Otimização não-linear

Prof. Tarcisio F. Maciel, Dr.-Eng.
Colaboradores: Diego A. Sousa, M.-Eng., José Mairton B. da Silva Jr., M.-Eng., Francisco Hugo C. Neto, M.-Eng., e Yuri Victor L. de Melo, M.-Eng.

Universidade Federal do Ceará
Centro de Tecnologia
Programa de Pós-Graduação em Engenharia de Teleinformática

19 de agosto de 2022

Introdução

Conteúdo

Conteúdo

Introdução [?]

- Pessoas otimizam
  - Investidores ➜ criam portfólios que minimizam riscos e atingem uma certa taxa de retorno
  - Fabricantes ➜ maximizam eficiência no projeto e operação de suas plantas produtivas
  - Engenheiros ➜ ajustam parâmetros para reduzir custos e aumentar a eficiência de seus projetos
- A natureza otimiza
  - Sistemas físicos ➜ tendem ao estado de mínima energia
  - Raios de luz ➜ percorrem o caminho de menor tempo de percurso
  - Moléculas ➜ se acomodam para minimizar a energia potencial dos elétrons
- Otimização é uma ferramenta importante para tomada de decisões e análise de sistemas físicos

Introdução [?]

- O que é otimização?
  - Dar a algo um rendimento ótimo, criando-lhe as condições mais favoráveis ou tirando o melhor partido possível; tornar algo ótimo ou ideal [**?**]
- Porque otimizar?
  - Com otimização é possível melhorar o desempenho de um sistema, ou seja, deixar o sistema mais rápido e eficiente [**?**]
- Como otimizar?
  - Para otimizar é preciso definir o objetivo, uma medida que quantifica o desempenho do sistema
  - O objetivo depende de certos de certas características do sistema, chamadas de variáveis que otimizam o sistema
  - Por fim é frequente o uso de restrições que descrevem situações do sistema consideradas não-desejáveis [**?**]

Conteúdo

Elementos de um problema de otimização

- Uso da ferramenta de otimização ➜ identificar alguns elementos
  - Objetivo ➜ medida quantitativa do desempenho do sistema em estudo ➜ lucro, tempo, energia, entre outros, ou uma combinação de fatores resultando em um escalar
  - Variáveis ➜ parâmetros cujos valores podem ser ajustados e dos quais depende o desempenho do sistema
  - Restrições ➜ condições às quais os valores da variáveis estão sujeitos
- Modelagem de problemas de otimização ➜ Processo de identificação do objetivo, variáveis e restrições
  - Construção de um modelo apropriado ➜ algumas vezes é o passo mais importante
  - Excessivamente simplista ➜ não provê informação suficiente
  - Complexo demais ➜ difícil de resolver
- Após modelado ➜ solucionado usando um algoritmo de otimização
- Não há algoritmo universal ➜ coleção de algoritmos especializados para cada tipo de problema
- Maiores benefícios ➜ surgem quando o tipo do problema é conhecido
- Resultado do algoritmo ➜ validado utilizando as condições de optimalidade

Modelo matématico de um problema de otimização

- Matematicamente, otimização é a maximização ou minimização de uma função objetivo sujeita a restrições sobre suas variáveis de otimização
- Em nossos modelos matemáticos, tipicamente:
  - $\boldsymbol{x} \in \mathbb{R}^n$ ➜ vetor de variáveis de otimização
  - $f(\boldsymbol{x}) : \mathbb{R}^n \to \mathbb{R}$ ➜ função objetivo que se deseja maximizar ou minimizar
  - $f_i(\boldsymbol{x}) : \mathbb{R}^n \to \mathbb{R}$ ➜ as restrições de igualdade e desigualdade que $\boldsymbol{x}$ deve satisfazer
- Logo, pode-se escrever um problema de otimização como

$$\boldsymbol{x}^\star = \underset{\boldsymbol{x}\in\mathbb{R}^n}{\text{minimize}}\ f(\boldsymbol{x}) \tag{1a}$$

$$\text{sujeito a} \quad f_i(\boldsymbol{x}) = 0, \quad i = 1, 2, \ldots, I, \tag{1b}$$

$$f_j(\boldsymbol{x}) \geq 0, \quad j = 1, 2, \ldots, J, \tag{1c}$$

onde i e j são os índices para as restrições de igualdade e desigualdade, respectivamente, e $\boldsymbol{x}^\star$ é uma solução ótima

Exemplo de problema de otimização: transporte de produtos [?]

- Uma companhia possui 02 fábricas F_1 e F_2 e 12 lojas $R_1, R_2, \ldots, R_{12}$. Cada fábrica F_i pode produzir a_i toneladas de um produto (a_i é a capacidade de produção da planta) e cada loja possui uma demanda semanal de b_j toneladas do produto. O custo de transporte da fábrica F_i para a loja R_j de uma tonelada do produto é $c_{i,j}$. O problema é determinar as quantidades $x_{i,j} \in \mathbb{R}_+$ do produto que devem ser transportadas de cada fábrica para cada loja de modo a atender a todos os requisitos e minimizar o custo total.

Exemplo de problema de otimização: transporte de produtos [?]

- Uma companhia possui 02 fábricas F_1 e F_2 e 12 lojas $R_1, R_2, \ldots, R_{12}$. Cada fábrica F_i pode produzir a_i toneladas de um produto (a_i é a capacidade de produção da planta) e cada loja possui uma demanda semanal de b_j toneladas do produto. O custo de transporte da fábrica F_i para a loja R_j de uma tonelada do produto é $c_{i,j}$. O problema é determinar as quantidades $x_{i,j} \in \mathbb{R}_+$ do produto que devem ser transportadas de cada fábrica para cada loja de modo a atender a todos os requisitos e minimizar o custo total. Esse problema pode ser formulado como segue:

$$\{x^\star_{i,j}\} = \underset{\{x_{i,j}\}}{\text{minimize}} \sum_{i=1}^{2}\sum_{j=1}^{12} c_{i,j}x_{i,j} \tag{2a}$$

$$\text{sujeito a} \quad \sum_{j=1}^{12} x_{i,j} \leq a_i, \quad i = 1, 2 \tag{2b}$$

$$\sum_{i=1}^{2} x_{i,j} \geq b_j, \quad j = 1, 2, \ldots, 12 \tag{2c}$$

$$x_{i,j} \geq 0, \quad i = 1, 2 \quad \text{e} \quad j = 1, 2, \ldots, 12 \tag{2d}$$

- Este problema é um problema de otimização linear ➜ função custo e todas as restrições são funções lineares das variáveis do problema

Reescreva o problema acima em forma vetorial/matricial

Exemplo de problema de otimização: minimização da soma das correlações espaciais [?]

- A correlação espacial $\rho_{i,j}$ entre os canais $\boldsymbol{h}_i = \begin{bmatrix} h_{i,1} & h_{i,2} & \dots & h_{i,N} \end{bmatrix}$ e $\boldsymbol{h}_j = \begin{bmatrix} h_{j,1} & h_{j,2} & \dots & h_{j,N} \end{bmatrix}$, com $\boldsymbol{h}_i, \boldsymbol{h}_j \in \mathbb{C}^N$ do enlace direto de uma estação rádio base com N antenas para os terminais móveis i, j é dada por $\rho_{i,j} = \dfrac{|\boldsymbol{h}_i \boldsymbol{h}_j^{\mathrm{H}}|}{\|\boldsymbol{h}_i\|_2 \|\boldsymbol{h}_j\|_2}$. Sabendo que existem K terminais móveis, selecione $G \leq N$ terminais móveis tal que a soma das correlações entre eles dois-a-dois seja mínima, ou seja, selecione os G terminais móveis com os canais menos correlacionados.

Exemplo de problema de otimização: minimização da soma das correlações espaciais [?]

- A correlação espacial $\rho_{i,j}$ entre os canais $\boldsymbol{h}_i = \begin{bmatrix} h_{i,1} & h_{i,2} & \dots & h_{i,N} \end{bmatrix}$ e $\boldsymbol{h}_j = \begin{bmatrix} h_{j,1} & h_{j,2} & \dots & h_{j,N} \end{bmatrix}$, com $\boldsymbol{h}_i, \boldsymbol{h}_j \in \mathbb{C}^N$ do enlace direto de uma estação rádio base com N antenas para os terminais móveis i, j é dada por $\rho_{i,j} = \frac{|\boldsymbol{h}_i \boldsymbol{h}_j^{\mathrm{H}}|}{\|\boldsymbol{h}_i\|_2 \|\boldsymbol{h}_j\|_2}$. Sabendo que existem K terminais móveis, selecione $G \leq N$ terminais móveis tal que a soma das correlações entre eles dois-a-dois seja mínima, ou seja, selecione os G terminais móveis com os canais menos correlacionados. Esse problema pode ser formulado como segue:

$$\boldsymbol{x}^{\star} = \underset{\boldsymbol{x}}{\text{minimize}} \ \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{R}\boldsymbol{x}, \tag{3a}$$

$$\text{sujeito a} \quad \mathbf{1}^{\mathrm{T}}\boldsymbol{x} = G, \tag{3b}$$

$$\boldsymbol{x} \in \mathbb{B}^K, \tag{3c}$$

  onde $\boldsymbol{R} = [\rho_{i,j}]_{i,j}, \quad i, j \in \{1, 2, \dots, K\}$.

- Este problema é um problema de otimização binário quadrático ➜ função custo quadrática com variáveis de otimização binárias

Reescreva o problema acima utilizando somatórios

Conteúdo

Classes de problemas de otimização

- Natureza das variáveis de otimização, da função objetivo, e das restrições ➜ diferentes tipos de problemas de otimização e algoritmos de otimização

- Variáveis de otimização
  - $\boldsymbol{x} \in \mathbb{R}^n$ ➜ otimização contínua (mais fácil de resolver)
  - $\boldsymbol{x} \in \mathbb{Z}^n$ ➜ otimização inteira (pode requerer relaxações contínuas)
  - $x_1, \ldots, x_k \in \mathbb{R}$ e $x_{k+1}, \ldots, x_n \in \mathbb{Z}$ ➜ otimização inteira mista

- Função objetivo e restrições
  - $f(\boldsymbol{x})$, $f_i(\boldsymbol{x})$ e $f_j(\boldsymbol{x})$ lineares ➜ Otimização linear
  - $f(\boldsymbol{x})$ e $f_i(\boldsymbol{x})$ convexas, $f_j(\boldsymbol{x})$ lineares ➜ Otimização convexa

- Máximo/mínimo local ➜ $f(\boldsymbol{x}^\star)$ é um máximo/mínimo local de $f(\boldsymbol{x})$ se existe um sub-espaço aberto $\mathbb{A} \subset \mathbb{R}^n$ contendo $\boldsymbol{x}^\star$ tal que $f(\boldsymbol{x}) \lesseqgtr f(\boldsymbol{x}^\star), \forall \boldsymbol{x} \in \mathbb{A}$.

- Máximo/mínimo global ➜ $f(\boldsymbol{x}^\star)$ é um máximo/mínimo global de $f(\boldsymbol{x})$ se $f(\boldsymbol{x}) \lessgtr f(\boldsymbol{x}^\star), \forall \boldsymbol{x} \in \mathbb{R}^n$

Classes de problemas de otimização

- Algoritmos de otimização ➜ especializados para cada tipo de problema
  - Otimização linear ➜ método Simplex
  - Otimização convexa ➜ método dos pontos interiores
  - Otimização linear inteira ➜ método *branch-and-bound*
  - Otimização sem restrições ➜ método de busca direta e métodos do gradiente
  - Otimização com restrições ➜ métodos dos pontos interiores
  - Otimização determinística ➜ restrições e parâmetros dados por funções bem definidas
  - Otimização estocástica ➜ restrições ou parâmetros dependem de variáveis aleatórias

Exemplo de problema de otimização: mínimos quadrados [?]

- Considere o problema de minimizar a soma dos erros quadráticos entre as componentes de um vetor $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ e um vetor de referência $\boldsymbol{b}$. Esse é um problema de mínimos quadrados sem restrições que pode ser escrito como

$$\boldsymbol{x}^\star = \underset{\boldsymbol{x}}{\text{minimize}} \; \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2^2 \tag{4}$$

Exemplo de problema de otimização: mínimos quadrados [?]

- Considere o problema de minimizar a soma dos erros quadráticos entre as componentes de um vetor $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ e um vetor de referência $\boldsymbol{b}$. Esse é um problema de mínimos quadrados sem restrições que pode ser escrito como

$$\boldsymbol{x}^\star = \underset{\boldsymbol{x}}{\text{minimize}} \; \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2^2 \tag{4}$$

- Note que

$$\begin{aligned} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2^2 &= (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}) = (\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}} - \boldsymbol{b}^{\mathrm{T}})(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}) \\ &= \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} - \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b} - \boldsymbol{b}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}^{\mathrm{T}}\boldsymbol{b} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} - 2\boldsymbol{b}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}^{\mathrm{T}}\boldsymbol{b} \end{aligned} \tag{5}$$

- Derivando a equação acima em relação a $\boldsymbol{x}$ e igualando a $\mathbf{0}$ temos

$$\frac{d}{d\boldsymbol{x}}\left(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} - 2\boldsymbol{b}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}^{\mathrm{T}}\boldsymbol{b}\right) = 0 \Rightarrow 2\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} - 2\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b} = 0 \Rightarrow \boxed{\boldsymbol{x}^\star = (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}} \tag{6}$$

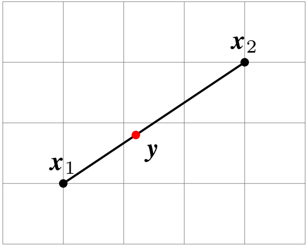
Liste as classes de problemas de otimização às quais esse problema pertence

Conjuntos convexos e funções convexas

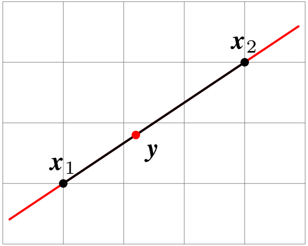
Conteúdo

Conteúdo

Conjuntos afins



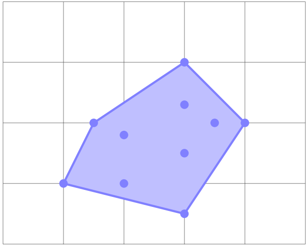
- Linha e segmentos
  - Sejam os pontos $x_1, x_2 \in \mathbb{R}^n$ com $x_1 \neq x_2$ ➜ pontos y da forma $y = \theta x_1 + (1 - \theta) x_2$ com $\theta \in \mathbb{R}, 0 \leq \theta \leq 1$, formam um segmento fechado ligando x_1 a x_2
  - Representando $y = x_2 + \theta(x_1 - x_2)$ ➜ Ponto base x_2 e direção $(x_1 - x_2)$ apontando de x_2 para x_1 e escalonada por θ



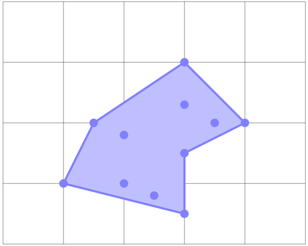
- Conjuntos afins
  - Um conjunto $\mathbb{A} \subset \mathbb{R}^n$ é afim ➜ para qualquer $x_1, x_2 \in \mathbb{A} \subset \mathbb{R}$ e $\theta \in \mathbb{R}$, o segmento de reta $y = \theta x_1 + (1 - \theta) x_2 \in \mathbb{A}$
  - Combinação afim ➜ $y = \theta_1 x_1 + \theta_2 x_2 + \ldots + \theta_k x_k \in \mathbb{A} \subset \mathbb{R}$, com $\theta_1 + \theta_2 + \ldots + \theta_k = 1$

Conjuntos convexos

- Conjuntos convexos
  - Um conjunto $\mathbb{A} \subset \mathbb{R}^n$ é convexo caso um segmento entre $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{A}$ e $0 \leq \theta \leq 1$, pertença a $\mathbb{A}$, ou seja, $\theta \boldsymbol{x}_1 + (1-\theta)\boldsymbol{x}_2 \in \mathbb{A}$
  - Em termos geométricos, um conjunto convexo é um conjunto sem buracos ou reentrâncias



- Invólucro convexo (*convex hull*)
  - Dado um conjunto $\mathbb{A}$, o invólucro convexo de $\mathbb{A}$ é o menor conjunto convexo que engloba $\mathbb{A}$, sendo denotado por $\operatorname{conv}\{\mathbb{A}\} = \{\theta_1 \boldsymbol{x}_1 + \cdots + \theta_k \boldsymbol{x}_k \mid \boldsymbol{x}_i \in \mathbb{A}, \theta_i \geq 0, i = 1, \cdots, k, \theta_1 + \cdots + \theta_k = 1\}$



Conteúdo

Operações sobre conjuntos que preservam convexidade

- Interseção ➜ se $\mathbb{A}_\alpha$ é convexo para todo $\alpha \in \mathbb{R}$, então

$$\bigcap_{\alpha \in \mathbb{R}} \mathbb{A}_\alpha \qquad \text{também é convexo} \tag{7}$$

- Multiplicação por escalar ➜ se $\mathbb{A} \subseteq \mathbb{R}^n$ é convexo e $\alpha \in \mathbb{R}$, então

$$\alpha\mathbb{A} = \{\alpha\boldsymbol{x} \mid \boldsymbol{x} \in \mathbb{A}\} \qquad \text{também é convexo} \tag{8}$$

- Translação ➜ se $\mathbb{A} \subseteq \mathbb{R}^n$ é convexo e $\boldsymbol{\alpha} \in \mathbb{R}^n$, então

$$\mathbb{A} + \boldsymbol{\alpha} = \{\boldsymbol{x} + \boldsymbol{\alpha} \mid \boldsymbol{x} \in \mathbb{A}\} \qquad \text{também é convexo} \tag{9}$$

- Soma ➜ se $\mathbb{A}_1$ e $\mathbb{A}_2$ são convexos, então

$$\mathbb{A}_1 + \mathbb{A}_2 = \{\boldsymbol{x} + \boldsymbol{y} \mid \boldsymbol{x} \in \mathbb{A}_1, \boldsymbol{y} \in \mathbb{A}_2\} \qquad \text{também é convexo} \tag{10}$$

- Produto cartesiano ➜ se $\mathbb{A}_1$ e $\mathbb{A}_2$ são convexos, então

$$\mathbb{A}_1 \times \mathbb{A}_2 = \{(\boldsymbol{x}, \boldsymbol{y}) \mid \boldsymbol{x} \in \mathbb{A}_1, \boldsymbol{y} \in \mathbb{A}_2\} \qquad \text{também é convexo} \tag{11}$$

Conteúdo

Hiperplanos, cones e desigualdades generalizadas

- Hiperplano ➜ conjunto de pontos, podendo ser escrito como

$$\{\boldsymbol{x} \mid \boldsymbol{a}^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{x}_0) = 0)\} \tag{12}$$

  onde $\boldsymbol{a} \in \mathbb{R}^n$, $\boldsymbol{a} \neq \mathbf{0}$, e $\boldsymbol{x}_0$ determina o offset do hiperplano. Um hiperplano divide o espaço em dois semi-espaços

- Semi-espaço ➜ conjunto da forma

$$\{\boldsymbol{x} \mid \boldsymbol{a}^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{x}_0) \leq 0)\} \tag{13}$$

  onde $\boldsymbol{a} \in \mathbb{R}^n$, $\boldsymbol{a} \neq \mathbf{0}$

TODO

Adicionar figuras

Hiperplanos, cones e desigualdades generalizadas

- Cone é conjunto de pontos tais que $\forall \boldsymbol{x} \in \mathbb{A} \subset \mathbb{R}^n$ e $\theta \geq 0, \theta \in \mathbb{R}_+$

$$\theta \boldsymbol{x} \in \mathbb{A}, \forall \boldsymbol{x} \in \mathbb{A} \tag{14}$$

- Cone Convexo é um conjunto que é simultaneamente um cone e convexo, ou seja, para qualquer $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{A} \subset \mathbb{R}^n$ e $\theta_1, \theta_2 \geq 0, \theta_1$

$$\theta_1 \boldsymbol{x}_1 + \theta_2 \boldsymbol{x}_2 \in \mathbb{A} \tag{15}$$

TODO

Adicionar figuras

Conjuntos Convexos

- A bola Euclidiana com o centro em $\boldsymbol{x}_c$ e raio r é representada por

$$B(\boldsymbol{x}_c, r) = \{\boldsymbol{x}_c + r\boldsymbol{u} \mid \|\boldsymbol{u}\|_2 \leq 1\} \tag{16}$$

- O elipsoide com o centro em $\boldsymbol{x}_c$ é um conjunto representado na forma

$$\mathcal{E} = \{\boldsymbol{x} \mid (\boldsymbol{x} - \boldsymbol{x}_c)^{\mathrm{T}}\boldsymbol{P}^{-1}(\boldsymbol{x} - \boldsymbol{x}_c) \leq 1)\} \tag{17}$$

onde $\boldsymbol{P} \in \mathbb{A} \subset \mathbb{S}^n_{++}$, ou seja, $\boldsymbol{P}$ é uma matriz simétrica positiva definida

TODO

Adicionar figuras

Conjuntos Convexos

- Poliedro
  - O poliedro é definido como um conjunto de igualdades lineares e inequações.

$$Ax \preceq b, Cx = d \tag{18}$$

  onde $\preceq$ é o símbolo que representa desiguadade componente a componente entre vetores.
  - Dessa forma, a representação do poliedro pode escrita como:

$$P = \{x \mid Ax \preceq b, Cx = d\} \tag{19}$$

Conjuntos Convexos

- Norma da Bola com centro x_c e raio r é representado por:

$$\{x \mid \|x - x_c\| \leq r\} \tag{20}$$

  onde $\|\cdot\|$ é a norma.

