

Fundamentos Básicos de Álgebra Linear e Otimização

Índice Geral

1	Escalar.....	3
2	Vetor	3
3	Matriz.....	5
4	Conjuntos e operações com conjuntos.....	6
4.1	Conjuntos especiais de números reais	8
5	Espaço Vetorial Linear	9
5.1	Axiomas.....	9
5.2	Propriedades adicionais.....	10
5.3	Exemplos	11
5.4	Produto cartesiano	11
5.5	Subespaço vetorial linear.....	12
5.6	Conjuntos convexos.....	13
5.7	Combinação linear e combinação convexa.....	14
5.8	Dependência linear e dimensão de um espaço vetorial	15
5.9	Produto externo	16
5.10	Produto interno	17
5.11	Norma, semi-norma e quase-norma.....	18
5.12	Ângulo entre dois vetores.....	21
5.13	Ortogonalidade e ortonormalidade entre dois vetores	21
5.14	Espaços ortogonais	22
5.15	Projeção de um vetor em uma determinada direção	22
5.16	Vetores ortonormais gerados a partir de vetores linearmente independentes.....	23
6	Transformações e funcionais	25

6.1	Transformações lineares	25
6.2	Operadores lineares	26
6.3	Posto de uma matriz	27
6.4	Matrizes idempotentes.....	28
6.5	Definições adicionais para matrizes	28
6.6	Matrizes singulares e não-singulares	31
6.7	Autovalores e autovetores	31
6.8	Formas Quadráticas.....	32
6.8.1	Cálculo diferencial aplicado a formas quadráticas.....	33
6.8.2	Normas ponderadas e a medida de distância de Mahalanobis	34
6.9	Matrizes simétricas: positividade e autovalores	35
6.10	A inversa de uma matriz	39
6.11	O lema de inversão de matrizes	40
6.12	A pseudo-inversa de uma matriz	40
6.12.1	Exemplos de pseudo-inversas	41
6.12.2	Uso de pseudo-inversão para a solução de sistemas lineares.....	42
6.13	Operadores de projeção ortogonal.....	43
6.13.1	Um exemplo de operador simétrico e idempotente.....	44
6.14	Decomposição em valores singulares.....	46
6.15	Transformações contínuas	50
6.16	Funcional	51
6.17	Funcional convexo.....	51
6.18	Funcional convexo diferenciável.....	53
7	Mínimos Locais	54
8	Expansão em Série de Taylor	55
9	Condição Necessária de Otimalidade	56
10	Condição Suficiente de Otimalidade	57
11	Referências bibliográficas.....	60

1 Escalar

- uma variável que assume valores no eixo dos números reais é denominada escalar. Os escalares são descritos por letras minúsculas do alfabeto romano expressas em itálico, ou do alfabeto grego. O conjunto de todos os escalares reais é representado por \Re ou \Re^1 .
- o módulo de um escalar real x é dado na forma: $|x| = \begin{cases} x & \text{se } x \geq 0 \\ -x & \text{se } x < 0 \end{cases}$

2 Vetor

- um arranjo ordenado de n escalares $x_i \in \Re$ ($i=1,2,\dots,n$) é denominado vetor de dimensão n . Os vetores são descritos por letras minúsculas do alfabeto romano expressas em negrito, e assumem a forma de vetores-coluna ou vetores-linha. Neste estudo, todos os vetores são representados por vetores-coluna, na forma:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad \text{ou} \quad \mathbf{x} = [x_1 \quad x_2 \quad \cdots \quad x_n]^T.$$

- o conjunto de todos os vetores de dimensão n com elementos reais é representado por \Re^n . Diz-se então que $\mathbf{x} \in \Re^n$.
- um escalar é um vetor de dimensão 1.
- vetor $\mathbf{0}_n$: é o vetor nulo de dimensão n , com todos os elementos iguais a zero. O subscrito n é suprimido quando não há margem à dúvida.
- vetor $\mathbf{1}_n$: é o vetor de dimensão n com todos os elementos iguais a 1.
- vetor \mathbf{e}_i : é o vetor normal unitário de dimensão n (a dimensão deve ser indicada pelo contexto) com todos os elementos iguais a 0, exceto o i -ésimo elemento que é igual a 1. Neste caso, $1 \leq i \leq n$.

3 Matriz

- um arranjo ordenado de $m.n$ escalares x_{ij} ($i=1,2,\dots,m; j=1,2,\dots,n$) é denominado matriz de dimensão $m \times n$. As matrizes são descritas por letras maiúsculas do alfabeto romano expressas em itálico, e assumem a forma:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix}.$$

- o conjunto de todas as matrizes $m \times n$ com elementos reais é representado por $\Re^m \times \Re^n$ ou $\Re^{m \times n}$. Diz-se então que $X \in \Re^{m \times n}$.

- as colunas da matriz X são vetores-coluna descritos por $\mathbf{x}_i = \begin{bmatrix} x_{1i} \\ x_{2i} \\ \vdots \\ x_{mi} \end{bmatrix}$, $i=1,\dots,n$.

- as linhas da matriz X são vetores-linha descritos por $\mathbf{x}_{(j)} = [x_{j1} \quad x_{j2} \quad \cdots \quad x_{jn}]$, $j=1,\dots,m$.
- um vetor é uma matriz com número unitário de linhas e/ou colunas.

4 Conjuntos e operações com conjuntos

- um conjunto pode ser definido como uma agregação de objetos. Os conjuntos são descritos por letras maiúsculas do alfabeto romano expressas em itálico. Por conveniência de notação, alguns conjuntos especiais são descritos por símbolos específicos. Exemplos:
- \mathbb{N} : conjunto dos números naturais
- \mathbb{R} : conjunto dos números reais
- \mathbb{C} : conjunto dos números complexos
- o estado lógico ou associação de um elemento \mathbf{x} a um conjunto X qualquer é representado por

$\mathbf{x} \in X$: \mathbf{x} pertence a X

$\mathbf{x} \notin X$: \mathbf{x} não pertence a X

- um conjunto pode ser especificado listando-se seus elementos entre colchetes

$$X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$$

ou evidenciando uma ou mais propriedades comuns aos seus elementos

$$X_2 = \{\mathbf{x} \in X_1 \text{ tal que } P(\mathbf{x}) \text{ é verdade}\} \text{ ou } X_2 = \{\mathbf{x} \in X_1 : P(\mathbf{x})\}$$

- as principais operações entre conjuntos são:
 - União: $X_1 \cup X_2 = \{\mathbf{x} : \mathbf{x} \in X_1 \text{ ou } \mathbf{x} \in X_2\}$;
 - Interseção: $X_1 \cap X_2 = \{\mathbf{x} : \mathbf{x} \in X_1 \text{ e } \mathbf{x} \in X_2\}$;
- $X_1 \cap X_2 = \emptyset$ (conjunto vazio) se X_1 e X_2 são conjuntos disjuntos.
- o complemento de um conjunto X é representado por \bar{X} e é definido na forma:

$$\bar{X} = \{\mathbf{x} : \mathbf{x} \notin X\}.$$

- S é um subconjunto de X , se $\mathbf{x} \in S$ implica $\mathbf{x} \in X$. Neste caso, diz-se que S está contido em X ($S \subset X$) ou que X contém S ($X \supset S$). Se $S \subset X$ e S não é igual a X , então S é um subconjunto próprio de X .

4.1 Conjuntos especiais de números reais

- se a e b são números reais ($a, b \in \mathfrak{R}$), define-se:

$$[a, b] = \{x : a \leq x \leq b\}$$

$$(a, b] = \{x : a < x \leq b\}$$

$$[a, b) = \{x : a \leq x < b\}$$

$$(a, b) = \{x : a < x < b\}$$

- se X é um conjunto de números reais, então o menor limitante superior de X , dado por

$$\bar{x} = \sup x = \sup_{x \in X} \{x : x \in X\},$$

é o supremo de X , e o maior limitante inferior de X , dado por

$$\underline{x} = \inf_{x \in X} x = \inf\{x : x \in X\},$$

é o ínfimo de X . Se $\bar{x} = +\infty$ então X não é limitado superiormente. De forma análoga, se $\underline{x} = -\infty$ então X não é limitado inferiormente.

5 Espaço Vetorial Linear

- um espaço vetorial X , associado a um campo F , consiste de um conjunto de elementos (vetores) sobre os quais estão definidas 2 operações:

1. Adição ($X \times X \rightarrow X$): $(\mathbf{x} + \mathbf{y}) \in X, \forall \mathbf{x}, \mathbf{y} \in X$;
2. Multiplicação por escalar ($F \times X \rightarrow X$): $(\alpha \cdot \mathbf{x}) \in X, \forall \mathbf{x} \in X$ e $\alpha \in F$.

