$$

Portanto, $\nabla F(0,1) = \begin{bmatrix} -2 \\ 20 \end{bmatrix}$.

Para determinar a matriz Hessiana, temos de calcular as segundas derivadas parciais:

$$\frac{\partial^2 F(w_1, w_2)}{\partial w_1^2} = -40w_2 + 120w_1^2 + 2,$$

$$\frac{\partial^2 F(w_1, w_2)}{\partial w_1 w_2} = \frac{\partial^2 F(w_1, w_2)}{\partial w_2 w_1} = -40w_1,$$

$$\frac{\partial^2 F(w_1, w_2)}{\partial w_2^2} = 20.$$

Portanto, $\nabla^2 F(0,1) = \begin{bmatrix} \frac{\partial^2 F(0,1)}{\partial w_1^2} & \frac{\partial^2 F(0,1)}{\partial w_1 \partial w_2} \\ \frac{\partial^2 F(0,1)}{\partial w_2 \partial w_1} & \frac{\partial^2 F(0,1)}{\partial w_2^2} \end{bmatrix} = \begin{bmatrix} -38 & 0 \\ 0 & 20 \end{bmatrix}$