- A norma do cone pode ser definido como:

$$\{(x,t) \in \mathbb{R}^n + 1 \mid \|x\| \leq t\} \tag{21}$$

Conjuntos Convexos

- PSD (Cone positivo semidefinido) é um cone formado a parti das matrizes positivas semidefinidas que utiliza a notação:

$$\mathbb{A}^n_+ = \{X \in \mathbb{A}^n \mid X \succ 0\} \tag{22}$$

onde $\mathbb{A}^n$ denota o conjunto de matrizes simetricas. Dessa forma o conjunto $\mathbb{A}^n_+$ é um cone positivo semidefinido, se $\theta_1, \theta_2 \geqslant 0$ e $A, B \in \mathbb{A}^n_+$, assim $\theta_1 A + \theta_2 B \in \mathbb{A}^n_+$.

Conjuntos Convexos

- Desigualdade generalizadas
  - Um cone pode ser usado para definir a desigualdade generalizada, que é similar a relação de ordem apresentada em $\mathbb{R}$. Desta forma, para um cone K onde $K \subseteq \mathbb{R}^n$, as desigualdades não estritas e estritas $\preceq_K$ e $\prec_K$, respectivamente são definidas da seguinte maneira:

$$x \preceq_K y \Leftrightarrow y - x \in K \tag{23a}$$

$$x \prec_K y \Leftrightarrow y - x \in \textbf{int}(K) \tag{23b}$$

  onde **int**(K) representa o interior do conjunto K.
  - Propriedades
    - Se $x \preceq_K y$ e $u \preceq_K v$, então $x + u \preceq_K y + v$
    - Se $x \preceq_K y$ e $y \preceq_K u$, então $x \preceq_K u$
    - Se $x \preceq_K y$ e $y \preceq_K x$, então $x = y$
    - Se $x \prec_K y$ e $u \preceq_K v$, então $x + u \prec_K y + v$

Conteúdo

Conteúdo

Propriedades básicas

- Uma função $f : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ é convexa somente se o domínio de f $(\mathrm{dom}\ f)$ for convexo e $\forall x, y \in \mathrm{dom}\ f$, e $0 \leq \theta \leq 1$ (24) for verdade.

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta) f(y) \tag{24}$$

- Observando a condição (24), temos que em uma função $f(\cdot)$ convexa, qualquer segmento de reta entre os pontos $(x, f(x))$ e $(y, f(y))$ estará acima de $f(\cdot)$ entre os valores x e y, como mostrado na figura abaixo.

Propriedades básicas

- Dizemos que uma função $f(\cdot)$ é estritamente convexa se a desigualdade ocorrer sempre para $0 < \theta < 1$ e $x \neq y$, ou seja,
$$f(\theta x + (1-\theta)y) < \theta f(x) + (1-\theta)f(y) \tag{25}$$
- Define-se uma função $f(\cdot)$ como côncava, se $-f(\cdot)$ for convexa.
- Funções afins são convexas e côncavas, visto que
$$f(\theta x + (1-\theta)y) = \theta f(x) + (1-\theta)f(y) \tag{26}$$
para $0 \leq \theta \leq 1$.

Propriedades básicas

- Uma função $f(\cdot)$ é convexa se e somente se ela é convexa quando restrita a qualquer linha que intercepta seu domínio. Ou seja, $f(x)$ é convexa se e somente se para todo $x \in \mathrm{dom}\, f$ e para todo v, a função $g(t) = f(x + tv)$ é convexa no domínio $\{t | x + tv \in \mathrm{dom}\, f\}$.
- Esta propriedade é muito importante, pois pode-se avaliar a convexidade de uma função, restringindo-a a uma linha.

Extensão

- As vezes é conveniente que o domínio da função $f(\cdot)$ seja estendido para todo $\mathbb{R}^n$, atribuindo o valor ∞ para $f(x)$ quando $x \notin \operatorname{dom} f$.
- Assim, a extensão de $f(x)$, $\tilde{f} : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \infty$ é definida como:

$$\tilde{f}(x) = \begin{cases} f(x) & x \in \operatorname{dom} f \\ \infty & x \notin \operatorname{dom} f \end{cases} \tag{27}$$

- Essa extensão é útil pelo fato de simplificar a notação e de não haver a necessidade de se explicitar o seu domínio.
- De modo análogo, podemos estender funções côncavas adicionando $-\infty$ a seu domínio.

Condições de primeira ordem

- Seja $f(\cdot)$ uma função diferenciável, então $f(\cdot)$ é convexo se e somente se o $\mathrm{dom}\, f$ for convexo e

$$f(y) \geq f(x) + \nabla f(x)^{\mathrm{T}}(y - x) \quad \forall x, y \in \mathrm{dom}\, f \tag{28}$$

- Note que o segundo termo de (28) é uma função afim em relação a y e que é igual a aproximação de primeira ordem de Taylor de $f(y)$ em torno de x.

Condições de primeira ordem

- Assim, se a aproximação de primeira ordem de Taylor de uma função sempre representar um subestimador global, a função $f(\cdot)$ é convexa.
- De (28), podemos ver que a partir da informação local (x e $\nabla f(x)$) nós podemos derivar a informação global (subestimador global) de $f(\cdot)$.
- Podemos ainda deduzir de (28) que se $\nabla f(x) = 0$, então $f(y) \geq f(x)$, o que implica dizer que o ponto $(x, f(x))$ representa o mínimo global de $f(\cdot)$.
- Para que $f(\cdot)$ seja estritamente convexo, temos que a desigualdade em (28) tem que ser estrita, ou seja,

$$f(y) > f(x) + \nabla f(x)^{\mathrm{T}}(y - x) \quad \forall x, y \in \mathrm{dom}\, f \text{ e } x \neq y \tag{29}$$

Condições de segunda ordem

- Seja $f(\cdot)$ duas vezes diferenciável. $f(\cdot)$ é convexo se e somente se o $\mathrm{dom}\, f$ é convexo e sua Hessiana é semi-positiva definida $\forall x \in \mathrm{dom}\, f$, ou seja,

$$\nabla^2 f(x) \succeq 0. \tag{30}$$

- No caso das funções reais (30) se reduz a $f''(x) \geq 0$, o que implica dizer que $f(\cdot)$ é não decrescente.
- Geograficamente, a condição (30) pode ser interpretada como o gráfico da função possuir curvatura sempre positiva.
- De modo análogo, temos que $f(\cdot)$ é côncava se e somente se

$$\nabla^2 f(x) \preceq 0. \tag{31}$$

- Se $\nabla^2 f(x) \succ 0$, então $f(\cdot)$ é estritamente convexa, contudo, a recíproca não é verdadeira!

Conjunto de subníveis

- O conjunto de subnível α de uma função $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ é definido como:

$$C_\alpha = \{x \in \mathrm{dom}\, f | f(x) \leq \alpha\} \tag{32}$$

- Conjuntos de subnível de uma função convexa são convexos para qualquer α.
- Prova: Seja $x, y \in \mathbb{A}_\alpha$, então $f(x) \leq \alpha$ e $f(y) \leq \alpha$. Assim, de (24), temos que:

$$\begin{aligned} f(\theta x + (1-\theta)y) &\leq \theta f(x) + (1-\theta) f(y) \\ &\leq \theta\alpha + (1-\theta)\alpha \leq \alpha \end{aligned}$$

  Logo, $\theta x + (1-\theta)y \in \mathbb{A}_\alpha$
- A recíproca não é verdadeira! Uma função pode ter todos seus subníveis convexos e não ser convexa.

Epígrafo e Hipografo

- O grafo de uma função $f : \mathbb{R}^n \to \mathbb{R}$ é definido como

$$\{(x, f(x))|x \in \operatorname{dom} f\}, \tag{33}$$

  que é um subconjunto de $\mathbb{R}^{n+1}$.
- O epígrafo de $f(\cdot)$ é definido como:

$$\operatorname{epi} f = \{(x, t)|x \in \operatorname{dom} f, f(x) \leq t\} \tag{34}$$

  que é um subconjunto de $\mathbb{R}^{n+1}$.

Epígrafo e Hipografo

- O epígrafo faz a ligação entre funções e conjuntos convexos.
- Uma função $f(\cdot)$ é convexa se e somente se seu epígrafo representar um conjunto convexo.
- O hipografo de uma função é definido como sendo

$$\text{hypo}\, f = \{(x, t) | x \in \text{dom}\, f, t \leq f(x)\} \tag{35}$$

- A função $f(\cdot)$ é côncava se e somente seu hipografo for convexo.

Desigualdade de Jensen e extensões

Conteúdo

Conteúdo

Conteúdo

Conteúdo

Conteúdo

Problemas de otimização convexa

Conteúdo

Conteúdo

Conteúdo

Conteúdo

Conteúdo

Conteúdo

Conteúdo

Conteúdo

Aplicações

Conteúdo

Conteúdo

Algoritmo de *water filling*

- O objetivo do water filling é distribuir potências entre os links de uma sistema telecomunicação objetivando a maximizaçao da vazão de dados, uma vez que a qualidade do link dependente da potência utilizada na comunicação, dessa forma:

$$\min -\sum_{i=1}^{n} \log(\alpha_i + x_i) \text{sujeito a} \begin{cases} x \geq 0, \\ \mathbf{1}^T x = 1 \end{cases} \tag{36}$$

onde $\alpha_i > 0, x_i$ representa a potência alocada ao canal i.

Algoritmo de *water filling*

- Adotando o multiplicador lagrangeano $\lambda^* \in \mathbb{R}^n$ para $x \geq 0$ e o multiplicador $v^* \in \mathbb{R}$ para $\mathbf{1}^T x = 1$, o lagrangeano apresenta a seguinte forma:

$$H(x^*,\lambda^*,v^*) = -\sum_{i=1}^{n} \log(\alpha_i + x_i) - \lambda^* x + v^*(\mathbf{1}^T x - 1) \tag{37}$$

- Derivando em seguida utilizando as KKT:

$$\frac{H(x^*,\lambda^*,v^*)}{\partial x^*} = -\frac{1}{(\alpha_i + x_i)} - \lambda^* + v^* = 0 \tag{38a}$$

$$\frac{H(x^*,\lambda^*,v^*)}{\partial \lambda^*} = x = 0 \tag{38b}$$

$$\frac{H(x^*,\lambda^*,v^*)}{\partial v^*} = \mathbf{1}^T x - 1 = 0 \tag{38c}$$

Algoritmo de *water filling*

- Utilizando as KKT:

$$
\begin{aligned}
x_* &\geq 0, && (39a)\\
\mathbf{1}^T x^* &= 1, && (39b)\\
\lambda^* &\geq 0, && (39c)\\
\lambda_i^* x_i^* &= 0, && (39d)\\
-\frac{1}{(\alpha_i + x_i^*)} - \lambda_i^* + \nu^* &= 0 && (39e)
\end{aligned}
$$

- Isolando λ_i^* e substituindo em $\lambda_i^* x_i^*$:

$$
\begin{aligned}
x_i^*(\nu^* - \frac{1}{\alpha_i + x_i^*}) &= 0, && (40a)\\
\nu^* - \frac{1}{\alpha_i + x_i^*} &\geq 0 && (40b)
\end{aligned}
$$

Algoritmo de *water filling*

- Isolando x_i^* da expressão $v^* \geq \frac{1}{\alpha_i + x_i^*}$ temos:

$$x_i^* \geq \frac{1}{v_*} - \alpha_i \tag{41}$$

- Entretando x_i^* não pode ser negativo no caso $v^* > \frac{1}{\alpha_i}$, assim:

$$x_i^* = \begin{cases} \frac{1}{v^*} - \alpha_i, & v^* < \frac{1}{\alpha_i} \\ 0, & v^* \geq \frac{1}{\alpha_i} \end{cases} \tag{42}$$

Conteúdo

Regressão

- Sejam dois vetores $\boldsymbol{x} = [x_1 \; x_2 \; \ldots \; x_N]^\mathrm{T}$ e $\boldsymbol{y} = [y_1 \; y_2 \; \ldots \; y_N]^\mathrm{T}$, que representam respectivamente a entrada e a saída de um determinado sistema, em que $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^{N\times 1}$;
- A análise de regressão visa estimar relacionamento entre $\boldsymbol{x}$ e $\boldsymbol{y}$.
- Em um caso geral, a análise de regressão visa encontrar os M parâmetros $\boldsymbol{\alpha} = [\alpha_1 \; \alpha_2 \; \ldots \; \alpha_M]^\mathrm{T}$ de uma função $f(\cdot)$ qualquer, de modo que $f(\boldsymbol{x}, \boldsymbol{\alpha})$ seja o mais próximo de $\boldsymbol{y}$, ou seja,

$$\boldsymbol{y} \approx f(\boldsymbol{x}, \boldsymbol{\alpha})$$

- Em outras palavras, para se estimar os parâmetros $\boldsymbol{\alpha}$, podemos minimizar erro quadrático médio entre $\boldsymbol{y}$ e $f(\boldsymbol{x}, \boldsymbol{\alpha})$;
- Assim, temos que:

$$e = \frac{1}{N} \|\boldsymbol{y} - f(\boldsymbol{x}, \boldsymbol{\alpha})\|^2$$

Regressão

- Assim, escrevendo o problema de otimização, temos que:

$$\begin{aligned}\dot{\boldsymbol{\alpha}} &= \underset{\boldsymbol{\alpha}}{\arg\min.}\{e\} \\ &= \underset{\boldsymbol{\alpha}}{\arg\min.}\left\{\frac{1}{N}\|\boldsymbol{y} - f(\boldsymbol{x}, \boldsymbol{\alpha})\|^2\right\}\end{aligned}$$

- Como N é uma constante positiva, então o problema de otimização pode ser simplificado como

$$\dot{\boldsymbol{\alpha}} = \underset{\boldsymbol{\alpha}}{\arg\min.}\left\{\|\boldsymbol{y} - f(\boldsymbol{x}, \boldsymbol{\alpha})\|^2\right\}$$

- Como a norma ao quadrado é uma função convexa, então, desde que a função $f(\boldsymbol{x}, \boldsymbol{\alpha})$ seja convexa, o problema de otimização em questão pode ser resolvido pelo método *Least-squares*.
- Na literatura existem vários modelos de regressão, para diversos casos.

Conteúdo

Regressão polinomial

- Um dos modelos de regressão mais utilizados é a polinomial;
- É um caso geral da regressão linear;
- Neste caso, seja um polinômio de ordem M, então

$$\boldsymbol{\alpha} = [\alpha_0 \ \alpha_1 \ \ldots \ \alpha_M]^{\mathrm{T}} \tag{43}$$

é o vetor de coeficientes com $M + 1$ elementos;
- Temos que a função $f(\boldsymbol{x}, \boldsymbol{\alpha})$ é definida como:

$$f(\boldsymbol{x}, \boldsymbol{\alpha}) = \boldsymbol{V}\boldsymbol{\alpha}$$

em que $\boldsymbol{V} \in \mathbb{R}^{N \times M}$ é uma matriz de Vandermonde, gerada pelos elementos de $\boldsymbol{x}$

$$\boldsymbol{V} = \begin{bmatrix} 1 & x_1 & x_1^2 & \ldots & x_1^M \\ 1 & x_2 & x_2^2 & \ldots & x_2^M \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_N & x_N^2 & \ldots & x_N^M \end{bmatrix} \tag{44}$$

Regressão polinomial

- Nesse caso, $V\alpha$ é convexo.
- Assim, denotando o resultado da função objetivo por $L(\alpha)$, temos que:

$$
\begin{aligned}
L(\boldsymbol{\alpha}) &= \|\boldsymbol{y} - \boldsymbol{V}\boldsymbol{\alpha}\|^2 \\
&= (\boldsymbol{y} - \boldsymbol{V}\boldsymbol{\alpha})^{\mathrm{T}}(\boldsymbol{y} - \boldsymbol{V}\boldsymbol{\alpha}) \\
&= (\boldsymbol{y}^{\mathrm{T}} - \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{V}^{\mathrm{T}})(\boldsymbol{y} - \boldsymbol{V}\boldsymbol{\alpha}) \\
&= \boldsymbol{y}^{\mathrm{T}}\boldsymbol{y} - \boldsymbol{y}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{\alpha} - \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{y} + \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{\alpha}
\end{aligned}
$$

- Assim, para encontrar o valor de α que minimiza o valor de $L(\alpha)$, temos que $\frac{\partial L(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} = 0$, logo:

$$
\frac{\partial L(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} = -\boldsymbol{V}^{\mathrm{T}}\boldsymbol{y} - \boldsymbol{V}^{\mathrm{T}}\boldsymbol{y} + (\boldsymbol{V}\boldsymbol{V}^{\mathrm{T}} + \boldsymbol{V}^{\mathrm{T}}\boldsymbol{V})\boldsymbol{\alpha} = 0
$$

Regressão polinomial

- Como $\boldsymbol{V}\boldsymbol{V}^{\mathrm{T}}$ é simétrico, então, temos que:

$$2\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{\alpha} - 2\boldsymbol{V}^{\mathrm{T}}\boldsymbol{y} = 0$$

$$\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{\alpha} = \boldsymbol{V}^{\mathrm{T}}\boldsymbol{y}$$

$$\boldsymbol{\alpha} = \left(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}\right)^{-1}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{y}$$

$$\boxed{\boldsymbol{\alpha} = \boldsymbol{V}^{\dagger}\boldsymbol{y}}$$

Conteúdo

Regressão polinomial multivariada

- Em alguns casos, a variável de saída depende de uma associação de K de valores de entrada;
- Matematicamente, podemos escrever que cada elemento do vetor de saída $\boldsymbol{y} = [y_1 \;\; y_2 \;\; \dots \;\; y_N]^{\mathrm{T}}$ depende de uma linha da matriz de dados de entrada

$$\boldsymbol{x} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1K} \\ x_{21} & x_{22} & \dots & x_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \dots & x_{NK} \end{bmatrix}$$

- Podemos reescrever nosso problema como um problema de regressão polinomial normal, alterando apenas a matriz $\boldsymbol{V}$.

Regressão polinomial multivariada

- No caso da regressão polinomial multivariada, seja um polinômio de grau M, o vetor de coeficientes de $KM + 1$ elementos e é dado por:

$$\boldsymbol{\alpha} = [\alpha_0 \; \alpha_{11} \; \dots \; \alpha_{M1} \; \dots \; \alpha_{12} \; \dots \; \alpha_{M2} \; \dots \; \alpha_{1K} \; \dots \; \alpha_{MK}]^{\mathrm{T}} \tag{45}$$

- Temos que a matriz $\boldsymbol{V}$ é dada por:

$$\boldsymbol{V} = \begin{bmatrix} 1 & x_{11} & x_{11}^2 & \dots & x_{11}^M & x_{12} & x_{12}^2 & \dots & x_{12}^M & \dots & x_{1K} & \dots & x_{1K}^M \\ 1 & x_{21} & x_{21}^2 & \dots & x_{21}^M & x_{22} & x_{22}^2 & \dots & x_{22}^M & \dots & x_{2K} & \dots & x_{2K}^M \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N1} & x_{N1}^2 & \dots & x_{N1}^M & x_{N2} & x_{N2}^2 & \dots & x_{N2}^M & \dots & x_{NK} & \dots & x_{NK}^M \end{bmatrix} \tag{46}$$

- A solução do problema de otimização se dá da mesma forma que no caso monovariável.