5.1 Axiomas

- $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$ (propriedade comutativa)
- $\left. \begin{array}{l} \mathbf{x} + (\mathbf{y} + \mathbf{z}) = (\mathbf{x} + \mathbf{y}) + \mathbf{z} \\ \alpha \cdot (\beta \cdot \mathbf{x}) = (\alpha \cdot \beta) \cdot \mathbf{x} \end{array} \right\}$ (propriedade associativa)

- $\left. \begin{array}{l} \alpha \cdot (\mathbf{x} + \mathbf{y}) = \alpha \cdot \mathbf{x} + \alpha \cdot \mathbf{y} \\ (\alpha + \beta) \cdot \mathbf{x} = \alpha \cdot \mathbf{x} + \beta \cdot \mathbf{x} \end{array} \right\}$ (propriedade distributiva)
- $\mathbf{x} + \mathbf{0} = \mathbf{x}$ (vetor nulo)
- $1 \cdot \mathbf{x} = \mathbf{x}$ (elemento neutro)

5.2 Propriedades adicionais

- $0 \cdot \mathbf{x} = \mathbf{0}$
- $\alpha \cdot \mathbf{0} = \mathbf{0}$
- $\mathbf{x} + \mathbf{y} = \mathbf{x} + \mathbf{z} \Rightarrow \mathbf{y} = \mathbf{z}$
- $\alpha \cdot \mathbf{x} = \alpha \cdot \mathbf{y}, \alpha \neq 0 \Rightarrow \mathbf{x} = \mathbf{y}$
- $\alpha \cdot \mathbf{x} = \beta \cdot \mathbf{x}, \mathbf{x} \neq \mathbf{0} \Rightarrow \alpha = \beta$
- $\left. \begin{array}{l} \alpha \cdot (\mathbf{x} - \mathbf{y}) = \alpha \cdot \mathbf{x} - \alpha \cdot \mathbf{y} \\ (\alpha - \beta) \cdot \mathbf{x} = \alpha \cdot \mathbf{x} - \beta \cdot \mathbf{x} \end{array} \right\}$ (propriedade distributiva)

5.3 Exemplos

- considere o campo F como sendo o conjunto dos números reais (\mathbb{R}):
 1. conjunto dos números reais: $X \equiv \mathbb{R}$
 2. conjunto dos vetores n -dimensionais, com elementos reais: $X \equiv \mathbb{R}^n$
 3. conjunto das matrizes $m \times n$ com elementos reais: $X \equiv \mathbb{R}^m \times \mathbb{R}^n$ ou $X \equiv \mathbb{R}^{m \times n}$

5.4 Produto cartesiano

- o produto cartesiano de dois espaços vetoriais X e Y (os dois espaços vetoriais devem estar associados ao mesmo campo) é dado por $X \times Y$ e é definido como o conjunto de pares ordenados (\mathbf{x}, \mathbf{y}) , com $\mathbf{x} \in X$ e $\mathbf{y} \in Y$. As operações de adição e multiplicação são definidas na forma:

$$\triangleright (\mathbf{x}_1, \mathbf{y}_1) + (\mathbf{x}_2, \mathbf{y}_2) = (\mathbf{x}_1 + \mathbf{x}_2, \mathbf{y}_1 + \mathbf{y}_2)$$

$$\triangleright \alpha \cdot (\mathbf{x}, \mathbf{y}) = (\alpha \cdot \mathbf{x}, \alpha \cdot \mathbf{y})$$

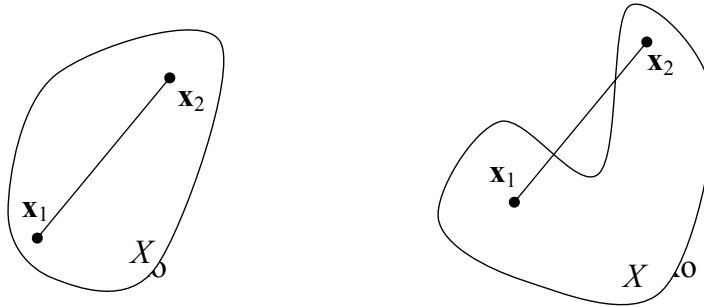
- por convenção, $X^n \equiv \underbrace{X \times X \times \cdots \times X}_{n \text{ vezes}}$

5.5 Subespaço vetorial linear

- um subconjunto não-vazio S de um espaço vetorial linear X é um subespaço de X se $\alpha \cdot \mathbf{x} + \beta \cdot \mathbf{y} \in S$ sempre que \mathbf{x} e $\mathbf{y} \in S$.
- dito de outra forma, seja (X, F) um espaço vetorial linear e S um subconjunto de X . Diz-se então que (S, F) é um subespaço vetorial de (X, F) se S forma um espaço vetorial sobre F através das mesmas operações definidas sobre (X, F) .
- exemplos:
 1. $S \equiv \{0\}$
 2. $S \equiv \mathbb{R}^n$ é subespaço de $X \equiv \mathbb{R}^n$
- se M e N são subespaços de X , então $M \cap N \neq \emptyset$ também é um subespaço de X .
- todo espaço é um subespaço de si mesmo.
- subespaço próprio é um subespaço que não é igual ao espaço inteiro.

5.6 Conjuntos convexos

- um conjunto X de elementos de um espaço vetorial é convexo se, dados $\mathbf{x}_1, \mathbf{x}_2 \in X$, todos os pontos na forma $\alpha \cdot \mathbf{x}_1 + (1-\alpha) \cdot \mathbf{x}_2$, com $\alpha \in [0,1]$, pertencem também a X .



- exemplos:
 - normalmente, (sub-)espaços vetoriais lineares são convexos.
 - o conjunto vazio \emptyset é convexo, por definição.
 - dados X e Y convexos, então $X \cap Y$ é convexo.

5.7 Combinação linear e combinação convexa

- seja $S = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ um conjunto de vetores de um espaço vetorial linear (X, \mathfrak{R}) . Combinações lineares de elementos de S são formadas através de

$$a_1 \mathbf{x}_1 + a_2 \mathbf{x}_2 + \dots + a_n \mathbf{x}_n,$$

onde $a_1, a_2, \dots, a_n \in \mathfrak{R}$.

- se os escalares $a_1, a_2, \dots, a_n \in \mathfrak{R}$ são tais que $a_i \geq 0$ ($i=1,2,\dots,n$) e $\sum_{i=1}^n a_i = 1$, então a combinação linear é chamada combinação convexa dos elementos $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n \in X$.
- combinação cônica: $a_i \geq 0$ ($i=1,2,\dots,n$) e $\sum_{i=1}^n a_i$ qualquer
- combinação afim: a_i ($i=1,2,\dots,n$) quaisquer e $\sum_{i=1}^n a_i = 1$

5.8 Dependência linear e dimensão de um espaço vetorial

- considere $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ um conjunto de vetores pertencentes a X . O conjunto $S \equiv [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]$, chamado subespaço gerado pelos vetores $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$, consiste de todos os vetores em X escritos como combinação linear de vetores em $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$. Neste caso, S é um subespaço vetorial de X .
- um vetor \mathbf{x} é linearmente dependente em relação a um conjunto de vetores $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ se $\mathbf{x} \in S \equiv [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]$.
- um vetor \mathbf{x} é linearmente independente em relação a um conjunto de vetores $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ se $\mathbf{x} \notin S \equiv [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]$.
- um conjunto de vetores $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ é linearmente independente se e somente se $\sum_{i=1}^n a_i \mathbf{x}_i = 0$ implica $a_i = 0, i=1, \dots, n$.
- um conjunto de vetores linearmente independentes $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ forma uma base para X se $X \equiv [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]$.
- neste caso, diz-se que X tem dimensão n . Se n é finito, então X é um espaço vetorial de dimensão finita.

5.9 Produto externo

- o produto externo entre dois vetores $\mathbf{x} \in \mathbb{R}^n$ e $\mathbf{y} \in \mathbb{R}^m$ é uma matriz de dimensão $n \times m$ de posto unitário. Sendo $\mathbf{x} = [x_1 \ x_2 \ \dots \ x_n]^T$ e $\mathbf{y} = [y_1 \ y_2 \ \dots \ y_m]^T$, o produto externo assume a forma:

$$\mathbf{xy}^T = \begin{bmatrix} x_1 y_1 & x_1 y_2 & \dots & x_1 y_m \\ x_2 y_1 & x_2 y_2 & & \vdots \\ \vdots & & \ddots & \\ x_n y_1 & \dots & & x_n y_m \end{bmatrix}$$

- em contraste com o caso do produto interno, os vetores \mathbf{x} e \mathbf{y} podem ter dimensões distintas.
- mesmo quando as dimensões são as mesmas, ou seja, $n = m$, a matriz quadrada resultante pode não ser simétrica.