Conteúdo

Regressão logística

- Considere agora que a variável de saída $\boldsymbol{y}$ é qualitativa dicotômica (sucesso (1) ou falha (0));
- Nosso objetivo é saber qual a probabilidade de sucesso de y_i dado um conjunto de K dados de entrada $\boldsymbol{x}_i = [x_{i1} \; x_{2i} \; \ldots \; x_{iK}]$, para $i \in (1, N)$;
- Como y_i possui uma distribuição binomial de probabilidades, podemos calcular o valor da probabilidade de sucesso $(y_i = 1)$ a partir da média de várias amostras de y_i sob as mesmas condições $\boldsymbol{x}_i$.
- Assim temos que o vetor de probabilidades de sucesso é dado por:

$$\boldsymbol{p}_y = \begin{bmatrix} P(y_1|\boldsymbol{x}_1) \\ P(y_2|\boldsymbol{x}_2) \\ \vdots \\ P(y_N|\boldsymbol{x}_N) \end{bmatrix} = \begin{bmatrix} \mathcal{E}\left\{y_1|\boldsymbol{x}_1\right\} \\ \mathcal{E}\left\{y_2|\boldsymbol{x}_2\right\} \\ \vdots \\ \mathcal{E}\left\{y_N|\boldsymbol{x}_N\right\} \end{bmatrix}$$

Regressão logística

- Da maneira como os dados de saída estão, o modelo de regressão polinomial não se mostra adequado, visto que o mesmo pode inferir valores de probabilidade maiores que 1 ou menores que 0, o que seria incorreto!
- Para utilizar a regressão polinomial seria necessário aplicar uma função de mapeamento em $\boldsymbol{p}_y$ de forma a gerar $\boldsymbol{q}_y$, que pudesse admitir quaisquer valores reais. Ou seja,

$$\boldsymbol{p}_y \xrightarrow{f} \boldsymbol{q}_y, \quad \boldsymbol{p}_y \in (0,1) \text{ e } \boldsymbol{q}_y \in \mathbb{R} \tag{47}$$

- Uma função que resolve este problema é a função logística, dada por:

$$f(p) = \ln\left(\frac{p}{1-p}\right) \tag{48}$$

- Assim, temos que:

$$\boldsymbol{q}_y = \ln\left(\frac{\boldsymbol{p}_y}{1-\boldsymbol{p}_y}\right) \tag{49}$$

Regressão logística

- Agora, podemos aplicar uma regressão polinomial considerando os dados de entrada $\boldsymbol{X} = [\boldsymbol{x}_1 \ \boldsymbol{x}_2 \ \ldots \ \boldsymbol{x}_K]^{\mathrm{T}}$ e a saída $\boldsymbol{q}_y$. Logo,
$$\boldsymbol{q}_y = \boldsymbol{V}\boldsymbol{\alpha}, \tag{50}$$
em que $\boldsymbol{\alpha}$ e $\boldsymbol{V}$ são definidos de acordo com (43) e (44) se for utilizada a regressão polinomial monovariada $(K = 1)$ ou (45) e (46) se for utilizada a regressão polinomial multivariada.
- Assim, temos que os coeficientes da regressão polinominal são dados por:

$$\boldsymbol{\alpha} = \boldsymbol{V}^{\dagger}\boldsymbol{q}_y \quad \Longrightarrow \quad \boldsymbol{\alpha} = \boldsymbol{V}^{\dagger} \ln\left(\frac{\boldsymbol{p}_y}{1 - \boldsymbol{p}_y}\right) \tag{51}$$

- Assim, temos que a função de regressão que relaciona $\boldsymbol{X}$ com $\boldsymbol{p}_y$ é dada por:

$$\boxed{\boldsymbol{p}_y = \frac{\exp\left(\boldsymbol{V}\boldsymbol{\alpha}\right)}{1 + \exp\left(\boldsymbol{V}\boldsymbol{\alpha}\right)}} \tag{52}$$

Conteúdo

Conteúdo

Motivação

- Em um arranjo de antenas, as saídas de inúmeros elementos emissores são linearmente combinadas de modo a gerar um padrão de radiação resultante.
- O arranjo resultante tem um padrão direcional que depende dos pesos relativos (fatores de escala) usados no processo de combinação.
- O objetivo do projeto de fatores de escala é escolher os pesos de modo a gerar um diagrama de irradiação desejado.
- O diagrama de irradiação pode ser ajustado de modo a aumentar o ganho na direção de um usuário ou reduzir o ganho na direção de maior interferência.

Conteúdo

Osciladores Harmônicos

- Osciladores harmônicos isotrópicos constituem a unidade básica das antenas transmissoras. Estes elementos emitem ondas eletromagnéticas com frequência ω e comprimento de onda λ.
- O campo elétrico gerado por um oscilador em um ponto P do espaço, localizado a uma distância d da antena, é dado por

$$\frac{1}{d} \cdot \mathbf{Re} \left[z \cdot \exp\left(j\omega t - \frac{2\pi d}{\lambda} \right) \right] \tag{53}$$

- Observe que $z \in \mathbb{C}$, é um parâmetro de projeto denominado peso da antena. Este fator dimensiona a magnitude e a fase do campo elétrico.

Arranjo de osciladores

- Arranjo de osciladores é uma combinação de vários elementos de irradiação, com uma determinada distribuição espacial e interconectados entre si.
- Considere que colocamos n osciladores nas posições $p_k \in \mathbb{R}^3$, $k = 1, \ldots, n$. Cada oscilador é associado a um peso complexo z_k. Assim, o campo elétrico total recebido em um ponto $p \in \mathbb{R}^3$ é dado por

$$E = \Re\mathfrak{e}\left[\exp\left(j\omega t\right) \cdot \sum_{k=1}^{n} \frac{1}{d_k} z_k \cdot \exp\left(-\frac{2\pi j d_k}{\lambda}\right)\right] \tag{54}$$

- onde $d_k = \|p - p_k\|$ é a distância entre os pontos p e p_k, onde $k = 1, \ldots, n$.

Diagrama de um Arranjo Linear

- Considera-se que o ponto p em análise está muito distante do arranjo de osciladores. Desse modo, sua posição será dada por $p = ru$, onde $u \in \mathbb{R}^3$ é um vetor unitário que determina a direção e r é um escalar que especifica a distância da origem.
- Para um arranjo linear, o campo elétrico E depende apenas do ângulo ϕ entre o vetor e o ponto distante p

$$E \approx \mathfrak{Re}\left(\frac{1}{r} \cdot \exp\left(j\omega t - \frac{2\pi r}{\lambda}\right)\right) \cdot D_z(\phi) \tag{55}$$

- observe que ϕ é o â ngulo entre os vetores u e e_1.
- A função $D_z : [0, 2\pi] \longrightarrow \mathbb{C}$ é chamada de diagrama da antena. Esta função depende da escolha do vetor de pesos complexos $z = (z_1, \ldots, z_n)$

$$D_z(\phi) = \sum_{k=1}^{n} z_k \cdot \exp\left(\frac{2\pi jk \cos\phi}{\lambda}\right) \tag{56}$$

Conteúdo

Modelagem do Diagrama de Antenas

- O quadrado do módulo do diagrama de antena, $|D_z(\phi)|$, é proporcional a direção da densidade de energia emitida pela antena.
- É de grande interesse modelar a magnitude do diagrama $\||D_z(\cdot)|$, por meio da escolha do vetor de pesos, z, de modo a atender os requisitos de diretividade do sistema.
- Um requerimento típico estabelece que a antena transmite bem ao longo de uma determinada direção. Ou seja, a energia é concentrada ao longo de direção alvo, ϕ_{alvo}, enquanto é reduzida numa outra região.
- Outro requerimento clássico considera a minimização da potência de ruiído térmico gerado pela antena.

Normalização

- A energia enviada ao longo da direção alvo deve ser normalizada.

$$\mathfrak{Re}\left(D_z\left(\phi_{alvo}\right)\right) \geq 1 \tag{57}$$

- Não se modifica a direção da distribuição de energia ao multiplicar-se todos os pesos por um número complexo não nulo.
- Esta é uma restrição afim nas partes reais e imaginárias da variável $z \in \mathbb{C}^n$

Atenuação do Lóbulo Lateral

- Define-se como banda de passagem o intervalo angular $[-\Phi, \Phi]$ onde se pretende concentrar a energia; a banda de parada, corresponde aos pontos fora deste intervalo.
- Para garantir o cumprimento do requerimento de concentração de energia, estabelece-se que

$$|\phi| \geq \Phi \iff |D_z(\phi)| \leq \delta \tag{58}$$

- onde δ é o nível desejado de atenuação na banda de parada ou como o nível do lóbulo lateral.
- Em vez de considerar um intervalo contínuo, é feita uma discretização dessa restrição

$$|D_z(\phi_i)| \leq \delta, i = 1, \ldots, N \tag{59}$$

- os angulos $\phi_1, \ldots, \phi_N$ são regularmente espaçados na banda de parada.

Limitação da Potência do Ruído Térmico

- Em algumas situações é importante controlar a potência do ruído térmico gerado pelas antenas.
- Verifica-se que esta potência é proporcional à norma Euclidiana do vetor complexo z

$$\Gamma \propto \sqrt{\sum_{i=1}^{n} |z_i|^2} \tag{60}$$

Dilema entre Atenuação do Lóbulo Lateral e a Potência do Ruído Térmico

Um problema de otimização típico envolveria

- uma restrição de normalização, que determina um valor unitário no diagrama de magnitude em uma direção específica :

$$\mathfrak{Re}\left(D_z\left(\phi_{alvo}\right) \geq 1\right) \tag{61}$$

- uma restrição quanto ao nível de atenuação do lóbulo lateral :

$$\left|D_z\left(\phi_z\right)\right| \leq \delta, i = 1, \ldots, N \tag{62}$$

- uma restrição na potência do ruído térmico:

$$\|z\|_2 \leq \gamma \tag{63}$$

Conteúdo

Formulação do problema de otimização

- Uma curva de trade-off típica pode ser obtida comparando-se o nível de ruído térmico γ para um dado valor de atenuação do lóbulo lateral δ.
- Cada ponto da curva (δ, γ) pode ser obtido solucionando-se o problema de otimização

$$\begin{array}{ll} \underset{z}{\text{minimize}} & \|z\|_2 \\ \text{subject to} & \mathfrak{Re}(D_z(\phi_{alvo})) \geq 1 \\ & |D_z(\phi_i)| \leq \delta, i = 1, \ldots, N \end{array}$$

Projeto de Arranjo de Antenas usando o Método dos Mínimos Quadrados

- A ideia básica para a resolução deste problema consiste em penalizar as restrições, isto é, definir um parâmetro de trade-off, μ e se reescrever o problema como

$$\begin{array}{ll} \underset{z}{\text{minimize}} & \|z\|_2^2 + \mu \sum_{i=1}^{n} |D_z(\phi_i)|^2 \\ \text{subject to} & \Re\mathfrak{e}(D_z(\phi_{alvo})) \geq 1 \end{array}$$

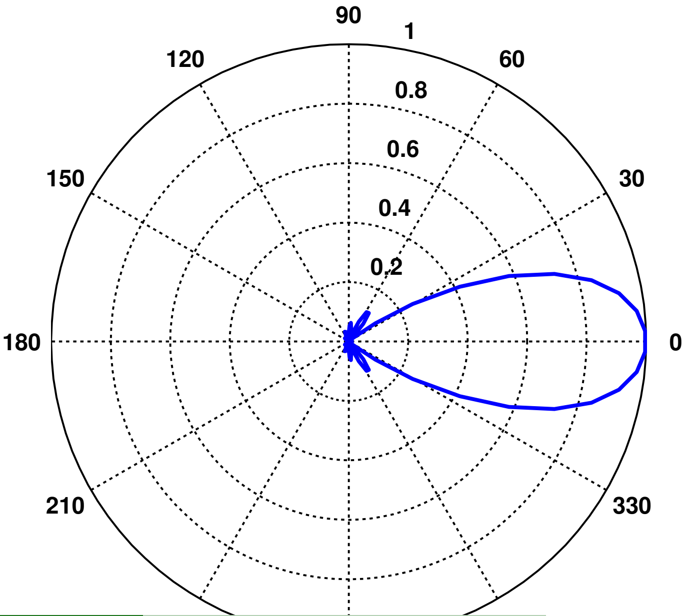
- Lembrando que a função é linear em z, pode-se afirmar que se trata de um problema de mínimos quadrados com restrições lineares.

Implementação no CVX

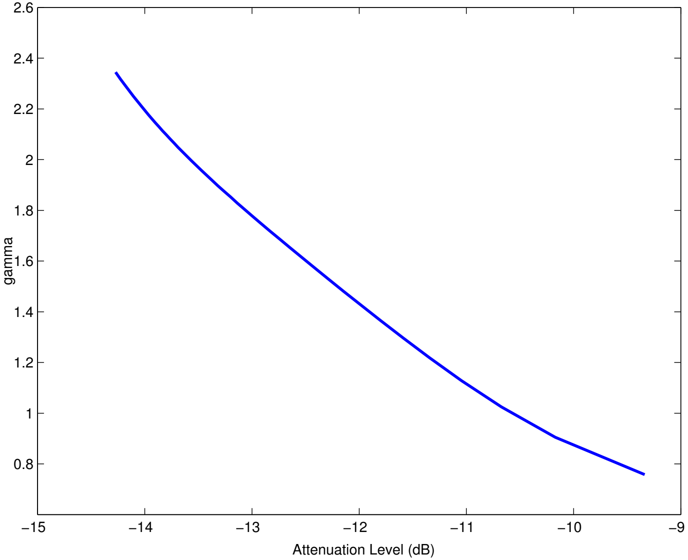
Tabela: Parâmetros de Simulação

Parâmetro	**Valor**
Número de antenas	10
Comprimento de onda	8
Ângulo alvo	0
Banda de Passagem	$[-\pi/6, \pi/6]$
Parâmetro de Discretização	90
Parâmetro de Trade-off	0,5

Diagrama de Irradiação



Curva (δ, γ)



Projeto de Arranjo de Antenas usando SOCP

O método Second Order Cone Programming (SOCP) permite realizar o projeto de arranjo de antenas de dois modos diferentes

- Minimização do ruído térmico para um dado nível dos lóbulos laterais
- Minimização da atenuação do lóbulo lateral

Minimização do ruído térmico para um dado nível dos lóbulos laterais

O problema de minimização da potência do ruído térmico submetido a uma restrição do nível dos lóbulos laterais pode ser escrito como

$$
\begin{array}{ll}
\underset{z \in \mathbf{C}^n,\, \delta}{\text{minimize}} & \sum_{i=1}^{n} \|z_i\|_2 \\
\text{subject to} & \Re\mathfrak{e}(D_z(\phi_{alvo})) \geq 1 \\
 & |D_z(\phi_i)| \leq \delta, i = 1, \ldots, m
\end{array}
$$

As N restrições representam cones de segunda ordem em função das variáveis de decisão, uma vez que eles envolvem restrição de magnitude em um vetor complexo que depende de modo afim dessas variáveis.

Restrição de Magnitude em Vetores Complexos Afins

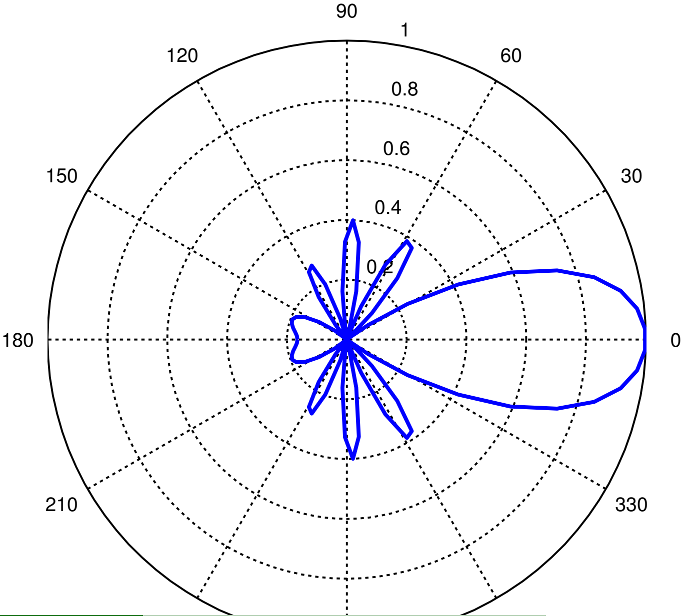
- Muitos problemas que envolvem variáveis complexas e restrições de magnitude podem ser solucionados usando SOCP.
- A ideia básica consiste em escrever a magnitude de um número complexo como uma norma euclidiana

$$|z| = \sqrt{z_R^2 + z_I^2} = \left\| \begin{matrix} z_R \\ z_I \end{matrix} \right\|_2$$

- Por exemplo, considere o número complexo f(x), onde $x \in \mathbf{R}^n$ é uma variável de projeto e que a função $f : \mathbf{R}^n \to \mathbf{C}$ é afim. Os valores dessa função podem ser escritos como $f(x) = (a_R^T x + b_R) + j(a_I^T x + b_I)$.
- Uma restrição de magnitude da forma $|f(x)| \leq t$ pode ser escrita como um cone de segunda ordem em (x,t)

$$\left\| \begin{matrix} a_R^T x + b_R \\ a_I^T x + b_I \end{matrix} \right\|_2 \leq t$$

Resultado

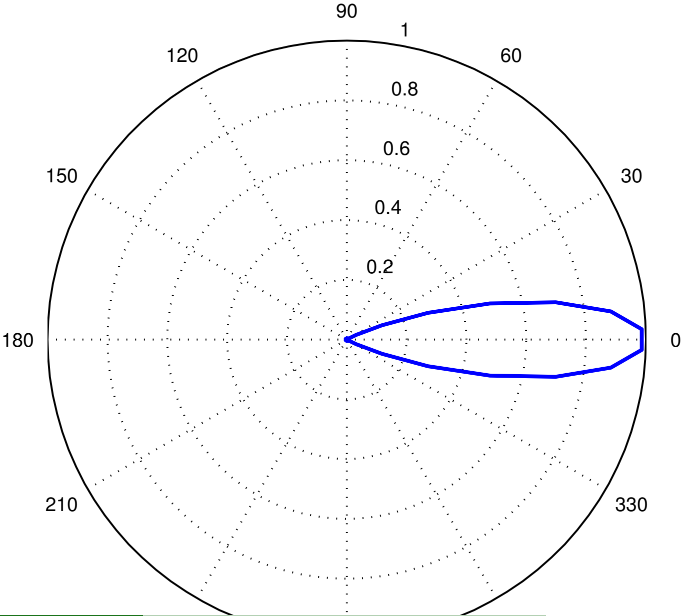


Minimização da atenuação do lóbulo lateral

Este problema tem como objetivo minimizar o nível de atenuação dos lóbulos laterais, δ, dado o requerimento de normalização $\mathfrak{Re}(D_z(0)) \geq 1$.
Esta análise pode ser realizada a partir do seguinte formulaąão de um SOCP

$$\begin{array}{ll} \underset{z \in \mathbf{C}^n, \delta}{\text{minimize}} & \delta \\ \text{subject to} & \mathfrak{Re}(D_z(0)) \geq 1 \\ & |D_z(\phi_i)| \leq \delta, i = 1, \ldots, m \end{array}$$

Resutado



Revisão de ferramentas matemática

Conteúdo

Conteúdo

Campo de um espaço vetorial [?]