5.10 Produto interno

- considere $\mathbf{x}, \mathbf{y} \in X$ e $\alpha \in \mathfrak{R}$. O produto interno $\langle \mathbf{x}, \mathbf{y} \rangle$ é um número real que satisfaz:

- $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$
- $\langle \mathbf{x} + \mathbf{y}, \mathbf{z} \rangle = \langle \mathbf{x}, \mathbf{z} \rangle + \langle \mathbf{y}, \mathbf{z} \rangle$
- $\langle \alpha \cdot \mathbf{x}, \mathbf{y} \rangle = \alpha \cdot \langle \mathbf{x}, \mathbf{y} \rangle$
- $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0 \quad \forall \mathbf{x} \in X$, e $\langle \mathbf{x}, \mathbf{x} \rangle = 0 \Leftrightarrow \mathbf{x} = \mathbf{0}$

- para $X \equiv \mathfrak{R}^n$ e $\mathbf{x}, \mathbf{y} \in X$, tem-se que $\mathbf{x} = [x_1 \ x_2 \ \cdots \ x_n]^T$ e $\mathbf{y} = [y_1 \ y_2 \ \cdots \ y_n]^T$, e o produto interno assume a forma:

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^n x_i y_i = \mathbf{x}^T \mathbf{y}$$

5.11 Norma, semi-norma e quase-norma

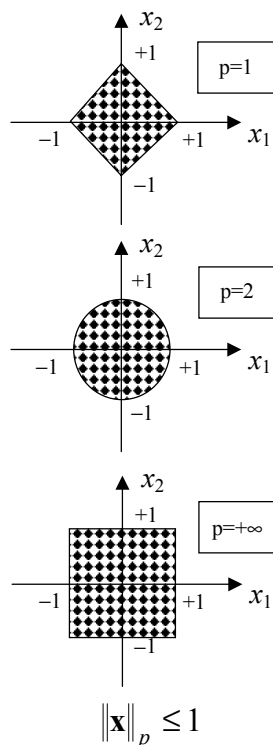
- as definições até agora apresentadas permitem relacionar propriedades algébricas. Introduzindo-se a noção de norma (medida de distância), podemos então tratar propriedades topológicas, como continuidade e convergência.
- norma é uma função $\|\cdot\|$ que associa a cada elemento $\mathbf{x} \in X$ um número real $\|\mathbf{x}\|$, obedecendo aos seguintes axiomas:
 1. $\|\mathbf{x}\| \geq 0, \forall \mathbf{x} \in X; \quad \|\mathbf{x}\| = 0 \Leftrightarrow \mathbf{x} = \mathbf{0};$
 2. $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|, \forall \mathbf{x}, \mathbf{y} \in X$ (desigualdade triangular);
 3. $\|\alpha \cdot \mathbf{x}\| = |\alpha| \cdot \|\mathbf{x}\|, \forall \mathbf{x} \in X, \forall \alpha \in \mathfrak{R}.$
- toda vez que se associa uma norma a um espaço vetorial (sendo que a este espaço já está associado um campo), diz-se que se tem um espaço vetorial normado.
- uma semi-norma satisfaz todas as propriedades de norma, com exceção do primeiro axioma. Para $X \equiv \mathfrak{R}^n$, o subespaço linear $X_0 \subset \mathfrak{R}^n$, cujos elementos obedecem $\|\mathbf{x}\| = 0$, é denominado espaço nulo da semi-norma.

- uma quase-norma satisfaz todas as propriedades de norma, com exceção do segundo axioma (desigualdade triangular), o qual assume a forma:

$$\|\mathbf{x} + \mathbf{y}\| \leq b \cdot (\|\mathbf{x}\| + \|\mathbf{y}\|), \quad \forall \mathbf{x}, \mathbf{y} \in X, \text{ com } b \in \mathfrak{R}.$$

- Exemplos de normas e relações entre normas: $X \equiv \mathfrak{R}^n$

- $\|\mathbf{x}\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}}, p \geq 1$ é um número real.
- $\|\mathbf{x}\|_2 \leq \|\mathbf{x}\|_1 \leq \sqrt{n} \|\mathbf{x}\|_2$
- $\|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_2 \leq \sqrt{n} \|\mathbf{x}\|_\infty$
- $\|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_1 \leq n \|\mathbf{x}\|_\infty$.
- $\langle \mathbf{x}, \mathbf{x} \rangle^{\frac{1}{2}}$ é a conhecida norma euclidiana, pois $\|\mathbf{x}\|_2 = \left(\sum_{i=1}^n x_i^2 \right)^{\frac{1}{2}} = \langle \mathbf{x}, \mathbf{x} \rangle^{\frac{1}{2}}$.
- relação entre produto interno e norma euclidiana (desigualdade de Cauchy-Schwartz-Buniakowsky): $|\langle \mathbf{x}, \mathbf{y} \rangle|^2 \leq \langle \mathbf{x}, \mathbf{x} \rangle \cdot \langle \mathbf{y}, \mathbf{y} \rangle \Rightarrow |\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\|_2 \cdot \|\mathbf{y}\|_2$



$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}}$$

$$p = 1 \Rightarrow \|\mathbf{x}\|_1 = \sum_{i=1}^n |x_i|$$

$$p = 2 \Rightarrow \|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^n |x_i|^2}$$

$$p = +\infty \Rightarrow \|\mathbf{x}\|_\infty = \max_i |x_i|$$

5.12 Ângulo entre dois vetores

- para qualquer inteiro $n \geq 2$, dados dois vetores $\mathbf{x}, \mathbf{y} \in X \subset \mathfrak{R}^n$, $\mathbf{x}, \mathbf{y} \neq 0$, o co-seno do ângulo θ formado pelos vetores \mathbf{x} e \mathbf{y} é dado na forma:

$$\cos(\theta) = \frac{\mathbf{x}^T \mathbf{y}}{\|\mathbf{x}\|_2 \cdot \|\mathbf{y}\|_2}.$$

5.13 Ortogonalidade e ortonormalidade entre dois vetores

- se $\cos(\theta) = 0$, isto implica que $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y} = 0$. Então diz-se que \mathbf{x} e \mathbf{y} são ortogonais entre si, condição representada na forma: $\mathbf{x} \perp \mathbf{y}$.
- além disso, se $\langle \mathbf{x}, \mathbf{x} \rangle = \langle \mathbf{y}, \mathbf{y} \rangle = 1$, então os vetores $\mathbf{x}, \mathbf{y} \in X \subset \mathfrak{R}^n$ são ortonormais entre si.

5.14 Espaços ortogonais

- um vetor \mathbf{x} é ortogonal a um espaço vetorial Y se for ortogonal a todos os vetores pertencentes a Y , condição representada na forma: $\mathbf{x} \perp Y$.
- os espaços vetoriais X e Y são ortogonais entre si se cada vetor pertence a X for ortogonal a todos os vetores pertencentes a Y , condição representada na forma: $X \perp Y$.

5.15 Projeção de um vetor em uma determinada direção

- dado um espaço vetorial linear X , seja $\mathbf{y} \in X$ um vetor que fornece uma determinada direção. A projeção de qualquer vetor $\mathbf{x} \in X$ na direção de \mathbf{y} é dada na forma:

$$\text{proj}_{\mathbf{y}}(\mathbf{x}) = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\langle \mathbf{y}, \mathbf{y} \rangle^{1/2}} \cdot \frac{\mathbf{y}}{\langle \mathbf{y}, \mathbf{y} \rangle^{1/2}} = \frac{\mathbf{x}^T \mathbf{y}}{\mathbf{y}^T \mathbf{y}} \cdot \mathbf{y}$$

5.16 Vetores ortonormais gerados a partir de vetores linearmente independentes

- apesar de todo conjunto de vetores ortonormais não-nulos ser linearmente independente, nem todo conjunto de vetores linearmente independentes é ortonormal, mas pode passar por um processo de ortonormalização, como segue.
- dado um conjunto de m vetores n -dimensionais ($m \leq n$) linearmente independentes $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m\}$, é sempre possível estabelecer uma combinação linear adequada destes vetores que produza m vetores n -dimensionais mutuamente ortogonais $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m\}$ que geram o mesmo espaço. Além disso, se os vetores \mathbf{u}_i ($i=1, \dots, m$) apresentarem norma unitária, eles são mutuamente ortonormais.
- um conjunto de vetores ortonormais $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m\}$ pode ser obtido a partir de um conjunto de vetores linearmente independentes $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m\}$ através do processo de ortogonalização de Gram-Schmidt, o qual é dividido em duas etapas:

Etapa (1) $\mathbf{y}_1 = \mathbf{x}_1$

$$\mathbf{y}_i = \mathbf{x}_i - \sum_{j=1}^{i-1} \frac{\langle \mathbf{x}_i, \mathbf{y}_j \rangle}{\langle \mathbf{y}_j, \mathbf{y}_j \rangle} \cdot \mathbf{y}_j, \quad i = 2, \dots, m$$

Etapa (2) $\mathbf{u}_i = \frac{\mathbf{y}_i}{\langle \mathbf{y}_i, \mathbf{y}_i \rangle^{1/2}}, \quad i = 1, \dots, m$

- com isso, resulta:

$$\mathbf{u}_1 = a_{11} \cdot \mathbf{x}_1$$

$$\mathbf{u}_2 = a_{21} \cdot \mathbf{x}_1 + a_{22} \cdot \mathbf{x}_2$$

$$\mathbf{u}_3 = a_{31} \cdot \mathbf{x}_1 + a_{32} \cdot \mathbf{x}_2 + a_{33} \cdot \mathbf{x}_3$$

$$\vdots \quad \vdots \quad \vdots \quad \ddots$$

onde $a_{ii} > 0$ ($i=1, \dots, m$). Certamente existem outros processos de ortonormalização mais gerais, que não impõem qualquer tipo de restrição aos coeficientes a_{ij} ($i \geq j$; $i, j=1, \dots, m$).

6 Transformações e funcionais

- sejam X e Y espaços vetoriais lineares, e seja D um subconjunto de X . A regra que associa cada elemento $\mathbf{x} \in D \subseteq X$ a um elemento $\mathbf{y} \in Y$ é chamada de transformação de X para Y com domínio D . Notação: $T: D \subseteq X \rightarrow Y$.
- $\mathbf{y} = T(\mathbf{x})$ é a imagem de \mathbf{x} sob a transformação $T(\cdot)$.
- a coleção de vetores $\mathbf{y} \in Y$ para os quais existe um $\mathbf{x} \in D$ tal que $\mathbf{y} = T(\mathbf{x})$ é chamada de range de T .