- O campo escalar subjacente a um espaço vetorial é o conjunto de escalares onde os elementos do vetor são definidos
  - Normalmente o campo dos números reais $\mathbb{R}$ ou complexos $\mathbb{C}$
  - Alternativamente poderia ser o dos racionais, inteiros módulo algum número primo, etc.
- Um campo precisa ser fechado sob duas operações binárias (i.e., que recebem dois operandos)
  - Adição
  - Multiplicação
- As operações precisam ser associativas, comutativas e possuir um elemento neutro
- Elementos inversos precisam existir para todos elemento sob a adição e multiplicação, exceto para a identidade sobre a multiplicação
- Multiplicação precisa ser distributiva sobre adição

Vetores e (sub)espaços vetoriais

- O **espaço vetorial linear** $\mathbb{R}^n$ é o conjunto de todos os vetores $\boldsymbol{x}$ de dimensão $n \times 1$ juntamente com as operações de adição de vetores e multiplicação por um escalar [**?**, cap. 2]
- Um **vetor** $\boldsymbol{x} \in \mathbb{R}^n$ é normalmente representado como

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \text{ com } \boldsymbol{x}^\mathrm{T} = \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix} \tag{64}$$

- Embora a maior parte dos conceitos se estenda facilmente para vetores complexos, consideraremos apenas vetores reais (exceto se explicitamente mencionado)

Vetores e (sub)espaços vetoriais

- Há oito propriedades que precisam ser satisfeitas por um espaço vetorial[1] [**?**, Ex. 2.1.5]

$$\boldsymbol{x} + \boldsymbol{y} = \boldsymbol{y} + \boldsymbol{x} \qquad \text{(Comutatividade da adição)} \tag{65a}$$

$$\boldsymbol{x} + (\boldsymbol{y} + \boldsymbol{z}) = (\boldsymbol{x} + \boldsymbol{y}) + \boldsymbol{z} \qquad \text{(Associatividade da adição)} \tag{65b}$$

$$\exists \mathbf{0} \Rightarrow \boldsymbol{x} + \mathbf{0} = \boldsymbol{x} \qquad \text{(Elemento neutro da adição)} \tag{65c}$$

$$\exists - \boldsymbol{x} \Rightarrow \boldsymbol{x} + (-\boldsymbol{x}) = \mathbf{0} \qquad \text{(Elemento simétrico)} \tag{65d}$$

$$1 \cdot \boldsymbol{x} = \boldsymbol{x} \qquad \text{(Elemento neutro da multiplicação)} \tag{65e}$$

$$(\alpha_1 \cdot \alpha_2) \cdot \boldsymbol{x} = \alpha_1 \cdot (\alpha_2 \cdot \boldsymbol{x}) \qquad \text{(Associatividade da multiplicação)} \tag{65f}$$

$$\alpha_1 \cdot (\boldsymbol{x} + \boldsymbol{y}) = \alpha_1 \cdot \boldsymbol{x} + \alpha_1 \cdot \boldsymbol{y} \qquad \text{(Distributividade)} \tag{65g}$$

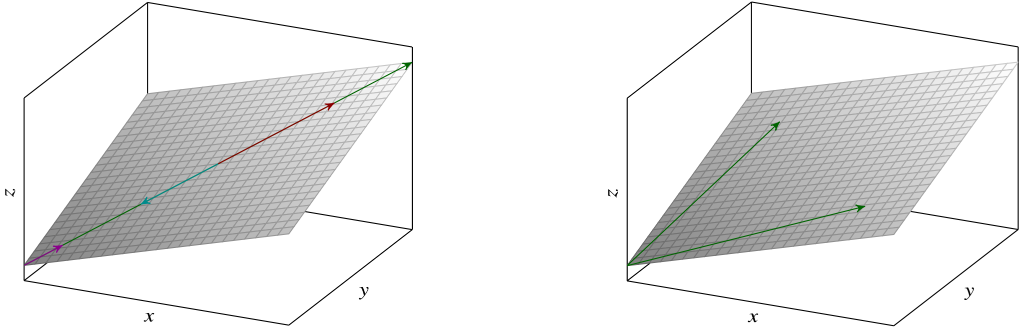
$$(\alpha_1 + \alpha_2) \cdot \boldsymbol{x} = \alpha_1 \cdot \boldsymbol{x} + \alpha_2 \cdot \boldsymbol{x} \qquad \text{(Distributividade)} \tag{65h}$$

- Um **subespaço vetorial linear** $\mathbb{S} \subset \mathbb{R}^n$ é um conjunto não-vazio que satisfaz:
  - $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{S} \Rightarrow \boldsymbol{x} + \boldsymbol{y} \in \mathbb{S}$
  - $\forall \boldsymbol{x} \in \mathbb{S}$ e $\alpha \in \mathbb{R} \Rightarrow \alpha \boldsymbol{x} \in \mathbb{S}$
- Um subespaço é portanto um subconjunto fechado sob as operações de adição e multiplicação por escalar

[1] Relações similares podem ser definidas para números complexos em $\mathbb{C}^n$

Vetores e (sub)espaços vetoriais

- Considerando o espaço vetorial $\mathbb{R}^3$, alguns exemplos simples de subespaços vetoriais $\mathbb{S} \subset \mathbb{R}^3$ são retas e planos
- Um vetor $\boldsymbol{v} \in \mathbb{R}^3$ age como suporte de uma reta que representa um subespaço $\mathbb{S}_1 \subset \mathbb{R}^3$
  - Quaisquer vetores $\boldsymbol{v}_1, \boldsymbol{v}_2 \in \mathbb{S}_1$ encontram-se sobre a reta e, portanto, podem ser escritos como $\boldsymbol{v}_1 = \alpha\boldsymbol{v}$ e $\boldsymbol{v}_2 = \beta\boldsymbol{v}$, para algum $\alpha, \beta \in \mathbb{R}$
  - Além disso, $\boldsymbol{v}_3 = \boldsymbol{v}_1 + \boldsymbol{v}_2 = (\alpha + \beta)\boldsymbol{v}$ está também sobre a reta



Norma de vetores

- A **norma** de um vetor é uma generalização do conceito de comprimento ou magnitude de um vetor
- A **norma** $\|\boldsymbol{x}\|$ de $\boldsymbol{x} \in \mathbb{R}^n$ é uma função $f : \mathbb{R}^n \rightarrow \mathbb{R}$ que satisfaz [**?**, pág. 46]:

$$\|\boldsymbol{x}\| \geq 0 \quad \text{(Não-negatividade)} \tag{66a}$$

$$\|\boldsymbol{x}\| = 0 \Leftrightarrow \boldsymbol{x} = \boldsymbol{0} \quad \text{(Elemento neutro)} \tag{66b}$$

$$\|\alpha \boldsymbol{x}\| = |\alpha| \|\boldsymbol{x}\|, \forall \alpha \in \mathbb{R} \quad \text{(Escalabilidade)} \tag{66c}$$

$$\|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|, \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n \quad \text{(Desigualdade triangular)} \tag{66d}$$

- Uma família de normas (norma-p) que atende às propriedades é dada por

$$\|\boldsymbol{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}}, \quad p \in \mathbb{N}_+ \tag{67}$$

- Um caso caso de interesse é **norma-1** (ou norma ℓ_1) onde $p = 1$ em (67)
- Outro caso de interesse é **norma euclidiana** (ou norma-2 ou ainda norma ℓ_2) onde $p = 2$ em (67)

Norma euclidiana e produto interno [?, cap. 1]

- A norma euclidiana atende ainda as seguintes propriedades:
  - $|\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}| \leq \|\boldsymbol{x}\|_2\|\boldsymbol{y}\|_2$ (Desigualdade de Cauchy-Schwartz)
  - $\max\limits_{1\leq i\leq n} |x_i| \leq \|\boldsymbol{x}\|_2 \leq \sqrt{n} \max\limits_{1\leq i\leq n} |x_i|$
- O **produto interno** ou **produto escalar** $\langle \boldsymbol{x},\boldsymbol{y}\rangle$ entre dois vetores $\boldsymbol{x}$ e $\boldsymbol{y} \in \mathbb{R}^n$ é escrito como

$$\langle \boldsymbol{x},\boldsymbol{y}\rangle = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} = \sum_{i=1}^{n} x_i y_i = \boldsymbol{y}^{\mathrm{T}}\boldsymbol{x} = \langle \boldsymbol{y},\boldsymbol{x}\rangle \tag{68}$$

- A norma euclidiana guarda as seguintes relações com o produto interno

$$\langle \boldsymbol{x},\boldsymbol{x}\rangle = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} = \sum_{i=1}^{n} x_i x_i = \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} |x_i|^2 = \|\boldsymbol{x}\|_2^2 \tag{69a}$$

$$\langle \boldsymbol{x},\boldsymbol{y}\rangle = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} = \|\boldsymbol{x}\|_2\|\boldsymbol{y}\|_2 \cos\theta_{\angle_{\boldsymbol{y}}^{\boldsymbol{x}}} \tag{69b}$$

- Pare vetores pertencentes a $\mathbb{C}$, $(\cdot)^{\mathrm{T}}$ é substituído por $(\cdot)^{\mathrm{H}}$ que representa o conjugado-transposto de um vetor com a conjugação denotada por $(\cdot)^*$

Conteúdo

Independência linear [?, cap. 3]

- Os vetores $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m \in \mathbb{R}^n$ são ditos **linearmente independentes** (L.I.) se e somente se, para $\alpha_1, \alpha_2, \ldots, \alpha_m \in \mathbb{R}$,

$$\alpha_1 \boldsymbol{x}_1 + \alpha_2 \boldsymbol{x}_2 + \ldots + \alpha_m \boldsymbol{x}_m = \mathbf{0} \Leftrightarrow \alpha_1 = \alpha_2 = \ldots = \alpha_m = 0, \tag{70}$$

  caso contrário $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m$ são ditos **linearmente dependentes** (L.D.)
- Se $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m$ são L.D., então existe pelo menos um $\alpha_i \neq 0$, tal que é possível escrever

$$\boldsymbol{x}_i = -\frac{1}{\alpha_i}\left[\alpha_1 \boldsymbol{x}_1 + \alpha_2 \boldsymbol{x}_2 + \ldots + \alpha_{i-1} \boldsymbol{x}_{i-1} + \alpha_{i+1} \boldsymbol{x}_{i+1} + \ldots + \alpha_m \boldsymbol{x}_m\right] \tag{71}$$

- A **dimensão** de um (sub)espaço vetorial linear é dado pelo máximo número de vetores L.I. neste (sub)espaço

Bases, representações e ortonormalização

- Um conjunto de n vetores L.I. pertencentes a um (sub)espaço vetorial é chamado uma **base** para esse (sub)espaço
- Dada uma base para um (sub)espaço vetorial, todo vetor desse subespaço pode ser escrito como uma combinação linear única dos vetores que formam a base
- Sejam $\{\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_n\} \in \mathbb{R}^n$ um conjunto de vetores L.I. que formam uma base para $\mathbb{R}^n$
- Todo vetor $\boldsymbol{x} \in \mathbb{R}^n$ pode ser representado como

$$\boldsymbol{x} = \alpha_1\boldsymbol{b}_1 + \alpha_2\boldsymbol{b}_2 + \ldots + \alpha_n\boldsymbol{b}_n = \underbrace{\begin{bmatrix}\boldsymbol{b}_1 & \boldsymbol{b}_2 & \ldots & \boldsymbol{b}_n\end{bmatrix}}_{\boldsymbol{B}} \underbrace{\begin{bmatrix}\alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n\end{bmatrix}}_{\boldsymbol{\alpha}} = \boldsymbol{B}\boldsymbol{\alpha} \tag{72}$$

Bases, representações e ortonormalização

- O vetor $\boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 & \alpha_2 & \dots & \alpha_n \end{bmatrix}^{\mathrm{T}}$ é chamado de **representação** do vetor $\boldsymbol{x}$ na base $\{\boldsymbol{b}_1, \boldsymbol{b}_2, \dots, \boldsymbol{b}_n\}$ (ou ainda na base $\boldsymbol{B}$)
- A cada $\mathbb{R}^n$ é associada uma **base canônica** $\{\boldsymbol{i}_1, \boldsymbol{i}_2, \dots, \boldsymbol{i}_n\}$ onde $\boldsymbol{i}_i$ é a i-ésima coluna de uma matriz identidade $\boldsymbol{I}_n$ de dimensão $n \times n$
- Note que, na base canônica, um vetor $\boldsymbol{x} \in \mathbb{R}^n$ é representado como

$$\boldsymbol{x} = x_1\boldsymbol{i}_1 + x_2\boldsymbol{i}_2 + \dots + x_n\boldsymbol{i}_n = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \tag{73}$$

Vetores normalizados, ortogonais e ortonormais

- Um vetor $\boldsymbol{x}$ é dito **normalizado** se sua norma euclidiana $\|\boldsymbol{x}\|_2 = 1$
- Um vetor unitário $\boldsymbol{u}$ na direção de um vetor $\boldsymbol{x}$ é obtido normalizando o vetor $\boldsymbol{x}$ como

$$\boldsymbol{u}_{\boldsymbol{x}} = \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|_2} \tag{74}$$

- Dois vetores $\boldsymbol{x}_1$ e $\boldsymbol{x}_2$ são ditos **ortogonais** se o produto interno entre eles é nulo, i.e., se $\boldsymbol{x}_1^{\mathrm{T}}\boldsymbol{x}_2 = \boldsymbol{x}_2^{\mathrm{T}}\boldsymbol{x}_1 = 0$
- Um conjunto de vetores $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ é dito **ortonormal** se

$$\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \tag{75}$$

- Um conjunto de vetores L.I. $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\} \in \mathbb{R}^n$ que atende a (75) formam uma **base ortonormal**

Componente ortogonal e paralela

- O comprimento $|P_{\|}(\boldsymbol{x}_i, \boldsymbol{x}_j)|$ da componente paralela $P_{\|}(\boldsymbol{x}_i, \boldsymbol{x}_j)$ de um vetor $\boldsymbol{x}_i$ na direção de um vetor $\boldsymbol{x}_j$ é dado pelo produto escalar do primeiro vetor com o vetor unitário na direção do segundo vetor, i.e.,

$$|P_{\|}(\boldsymbol{x}_i, \boldsymbol{x}_j)| = \frac{\boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{x}_j}{\|\boldsymbol{x}_j\|_2} = \boldsymbol{x}_i^{\mathrm{T}} \frac{\boldsymbol{x}_j}{\|\boldsymbol{x}_j\|_2} = \boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{u}_j, \text{ onde } \boldsymbol{u}_j = \frac{\boldsymbol{x}_j}{\|\boldsymbol{x}_j\|_2} \tag{76}$$

- Usando (76), a **componente paralela** $P_{\|}(\boldsymbol{x}_i, \boldsymbol{x}_j)$ e a **componente ortogonal** $P_{\perp}(\boldsymbol{x}_i, \boldsymbol{x}_j)$ de um vetor $\boldsymbol{x}_i$ em relação a um vetor $\boldsymbol{x}_j$ são dadas respectivamente por

$$P_{\|}(\boldsymbol{x}_i, \boldsymbol{x}_j) = (\boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{u}_j)\boldsymbol{u}_j \tag{77a}$$

$$P_{\perp}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \boldsymbol{x}_i - P_{\|}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \boldsymbol{x}_i - (\boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{u}_j)\boldsymbol{u}_j \tag{77b}$$

Processo de ortonormalização de Gram-Schmidt

- O **processo de ortonormalização de Gram-Schmidt** permite construir uma base ortonormal a partir de um conjunto de vetores L.I.
- Segundo esse processo, um conjunto $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n\} \in \mathbb{R}^n$ de vetores L.I. produz a base ortonormal $\{\boldsymbol{b}_1, \boldsymbol{b}_2, \dots, \boldsymbol{b}_n\}$ como

$$
\begin{aligned}
\boldsymbol{v}_1 &= \boldsymbol{x}_1, & \boldsymbol{b}_1 &= \boldsymbol{v}_1/\|\boldsymbol{v}_1\|_2 \\
\boldsymbol{v}_2 &= \boldsymbol{x}_2 - (\boldsymbol{x}_2^{\mathrm{T}}\boldsymbol{b}_1)\boldsymbol{b}_1, & \boldsymbol{b}_2 &= \boldsymbol{v}_2/\|\boldsymbol{v}_2\|_2 \\
\boldsymbol{v}_3 &= \boldsymbol{x}_3 - (\boldsymbol{x}_3^{\mathrm{T}}\boldsymbol{b}_1)\boldsymbol{b}_1 - (\boldsymbol{x}_3^{\mathrm{T}}\boldsymbol{b}_2)\boldsymbol{b}_2, & \boldsymbol{b}_3 &= \boldsymbol{v}_3/\|\boldsymbol{v}_3\|_2 \\
\boldsymbol{v}_4 &= \boldsymbol{x}_4 - (\boldsymbol{x}_4^{\mathrm{T}}\boldsymbol{b}_1)\boldsymbol{b}_1 - (\boldsymbol{x}_4^{\mathrm{T}}\boldsymbol{b}_2)\boldsymbol{b}_2 - (\boldsymbol{x}_4^{\mathrm{T}}\boldsymbol{b}_3)\boldsymbol{b}_3, & \boldsymbol{b}_4 &= \boldsymbol{v}_4/\|\boldsymbol{v}_4\|_2 \\
&\dots \\
\boldsymbol{v}_n &= \boldsymbol{x}_n - \sum_{i=1}^{n-1}(\boldsymbol{x}_n^{\mathrm{T}}\boldsymbol{b}_i)\boldsymbol{b}_i, & \boldsymbol{b}_n &= \boldsymbol{v}_n/\|\boldsymbol{v}_n\|_2
\end{aligned}
$$

Conteúdo

Matrizes

- Uma matriz $\boldsymbol{A}$ de dimensão $m \times n$ pertencente ao $\mathbb{R}^{m \times n}$ e sua transposta $\boldsymbol{A}^{\mathrm{T}} \in \mathbb{R}^{n \times m}$ são denotadas por

$$\begin{aligned}\boldsymbol{A} &= \begin{bmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \dots & a_{m,n} \end{bmatrix} = \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \dots & \boldsymbol{a}_n \end{bmatrix} \text{ e} \\ \boldsymbol{A}^{\mathrm{T}} &= \begin{bmatrix} a_{1,1} & a_{2,1} & \dots & a_{m,1} \\ a_{1,2} & a_{2,2} & \dots & a_{m,2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1,n} & a_{2,n} & \dots & a_{m,n} \end{bmatrix} = \begin{bmatrix} \boldsymbol{a}_1^{\mathrm{T}} \\ \boldsymbol{a}_2^{\mathrm{T}} \\ \vdots \\ \boldsymbol{a}_n^{\mathrm{T}} \end{bmatrix},\end{aligned} \tag{78}$$

  onde $\boldsymbol{a}_i$, $1 \leq i \leq n$ denota a i-ésima coluna de $\boldsymbol{A}$
- Assume-se aqui conhecimento sobre as operações de adição de matrizes, multiplicação por escalar, e multiplicação de matrizes

Algumas operações com matrizes

- Para uma matriz $\boldsymbol{A} \in \mathbb{C}^{m \times n}$, a **matrix conjugada** $\boldsymbol{A}^*$ é obtida conjugando cada elemento $a_{i,j}$ de $\boldsymbol{A}$
- De forma similar, a **matriz conjugada-transposta** $\boldsymbol{A}^{\mathrm{H}}$ de $\boldsymbol{A}$ é obtida conjugando a matriz transposta $\boldsymbol{A}^{\mathrm{T}}$ de $\boldsymbol{A}$, i.e., $\boldsymbol{A}^{\mathrm{H}} = \left(\boldsymbol{A}^{\mathrm{T}}\right)^*$
- A **matriz inversa** $\boldsymbol{A}^{-1}$ de uma matriz $\boldsymbol{A}$ de dimensão $n \times n$ é a matriz que satisfaz $\boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}_n$
- Algumas propriedades relevantes envolvendo matrizes são [**?**]:

$$(\boldsymbol{A}\boldsymbol{B})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1} \tag{79a}$$

$$(\boldsymbol{A}\boldsymbol{B})^{\mathrm{T}} = \boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}} \tag{79b}$$

$$(\boldsymbol{A}\boldsymbol{B})^{\mathrm{H}} = \boldsymbol{B}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}} \tag{79c}$$

$$\left(\boldsymbol{A}^{\mathrm{T}}\right)^{-1} = \left(\boldsymbol{A}^{-1}\right)^{\mathrm{T}} \tag{79d}$$

$$\left(\boldsymbol{A}^{\mathrm{H}}\right)^{-1} = \left(\boldsymbol{A}^{-1}\right)^{\mathrm{H}} \tag{79e}$$

$$(\boldsymbol{A}+\boldsymbol{B})^{\mathrm{T}} = \boldsymbol{A}^{\mathrm{T}}+\boldsymbol{B}^{\mathrm{T}} \tag{79f}$$

$$(\boldsymbol{A}+\boldsymbol{B})^{\mathrm{H}} = \boldsymbol{A}^{\mathrm{H}}+\boldsymbol{B}^{\mathrm{H}} \tag{79g}$$

Traço de uma matriz

- O **traço** $\operatorname{tr}(\boldsymbol{A})$ de uma matriz $\boldsymbol{A}$ de dimensão $n \times n$ é definido como a soma dos elementos da diagonal da mesma, i.e.,

$$\operatorname{tr}(\boldsymbol{A}) = \sum_{i=1}^{n} a_{i,i} \tag{80}$$

- Entre as propriedades do $\operatorname{tr}(\cdot)$ temos [?]:

$$\operatorname{tr}(\boldsymbol{A}) = \operatorname{tr}\left(\boldsymbol{A}^{\mathrm{T}}\right) \tag{81a}$$

$$\operatorname{tr}(\boldsymbol{A}\boldsymbol{B}) = \operatorname{tr}(\boldsymbol{B}\boldsymbol{A}) \tag{81b}$$

$$\operatorname{tr}(\boldsymbol{A}\boldsymbol{B}\boldsymbol{C}) = \operatorname{tr}(\boldsymbol{C}\boldsymbol{A}\boldsymbol{B}) = \operatorname{tr}(\boldsymbol{B}\boldsymbol{C}\boldsymbol{A}) \tag{81c}$$

$$\operatorname{tr}(\alpha\boldsymbol{A} + \beta\boldsymbol{B}) = \alpha \operatorname{tr}(\boldsymbol{A}) + \beta \operatorname{tr}(\boldsymbol{B}) \tag{81d}$$

Determinante uma matriz

- O **determinante** $\det(\boldsymbol{A})$ de uma matriz $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ pode ser escrito através da expansão de Laplace sobre uma linha ou uma coluna da matriz como

$$\det(\boldsymbol{A}) = \sum_{i=1}^{n} a_{i,j} c_{i,j} = \sum_{j=1}^{n} a_{i,j} c_{i,j}, \tag{82}$$

onde $c_{i,j}$ é o **cofator** associado a $a_{i,j}$, o qual é dado por

$$c_{i,j} = (-1)^{i+j} \det\left(\tilde{\boldsymbol{A}}_{i,j}\right), \tag{83}$$

onde a matriz $\tilde{\boldsymbol{A}}_{i,j} \in \mathbb{R}^{(n-1)\times(n-1)}$ é formada a partir de $\boldsymbol{A}$ pela exclusão de sua i-ésima linha e j-ésima coluna

- A **matriz adjunta** $\operatorname{adj}(\boldsymbol{A})$ de $\boldsymbol{A}$ é a matriz transposta dos cofatores de $\boldsymbol{A}$
- Uma matriz $\boldsymbol{A}$ é dita **não-singular** e possui inversa $\boldsymbol{A}^{-1}$ se $\det(\boldsymbol{A}) \neq 0$
- A matriz inversa pode ser calculada como

$$\boldsymbol{A}^{-1} = \frac{1}{\det(\boldsymbol{A})} \operatorname{adj}(\boldsymbol{A}) \tag{84}$$

Exemplo para inversa de uma matriz

- Se a matriz $\boldsymbol{A}$ é dada por

$$\boldsymbol{A} = \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix}$$

  temos que

$$\begin{aligned} \boldsymbol{A}^{-1} &= \frac{1}{\det(\boldsymbol{A})} \operatorname{adj}(\boldsymbol{A}) = \frac{1}{a_{1,1}a_{2,2} - a_{1,2}a_{2,1}} \begin{bmatrix} c_{1,1} & c_{2,1} \\ c_{1,2} & c_{2,2} \end{bmatrix} \\ &= \frac{1}{a_{1,1}a_{2,2} - a_{1,2}a_{2,1}} \begin{bmatrix} a_{2,2} & -a_{1,2} \\ -a_{2,1} & a_{1,1} \end{bmatrix} \end{aligned}$$

Conteúdo

Transformação linear

- Um função $f(\cdot) : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ é um **operador linear** se e somente se

$$f(\alpha_1 \boldsymbol{x}_1 + \alpha_2 \boldsymbol{x}_2) = \alpha_1 f(\boldsymbol{x}_1) + \alpha_2 f(\boldsymbol{x}_2), \forall \boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}^n, \alpha_1, \alpha_2 \in \mathbb{R} \tag{85}$$

- Sejam $\mathbb{X} \subset \mathbb{R}^n$ e $\mathbb{Y} \subset \mathbb{R}^m$ são dois espaços vetoriais e sejam:
  - $L(\cdot) : \mathbb{X} \rightarrow \mathbb{Y}$ um operador linear
  - $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ um conjunto de n vetores L.I. em $\mathbb{X}$
  - $\boldsymbol{y}_i = L(\boldsymbol{x}_i), i = 1, 2, \ldots, n$ um conjunto de n vetores em $\mathbb{Y}$
- Então, podemos afirmar que:
  - O operador linear $L(\cdot)$ é unicamente determinado pelos n pares $(\boldsymbol{x}_i, \boldsymbol{y}_i), i = 1, 2, \ldots, n$
  - Se $\{\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_n\}$ e $\{\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_m\}$ são bases para $\mathbb{X}$ e $\mathbb{Y}$, respectivamente, então o operador linear $L(\cdot)$ pode ser representado por uma matriz $\boldsymbol{T}$ de dimensão $m \times n$
  - A i-ésima coluna da matriz $\boldsymbol{T}$ é a representação de $\boldsymbol{y}_i$ na base $\boldsymbol{P} = \begin{bmatrix} \boldsymbol{p}_1 & \boldsymbol{p}_2 & \ldots & \boldsymbol{p}_n \end{bmatrix}$

Transformações lineares: mudança de base

- Exemplos de transformações lineares comuns no $\mathbb{R}^2$
  - **Mudança de escala**: $\boldsymbol{T} = \alpha \boldsymbol{I}$ (mesma escala nos eixos x e y) ou $\boldsymbol{T} = \begin{bmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{bmatrix}$ (escalas diferentes para os eixos x e y)
  - **Rotação**: $\boldsymbol{T} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$
- Outra transformação linear de interesse refere-se à mudança de base em um espaço vetorial
- Se $\boldsymbol{\alpha}$ e $\boldsymbol{\beta}$ são respectivamente as representações de $\boldsymbol{x} \in \mathbb{X} \subset \mathbb{R}^n$ nas bases $\boldsymbol{A} = \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \dots & \boldsymbol{a}_n \end{bmatrix}$ e $\boldsymbol{B} = \begin{bmatrix} \boldsymbol{b}_1 & \boldsymbol{b}_2 & \dots & \boldsymbol{b}_n \end{bmatrix}$, temos

$$\boldsymbol{x} = \underbrace{\begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \dots & \boldsymbol{a}_n \end{bmatrix}}_{\boldsymbol{A}} \underbrace{\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix}}_{\boldsymbol{\alpha}} = \underbrace{\begin{bmatrix} \boldsymbol{b}_1 & \boldsymbol{b}_2 & \dots & \boldsymbol{b}_n \end{bmatrix}}_{\boldsymbol{B}} \underbrace{\begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix}}_{\boldsymbol{\beta}} \tag{86}$$

Transformações lineares: mudança de base

- Se $\boldsymbol{p}_i$ é a representação de $\boldsymbol{a}_i$ na base $\boldsymbol{B}$, então temos que

$$\boldsymbol{a}_i = \begin{bmatrix} \boldsymbol{b}_1 & \boldsymbol{b}_2 & \dots & \boldsymbol{b}_n \end{bmatrix} \boldsymbol{p}_i = \boldsymbol{B}\boldsymbol{p}_i, \quad i = 1, 2, \dots, n \tag{87}$$

- Juntando as n equações acima em expressão matricial obtemos

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{b}_1 & \boldsymbol{b}_2 & \dots & \boldsymbol{b}_n \end{bmatrix} \begin{bmatrix} \boldsymbol{p}_1 & \boldsymbol{p}_2 & \dots & \boldsymbol{p}_n \end{bmatrix} = \boldsymbol{B}\boldsymbol{P} \tag{88}$$

- Substituindo (88) em (86), obtemos

$$\boldsymbol{A}\boldsymbol{\alpha} = \boldsymbol{B}\boldsymbol{\beta} \Rightarrow \boldsymbol{B}\boldsymbol{P}\boldsymbol{\alpha} = \boldsymbol{B}\boldsymbol{\beta} \Rightarrow \boldsymbol{B}^{-1}\boldsymbol{B}\boldsymbol{P}\boldsymbol{\alpha} = \boldsymbol{B}^{-1}\boldsymbol{B}\boldsymbol{\beta} \Rightarrow \boxed{\boldsymbol{P}\boldsymbol{\alpha} = \boldsymbol{\beta}}$$

- Logo, temos que $\boldsymbol{P}$ é a transformação linear que leva a representação $\boldsymbol{\alpha}$ de $\boldsymbol{x}$ na base $\boldsymbol{A}$ para sua representação $\boldsymbol{\beta}$ na base $\boldsymbol{B}$

Transformações lineares: mudança de base

- Um desenvolvimento análogo ao anterior provê a transformação linear $\boldsymbol{Q}$ que leva a representação $\boldsymbol{\beta}$ de $\boldsymbol{x}$ na base $\boldsymbol{B}$ para sua representação $\boldsymbol{\alpha}$ na base $\boldsymbol{A}$
- Se $\boldsymbol{P}$ e $\boldsymbol{Q}$ são conhecidas, então para um vetor $\boldsymbol{x}$ qualquer com representações $\boldsymbol{\alpha}$ na base $\boldsymbol{A}$ e $\boldsymbol{\beta}$ na base $\boldsymbol{B}$ temos

$$\boldsymbol{\beta} = \boldsymbol{P}\boldsymbol{\alpha} \text{ e } \boldsymbol{\alpha} = \boldsymbol{Q}\boldsymbol{\beta} \Rightarrow \boldsymbol{\beta} = \boldsymbol{P}\boldsymbol{Q}\boldsymbol{\beta} \Rightarrow \boldsymbol{Q}\boldsymbol{P} = \boldsymbol{I} \Rightarrow \boxed{\boldsymbol{P}^{-1} = \boldsymbol{Q}}$$

Transformações lineares: transformações de similaridade

- Considere que:
  - $\boldsymbol{U} = \begin{bmatrix}\boldsymbol{u}_1 & \boldsymbol{u}_2 & \dots & \boldsymbol{u}_n\end{bmatrix}$ e $\bar{\boldsymbol{U}} = \left[\bar{\boldsymbol{u}}_1, \bar{\boldsymbol{u}}_2, \dots, \bar{\boldsymbol{u}}_n\right]$ são duas bases para um subespaço vetorial $\mathbb{X}$
  - $L(\cdot)$ é um operador linear tal que $\boldsymbol{y}_i = L(\boldsymbol{u}_i)$ e $\bar{\boldsymbol{y}}_i = L(\bar{\boldsymbol{u}}_i), i = 1, 2, \dots, n$
  - $\boldsymbol{V} = \left[\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_n\right]$ e $\bar{\boldsymbol{V}} = \left[\bar{\boldsymbol{v}}_1, \bar{\boldsymbol{v}}_2, \dots, \bar{\boldsymbol{v}}_n\right]$ são duas uma base para o espaço gerado por $\boldsymbol{y}_i$
  - $\boldsymbol{A} = \left[\boldsymbol{a}_1, \boldsymbol{a}_2, \dots, \boldsymbol{a}_n\right]$ e $\bar{\boldsymbol{A}} = \left[\bar{\boldsymbol{a}}_1, \bar{\boldsymbol{a}}_2, \dots, \bar{\boldsymbol{a}}_n\right]$ são as representação de $\boldsymbol{y}_i$ e $\bar{\boldsymbol{y}}_i$ nas bases $\boldsymbol{V}$ e $\bar{\boldsymbol{V}}$, respectivamente
- Como $\boldsymbol{U}$ e $\bar{\boldsymbol{U}}$ são base para $\mathbb{X}$, um vetor $\boldsymbol{x} \in \mathbb{X}$ pode ser escrito como combinação linear das colunas de $\boldsymbol{U}$ ou de $\bar{\boldsymbol{U}}$ e como $L(\boldsymbol{x})$ é um operador linear temos

$$L(\boldsymbol{x}) = L\left(\sum_{i=1}^{n} \alpha_i \boldsymbol{u}_i\right) = \sum_{i=1}^{n} \alpha_i L\left(\boldsymbol{u}_i\right) = \sum_{i=1}^{n} \alpha_i \boldsymbol{y}_i \tag{89a}$$

$$L(\boldsymbol{x}) = L\left(\sum_{i=1}^{n} \bar{\alpha}_i \bar{\boldsymbol{u}}_i\right) = \sum_{i=1}^{n} \bar{\alpha}_i L\left(\bar{\boldsymbol{u}}_i\right) = \sum_{i=1}^{n} \bar{\alpha}_i \bar{\boldsymbol{y}}_i \tag{89b}$$

Transformações lineares: transformações de similaridade

- Podemos escrever ainda que

$$
\begin{aligned}
L\left(\begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \dots & \boldsymbol{u}_n \end{bmatrix}\right) &= \begin{bmatrix} \boldsymbol{y}_1 & \boldsymbol{y}_2 & \dots & \boldsymbol{y}_n \end{bmatrix} \\
&= \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \dots & \boldsymbol{v}_n \end{bmatrix} \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \dots & \boldsymbol{a}_n \end{bmatrix} = \boldsymbol{V}\boldsymbol{A} \qquad (90a) \\
L\left(\begin{bmatrix} \bar{\boldsymbol{u}}_1 & \bar{\boldsymbol{u}}_2 & \dots & \bar{\boldsymbol{u}}_n \end{bmatrix}\right) &= \begin{bmatrix} \bar{\boldsymbol{y}}_1 & \bar{\boldsymbol{y}}_2 & \dots & \bar{\boldsymbol{y}}_n \end{bmatrix} \\
&= \begin{bmatrix} \bar{\boldsymbol{v}}_1 & \bar{\boldsymbol{v}}_2 & \dots & \bar{\boldsymbol{v}}_n \end{bmatrix} \begin{bmatrix} \bar{\boldsymbol{a}}_1 & \bar{\boldsymbol{a}}_2 & \dots & \bar{\boldsymbol{a}}_n \end{bmatrix} = \bar{\boldsymbol{V}}\bar{\boldsymbol{A}} \qquad (90b)
\end{aligned}
$$

- Usando (89) e (90), para um vetor $\boldsymbol{x}$ com representação α na base $\boldsymbol{U}$ e um vetor $\boldsymbol{y} = L(\boldsymbol{x})$ com representação $\boldsymbol{\beta}$ na base $\boldsymbol{V}$ temos

$$
\begin{aligned}
&\boldsymbol{y} = L(\boldsymbol{x}) \Rightarrow \boldsymbol{V}\boldsymbol{\beta} = L(\boldsymbol{U}\alpha) \Rightarrow \boldsymbol{V}\boldsymbol{\beta} = \sum_{i=1}^{n} \alpha_i L\left(\boldsymbol{u}_i\right) \Rightarrow \boldsymbol{V}\boldsymbol{\beta} = \sum_{i=1}^{n} \alpha_i \boldsymbol{y}_i \\
&\Rightarrow \boldsymbol{V}\boldsymbol{\beta} = \boldsymbol{V}\boldsymbol{A}\alpha \Rightarrow \boxed{\boldsymbol{\beta} = \boldsymbol{A}\alpha}
\end{aligned} \qquad (91)
$$

- Analogamente para um vetor $\boldsymbol{x}$ com representação $\bar{\alpha}$ na base $\bar{\boldsymbol{U}}$ e um vetor $\bar{\boldsymbol{y}} = L(\bar{\boldsymbol{x}})$ com representação $\bar{\boldsymbol{\beta}}$ na base $\bar{\boldsymbol{V}}$ obtemos

$$
\boxed{\bar{\boldsymbol{\beta}} = \bar{\boldsymbol{A}}\bar{\alpha}} \qquad (92)
$$

Transformações lineares: transformações de similaridade

- Seja $\boldsymbol{P}$ a transformação linear que leva a representação $\boldsymbol{\alpha}$ de $\boldsymbol{x}$ da base $\boldsymbol{U}$ para $\bar{\boldsymbol{\alpha}}$ na base $\bar{\boldsymbol{U}}$ e seja $\boldsymbol{Q} = \boldsymbol{P}^{-1}$ a transformação linear que traz a representação $\boldsymbol{\alpha}$ de $\boldsymbol{x}$ na base $\bar{\boldsymbol{U}}$ de volta para a base $\boldsymbol{U}$
- Como há um representação única para o operador linear $L(\boldsymbol{x})$ e $\boldsymbol{y} = L(\boldsymbol{x})$, as transformações $\boldsymbol{P}$ e $\boldsymbol{Q} = \boldsymbol{P}^{-1}$ também realizam as mudança de base de $\boldsymbol{V}$ para $\bar{\boldsymbol{V}}$ e vice-versa para um vetor $\boldsymbol{y}$

$$\begin{tikzcd} \boldsymbol{x} \arrow[r, "L"] & \boldsymbol{y} = L(\boldsymbol{x}) \\ \boldsymbol{U}\boldsymbol{\alpha} \arrow[r, "\boldsymbol{A}"] \arrow[d, bend right, "\boldsymbol{P}"'] & \boldsymbol{\beta} = \boldsymbol{A}\boldsymbol{\alpha} \arrow[d, bend right, "\boldsymbol{P}"'] \\ \bar{\boldsymbol{U}}\bar{\boldsymbol{\alpha}} \arrow[r, "\bar{\boldsymbol{A}}"] \arrow[u, bend right, "\boldsymbol{P}^{-1}"'] & \bar{\boldsymbol{\beta}} = \bar{\boldsymbol{A}}\bar{\boldsymbol{\alpha}} \arrow[u, bend right, "\boldsymbol{P}^{-1}"'] \end{tikzcd}$$

- Com o auxílio do diagrama ao lado e usando (91) e (92) temos que

$$\begin{aligned} \bar{\boldsymbol{\beta}} = \bar{\boldsymbol{A}}\bar{\boldsymbol{\alpha}} \Rightarrow \bar{\boldsymbol{A}}\bar{\boldsymbol{\alpha}} &= \boldsymbol{P}\boldsymbol{\beta} \\ \Rightarrow \bar{\boldsymbol{A}}\bar{\boldsymbol{\alpha}} &= \boldsymbol{P}\boldsymbol{A}\boldsymbol{\alpha} \\ \Rightarrow \bar{\boldsymbol{A}}\bar{\boldsymbol{\alpha}} &= \boldsymbol{P}\boldsymbol{A}\boldsymbol{P}^{-1}\bar{\boldsymbol{\alpha}} \\ \Rightarrow & \boxed{\bar{\boldsymbol{A}} = \boldsymbol{P}\boldsymbol{A}\boldsymbol{P}^{-1}} \\ \Rightarrow & \boxed{\boldsymbol{A} = \boldsymbol{P}^{-1}\bar{\boldsymbol{A}}\boldsymbol{P}} \end{aligned} \tag{93}$$

Transformações lineares: transformações de similaridade

- As transformações em $\boldsymbol{P}\bar{\boldsymbol{A}}\boldsymbol{P}^{-1}$ e $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$ mostradas em (93) são chamadas **transformações de similaridade**
- Matrizes $\boldsymbol{A}$ e $\bar{\boldsymbol{A}}$ que se relacionam conforme (93) são ditas **matrizes similares**
- Em particular, as matrizes $\boldsymbol{A}$ e $\bar{\boldsymbol{A}}$ são representações de um mesmo operador linear $L(\cdot)$ em duas bases distintas
- Todas as representações de um mesmo operador linear são similares
- Como um operador linear em uma dada base pode ser representado por uma matriz $\boldsymbol{A}$, essa matriz pode ser vista como o operador linear propriamente dito
- Sendo o operador linear $L(\cdot) : \mathbb{R}^n \to \mathbb{R}^n$ descrito pela matriz $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, temos que

$$\begin{aligned} \boldsymbol{y} = L(\boldsymbol{x}) = \boldsymbol{A}\left(\sum_{i=1}^{n} \alpha_i \boldsymbol{u}_i\right) = \sum_{i=1}^{n} \alpha_i \boldsymbol{A}\boldsymbol{u}_i = \sum_{i=1}^{n} \alpha_i \boldsymbol{y}_i \\ \Rightarrow \boldsymbol{y}_i = \boldsymbol{A}\boldsymbol{u}_i, i = 1, 2, \ldots, n \end{aligned} \tag{94}$$