6.1 Transformações lineares

- uma transformação $T: X \rightarrow Y$ é linear se, $\forall \mathbf{x}, \mathbf{y} \in D \subseteq X$ e $\forall \alpha, \beta \in \mathbb{R}$, é válida a seguinte equação:

$$T(\alpha \cdot \mathbf{x} + \beta \cdot \mathbf{y}) = \alpha \cdot T(\mathbf{x}) + \beta \cdot T(\mathbf{y}).$$

6.2 Operadores lineares

- são as transformações que podem ser descritas por formas matriciais, de modo que $D \equiv X$. Um operador linear, que mapeia vetores do \mathbb{R}^n no \mathbb{R}^m , pode ser descrito por uma matriz $A \in \mathbb{R}^m \times \mathbb{R}^n$, ou $A \in \mathbb{R}^{m \times n}$, tal que

$$\mathbf{y} = A\mathbf{x}, \quad \mathbf{x} \in \mathbb{R}^n \text{ e } \mathbf{y} \in \mathbb{R}^m.$$

- a norma de um operador linear $A \in \mathbb{R}^{m \times n}$ é dada na forma:

$$\|A\| = \max_{\mathbf{x} \neq \mathbf{0}} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|}, \quad \text{com } \mathbf{x} \in \mathbb{R}^n.$$

- o range de um operador linear $A \in \mathbb{R}^{m \times n}$ é dado na forma:

$$\tau(A) = \{\mathbf{y} \in \mathbb{R}^m : \mathbf{y} = A\mathbf{x}, \text{ para algum } \mathbf{x} \in \mathbb{R}^n\}$$

correspondendo, portanto, ao espaço gerado pelas colunas de A .

- o espaço nulo de um operador linear $A \in \mathbb{R}^{m \times n}$ é dado na forma:

$$\eta(A) = \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = \mathbf{0}\}$$

- $\tau(A)$ e $\eta(A)$ são subespaços de \mathbb{R}^m e \mathbb{R}^n , respectivamente.

- $\dim(\tau(A)) + \dim(\eta(A)) = n$
- $\eta(A) \perp \tau(A^T)$ no \mathfrak{R}^n e $\eta(A^T) \perp \tau(A)$ no \mathfrak{R}^m

6.3 Posto de uma matriz

- o posto de uma matriz $A \in \mathfrak{R}^{m \times n}$ é dado pelo número de colunas (ou linhas) LI, de modo que $\text{posto}(A) \leq \min(m, n)$.
- se $\text{posto}(A) = \min(m, n)$, então diz-se que a matriz tem posto completo.
- uma matriz quadrada de posto completo é inversível (matrizes inversas serão discutidas mais adiante).
- $\text{posto}(A) = \dim(\tau(A))$
- $\text{posto}(A) = \text{posto}(A^T) = \text{posto}(A^T A) = \text{posto}(A A^T)$
- a matriz resultante do produto de duas matrizes quaisquer nunca vai ter um posto maior que o menor posto das matrizes que participam do produto.

6.4 Matrizes idempotentes

- uma matriz quadrada $A \in \mathfrak{R}^{n \times n}$ é dita ser idempotente se $A^r = \underbrace{A \cdot A \cdot \dots \cdot A}_{r \text{ vezes}} = A$, para qualquer potência inteira $r \geq 1$.
- se A é idempotente, então $I - A$ também será.

6.5 Definições adicionais para matrizes

Cofator

- dada uma matriz A de dimensão $n \times n$, o cofator do elemento a_{ij} ($i, j=1, 2, \dots, n$) é dado na forma:

$$c_{ij} = (-1)^{i+j} m_{ij},$$

onde m_{ij} é o determinante da matriz formada eliminando-se a i -ésima linha e a j -ésima coluna da matriz A .

Determinante

- dada uma matriz A de dimensão $n \times n$, o determinante de A é dado na forma:

$$|A| = \det(A) = \begin{cases} \sum_{i=1}^n a_{ij} c_{ij}, & \text{para qualquer } j \\ \text{ou} \\ \sum_{j=1}^n a_{ij} c_{ij}, & \text{para qualquer } i \end{cases}$$

onde c_{ij} é o cofator do elemento a_{ij} .

- com isso, se B é uma matriz obtida de A pela troca de duas de suas colunas, então $\det(B) = -\det(A)$.
- seja $A = [\mathbf{a}_1 \ \cdots \ \mathbf{a}_j \ \cdots \ \mathbf{a}_n]$ ($1 \leq j \leq n$). O determinante de A possui as seguintes propriedades:
 - Invariância: $\det([\mathbf{a}_1 \ \cdots \ \mathbf{a}_j \ \cdots \ \mathbf{a}_n]) = \det([\mathbf{a}_1 \ \cdots \ \mathbf{a}_j + \mathbf{a}_k \ \cdots \ \mathbf{a}_n])$,
 $j \neq k, 1 \leq j, k \leq n$;
 - Homogeneidade: $\det([\mathbf{a}_1 \ \cdots \ b\mathbf{a}_j \ \cdots \ \mathbf{a}_n]) = b \cdot \det([\mathbf{a}_1 \ \cdots \ \mathbf{a}_j \ \cdots \ \mathbf{a}_n])$

Traço

- dada uma matriz A de dimensão $n \times n$, o traço de A , representado por $\text{tr}(A)$, é a soma dos elementos da diagonal de A , ou seja:

$$\text{tr}(A) = \sum_{i=1}^n a_{ii}$$

Adjunta

- dada uma matriz A de dimensão $n \times n$, a adjunta de A , representada por $\text{adj}(A)$, é dada na forma:

$$\text{adj}(A) = \{a'_{ij}\}$$

onde $a'_{ij} = c_{ji}$, o cofator do elemento a_{ji} .

- são válidas as seguintes igualdades: $\begin{cases} A \cdot \text{adj}(A) = \det(A) \cdot I \\ \text{adj}(A) \cdot A = \det(A) \cdot I \end{cases} \Rightarrow A^{-1} = \frac{\text{adj}(A)}{\det(A)}$

6.6 Matrizes singulares e não-singulares

- uma matriz A de dimensão $n \times n$ é dita ser singular quando $\dim(\eta(A)) \neq 0$, ou seja, quando $\det(A) = 0$.
- $A \in \mathfrak{R}^{n \times n}$ é não-singular se e somente se $\dim(\eta(A)) = 0$.
- como $A^{-1} = \text{adj}(A)/\det(A)$, A admite inversa se e somente se $\det(A) \neq 0$, ou seja, quando A é não-singular.

6.7 Autovalores e autovetores

- seja uma matriz A de dimensão $n \times n$. Diz-se que um escalar $\lambda \in \mathbb{C}$ (conjunto dos números complexos) é um autovalor de A se existe um vetor não-nulo $\mathbf{x} \in \mathbb{C}^n$, chamado de autovetor associado a λ , tal que

$$A\mathbf{x} = \lambda\mathbf{x}.$$

- $A\mathbf{x} = \lambda\mathbf{x}$ pode ser reescrito como $(\lambda I - A)\mathbf{x} = \mathbf{0}$;

- $\exists \mathbf{x} \in \mathbb{C}^n, \mathbf{x} \neq \mathbf{0}$ tal que $(\lambda I - A)\mathbf{x} = \mathbf{0}$ se e somente se $\det(\lambda I - A) = 0$;
- $\Delta(\lambda) \triangleq \det(\lambda I - A)$ é o polinômio característico de A ;
- Como o grau de $\Delta(\lambda)$ é n , a matriz A possui n autovalores.

6.8 Formas Quadráticas

Definição: Para qualquer $\mathbf{y} \in \mathfrak{R}^n$ e $A = A^T \in \mathfrak{R}^{n \times n}$, $Q_A(\mathbf{y}) = \mathbf{y}^T A \mathbf{y}$ é uma forma quadrática associada a A .

Propriedades:

- $\nabla Q_A(\mathbf{y}) = 2A\mathbf{y}$
- $\nabla^2 Q_A(\mathbf{y}) = 2A$
- A e Q_A são chamadas de:
 - semi-definida positiva se $Q_A(\mathbf{y}) = \mathbf{y}^T A \mathbf{y} \geq 0, \quad \forall \mathbf{y} \in \mathfrak{R}^n$.
 - definida positiva se $Q_A(\mathbf{y}) = \mathbf{y}^T A \mathbf{y} > 0, \quad \forall \mathbf{y} \in \mathfrak{R}^n, \mathbf{y} \neq \mathbf{0}$.

- semi-definida negativa se $Q_A(\mathbf{y}) = \mathbf{y}^T A \mathbf{y} \leq 0, \quad \forall \mathbf{y} \in \mathfrak{R}^n$.
- definida negativa se $Q_A(\mathbf{y}) = \mathbf{y}^T A \mathbf{y} < 0, \quad \forall \mathbf{y} \in \mathfrak{R}^n, \mathbf{y} \neq \mathbf{0}$.