Conteúdo

Sistema de equações lineares

- Considere um sistema de m equações lineares e n variáveis $x_1, x_2, \ldots, x_n$,

$$\begin{aligned} a_{1,1} \cdot x_1 + a_{1,2} \cdot x_2 + \ldots + a_{1,n} \cdot x_n &= y_1 \\ a_{2,1} \cdot x_1 + a_{2,2} \cdot x_2 + \ldots + a_{2,n} \cdot x_n &= y_2 \\ &\ldots \\ a_{m,1} \cdot x_1 + a_{m,2} \cdot x_2 + \ldots + a_{m,n} \cdot x_n &= y_m \end{aligned} \tag{95}$$

  onde os coeficientes $a_{i,j}$ e y_i são dados e $i = 1, 2, \ldots, m$ e $j = 1, 2, \ldots, n$

- Usando notação matricial, podemos definir

$$\mathbf{A} = \begin{bmatrix} a_{1,1} & a_{1,2} & \ldots & a_{1,n} \\ a_{2,1} & a_{2,2} & \ldots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \ldots & a_{m,n} \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \text{e} \quad \boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}, \tag{96}$$

  e reescrever (95) como

$$\mathbf{A}\mathbf{x} = \boldsymbol{y} \tag{97}$$

Posto de uma matriz

- O **posto** ou **rank** de uma matriz $\boldsymbol{A} = \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \dots & \boldsymbol{a}_n \end{bmatrix}$ é denotado por $\operatorname{rank}(\boldsymbol{A})$ e pode ser definido como o número de colunas $\boldsymbol{a}_i$ de $\boldsymbol{A}$ que são L.I.
- O posto de uma matriz corresponde portanto à dimensão do espaço vetorial gerado pelas colunas da matriz
- Dada uma matriz $\boldsymbol{A}$ de dimensão $m \times n$, temos que

$$\operatorname{rank}(\boldsymbol{A}) = \operatorname{rank}\left(\boldsymbol{A}^{\mathrm{T}}\right) \Rightarrow \operatorname{rank}(\boldsymbol{A}) \leq \min\{m,n\} \tag{98}$$

  e portanto os espaço vetoriais gerados pelas colunas e pelas linhas de $\boldsymbol{A}$ têm a mesma dimensão
- Para $\boldsymbol{A}$ com dimensão $m \times n$ e $\boldsymbol{B}$ com dimensão $n \times p$, a **desigualdade de Sylvester** estabelece que

$$\operatorname{rank}(\boldsymbol{A}) + \operatorname{rank}(\boldsymbol{B}) - n \leq \operatorname{rank}(\boldsymbol{A}\boldsymbol{B}) \leq \min\{\operatorname{rank}(\boldsymbol{A}), \operatorname{rank}(\boldsymbol{B})\} \tag{99}$$

Subespaços fundamentais: espaço coluna

- O sistema (95) possui solução se $\boldsymbol{y}$ pode ser escrito como combinação linear das colunas de $\boldsymbol{A}$
- Nesse caso, cada vetor $\boldsymbol{x}$ (se existir) que satisfaz (95) é uma representação de $\boldsymbol{y}$ no subespaço gerado pelas colunas de $\boldsymbol{A}$
- Considere que $\{\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_r\}$ formam uma base para o subespaço gerado pelas colunas de $\boldsymbol{A}$, onde $r = \operatorname{rank}(\boldsymbol{A})$
- O sistema (95) possui solução se e somente se $\boldsymbol{y}$ pertence ao subespaço $\mathcal{R}(\boldsymbol{A})$ gerado pela base $\{\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_r\}$, ou seja, o subespaço gerado pelas colunas de $\boldsymbol{A}$
- O subespaço $\mathcal{R}(\boldsymbol{A})$ de $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ é chamado **espaço coluna** ou **espaço *range*** de $\boldsymbol{A}$

$$\mathcal{R}(\boldsymbol{A}) \triangleq \left\{\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} | \boldsymbol{x} \in \mathbb{R}^n\right\} \subset \mathbb{R}^m \tag{100}$$

- O $\operatorname{rank}(\boldsymbol{A})$ é a dimensão do espaço coluna $\mathcal{R}(\boldsymbol{A})$
- Um matriz $\boldsymbol{A}$ possui (pseudo-)inversa quando o seu $\operatorname{rank}(\boldsymbol{A})$ é máximo, i.e., $\operatorname{rank}(\boldsymbol{A}) = \min\{m,n\}$
- Para $\boldsymbol{A} \in \mathbb{R}^{m \times p}$ e $\boldsymbol{B} \in \mathbb{R}^{m \times q}$, temos que

$$\mathcal{R}(\boldsymbol{A}) + \mathcal{R}(\boldsymbol{B}) = \mathcal{R}\left(\begin{bmatrix}\boldsymbol{A} & \boldsymbol{B}\end{bmatrix}\right) \tag{101}$$

Subespaços fundamentais: espaço nulo

- Quando $\boldsymbol{y} = \boldsymbol{0}$ em (95) temos $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$
- O conjunto de soluções de $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ define por si só um subespaço vetorial: o **espaço nulo** $\mathcal{N}(\boldsymbol{A})$ de $\boldsymbol{A}$

$$\mathcal{N}(\boldsymbol{A}) \triangleq \{\boldsymbol{x} \in \mathbb{R}^n | \boldsymbol{A}\boldsymbol{x} = 0\} \subset \mathbb{R}^n \tag{102}$$

- A dimensão do espaço nulo $\mathcal{N}(\boldsymbol{A})$ é chamada de **nulidade** de $\boldsymbol{A}$ e é denotada por $\operatorname{nullity}(\boldsymbol{A})$
- Para $\boldsymbol{A} \in \mathbb{R}^{p\times n}$ e $\boldsymbol{B} \in \mathbb{R}^{q\times n}$, temos que

$$\mathcal{N}(\boldsymbol{A}) + \mathcal{N}(\boldsymbol{B}) = \mathcal{N}\left(\begin{bmatrix}\boldsymbol{A} \\ \boldsymbol{B}\end{bmatrix}\right) \tag{103}$$

- Note que o espaço coluna e o espaço nulo de uma matriz $\boldsymbol{A}$ são duais tal que o $\operatorname{rank}(\boldsymbol{A})$ é o dual da $\operatorname{nullity}(\boldsymbol{A})$
- De fato, para $\boldsymbol{A}$ com dimensão $m \times n$ temos que

$$\operatorname{rank}(\boldsymbol{A}) + \operatorname{nullity}(\boldsymbol{A}) = n \qquad \text{(Teorema do posto-nulidade)} \tag{104}$$

Subespaços fundamentais: espaço linha e espaço nulo à esquerda

- Associados a uma matriz $\boldsymbol{A}$ e seus subespaços $\mathcal{R}\left(\boldsymbol{A}\right)$ e $\mathcal{N}\left(\boldsymbol{A}\right)$ temos ainda dois outros subespaços:
  - O **espaço linha** de $\boldsymbol{A}$, denotado por $\mathcal{R}\left(\boldsymbol{A}^{\mathrm{T}}\right)$, que corresponde ao espaço coluna de $\boldsymbol{A}^{\mathrm{T}}$ e é gerado pelas linhas de $\boldsymbol{A}$
  - O **espaço nulo à esquerda** de $\boldsymbol{A}$, denotado $\mathcal{N}\left(\boldsymbol{A}^{\mathrm{T}}\right)$, que corresponde ao espaço nulo de $\boldsymbol{A}^{\mathrm{T}}$ e é o espaço que contém todos os vetores $\boldsymbol{z}$ que satisfazem $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{z} = 0$
- De maneira similar à anterior, temos para $\boldsymbol{A}$ com dimensão $m \times n$ que

$$\operatorname{rank}\left(\boldsymbol{A}^{\mathrm{T}}\right) + \operatorname{nullity}\left(\boldsymbol{A}^{\mathrm{T}}\right) = m \qquad \text{(Teorema do posto-nulidade)} \tag{105}$$

- Além disso, temos ainda de acordo com (98) que

$$\operatorname{rank}\left(\boldsymbol{A}\right) = \operatorname{rank}\left(\boldsymbol{A}^{\mathrm{T}}\right) = r \leq \min\left\{m,n\right\} \tag{106}$$

Caracterização do posto de uma matriz [?]

- Na caracterização do posto de uma matriz $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ são equivalentes
  1. $\operatorname{rank}(\boldsymbol{A}) = r$
  2. r, e não mais que r, linhas de $\boldsymbol{A}$ são linearmente independentes
  3. r, e não mais que r, colunas de $\boldsymbol{A}$ são linearmente independentes
  4. Alguma submatriz $r \times r$ de $\boldsymbol{A}$ tem determinante não-nulo e toda submatriz $(r+1) \times (r+1)$ de $\boldsymbol{A}$ tem determinante nulo
  5. $\dim(\mathcal{R}(\boldsymbol{A})) = r$
  6. Há r, e não mais que r, vetores $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_r$ tais que o sistema linear $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}_j$ é consistente para $j = 1, \ldots, r$
  7. $r = n - \dim(\mathcal{N}(\boldsymbol{A}))$ (teorema do posto-nulidade)
  8. $r = \min\left\{p : \boldsymbol{A} = \boldsymbol{X}\boldsymbol{Y}^{\mathrm{T}}\right\}$ para algum $\boldsymbol{X} \in \mathbb{R}^{m \times p}, \boldsymbol{Y} \in \mathbb{R}^{n \times p}$
  9. $r = \min\left\{p : \boldsymbol{A} = \boldsymbol{x}_1\boldsymbol{y}_1^{\mathrm{T}} + \ldots + \boldsymbol{x}_p\boldsymbol{y}_p^{\mathrm{T}}\right\}$ para algum $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p \in \mathbb{R}^m, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_p \in \mathbb{R}^n$

Desigualdades do posto de uma matriz [?]

1. Se $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, então $\operatorname{rank}(\boldsymbol{A}) \leq \min\{m, n\}$
2. Desigualdade de Sylvester: se $\boldsymbol{A} \in \mathbb{R}^{m \times k}$ e $\boldsymbol{B} \in \mathbb{R}^{k \times n}$, então $(\operatorname{rank}(\boldsymbol{A}) + \operatorname{rank}(\boldsymbol{B})) - k \leq \operatorname{rank}(\boldsymbol{A}\boldsymbol{B}) \leq \min\{\operatorname{rank}(\boldsymbol{A}), \operatorname{rank}(\boldsymbol{B})\}$ ()
3. Desigualdade soma-posto: se $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{m \times k}$ então $|\operatorname{rank}(\boldsymbol{A}) - \operatorname{rank}(\boldsymbol{B})| \leq \operatorname{rank}(\boldsymbol{A} + \boldsymbol{B}) \leq \operatorname{rank}(\boldsymbol{A}) + \operatorname{rank}(\boldsymbol{B})$ com igualdade na segunda desigualdade se e somente se $\mathcal{R}(\boldsymbol{A}) \cap \mathcal{R}(\boldsymbol{B}) = \emptyset$
4. Desigualdade de Frobenius: se $\boldsymbol{A} \in \mathbb{R}^{m \times k}$, $\boldsymbol{B} \in \mathbb{R}^{k \times p}$ e $\boldsymbol{C} \in \mathbb{R}^{p \times n}$, então $\operatorname{rank}(\boldsymbol{A}\boldsymbol{B}) + \operatorname{rank}(\boldsymbol{B}\boldsymbol{C}) \leq \operatorname{rank}(\boldsymbol{B}) + \operatorname{rank}(\boldsymbol{A}\boldsymbol{B}\boldsymbol{C})$ com igualdade se e somente se $\exists \boldsymbol{X}, \boldsymbol{Y} : \boldsymbol{B} = \boldsymbol{B}\boldsymbol{C}\boldsymbol{X} + \boldsymbol{Y}\boldsymbol{A}\boldsymbol{B}$

Igualdades do posto de uma matriz [?]

1. Se $\boldsymbol{A} \in \mathbb{C}^{m\times n}$, então $\operatorname{rank}(\boldsymbol{A}) = \operatorname{rank}\left(\boldsymbol{A}^{\mathrm{T}}\right) = \operatorname{rank}(\boldsymbol{A}^{*}) = \operatorname{rank}\left(\boldsymbol{A}^{\mathrm{H}}\right)$
2. Se $\boldsymbol{A} \in \mathbb{R}^{m\times m}$ e $\boldsymbol{C} \in \mathbb{R}^{n\times n}$ são matrizes não-singulares e $\boldsymbol{B} \in \mathbb{R}^{m\times n}$, então $\operatorname{rank}(\boldsymbol{AB}) = \operatorname{rank}(\boldsymbol{B}) = \operatorname{rank}(\boldsymbol{BC}) = \operatorname{rank}(\boldsymbol{ABC})$, ou seja, multiplicações à direita ou esquerda por matrizes não-singulares não afetam o posto
3. Se $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{m\times n}$ então $\operatorname{rank}(\boldsymbol{A}) = \operatorname{rank}(\boldsymbol{B})$ se e somente se $\exists \boldsymbol{X} \in \mathbb{R}^{m\times m}, \boldsymbol{Y} \in \mathbb{R}^{n\times n}$ não-singulares tais que $\boldsymbol{B} = \boldsymbol{XAY}$
4. Se $\boldsymbol{A} \in \mathbb{C}^{m\times n}$ então $\operatorname{rank}\left(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\right) = \operatorname{rank}(\boldsymbol{A})$
5. Fatorização de posto completo: se $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, então $\operatorname{rank}(\boldsymbol{A}) = k$ se e somente se $\boldsymbol{A} = \boldsymbol{XY}^{\mathrm{T}}$ onde $\boldsymbol{X} \in \mathbb{R}^{m\times k}$ e $\boldsymbol{Y} \in \mathbb{R}^{n\times k}$ têm colunas independentes cada
6. Se $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, $\boldsymbol{X} \in \mathbb{R}^{n\times k}$, e $\boldsymbol{Y} \in \mathbb{R}^{m\times k}$ e $\boldsymbol{W} = \boldsymbol{Y}^{\mathrm{T}}\boldsymbol{AX}$ é não-singular, então $\operatorname{rank}\left(\boldsymbol{A} - \boldsymbol{AXW}^{-1}\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{A}\right) = \operatorname{rank}(\boldsymbol{A}) - \operatorname{rank}\left(\boldsymbol{AXW}^{-1}\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{A}\right)$

Conteúdo

Definições básicas

- Sejam uma matriz $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, um vetor $\boldsymbol{v} \in \mathbb{R}^{n}$ e um escalar $\lambda \in \mathbb{R}$ tais que

$$\boldsymbol{A}\boldsymbol{v} = \lambda\boldsymbol{v} \tag{107}$$

- A transformação linear expressa por $\boldsymbol{A}$ aplicada ao vetor $\boldsymbol{v}$ resulta em uma versão escalonada de $\boldsymbol{v}$, i.e., $\lambda\boldsymbol{v}$
- Nesse caso, o escalar λ é denominado um autovalor de $\boldsymbol{A}$ e $\boldsymbol{v}$ é chamado o autovetor de $\boldsymbol{A}$ associado ao autovalor λ
- Autovalores e autovetores encontram diversas aplicações em engenharia
  - Esforços e direções principais em mecânica dos materiais
  - Estudos de momentos de inércia
  - Frequências naturais e modos de vibração
  - Alocação ótima de potência em comunicações co-canal
  - Formação de feixe em sistemas de comunicação com antenas inteligentes

Equação característica

- Sendo $\boldsymbol{I}$ uma matriz identidade e $\mathbf{0}$ um vetor de zeros, reescrevemos (107) como

$$(\boldsymbol{A} - \lambda \boldsymbol{I})\boldsymbol{v} = \mathbf{0} \tag{108}$$

- Se a matriz $(\boldsymbol{A} - \lambda \boldsymbol{I})$ não é singular ➜ $\exists\ (\boldsymbol{A} - \lambda \boldsymbol{I})^{-1}$ ➜ $\boldsymbol{v} = (\boldsymbol{A} - \lambda \boldsymbol{I})^{-1}\mathbf{0}$ (solução trivial)
- Se a matriz $(\boldsymbol{A} - \lambda \boldsymbol{I})$ é singular, temos que

$$\det(\boldsymbol{A} - \lambda \boldsymbol{I}) = 0, \tag{109}$$

- Expandindo (109) usando as regras para cálculo de determinantes ➜ equação polinomial de grau n
- De fato, (109) é chamada de equação característica da matriz $\boldsymbol{A}$ e suas raízes $\lambda_1, \lambda_2, \ldots, \lambda_n$ são os autovalores de $\boldsymbol{A}$
- O autovetor $\boldsymbol{v}_i$ associado a $\lambda_i,\ i = 1, 2, \ldots, n$, pode ser determinado substituindo-se λ_i em (107), ou seja,

$$\boldsymbol{A}\boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i, \quad i = 1, 2, \ldots, n \tag{110}$$

- Por convenção, assume-se que $|\lambda_{\max}| = |\lambda_1| \geq |\lambda_2| \geq \ldots \geq |\lambda_n| = |\lambda_{\min}|$ e que $\|\boldsymbol{v}_i\|_2 = 1, \quad i = 1, 2, \ldots, n$

Propriedades de autovalores e autovetores

Propriedade 1

Se $\lambda_1, \lambda_2, \ldots, \lambda_n$ são os autovalores da matriz $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, então os autovalores de $\boldsymbol{A}^k$, $k > 0$, são $\lambda_1^k, \lambda_2^k, \ldots, \lambda_n^k$.

Prova.

$$\boldsymbol{A}^k\boldsymbol{v}_i = \boldsymbol{A}^{k-1}\boldsymbol{A}\boldsymbol{v}_i = \lambda_i\boldsymbol{A}^{k-1}\boldsymbol{v}_i = \lambda_i^2\boldsymbol{A}^{k-2}\boldsymbol{v}_i = \ldots = \lambda_i^{k-1}\boldsymbol{A}\boldsymbol{v}_i = \lambda_i^k\boldsymbol{v}_i.$$ □

- Logo, todo autovetor $\boldsymbol{v}_i$ de $\boldsymbol{A}$ é autovetor de $\boldsymbol{A}^k$.

Propriedade 2

Sejam $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$ os autovetores da matriz $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, correspondentes a autovalores distintos $\lambda_1, \lambda_2, \ldots, \lambda_n$, então $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$ são linearmente independentes.

Propriedades de autovalores e autovetores

Prova.

- Se $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$, então existe α_i, $i = 1, 2, \ldots, n$, não todos nulos, tal que $\sum_{i=1}^{n} \alpha_i \boldsymbol{v}_i = 0$, que multiplicado repetidamente por $\boldsymbol{A}$ leva ao conjunto de n equações

$$\sum_{i=1}^{n} \alpha_i \lambda_i^k \boldsymbol{v}_i = 0, \quad k = 1, 2, \ldots, n, \tag{111}$$

o qual pode ser reescrito matricialmente como

$$\begin{bmatrix} \alpha_1 \boldsymbol{v}_1 & \alpha_2 \boldsymbol{v}_2 & \alpha_3 \boldsymbol{v}_3 & \ldots & \alpha_n \boldsymbol{v}_n \end{bmatrix} \underbrace{\begin{bmatrix} 1 & \lambda_1 & \lambda_1^2 & \ldots & \lambda_1^{n-1} \\ 1 & \lambda_2 & \lambda_2^2 & \ldots & \lambda_2^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \lambda_n & \lambda_n^2 & \ldots & \lambda_n^{n-1} \end{bmatrix}}_{\boldsymbol{S}} = \boldsymbol{0} \tag{112}$$

□

Propriedades de autovalores e autovetores

Prova.

- A matriz $\boldsymbol{S}$ em (112) é chamada matriz de Vandermonde e para λ_i distintos é não-singular. Logo

$$\begin{bmatrix}\alpha_1\boldsymbol{v}_1 & \alpha_2\boldsymbol{v}_2 & \dots & \alpha_n\boldsymbol{v}_n\end{bmatrix}\boldsymbol{S} = \boldsymbol{0} \Rightarrow \begin{bmatrix}\alpha_1\boldsymbol{v}_1 & \alpha_2\boldsymbol{v}_2 & \dots & \alpha_n\boldsymbol{v}_n\end{bmatrix} = \boldsymbol{0}\boldsymbol{S}^{-1} = \boldsymbol{0} \tag{113}$$

- Logo, α_i precisam ser todos nulos, já que $\boldsymbol{v}_i$ são não-nulos e, portanto, $\boldsymbol{v}_i$ são L.I.

□

Propriedade 3

Sejam $\lambda_1, \lambda_2, \dots, \lambda_n$ os autovalores da matriz $\boldsymbol{A} \in \mathbb{C}^{n\times n}$, $\boldsymbol{A} = \boldsymbol{A}^{\mathrm{H}}$, então $\lambda_1, \lambda_2, \dots, \lambda_n \in \mathbb{R}$ e $\lambda_i \geq 0$.

Prova.

Temos que $\boldsymbol{A}\boldsymbol{v}_i = \lambda_i\boldsymbol{v}_i \Rightarrow \boldsymbol{v}_i^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}} = \lambda_i^*\boldsymbol{v}_i^{\mathrm{H}} \Rightarrow \boldsymbol{v}_i^{\mathrm{H}}\boldsymbol{A} = \lambda_i^*\boldsymbol{v}_i^{\mathrm{H}}$. Logo
$\boldsymbol{A}\boldsymbol{v}_i = \lambda_i\boldsymbol{v}_i \Rightarrow \boldsymbol{v}_i^{\mathrm{H}}\boldsymbol{A}\boldsymbol{v}_i = \lambda_i\boldsymbol{v}_i^{\mathrm{H}}\boldsymbol{v}_i \geq 0 \Rightarrow \lambda_i^*\boldsymbol{v}_i^{\mathrm{H}}\boldsymbol{v}_i = \lambda_i\boldsymbol{v}_i^{\mathrm{H}}\boldsymbol{v}_i \geq 0 \Rightarrow \lambda_i = \lambda_i^* \geq 0$ □

Propriedades de autovalores e autovetores

Propriedade 4

Sejam $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$ *os autovetores da matriz* $\boldsymbol{A} \in \mathbb{C}^{n\times n}$, $\boldsymbol{A} = \boldsymbol{A}^{\mathrm{H}}$, *correspondentes a autovalores distintos* $\lambda_1, \lambda_2, \ldots, \lambda_n$, *então* $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$ *são ortogonais.*

Prova.

Temos que $\boldsymbol{A}\boldsymbol{v}_j = \lambda_j \boldsymbol{v}_j \Rightarrow \boldsymbol{v}_j^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}} = \lambda_j^* \boldsymbol{v}_j^{\mathrm{H}} \Rightarrow \boldsymbol{v}_j^{\mathrm{H}}\boldsymbol{A} = \lambda_j \boldsymbol{v}_j^{\mathrm{H}}$. Logo
$\boldsymbol{A}\boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i \Rightarrow \boldsymbol{v}_j^{\mathrm{H}}\boldsymbol{A}\boldsymbol{v}_i = \lambda_i \boldsymbol{v}_j^{\mathrm{H}}\boldsymbol{v}_i \Rightarrow \lambda_j \boldsymbol{v}_j^{\mathrm{H}}\boldsymbol{v}_i = \lambda_i \boldsymbol{v}_j^{\mathrm{H}}\boldsymbol{v}_i \Rightarrow \boldsymbol{v}_j^{\mathrm{H}}\boldsymbol{v}_i = 0$, pois $\lambda_i \neq \lambda_j$ □

- Os autovetores de uma matriz $\boldsymbol{A}$, são ortonormais se associados a autovalores distintos, i.e.,

$$\boldsymbol{v}_i^{\mathrm{T}}\boldsymbol{v}_i = 1, \quad \forall i, \quad i = 1,2,\ldots,n \tag{114a}$$

$$\boldsymbol{v}_i^{\mathrm{T}}\boldsymbol{v}_j = 0, \quad \forall i \neq j, \quad i,j = 1,2,\ldots,n \tag{114b}$$

- Formam uma base para o espaço coluna gerado pela matriz, i.e., qualquer vetor $\boldsymbol{x}$ nesse espaço pode ser escrito como uma combinação linear

$$\boldsymbol{x} = \alpha_1 \boldsymbol{v}_1 + \alpha_2 \boldsymbol{v}_1 + \cdots + \alpha_n \boldsymbol{v}_n \tag{115}$$

  onde $\alpha_i, \; i = 1,2,\ldots,n$, são constantes reais tais $\alpha_i = 0, \forall i \Leftrightarrow \boldsymbol{x} = \boldsymbol{0}$

Propriedades de autovalores e autovetores

Propriedade 5

Sejam $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$ *os autovetores da matriz* $\boldsymbol{A}$*, correspondentes a autovalores distintos* $\lambda_1, \lambda_2, \ldots, \lambda_n$*, então* $\boldsymbol{V}^{-1}\boldsymbol{A}\boldsymbol{V} = \boldsymbol{\Lambda}, \boldsymbol{A} \in \mathbb{R}^{n\times n}$ *e* $\boldsymbol{V}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{V} = \boldsymbol{\Lambda}, \boldsymbol{A} \in \mathbb{C}^{n\times n}$ *onde* $\boldsymbol{V} = \begin{bmatrix}\boldsymbol{v}_1 & \boldsymbol{v}_2 & \ldots & \boldsymbol{v}_n\end{bmatrix}$ *e* $\boldsymbol{\Lambda} = \mathrm{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$.

Prova.

A prova segue diretamente das propriedades de independência linear ($\exists \boldsymbol{V}^{-1}$) e de ortogonalidade ($\boldsymbol{V}^{\mathrm{H}}\boldsymbol{V} = \boldsymbol{I}$). □

Propriedade 6

Sejam $\lambda_1, \lambda_2, \ldots, \lambda_n$ *os autovalores da matriz* $\boldsymbol{A}$*, então o traço* $\mathrm{tr}\,(\boldsymbol{A})$ *de* $\boldsymbol{A}$ *é igual à soma dos autovalores de* $\boldsymbol{A}$.

Prova.

Temos que $\mathrm{tr}\,(\boldsymbol{A}) = \sum_{i=1}^{n} a_{i,i} = \mathrm{tr}\left(\boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{-1}\right) = \mathrm{tr}\left(\boldsymbol{\Lambda}\boldsymbol{V}^{-1}\boldsymbol{V}\right) = \mathrm{tr}\,(\boldsymbol{\Lambda}) = \sum_{i=1}^{n} \lambda_i$. □

Propriedades de autovalores e autovetores

Propriedade 7

Sejam $\lambda_1, \lambda_2, \ldots, \lambda_n$ os autovalores da matriz $\boldsymbol{A}$, então o determinante $\det(\boldsymbol{A})$ de $\boldsymbol{A}$ é igual ao produto dos autovalores de $\boldsymbol{A}$.

Prova.

Temos que $\det(\boldsymbol{A}) = \det\left(\boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{-1}\right) = \det(\boldsymbol{V})\det(\boldsymbol{\Lambda})\det\left(\boldsymbol{V}^{-1}\right) = \det(\boldsymbol{V})\det(\boldsymbol{\Lambda})\det(\boldsymbol{V})^{-1} = \det(\boldsymbol{\Lambda}) = \prod_{i=1}^{n}\lambda_i.$ □

Transformação de similaridade

- Transformações de similaridade são úteis para transformar uma matriz associada a um problema em uma forma similar e de mais simples manipulação
- Se agruparmos as n equações em (110) em uma única equação podemos escrever

$$\boldsymbol{A}\underbrace{\begin{bmatrix}\boldsymbol{v}_1 & \boldsymbol{v}_2 & \dots & \boldsymbol{v}_n\end{bmatrix}}_{\boldsymbol{V}} = \underbrace{\begin{bmatrix}\boldsymbol{v}_1 & \boldsymbol{v}_2 & \dots & \boldsymbol{v}_n\end{bmatrix}}_{\boldsymbol{V}}\underbrace{\begin{bmatrix}\lambda_1 & 0 & \dots & 0\\ 0 & \lambda_2 & \dots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \dots & \lambda_n\end{bmatrix}}_{\boldsymbol{\Lambda}} \tag{116}$$

$$\boldsymbol{A}\boldsymbol{V} = \boldsymbol{V}\boldsymbol{\Lambda} \Rightarrow \boxed{\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{-1}}$$

- As matrizes $\boldsymbol{\Lambda}$ e $\boldsymbol{V}$ são as matrizes dos autovalores e autovetores de $\boldsymbol{A}$, respectivamente

Transformação de similaridade

- Duas matrizes $\boldsymbol{A}$ e $\boldsymbol{B}$ de dimensão $n \times n$ são ditas **similares** se existe uma transformação $\boldsymbol{T}$ tal que

$$\boldsymbol{A} = \boldsymbol{T}\boldsymbol{B}\boldsymbol{T}^{-1} \Rightarrow \boldsymbol{B} = \boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T} \tag{117}$$

- Matrizes similares possuem os mesmos autovalores e encontram várias aplicações em engenharia
- A partir de (116) e (117), podemos observar que

$$\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{-1} \text{ e } \boldsymbol{\Lambda} = \boldsymbol{V}^{-1}\boldsymbol{A}\boldsymbol{V}, \tag{118}$$

  de modo que a matriz $\boldsymbol{A}$ é similar à matriz diagonal $\boldsymbol{\Lambda}$
- Observe ainda que há uma relação entre as inversas de matrizes similares dada por

$$\boldsymbol{A}^{-1} = \left(\boldsymbol{V}^{-1}\right)^{-1}\boldsymbol{\Lambda}^{-1}\boldsymbol{V}^{-1} = \boldsymbol{V}\boldsymbol{\Lambda}^{-1}\boldsymbol{V}^{-1}, \tag{119}$$

  de modo que a inversa de $\boldsymbol{A}$ pode ser facilmente obtida se $\boldsymbol{\Lambda}$ e $\boldsymbol{V}$ forem conhecidas, pois $\boldsymbol{\Lambda}$ é diagonal
- Note ainda que $\det\left(\boldsymbol{A}^{-1}\right) = \det\left(\boldsymbol{\Lambda}^{-1}\right) = \prod_{i=1}^{n} \lambda_i^{-1}$

Forma de Jordan

- ...

Conteúdo

Polinômios de matrizes quadradas

- Se $\boldsymbol{A}$ é uma matriz quadrada e k é um número não-negativo, então

$$\boldsymbol{A}^k = \underbrace{\boldsymbol{A}\cdot\boldsymbol{A}\cdot\boldsymbol{A}\cdots\boldsymbol{A}}_{k \text{ vezes}} \quad \text{e} \quad \boldsymbol{A}^0 = \boldsymbol{I} \tag{120}$$

- Se $f(\lambda)$ é um polinômio qualquer e, por exemplo, $f_1(\lambda) = \lambda^3 + 2\lambda^2 - 6$ e $f_2(\lambda) = (\lambda+2)(4\lambda-3)$ então

$$f_1(\boldsymbol{A}) = \boldsymbol{A}^3 + 2\boldsymbol{A}^2 - 6\boldsymbol{I} \quad \text{e} \quad f_2(\boldsymbol{A}) = (\boldsymbol{A} + 2\boldsymbol{I})(4\boldsymbol{A} - 3\boldsymbol{I}) \tag{121}$$

- Em outros termos, polinômios podem ser aplicados diretamente a matrizes constituindo assim uma classe de funções de matrizes

Teorema de Cayley-Hamilton

- Se $\boldsymbol{A}$ é uma matriz quadrada, então seu polinômio característico $\Delta(\lambda) = \det(\lambda \boldsymbol{I} - \boldsymbol{A})$ é dado por

$$\Delta(\lambda) = \lambda^n + \alpha_{n-1}\lambda^{n-1} + \alpha_{n-2}\lambda^{n-2} + \ldots + \alpha_1\lambda + \alpha_0 \tag{122}$$

- O teorema de Cayley-Hamilton estabelece que a função polinomial (122) aplicada à matriz $\boldsymbol{A}$ é identicamente nula, i.e.,

$$\Delta(\boldsymbol{A}) = \boldsymbol{A}^n + \alpha_{n-1}\boldsymbol{A}^{n-1} + \alpha_{n-2}\boldsymbol{A}^{n-2} + \ldots + \alpha_1\boldsymbol{A} + \alpha_0\boldsymbol{I} = \boldsymbol{0}, \tag{123}$$

- Logo, $\boldsymbol{A}$ satisfaz sua própria equação característica
- O teorema de Cayley-Hamilton é útil para calcular potências de $\boldsymbol{A}^k$, $k > n$ em função de $\boldsymbol{A}^n, \ldots, \boldsymbol{A}$, como por exemplo

$$\begin{aligned} \boldsymbol{A}^{n+1} &= \boldsymbol{A} \cdot (\boldsymbol{A}^n + \alpha_{n-1}\boldsymbol{A}^{n-1} + \alpha_{n-2}\boldsymbol{A}^{n-2} + \ldots + \alpha_1\boldsymbol{A} + \alpha_0\boldsymbol{I}) = \boldsymbol{0} \Rightarrow \\ \boldsymbol{A}^{n+1} &= -\alpha_{n-1}\boldsymbol{A}^n - \alpha_{n-2}\boldsymbol{A}^{n-1} - \ldots - \alpha_1\boldsymbol{A}^2 - \alpha_0\boldsymbol{A} \end{aligned} \tag{124}$$

Teorema de Cayley-Hamilton

- De acordo com (123), $\boldsymbol{A}^n$ pode ser escrita como combinação linear de $\left\{\boldsymbol{I}, \boldsymbol{A}, \ldots, \boldsymbol{A}^{n-1}\right\}$
- De acordo com (124), $\boldsymbol{A}^{n+1}$ pode ser escrita como combinação linear de $\left\{\boldsymbol{A}, \boldsymbol{A}^2, \ldots, \boldsymbol{A}^n\right\}$
- Logo, para qualquer função polinomial $f(\lambda)$ temos que $f(\boldsymbol{A})$ pode ser escrita como

$$f(\boldsymbol{A}) = \beta_0 \boldsymbol{I} + \beta_1 \boldsymbol{A} + \ldots + \beta_{n-2} \boldsymbol{A}^{n-2} + \beta_{n-1} \boldsymbol{A}^{n-1} \tag{125}$$

  para algum conjunto de valores $\beta_0, \beta_1, \ldots, \beta_{n-1}$
- Se $\boldsymbol{A} = \boldsymbol{T}\bar{\boldsymbol{A}}\boldsymbol{T}^{-1}$ então $\boldsymbol{A}^k = \boldsymbol{T}\bar{\boldsymbol{A}}^k\boldsymbol{T}^{-1}$ e, portanto $f(\boldsymbol{A}) = \boldsymbol{T} f(\bar{\boldsymbol{A}})\boldsymbol{T}^{-1}$, pois

$$\begin{aligned}
f(\boldsymbol{A}) = f(\boldsymbol{T}\bar{\boldsymbol{A}}\boldsymbol{T}^{-1}) \\
&= \beta_0 \boldsymbol{T}\boldsymbol{T}^{-1} + \beta_1 \boldsymbol{T}\bar{\boldsymbol{A}}\boldsymbol{T}^{-1} + \ldots + \beta_{n-2}\boldsymbol{T}\bar{\boldsymbol{A}}^{n-2}\boldsymbol{T}^{-1} + \beta_{n-1}\boldsymbol{T}\bar{\boldsymbol{A}}^{n-1}\boldsymbol{T}^{-1} \\
&= \boldsymbol{T}(\beta_0 \boldsymbol{I} + \beta_1 \boldsymbol{A} + \ldots + \beta_{n-2}\boldsymbol{A}^{n-2} + \beta_{n-1}\boldsymbol{A}^{n-1})\boldsymbol{T}^{-1} \\
&= \boldsymbol{T} f(\bar{\boldsymbol{A}})\boldsymbol{T}^{-1}
\end{aligned}$$

Conteúdo

Norma induzida

- A norma induzida $\|\boldsymbol{A}\|$ de uma matriz $\boldsymbol{A}$ pode ser definida através do problema de otimização

$$\|\boldsymbol{A}\| = \max \{\|\boldsymbol{A}\boldsymbol{x}\|_2\} = \max \left\{(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x})^{\frac{1}{2}}\right\}, \tag{126a}$$

$$\text{subject to: } \|\boldsymbol{x}\|_2 = 1 \tag{126b}$$

- Dado que o $\mathcal{R}(\boldsymbol{A}) = \mathcal{R}\left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\right)$ e que $\|\boldsymbol{x}\|_2 = 1$ conforme (126b), verificamos que a solução de (126) corresponde à raiz quadrada do maior autovalor de $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$
- A norma induzida acima pode ser definida em termos de outras normas que não a norma euclidiana
- Apenas a norma induzida pela norma euclidiana é também chamada norma espectral

Norma induzida

- As normas induzidas para matrizes satisfazem as mesmas propriedades que as normas de vetores, tais como:

$$
\begin{aligned}
\|\boldsymbol{A}\| = \|\boldsymbol{A}^{\mathrm{T}}\| &\geq 0 && \text{(Não-negatividade)} && (127a)\\
\|\boldsymbol{A}\boldsymbol{B}\| &\leq \|\boldsymbol{A}\|\|\boldsymbol{B}\| && \text{(Desigualdade de Cauchy-Schwarz)} && (127b)\\
\|\boldsymbol{A}\| = 0 &\Leftrightarrow \boldsymbol{A} = \boldsymbol{0} && \text{(Elemento neutro)} && (127c)\\
\|\alpha\boldsymbol{A}\| &= |\alpha|\|\boldsymbol{A}\|, \forall\alpha \in \mathbb{R} && \text{(Escalabilidade)} && (127d)\\
\|\boldsymbol{A}+\boldsymbol{B}\| &\leq \|\boldsymbol{A}\| + \|\boldsymbol{B}\|, \forall\boldsymbol{A},\boldsymbol{B} \in \mathbb{R}^{n\times n} && \text{(Desigualdade triangular)} && (127e)\\
\max_{\substack{1\leq i\leq m\\1\leq j\leq n}} |a_{i,j}| \leq \|\boldsymbol{A}\| &\leq \sqrt{mn}\max_{\substack{1\leq i\leq m\\1\leq j\leq n}} |a_{i,j}| && && (127f)
\end{aligned}
$$

- Para matrizes complexas, as matrizes transpostas são substituídas por matrizes hermitianas

Conteúdo

Cálculo com matrizes

- As regras de cálculo aplicadas a matrizes devem ser consistentes com as regras de cálculo utilizando escalares
- Os resultados sugerem apenas a reorganização dos termos entre formas escalares e matriciais
- Este princípio conduz à conclusão de que a derivada e a integral de matrizes pode ser definida elemento-a-elemento, i.e., o elemento i,j das matrizes

$$\int_0^t \boldsymbol{A}(\tau)d\tau \quad \text{e} \quad \frac{d\boldsymbol{A}(t)}{dt} \quad \text{são} \quad \int_0^t a_{i,j}(\tau)d\tau \quad \text{e} \quad \frac{da_{i,j}(t)}{dt}, \tag{128}$$

  respectivamente
- O teorema fundamental do cálculo aplicado a matrizes leva a

$$\frac{d}{dt}\left(\int_0^t \boldsymbol{A}(\tau)d\tau\right) = \boldsymbol{A}(t) \tag{129}$$

Cálculo com matrizes

- Similarmente, a regra do produto para matrizes $\boldsymbol{A}(t)$ e $\boldsymbol{B}(t)$ resulta em

$$\frac{d}{dt}\left(\boldsymbol{A}(t)\boldsymbol{B}(t)\right) = \dot{\boldsymbol{A}}(t)\boldsymbol{B}(t) + \boldsymbol{A}(t)\dot{\boldsymbol{B}}(t) \tag{130}$$

- No entanto, é importante observar que

$$\frac{d\boldsymbol{A}^2(t)}{dt} = \dot{\boldsymbol{A}}(t)\boldsymbol{A}(t) + \boldsymbol{A}(t)\dot{\boldsymbol{A}}(t) \tag{131}$$