6.8.1 Cálculo diferencial aplicado a formas quadráticas

Dados $\mathbf{x} \in \mathfrak{R}^n, \mathbf{y} \in \mathfrak{R}^n$ e $A \in \mathfrak{R}^{n \times n}$:

- $\frac{\partial}{\partial \mathbf{y}} (\mathbf{y}^T A \mathbf{x}) = A \mathbf{x}$
- $\mathbf{y}^T A \mathbf{x} = \mathbf{x}^T A^T \mathbf{y} \Rightarrow \frac{\partial}{\partial \mathbf{x}} (\mathbf{y}^T A \mathbf{x}) = \frac{\partial}{\partial \mathbf{x}} (\mathbf{x}^T A^T \mathbf{y}) = A^T \mathbf{y}$
- $\frac{\partial}{\partial \mathbf{x}} (\mathbf{x}^T A \mathbf{x}) = A^T \mathbf{x} + A \mathbf{x}$
- para $A^T = A, \frac{\partial}{\partial \mathbf{x}} (\mathbf{x}^T A \mathbf{x}) = 2 A \mathbf{x}$

6.8.2 Normas ponderadas e a medida de distância de Mahalanobis

- já foi apresentada a relação existente entre norma e produto interno. Com a introdução de formas quadráticas, é possível recorrer ao conceito de norma ponderada.
- sejam $\mathbf{x}, \mathbf{y} \in \mathfrak{R}^n$, então é possível expressar a distância ponderada entre \mathbf{x} e \mathbf{y} por:

$$\rho_{\Psi}(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_{\Psi} = \sqrt{(\mathbf{x} - \mathbf{y})^T \Psi (\mathbf{x} - \mathbf{y})} = \sqrt{\|\mathbf{x}\|_{\Psi}^2 + \|\mathbf{y}\|_{\Psi}^2 - 2\mathbf{x}^T \Psi \mathbf{y}}$$

onde $\Psi \in \mathfrak{R}^{n \times n}$ é uma matriz simétrica e semidefinida positiva, geralmente sendo uma matriz diagonal que pondera diferentemente cada coordenada.

- para $\Psi = I$, resulta a norma euclidiana.
- quando os elementos de \mathbf{x} e \mathbf{y} são variáveis aleatórias geradas a partir de uma distribuição normal, e sabendo haver uma dependência estatística entre os elementos de \mathbf{x} e \mathbf{y} , então tomando a matriz Ψ como sendo a inversa da matriz de covariância entre \mathbf{x} e \mathbf{y} resulta a medida de distância de Mahalanobis.

6.9 Matrizes simétricas: positividade e autovalores

- uma matriz quadrada $A \in \mathfrak{R}^{n \times n}$ é dita ser simétrica se $A^T = A$ ($\mathfrak{R}^{n \times n}$ ou $\mathfrak{V}^{n \times n}$).
- se a matriz quadrada é tal que $A \in \mathfrak{V}^{n \times n}$, então ela será hermitiana se $(A^*)^T = A$, ou seja, se A for idêntica ao transposto de seu complexo conjugado.
- matrizes simétricas só admitem autovalores reais.
- autovetores associados a autovalores distintos de uma matriz simétrica com elementos reais ($A \in \mathfrak{R}^{n \times n}$) são ortogonais.
- mesmo que os autovalores não sejam distintos, é possível obter autovetores ortogonais para matrizes simétricas $A \in \mathfrak{R}^n \times \mathfrak{R}^n$. Sendo assim, dados os autovetores ortogonais $\mathbf{v}_1, \dots, \mathbf{v}_n$, é possível construir a matriz T abaixo:

$$T = \begin{bmatrix} \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|} & \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|} & \dots & \frac{\mathbf{v}_n}{\|\mathbf{v}_n\|} \end{bmatrix}, \text{ onde } \|\cdot\| \equiv \|\cdot\|_2.$$

- a matriz T é ortogonal, pois $T^{-1} = T^T$, como pode ser verificado a seguir:

$$T^T T = \begin{bmatrix} \frac{\mathbf{v}_1^T}{\|\mathbf{v}_1\|} \\ \vdots \\ \frac{\mathbf{v}_n^T}{\|\mathbf{v}_n\|} \end{bmatrix} \begin{bmatrix} \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|} & \dots & \frac{\mathbf{v}_n}{\|\mathbf{v}_n\|} \end{bmatrix} = \begin{bmatrix} 1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1 \end{bmatrix} = I_{n \times n}$$

- com a matriz T , é possível obter uma matriz diagonal a partir da matriz A , tendo os autovalores de A na diagonal, como a seguir:

1. da definição de autovalores tem-se: $A\mathbf{v}_i = \lambda_i \mathbf{v}_i, i=1, \dots, n$.

$$2. \begin{cases} AT = A \begin{bmatrix} \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|} & \dots & \frac{\mathbf{v}_n}{\|\mathbf{v}_n\|} \end{bmatrix} = \begin{bmatrix} A\mathbf{v}_1 & \dots & A\mathbf{v}_n \end{bmatrix} = \begin{bmatrix} \lambda_1 \mathbf{v}_1 & \dots & \lambda_n \mathbf{v}_n \end{bmatrix} = \\ = \begin{bmatrix} \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|} & \dots & \frac{\mathbf{v}_n}{\|\mathbf{v}_n\|} \end{bmatrix} \begin{bmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n \end{bmatrix} = T\Lambda \end{cases}$$

$$3. AT = T\Lambda \Rightarrow \begin{cases} \Lambda = T^{-1}AT = T^T AT \\ A = T\Lambda T^{-1} = T\Lambda T^T \end{cases}$$

- este resultado é muito útil no caso da matriz A estar presente em formas quadráticas:

$$Q_A(\mathbf{x}) = \mathbf{x}^T A \mathbf{x} = \mathbf{x}^T T \Lambda T^T \mathbf{x} = (T^T \mathbf{x})^T \Lambda (T^T \mathbf{x})$$

- fazendo $\mathbf{y} = T^T \mathbf{x} \Rightarrow \mathbf{x} = T \mathbf{y}$, resulta:

$$Q_A(\mathbf{x})|_{\mathbf{x}=T\mathbf{y}} = \mathbf{y}^T \Lambda \mathbf{y} = \lambda_1 \mathbf{y}_1^2 + \lambda_2 \mathbf{y}_2^2 + \dots + \lambda_n \mathbf{y}_n^2$$

- como a matriz T tem posto completo, então $\mathbf{y} \in \mathbb{R}^n$ pode ser qualquer. Logo:
 - $A > 0 \Rightarrow \lambda_i > 0$ para $i=1, \dots, n$
 - $A \geq 0 \Rightarrow \lambda_i \geq 0$ para $i=1, \dots, n$
 - $A < 0 \Rightarrow \lambda_i < 0$ para $i=1, \dots, n$
 - $A \leq 0 \Rightarrow \lambda_i \leq 0$ para $i=1, \dots, n$
- se $A \in \mathbb{R}^{n \times n}$ é uma matriz simétrica definida positiva, então são condições equivalentes:
 - A é definida positiva;

- A^{-1} é definida positiva;
- todos os autovalores de A são reais positivos;
- $\mathbf{y}^T A \mathbf{y} > 0, \quad \forall \mathbf{y} \in \mathbb{R}^n, \mathbf{y} \neq \mathbf{0}$;
- é possível decompor A na forma: $A = B^T B$, com B não-singular;
- $\det(A_k) > 0, k=1, \dots, n$, onde A_k é a k -ésima sub-matriz principal líder.
- se ocorrer $\lambda_i > 0$ e $\lambda_j < 0$ para $i \neq j$ e $i, j \in \{1, \dots, n\}$, então a matriz é dita ser indefinida.
- se $A \in \mathbb{R}^{n \times n}$ é uma matriz simétrica com autovalores $\lambda_1, \lambda_2, \dots, \lambda_n$, então a matriz $A + \alpha I_n$ terá como autovalores $\lambda_1 + \alpha, \lambda_2 + \alpha, \dots, \lambda_n + \alpha$, e os autovetores de A e de $A + \alpha I_n$ serão os mesmos.
- sendo assim, é possível “corrigir” os autovalores de uma matriz simétrica indefinida ou singular, sem alterar seus autovetores, deixando a matriz definida positiva ou definida negativa, na forma:

$$\begin{cases} \alpha = \varepsilon - \min_{i \in \{1, \dots, n\}} \lambda_i, & \text{para deixar a matriz definida positiva} \\ \alpha = -\varepsilon - \max_{i \in \{1, \dots, n\}} \lambda_i, & \text{para deixar a matriz definida negativa} \end{cases}$$

- seja uma matriz A de dimensão $n \times m$ ($A \in \mathfrak{R}^{n \times m}$) e uma matriz B de dimensão $m \times n$ ($B \in \mathfrak{R}^{m \times n}$). Então, os autovalores não-nulos de AB e BA são os mesmos e têm as mesmas multiplicidades. Além disso, se \mathbf{x} é um autovetor de AB para algum autovalor $\lambda \neq 0$, então $\mathbf{y} = B\mathbf{x}$ é um autovetor de BA . Isto implica que AA^T e $A^T A$ têm os mesmos autovalores e todos são positivos.

6.10A inversa de uma matriz

- a inversa de uma matriz $A \in \mathfrak{R}^{n \times n}$ é uma matriz $M \in \mathfrak{R}^{n \times n}$ tal que

$$AM = MA = I$$

- a notação adotada para a matriz M é: $M = A^{-1}$.
- para que uma matriz seja inversível, ela tem que ser quadrada e não-singular.

- vale a seguinte propriedade: $(A^{-1})^T = (A^T)^{-1}$.

6.11O lema de inversão de matrizes

- assumamos que A e C são matrizes quadradas arbitrárias para as quais existe a inversa, e B é uma terceira matriz tal que BCB^T tem a mesma dimensão de A . Então o chamado lema de inversão de matrizes é dado na forma:

$$(A + BCB^T)^{-1} = A^{-1} - A^{-1}B(B^T A^{-1}B + C^{-1})^{-1}B^T A^{-1}$$

- a matriz C geralmente tem dimensões menores que a matriz A .

6.12A pseudo-inversa de uma matriz

- a pseudo-inversa de uma matriz $A \in \mathfrak{R}^{m \times n}$ é uma matriz $M \in \mathfrak{R}^{n \times m}$ tal que valem as seguintes propriedades:

- $AMA = A$
- $MAM = M$
- AM e MA são matrizes simétricas

- a notação adotada para a matriz M é: $M = A^+$.
- pode-se demonstrar que existe uma única pseudo-inversa para cada matriz.
- valem as seguintes propriedades:

$$\begin{aligned}
 &\triangleright \mathbf{0}^+ = \mathbf{0}^T && \triangleright (\alpha A)^+ = \alpha^{-1} A^+, \text{ se } \alpha \neq 0 \\
 &\triangleright (A^+)^+ = A && \triangleright A^+ = (A^T A)^+ A^T = A^T (A A^T)^+ \\
 &\triangleright (A^+)^T = (A^T)^+ && \triangleright A^+ = A^{-1}, \text{ se } A \text{ é quadrada e não-singular} \\
 &\triangleright (A A^+)^T = A A^+ \text{ e } (A^+ A)^T = A^+ A && \triangleright A^T A A^+ = A^T \text{ e } A A^T (A^+)^T = A
 \end{aligned}$$

6.12.1 Exemplos de pseudo-inversas

- caso escalar ($m = n = 1$): $A = a \Rightarrow \begin{cases} A^+ = a^{-1}, \text{ se } a \neq 0 \\ A^+ = 0, \text{ se } a = 0 \end{cases}$

- caso vetorial ($m > 1$ e $n = 1$): $A = \mathbf{a} \Rightarrow \begin{cases} A^+ = \frac{\mathbf{a}^T}{\mathbf{a}^T \mathbf{a}}, \text{ se } \mathbf{a} \neq \mathbf{0} \\ A^+ = \mathbf{0}^T, \text{ se } \mathbf{a} = \mathbf{0} \end{cases}$

6.12.2 Uso de pseudo-inversão para a solução de sistemas lineares

- considere o seguinte sistema linear de equações na forma matricial

$$A\mathbf{x} = \mathbf{b} \quad (1)$$

onde $A \in \Re^{m \times n}$, $\mathbf{x} \in \Re^n$ e $\mathbf{b} \in \Re^m$.

- supondo que $\text{posto}(A) = m \leq n$, então a seguinte expressão representa uma solução para a equação matricial (1): $\mathbf{x} = A^T (A A^T)^{-1} \mathbf{b} + (I - A^T (A A^T)^{-1} A) \mathbf{y}$, onde $\mathbf{y} \in \Re^n$ é um vetor arbitrário. Repare que, sob a condição $\text{posto}(A) = m \leq n$, $A^T (A A^T)^{-1}$ é a pseudo-inversa de A , de modo que é possível expressar a solução na forma:

$$\mathbf{x} = A^+ \mathbf{b} + (I - A^+ A) \mathbf{y}. \quad (2)$$

- quando múltiplas soluções são possíveis, como no caso acima, pode-se adotar aquela solução que otimiza algum critério. Por exemplo, a solução com norma euclidiana mínima é aquela que toma $\mathbf{y} = \mathbf{0}$.
- supondo que $\text{posto}(A) = n \leq m$, então a seguinte expressão representa uma solução para a equação matricial (1): $\mathbf{x} = (A^T A)^{-1} A^T \mathbf{b}$. Repare que, sob a condição $\text{posto}(A) = n \leq m$, $(A^T A)^{-1} A^T$ é a pseudo-inversa de A , de modo que é possível expressar a solução na forma:

$$\mathbf{x} = A^+ \mathbf{b} \quad (3)$$

- no entanto, para $\text{posto}(A) = n < m$, a equação matricial (1) nem sempre tem solução exata, e nestes casos o que se obtém é o mínimo de $\|A\mathbf{x} - \mathbf{b}\|_2$.

6.13 Operadores de projeção ortogonal

- seja $X \subset \mathfrak{R}^n$ um subespaço vetorial e seja $\mathbf{x} \in \mathfrak{R}^n$. Então é possível expressar \mathbf{x} na forma:

$$\mathbf{x} = \hat{\mathbf{x}} + \tilde{\mathbf{x}}$$

onde $\hat{\mathbf{x}} \in X$ e $\tilde{\mathbf{x}} \perp X$.

- neste caso, diz-se que $\hat{\mathbf{x}}$ é a projeção ortogonal de \mathbf{x} em X .
- de todas as decomposições na forma $\mathbf{x} = \mathbf{x}' + \mathbf{x}''$, onde $\mathbf{x}' \in X$, aquela em que $\mathbf{x}'' \perp X$ é tal que $\|\mathbf{x}''\|_2$ é mínima.
- existe sempre uma matriz simétrica $P \in \mathfrak{R}^{n \times n}$, chamado operador de projeção ortogonal em X , tal que

$$\hat{\mathbf{x}} = P\mathbf{x} \quad \text{e} \quad \tilde{\mathbf{x}} = (I - P)\mathbf{x}$$

- $(I - P)$ é o operador de projeção ortogonal em X^\perp (complemento ortogonal de X).

6.13.1 Um exemplo de operador simétrico e idempotente

- todo operador de projeção ortogonal é idempotente, sendo que a seguir iremos apresentar um exemplo de operador de projeção ortogonal que também é simétrico.

- as projeções ortogonais podem ser expressas através de transformações lineares, de modo que sempre é possível obter P .
- assumamos que $X \equiv [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k]$ é um subespaço do \mathfrak{R}^n , gerado pelos vetores $\mathbf{x}_i \in \mathfrak{R}^n$, $i = 1, \dots, k < n$. Seja A uma matriz que tem como colunas os vetores $\mathbf{x}_i \in \mathfrak{R}^n$, $i = 1, \dots, k < n$. O objetivo é obter o operador P de projeção ortogonal ao subespaço X , de modo que sua aplicação a um vetor qualquer $\mathbf{x} \in \mathfrak{R}^n$ produza $\hat{\mathbf{x}} = P\mathbf{x}$ e $\tilde{\mathbf{x}} = (I - P)\mathbf{x}$, onde $\hat{\mathbf{x}} \in X$ e $\tilde{\mathbf{x}} \perp X$.
- como, por definição, $\tilde{\mathbf{x}} \perp X$, então tem-se que $A^T \tilde{\mathbf{x}} = \mathbf{0}$.
- a solução desta equação matricial é dada pela expressão (2) acima, produzindo $\tilde{\mathbf{x}} = (I - (A^T)^+ A^T) \mathbf{y} = (I - AA^+) \mathbf{y}$, para um $\mathbf{y} \in \mathfrak{R}^n$ arbitrário e sabendo que a matriz $(A^T)^+ A^T$ é simétrica.
- $\mathbf{y} = \mathbf{x}$ é uma escolha possível, e como $\tilde{\mathbf{x}}$ é único, então a expressão

$$\tilde{\mathbf{x}} = (I - AA^+) \mathbf{x}$$

leva a que $I - P = I - AA^+$ e $P = AA^+$.

- conclusão:
 - AA^+ é o operador de projeção ortogonal ao subespaço gerado pelas colunas de A , ou seja, ao subespaço X .
 - as matrizes $I - AA^+$ e $I - A^+A$ são operadores de projeção ortogonal aos subespaços que são os complementos ortogonais dos subespaços gerados pelas colunas e linhas de A , respectivamente.

6.14 Decomposição em valores singulares

- a decomposição em valores singulares é uma poderosa ferramenta matemática para a solução de problemas de quadrados mínimos, pois fornece informações quantitativas importantes acerca da estrutura de um sistema de equações lineares do tipo: $A\mathbf{x} = \mathbf{b}$. Ela vale tanto para matrizes quadradas quanto retangulares.

- além disso, a matriz A pode ter elementos reais ou complexos. Neste estudo, iremos considerar apenas matrizes com elementos reais.
- em termos geométricos, os valores singulares de uma matriz A correspondem aos comprimentos dos semi-eixos do hiperelipsóide $E = \{A\mathbf{x} : \|\mathbf{x}\|_2 = 1\}$
- seja uma matriz A de dimensão $n \times m$ ($A \in \mathbb{R}^{n \times m}$) e de posto r , com $r \leq \min(n, m)$. Então, A pode ser expressa na forma:

$$A = U \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} V^T$$

onde $U \in \mathbb{R}^{n \times n}$ e $V \in \mathbb{R}^{m \times m}$ são matrizes unitárias, tais que $U^T U = U U^T = I_n$ e $V^T V = V V^T = I_m$, e $\Sigma \in \mathbb{R}^{r \times r}$ é uma matriz diagonal, com elementos $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$, denominados valores singulares da matriz A .

- repare que esta decomposição é sempre possível, independente de se ter $n = m$, $n < m$ ou $n > m$.

- note também que o número de valores singulares positivos coincide com o posto da matriz A , o que implica que a decomposição em valores singulares representa um método prático para se obter o posto da matriz A .
- é possível verificar também que $U^T A V = \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix}$ e que as colunas de U são autovetores de $A A^T$, enquanto que as colunas de V são autovetores de $A^T A$.
- como $U U^T = I_n$, então, $A V = U \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix}$ o que implica que:

$$\begin{cases} A \mathbf{v}_i = \sigma_i \mathbf{u}_i, & i = 1, \dots, r \\ A \mathbf{v}_i = 0, & i = r + 1, \dots, m \end{cases}$$

onde \mathbf{v}_i e \mathbf{u}_i são, respectivamente, as i -ésimas colunas de V e U .

- sendo assim, é possível expressar a matriz A na forma:

$$A = \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T$$

- se a matriz A for simétrica, então seus valores singulares correspondem aos valores absolutos de seus autovalores não-nulos (ver seção 6.9).
- para os propósitos deste curso, a principal motivação para o estudo de decomposição em valores singulares é a possibilidade de propor um método prático de cálculo da pseudo-inversa de uma matriz, independente de se ter $n < m$ ou $n > m$.
- seja uma matriz A de dimensão $n \times m$ ($A \in \Re^{n \times m}$) e de posto r , com $r \leq \min(n, m)$, que tenha uma decomposição em valores singulares tal que $U^T A V = \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix}$.

Então, a pseudo-inversa da matriz A pode ser obtida na forma:

$$A^+ = V \begin{bmatrix} \Sigma^{-1} & 0 \\ 0 & 0 \end{bmatrix} U^T$$

onde $\Sigma^{-1} = \text{diag}(\sigma_1^{-1}, \sigma_2^{-1}, \dots, \sigma_r^{-1})$.

- quando a matriz A tem posto completo, ou seja, quando $r = \min(n, m)$, então é possível mostrar que:

$$A^+ = V \begin{bmatrix} \Sigma^{-1} & 0 \\ 0 & 0 \end{bmatrix} U^T = (A^T A)^{-1} A^T, \text{ quando } n > m$$

$$A^+ = V \begin{bmatrix} \Sigma^{-1} & 0 \\ 0 & 0 \end{bmatrix} U^T = A^T (A A^T)^{-1}, \text{ quando } n < m$$

6.15 Transformações contínuas

- uma transformação $T: X \rightarrow Y$ é contínua em $\mathbf{x}_0 \in X$ se para todo $\varepsilon > 0$ existe um $\delta > 0$ tal que $\|\mathbf{x} - \mathbf{x}_0\| < \delta$ implica que $\|T(\mathbf{x}) - T(\mathbf{x}_0)\| < \varepsilon$. Obviamente assume-se que X e Y são espaços vetoriais sobre os quais está definida uma mesma norma.
- diz-se que T é contínua se ela for contínua para todo $\mathbf{x} \in X$.

6.16 Funcional

- uma transformação $T: X \rightarrow \Re$ é chamada de funcional sobre X .

6.17 Funcional convexo

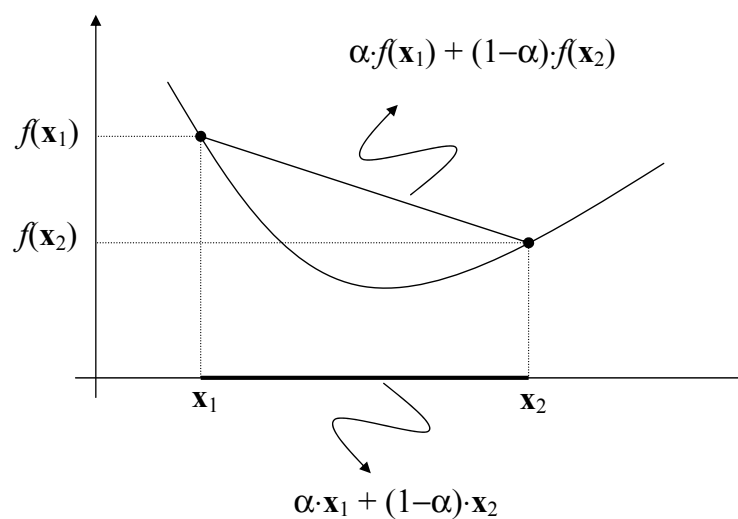
- um funcional $f: X \rightarrow \Re$ é convexo sobre um subconjunto convexo X de um espaço vetorial linear se e somente se

$$f(\alpha \cdot \mathbf{x}_1 + (1 - \alpha) \cdot \mathbf{x}_2) \leq \alpha \cdot f(\mathbf{x}_1) + (1 - \alpha) \cdot f(\mathbf{x}_2)$$

para todo $\mathbf{x}_1, \mathbf{x}_2 \in X$ e $\alpha \in [0, 1]$.

- Extensão 1: O funcional f é estritamente convexo se a desigualdade acima for estrita, com $\alpha \in (0, 1)$.
- Extensão 2: Um funcional f é (estritamente) côncavo se $-f$ é (estritamente) convexo, de modo que $\max f \equiv \min (-f)$.

Interpretação Geométrica



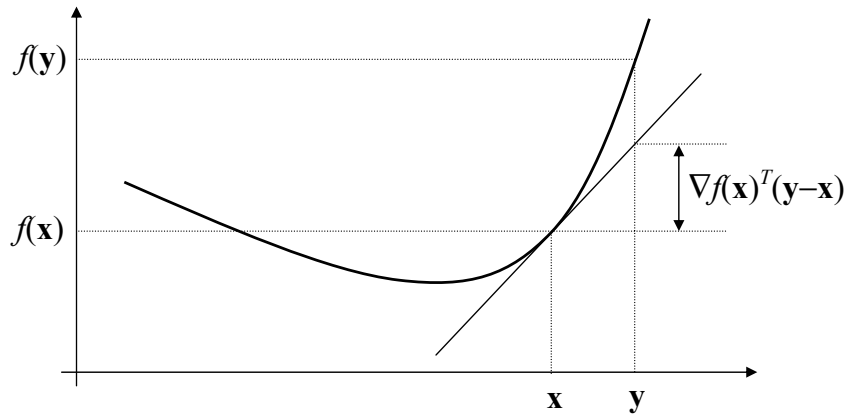
6.18 Funcional convexo diferenciável

- um funcional diferenciável $f: X \rightarrow \Re$ é convexo sobre um subconjunto convexo X de um espaço vetorial linear se e somente se

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T (\mathbf{y} - \mathbf{x})$$

para todo $\mathbf{x}, \mathbf{y} \in X$.

Interpretação Geométrica



7 Mínimos Locais

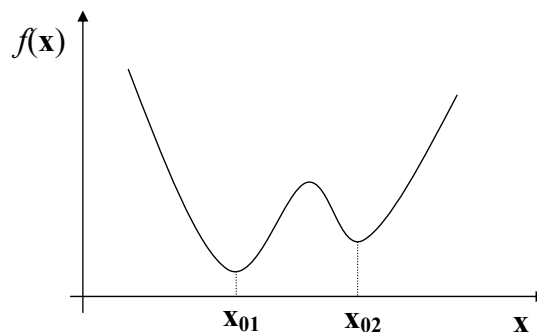
- seja f um funcional definido sobre $\Omega \subset X$. Um ponto $\mathbf{x}_0 \in \Omega$ é chamado MÍNIMO LOCAL de f sobre Ω se existe uma esfera

$$N(\mathbf{x}_0, \varepsilon) = \{\mathbf{x} : \|\mathbf{x} - \mathbf{x}_0\| < \varepsilon\}$$

tal que $f(\mathbf{x}_0) \leq f(\mathbf{x}), \forall \mathbf{x} \in \Omega \cap N(\mathbf{x}_0, \varepsilon)$.

- \mathbf{x}_0 é um MÍNIMO GLOBAL se $f(\mathbf{x}_0) \leq f(\mathbf{x}), \forall \mathbf{x} \in \Omega$.

Interpretação Geométrica



8 Expansão em Série de Taylor

- $f: \mathbb{R}^n \rightarrow \mathbb{R} \quad \mathbf{x} \in \mathbb{R}^n$
- expansão em série de Taylor em torno do ponto $\mathbf{x}^* \in \mathbb{R}^n$:

$$f(\mathbf{x}) = f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)^T (\mathbf{x} - \mathbf{x}^*) + \frac{1}{2} (\mathbf{x} - \mathbf{x}^*)^T \nabla^2 f(\mathbf{x}^*) (\mathbf{x} - \mathbf{x}^*) + O(3)$$

$$\nabla f(\mathbf{x}^*) = \begin{bmatrix} \frac{\partial f(\mathbf{x}^*)}{\partial x_1} \\ \frac{\partial f(\mathbf{x}^*)}{\partial x_2} \\ \vdots \\ \frac{\partial f(\mathbf{x}^*)}{\partial x_n} \end{bmatrix} \quad \nabla^2 f(\mathbf{x}^*) = \begin{bmatrix} \frac{\partial^2 f(\mathbf{x}^*)}{\partial x_1^2} & \frac{\partial^2 f(\mathbf{x}^*)}{\partial x_1 \partial x_2} & \dots & \frac{\partial^2 f(\mathbf{x}^*)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(\mathbf{x}^*)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(\mathbf{x}^*)}{\partial x_2^2} & & \vdots \\ & & \ddots & \\ \frac{\partial^2 f(\mathbf{x}^*)}{\partial x_n \partial x_1} & \dots & & \frac{\partial^2 f(\mathbf{x}^*)}{\partial x_n^2} \end{bmatrix}$$

vetor gradiente

matriz hessiana

9 Condição Necessária de Otimalidade

- Teorema: Assuma que $f: \mathbb{R}^n \rightarrow \mathbb{R}$ e $f \in C^2[\mathbb{R}^n]$ (conjunto das funções com derivadas contínuas até 2ª ordem). Se $\mathbf{x}^* \in \mathbb{R}^n$ é um mínimo local de $f(\mathbf{x})$, então $\nabla f(\mathbf{x}^*) = 0$.

Prova: Por absurdo, suponha que \mathbf{x}^* é mínimo local de $f(\mathbf{x})$ e que $\nabla f(\mathbf{x}^*) \neq 0$. Para $\varepsilon > 0$ suficientemente pequeno, é possível definir um $\mathbf{x} \in \mathbb{R}^n$ tal que $\mathbf{x} = \mathbf{x}^* - \varepsilon \nabla f(\mathbf{x}^*)$. Portanto:

$$\begin{aligned} f(\mathbf{x}) &\cong f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)^T (\mathbf{x} - \mathbf{x}^*) = f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)^T (-\varepsilon \nabla f(\mathbf{x}^*)) = \\ &= f(\mathbf{x}^*) - \varepsilon \nabla f(\mathbf{x}^*)^T \nabla f(\mathbf{x}^*) = f(\mathbf{x}^*) - \varepsilon \|\nabla f(\mathbf{x}^*)\|^2 \end{aligned}$$

Logo, em uma vizinhança de \mathbf{x}^* , $\exists \mathbf{x}$ tal que $f(\mathbf{x}) < f(\mathbf{x}^*)$. ← ABSURDO!

10 Condição Suficiente de Otimalidade

- Teorema: Assuma que $f: \mathfrak{R}^n \rightarrow \mathfrak{R}$ e $f \in C^2[\mathfrak{R}^n]$ (conjunto das funções com derivadas contínuas até 2ª ordem). Se $\mathbf{x}^* \in \mathfrak{R}^n$ é tal que $\nabla f(\mathbf{x}^*) = 0$ e $\nabla^2 f(\mathbf{x}^*) > 0$, então \mathbf{x}^* resolve $\min_{\mathbf{x}} f(\mathbf{x})$.

Prova: Para $\varepsilon > 0$ suficientemente pequeno, é possível definir um $\mathbf{x} \in \mathfrak{R}^n$ tal que $\mathbf{x} = \mathbf{x}^* - \varepsilon \mathbf{d}$, onde $\mathbf{d} \in \mathfrak{R}^n$ é uma direção arbitrária. Portanto:

$$\begin{aligned} f(\mathbf{x}) &\cong f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)^T (\mathbf{x} - \mathbf{x}^*) + \frac{1}{2} (\mathbf{x} - \mathbf{x}^*)^T \nabla^2 f(\mathbf{x}^*) (\mathbf{x} - \mathbf{x}^*) = \\ &= f(\mathbf{x}^*) + \frac{1}{2} (-\varepsilon \mathbf{d})^T \nabla^2 f(\mathbf{x}^*) (-\varepsilon \mathbf{d}) = f(\mathbf{x}^*) + \frac{\varepsilon^2}{2} \underbrace{\mathbf{d}^T \nabla^2 f(\mathbf{x}^*) \mathbf{d}}_{>0} \end{aligned}$$

Como $\mathbf{d} \in \mathfrak{R}^n$ é qualquer, então existe uma vizinhança de \mathbf{x}^* tal que $f(\mathbf{x}) > f(\mathbf{x}^*)$. Logo, \mathbf{x}^* é um mínimo local.

- Conclusão: Suponha que $\mathbf{x}^* \in \mathfrak{R}^n$ é tal que $\nabla f(\mathbf{x}^*) = 0$. Então \mathbf{x}^* é
 - (a) um mínimo global de $f(\mathbf{x})$ se $\nabla^2 f(\mathbf{x}) \geq 0 \quad \forall \mathbf{x} \in \mathfrak{R}^n$, ou seja, $f(\mathbf{x}^*) \leq f(\mathbf{x}) \quad \forall \mathbf{x} \in \mathfrak{R}^n$.
 - (b) um mínimo global estrito de $f(\mathbf{x})$ se $\nabla^2 f(\mathbf{x}) > 0 \quad \forall \mathbf{x} \in \mathfrak{R}^n$, ou seja, $f(\mathbf{x}^*) < f(\mathbf{x}) \quad \forall \mathbf{x} \in \mathfrak{R}^n, \mathbf{x} \neq \mathbf{x}^*$.
 - (c) um máximo global de $f(\mathbf{x})$ se $\nabla^2 f(\mathbf{x}) \leq 0 \quad \forall \mathbf{x} \in \mathfrak{R}^n$, ou seja, $f(\mathbf{x}^*) \geq f(\mathbf{x}) \quad \forall \mathbf{x} \in \mathfrak{R}^n$.
 - (d) um máximo global estrito de $f(\mathbf{x})$ se $\nabla^2 f(\mathbf{x}) < 0 \quad \forall \mathbf{x} \in \mathfrak{R}^n$, ou seja, $f(\mathbf{x}^*) > f(\mathbf{x}) \quad \forall \mathbf{x} \in \mathfrak{R}^n, \mathbf{x} \neq \mathbf{x}^*$.

Exemplo:

- Resolva o problema $\min_{\mathbf{x}} \frac{1}{2} \mathbf{x}^T A \mathbf{x} + \mathbf{b}^T \mathbf{x}$, onde $A = A^T > 0$ e $\mathbf{x} \in \Re^n$.

Solução:

- Definindo $f(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T A \mathbf{x} + \mathbf{b}^T \mathbf{x}$, a aplicação da condição necessária de otimalidade ao problema $\min_{\mathbf{x}} f(\mathbf{x})$ produz:

$$\nabla f(\mathbf{x}) = A\mathbf{x} + \mathbf{b} = \mathbf{0} \Rightarrow \mathbf{x}^* = -A^{-1}\mathbf{b}$$

- Como, por hipótese, $\nabla^2 f(\mathbf{x}) = A > 0$ (condição suficiente de otimalidade), então \mathbf{x}^* é um ponto de mínimo global, portanto solução do problema.

11 Referências bibliográficas

- ANDERSON, B.D.O. & MOORE, J.B. “Optimal Control – Linear Quadratic Methods”, Prentice-Hall, 1989.
- ATHANS, M. & FALB, P.L. “Optimal Control: An Introduction to the Theory and Its Application”, McGraw Hill, 1966.
- BAZARAA, M. S., SHERALY, H. D. & SHETTY, C. “Nonlinear Programming: Theory and Algorithms”, 2nd edition, John Wiley & Sons, 1992.
- BRONSON, R. “Theory and Problems of Matrix Operations”, Schaum’s Outline Series, McGraw Hill, 1989.
- FERREIRA, P.A.V. “Notas de Aula - Curso EA932: Sistemas de Controle II”, 1997.
- FRANKLIN, G.F., POWELL, J.D. & EMAMI-NAEINI, A. “Feedback Control of Dynamic Systems”, 3rd. edition, Addison-Wesley Publishing Company, 1994.
- GOLUB, G.H. & VAN LOAN, C.F. “Matrix Computations”, Johns Hopkins Series in the Mathematical Sciences, Johns Hopkins University Press, 3rd edition, 1996.
- HAYKIN, S. “Adaptive Filter Theory”, Prentice Hall, Third Edition, 1996.
- KIRK, D.E. “Optimal Control: An Introduction”, Prentice-Hall, 1970.
- KWAKERNAAK, H. & SIVAN, R. “Linear Optimal Control Systems”, John Wiley & Sons, 1972.
- LEVINE, W.S. (ed.) “The Control Handbook”, CRC Press, 1996.
- LEWIS, F.L. & SYRMOS, V.L. “Optimal Control”, 2nd edition, John Wiley & Sons, 1995.
- LUENBERGER, D. G. “Linear and Nonlinear Programming”, 2nd edition, Addison Wesley, 1984.
- LUENBERGER, D.G. “Optimization by Vector Space Methods”, John Wiley & Sons, 1969 (Paperback, 1997).
- LUENBERGER, D.G. “Introduction to Dynamic Systems – Theory, Models, and Applications”, John Wiley & Sons, 1979.
- MARDIA, K.V., KENT, J.T. & BIBBY, J.M. “Multivariate Analysis”, Academic Press, 1979.
- OGATA, K. “Modern Control Engineering”, Third Edition, Prentice Hall, 1997.
- PERES, P.L.D. “Notas de Aula - Curso IA600: Controle Ótimo”, 1993.
- SAGE, A.P. & WHITE III, C.C. “Optimum Systems Control”, 2nd edition, Prentice-Hall, 1977.
- SKELTON, R.E. “Dynamic Systems Control: Linear Systems Analysis and Synthesis”, John Wiley & Sons, 1988.
- STRANG, G. “Linear Algebra and Its Applications”, Harcourt Brace College Publishers, 1988 (4th edition, 2000).