- Nesse contexto, um relação importante é a desigualdade triangular

$$\|\int_{t_0}^{t} \boldsymbol{x}(\tau)d\tau\| \leq |\int_{t_0}^{t} \|\boldsymbol{x}(\tau)\|d\tau| \tag{132}$$

  onde para $t \geq t_0$ o módulo pode ser desconsiderado

Conteúdo

Forma quadrática e matriz positiva (semi)definida

- Seja $\boldsymbol{Q}$ uma matriz pertencente a $\mathbb{R}^{n\times n}$ e $\boldsymbol{x}$ um vetor pertencente a $\mathbb{R}^n$
- O produto $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x}$ é chamado forma quadrática em $\boldsymbol{x}$
- A matriz $\boldsymbol{Q}$ pode ser considerada simétrica (i.e., $\boldsymbol{Q} = \boldsymbol{Q}^{\mathrm{T}}$) no estudo de formas quadráticas pois

$$\boldsymbol{x}^{\mathrm{T}}\left(\boldsymbol{Q}+\boldsymbol{Q}^{\mathrm{T}}\right)\boldsymbol{x} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{x} = 2\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x} \tag{133}$$

  e a forma quadrática não muda ao se substituir $\boldsymbol{Q}$ por $\dfrac{\boldsymbol{Q}+\boldsymbol{Q}^{\mathrm{T}}}{2}$
- Um matriz $\boldsymbol{Q}$ é chamada:
  - Positiva definida se $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x} > 0, \forall \boldsymbol{x} \neq \mathbf{0}$
  - Positiva semidefinida se $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x} \geq 0, \forall \boldsymbol{x}$
- Um matriz $\boldsymbol{Q}$ é negativa definida ou semidefinida se $-\boldsymbol{Q}$ é positiva definida ou semidefinida, respectivamente

Forma quadrática e matriz positiva (semi)definida

- A notação $\boldsymbol{Q} \succ 0$ e $\boldsymbol{Q} \succeq 0$ é normalmente utilizada para indicar que $\boldsymbol{Q}$ é positiva definida ou semidefinida, respectivamente
- A notação $\boldsymbol{Q}_1 \succ \boldsymbol{Q}_2$ e $\boldsymbol{Q}_1 \succeq \boldsymbol{Q}_2$ implicam $\boldsymbol{Q}_1 - \boldsymbol{Q}_2 \succ 0$ e $\boldsymbol{Q}_1 - \boldsymbol{Q}_2 \succeq 0$, respectivamente
- Todos os autovalores de uma matriz simétrica são reais
  - Todos os autovalores de $\boldsymbol{Q}$ são reais e positivos se $\boldsymbol{Q} \succ 0$
  - Todos os autovalores de $\boldsymbol{Q}$ são reais e não-negativos se $\boldsymbol{Q} \succeq 0$
- Algumas relações importantes para formas quadráticas são:

$$\lambda_{\min}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} \leq \boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x} \leq \lambda_{\max}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}, \qquad \text{(Desigualdade de Rayleigh-Ritz)} \tag{134a}$$

$$\|\boldsymbol{Q}\| \leq \operatorname{tr}(\boldsymbol{Q}) \leq n\|\boldsymbol{Q}\| \tag{134b}$$

- Para matrizes e vetores complexos, a operação $(\cdot)^{\mathrm{T}}$ é substituída por $(\cdot)^{\mathrm{H}}$ e a forma quadrática em $\boldsymbol{x}$ torna-se $\boldsymbol{x}^{\mathrm{H}}\boldsymbol{Q}\boldsymbol{x}$

Menores principais dominantes

- Um menor principal de uma matriz $\boldsymbol{Q}$ simétrica em $\mathbb{R}^{n\times n}$ é o determinante da submatriz formada pela remoção de r linhas e colunas de mesmo índice da matriz
- Para a mesma matriz $\boldsymbol{Q}$, o p-ésimo menor principal dominante de $\boldsymbol{Q}$ é o determinante da submatriz $\boldsymbol{Q}_p$ superior esquerda compreendendo os elementos $q_{i,j}$ de $\boldsymbol{Q}$ para $i,j = 1, 2, \ldots, p$, de modo que o 1º, 2º e 3º menores principais dominantes $\boldsymbol{Q} \in \mathbb{R}^{n\times n}$, $n \geq 3$, são

$$\det\left(\boldsymbol{Q}_1\right) = \det\left(\begin{bmatrix} q_{1,1} \end{bmatrix}\right), \quad \det\left(\boldsymbol{Q}_2\right) = \det\left(\begin{bmatrix} q_{1,1} & q_{1,2} \\ q_{2,1} & q_{2,2} \end{bmatrix}\right), \text{ e}$$
$$\det\left(\boldsymbol{Q}_3\right) = \det\left(\begin{bmatrix} q_{1,1} & q_{1,2} & q_{1,3} \\ q_{2,1} & q_{2,2} & q_{2,3} \\ q_{3,1} & q_{3,2} & q_{3,3} \end{bmatrix}\right) \tag{135}$$

- A matriz $\boldsymbol{Q}$ é positiva definida se e somente se todos os seus menores principais dominantes são positivos, i.e., se $\det\left(\boldsymbol{Q}_p\right) > 0, \forall p = 1, 2, \ldots, n$
- A matriz $\boldsymbol{Q}$ é positiva semidefinida se e somente se todos os seus menores principais dominantes são não-negativos

Conteúdo

Fatoração QR [?, cap. 4]

- O método da fatoração QR faz uso de transformações de Householder e de transformações de similaridade para decompor uma matriz $\boldsymbol{A}$ em um produto de uma matriz ortogonal $\boldsymbol{Q}$ ($\boldsymbol{Q}\boldsymbol{Q}^{\mathrm{T}} = \boldsymbol{I}$), por uma matriz triangular superior $\boldsymbol{R}$
- A fatoração QR se inicia com a matriz $\boldsymbol{A}^{(1)} = \boldsymbol{A}$ cujos autovalores devem ser determinados
- A matriz $\boldsymbol{A}^{(1)}$ é fatorada como
$$\boldsymbol{A}^{(1)} = \boldsymbol{Q}^{(1)}\boldsymbol{R}^{(1)}, \tag{136}$$
onde $\boldsymbol{Q}^{(1)}$ é uma matriz ortogonal, i.e., $\boldsymbol{Q}^{(1)}\left(\boldsymbol{Q}^{(1)}\right)^{\mathrm{T}} = \boldsymbol{I}$, mas a matriz $\boldsymbol{R}^{(1)}$ não é ainda uma matriz triangular superior
- A matriz $\boldsymbol{A}^{(2)}$ é obtida multiplicando $\boldsymbol{R}^{(1)}$ à direita por $\boldsymbol{Q}^{(1)}$, i.e.,
$$\boldsymbol{A}^{(2)} = \boldsymbol{R}^{(1)}\boldsymbol{Q}^{(1)} \tag{137}$$
- Usando (136), temos que $\boldsymbol{R}^{(1)} = \left(\boldsymbol{Q}^{(1)}\right)^{-1}\boldsymbol{A}^{(1)}$, a qual substituida em (137) resulta em
$$\boldsymbol{A}^{(2)} = \left(\boldsymbol{Q}^{(1)}\right)^{-1}\boldsymbol{A}^{(1)}\boldsymbol{Q}^{(1)} \tag{138}$$
- Logo, $\boldsymbol{A}^{(1)}$ e $\boldsymbol{A}^{(2)}$ são matrizes similares, i.e., possuem os mesmos autovalores

Matriz de Householder

- Para obter $\boldsymbol{Q}$ e $\boldsymbol{R}$, o método de decomposição QR utiliza matrizes de transformação de Householder
- Dado um vetor

$$\boldsymbol{v} = \mathbf{c} + \|\mathbf{c}\|_2 \boldsymbol{e}, \tag{139}$$

a matriz de transformação de Householder $\mathbf{H}$ associada ao vetor $\boldsymbol{v}$ é definida como

$$\mathbf{H} = \boldsymbol{I} - 2\frac{\boldsymbol{v}\boldsymbol{v}^{\mathrm{T}}}{\boldsymbol{v}^{\mathrm{T}}\boldsymbol{v}}, \tag{140}$$

onde $\boldsymbol{I}$ é a matriz identidade e $\boldsymbol{e}$ é um vetor com uma única componente igual a ± 1 e todas as demais componentes igual a zero.

- Por conveniência, costuma-se definir ainda o vetor $\boldsymbol{e}$ como $\pm \boldsymbol{i}_i$, i.e., em termos da i-ésima coluna $\boldsymbol{I}$
- Note que a matriz de transformação de Householder é ortogonal, i.e., $\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H} = \boldsymbol{I}$

Algoritmo de fatoração QR

- O **passo 1** do algoritmo QR consiste em determinar $\boldsymbol{Q}^{(1)}$ e $\boldsymbol{R}^{(1)}$
- Para tanto, os vetores $\mathbf{c}^{(1)}$ e $\boldsymbol{e}^{(1)}$ geradores da matriz de Householder $\mathbf{H}^{(1)}$ são definidos como

$$
\begin{aligned}
\mathbf{c}^{(1)} &= \mathbf{a}_1, \text{ onde } \boldsymbol{A} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \dots & \mathbf{a}_n \end{bmatrix} = \begin{bmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \dots & a_{n,n} \end{bmatrix}, \text{ e} \\
\boldsymbol{e}^{(1)} &= \begin{cases} \boldsymbol{i}_1, & a_{1,1} \geq 0 \\ -\boldsymbol{i}_1, & a_{1,1} < 0 \end{cases}, \text{ onde } \boldsymbol{I} = \begin{bmatrix} \boldsymbol{i}_1 & \boldsymbol{i}_2 & \dots & \boldsymbol{i}_n \end{bmatrix}
\end{aligned}
\tag{141}
$$

- Usando (141) em (140), as matrizes $\boldsymbol{Q}^{(1)}$ e $\boldsymbol{R}^{(1)}$ são definidas como

$$
\boldsymbol{Q}^{(1)} = \mathbf{H}^{(1)} \quad \text{e} \quad \boldsymbol{R}^{(1)} = \mathbf{H}^{(1)} \boldsymbol{A}^{(1)} \tag{142}
$$

Algoritmo de fatoração QR

- A matriz $\boldsymbol{R}^{(1)}$ obtida no **passo 1** é da forma

$$\boldsymbol{R}^{(1)} = \begin{bmatrix} \mathbf{r}^{(1)}_{1,1} & \mathbf{r}^{(1)}_{1,2} & \mathbf{r}^{(1)}_{1,3} & \dots & \mathbf{r}^{(1)}_{1,n} \\ 0 & \mathbf{r}^{(1)}_{2,2} & \mathbf{r}^{(1)}_{2,3} & \dots & \mathbf{r}^{(1)}_{2,n} \\ 0 & \mathbf{r}^{(1)}_{3,2} & \mathbf{r}^{(1)}_{3,3} & \dots & \mathbf{r}^{(1)}_{3,n} \\ \vdots & \vdots & \ddots & \vdots & \\ 0 & \mathbf{r}^{(1)}_{n,2} & \mathbf{r}^{(1)}_{n,3} & \dots & \mathbf{r}^{(1)}_{n,n} \end{bmatrix} \tag{143}$$

- No **passo 2** do algoritmo QR, aproximadamente o mesmo processo do **passo 1** é repetido
- No entanto, os vetores $\mathbf{c}^{(2)}$ e $\boldsymbol{e}^{(2)}$ geradores da matriz de Householder $\mathbf{H}^{(2)}$ são definidos como

$$\mathbf{c}^{(2)} = \begin{bmatrix} 0 & r^{(1)}_{2,2} & r^{(1)}_{3,2} & \dots & r^{(1)}_{n,2} \end{bmatrix}^T, \quad \text{e} \quad \boldsymbol{e}^{(2)} = \begin{cases} \boldsymbol{i}_2, & r^{(1)}_{2,2} \geq 0 \\ -\boldsymbol{i}_2, & r^{(1)}_{2,2} < 0 \end{cases} \tag{144}$$

Algoritmo de fatoração QR

- Usando (144) e (114), as matrizes $\boldsymbol{Q}^{(2)}$ e $\boldsymbol{R}^{(2)}$ são obtidas como

$$\boldsymbol{Q}^{(2)} = \boldsymbol{Q}^{(1)}\mathbf{H}^{(2)} \quad \text{e} \quad \boldsymbol{R}^{(2)} = \mathbf{H}^{(2)}\boldsymbol{R}^{(1)} \tag{145}$$

onde a matriz $\boldsymbol{R}^{(2)}$ tem a forma

$$\boldsymbol{R}^{(2)} = \begin{bmatrix} \mathbf{r}_{1,1}^{(2)} & \mathbf{r}_{1,2}^{(2)} & \mathbf{r}_{1,3}^{(2)} & \dots & \mathbf{r}_{1,n}^{(2)} \\ 0 & \mathbf{r}_{2,2}^{(2)} & \mathbf{r}_{2,3}^{(2)} & \dots & \mathbf{r}_{2,n}^{(2)} \\ 0 & 0 & \mathbf{r}_{3,3}^{(2)} & \dots & \mathbf{r}_{3,n}^{(2)} \\ \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & \mathbf{r}_{n,3}^{(2)} & \dots & \mathbf{r}_{n,n}^{(2)} \end{bmatrix} \tag{146}$$

- No **passo 3** aproximadamente o mesmo processo do **passo 2** é repetido, exceto que os vetores $\mathbf{c}^{(3)}$ e $\boldsymbol{e}^{(3)}$ geradores $\mathbf{H}^{(3)}$ são dados por

$$\mathbf{c}^{(3)} = \begin{bmatrix} 0 & 0 & r_{3,3}^{(2)} & \dots & r_{n,3}^{(2)} \end{bmatrix}^T, \quad \text{e} \quad \boldsymbol{e}^{(3)} = \begin{cases} \boldsymbol{i}_3, & r_{3,3}^{(2)} \geq 0 \\ -\boldsymbol{i}_3, & r_{3,3}^{(2)} < 0 \end{cases} \tag{147}$$

Algoritmo de fatoração QR

- Usando (147) e (140), as matrizes $\boldsymbol{Q}^{(3)}$ e $\boldsymbol{R}^{(3)}$ são obtidas como

$$\boldsymbol{Q}^{(3)} = \boldsymbol{Q}^{(2)}\mathbf{H}^{(3)} \quad \text{e} \quad \boldsymbol{R}^{(3)} = \mathbf{H}^{(3)}\boldsymbol{R}^{(2)} \tag{148}$$

- Em cada passo, o processo descrito anteriormente é repetido
- Observe ainda que a cada passo i, os elementos da i-ésima coluna da matriz $\boldsymbol{R}^{(i)}$ são zerados
- De forma geral, as iterações descritas anteriormente podem ser escritas como

$$\boldsymbol{Q}^{(i)} = \boldsymbol{Q}^{(i-1)}\mathbf{H}^{(i)} \quad \text{e} \quad \boldsymbol{R}^{(i)} = \mathbf{H}^{(i)}\boldsymbol{R}^{(i-1)} \tag{149}$$

  onde $\boldsymbol{Q}^{(0)} = \boldsymbol{I}$ e $\boldsymbol{R}^{(0)} = \boldsymbol{A}$
- Um total de $n-1$ passos é realizado até que a matriz $\boldsymbol{R}^{(n-1)} = \mathbf{H}^{(n-1)}\boldsymbol{R}^{(n-2)}$ obtida é triangular superior
- Esse processo pode ser repetido até que a matriz $\boldsymbol{A}^{(n)} = \boldsymbol{R}^{(n-1)}\boldsymbol{Q}^{(n-1)}$ seja triangular. Nesse caso, os autovalores de $\boldsymbol{A}$ serão os elementos da diagonal de $\boldsymbol{A}^{(n)}$

Conteúdo

Conteúdo

Derivadas Multivariável

- Na resolução de problemas de otimização é bastante importante que se tenha uma base sólida sobre derivadas multivariável.
- Nesta seção, serão apresentados alguns casos, que são comumente encontrados na literatura.

Derivada Vetor-Escalar

- Seja um vetor $\boldsymbol{y} = \begin{bmatrix} y_1 & y_2 & \dots & y_N \end{bmatrix}^{\mathrm{T}} \in \mathbb{C}^{N \times 1}$ e um escalar x, temos que:

$$\frac{\partial \boldsymbol{y}}{\partial x} = \begin{bmatrix} \dfrac{\partial y_1}{\partial x} \\ \dfrac{\partial y_2}{\partial x} \\ \vdots \\ \dfrac{\partial y_N}{\partial x} \end{bmatrix}$$

- Como se pode perceber, para se derivar um vetor em relação a um escalar, basta que derivar cada elemento do vetor $\boldsymbol{y}$ pelo escalar x.
- Neste caso, a dimensão de $\dfrac{\partial \boldsymbol{y}}{\partial x}$ é a mesma de $\boldsymbol{y}$.

Derivada Escalar-Vetor

- Sejam agora o vetor $\boldsymbol{x} = \begin{bmatrix} x_1 & x_2 & \dots & x_N \end{bmatrix}^{\mathrm{T}} \in \mathbb{C}^{N\times 1}$ e um escalar y, temos que:

$$\frac{\partial y}{\partial \boldsymbol{x}} = \begin{bmatrix} \dfrac{\partial y}{\partial x_1} \\ \dfrac{\partial y}{\partial x_2} \\ \vdots \\ \dfrac{\partial y}{\partial x_N} \end{bmatrix}$$

- De modo análogo ao caso anterior, para se derivar um escalar em relação a um vetor, basta que derivar o escalar y em relação a cada elemento do vetor $\boldsymbol{x}$.
- A dimensão de $\dfrac{\partial y}{\partial \boldsymbol{x}}$ é a mesma de $\boldsymbol{x}$.

Derivada Vetor-Vetor

- Sejam os vetores $\boldsymbol{x} = \begin{bmatrix} x_1 & x_2 & \dots & x_N \end{bmatrix}^{\mathrm{T}}$ e $\boldsymbol{y} = \begin{bmatrix} y_1 & y_2 & \dots & y_M \end{bmatrix}^{\mathrm{T}}$, em que $\boldsymbol{x} \in \mathbb{C}^{N\times 1}$ e $\boldsymbol{y} \in \mathbb{C}^{M\times 1}$.
- Existem na literatura duas formas distintas de se representar a derivada de vetor em relação a outro vetor.
  - "*Denominator Layout*" ou formulação ~~Jacobiana~~; Hessiana

$$\left[\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}}\right]_{\mathrm{Den}} = \begin{bmatrix} \dfrac{\partial \boldsymbol{y}}{\partial x_1} \\ \dfrac{\partial \boldsymbol{y}}{\partial x_2} \\ \vdots \\ \dfrac{\partial \boldsymbol{y}}{\partial x_N} \end{bmatrix}$$

  - "*Numerator Layout*" ou formulação ~~Hessiana~~; Jacobiana

$$\left[\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}}\right]_{\mathrm{Num}} = \begin{bmatrix} \dfrac{\partial \boldsymbol{y}}{\partial x_1} & \dfrac{\partial \boldsymbol{y}}{\partial x_2} & \dots & \dfrac{\partial \boldsymbol{y}}{\partial x_N} \end{bmatrix}$$

Derivada Vetor-Vetor

- Como se pode perceber

$$\left[\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}}\right]_{\text{Den}} = \left[\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}}\right]_{\text{Num}}^{\text{T}}$$

- Para nosso estudo, iremos utilizar a formulação Hessiana, por ser a mais comumente utilizada na literatura.
- Assim,

$$\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_2}{\partial x_1} & \cdots & \frac{\partial y_M}{\partial x_1} \\ \frac{\partial y_1}{\partial x_2} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_M}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_1}{\partial x_N} & \frac{\partial y_2}{\partial x_N} & \cdots & \frac{\partial y_M}{\partial x_N} \end{bmatrix}$$

Derivada Matriz-Escalar

- Segue a mesma lógica que a derivada vetor-escalar.
- Desta forma, seja uma matriz $\boldsymbol{Y} \in \mathbb{C}^{N \times M}$ e um escalar x
- Assim,

$$\frac{\partial \boldsymbol{Y}}{\partial x} = \begin{bmatrix} \frac{\partial y_{11}}{\partial x} & \frac{\partial y_{12}}{\partial x} & \cdots & \frac{\partial y_{1M}}{\partial x} \\ \frac{\partial y_{21}}{\partial x} & \frac{\partial y_{22}}{\partial x} & \cdots & \frac{\partial y_{2M}}{\partial x} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_{N1}}{\partial x} & \frac{\partial y_{N2}}{\partial x} & \cdots & \frac{\partial y_{NM}}{\partial x} \end{bmatrix}$$

Derivada Escalar-Matriz

- Segue a mesma lógica que a derivada escalar-vetor.
- Desta forma, seja uma matriz $\boldsymbol{X} \in \mathbb{C}^{N \times M}$ e um escalar y
- Assim,

$$\frac{\partial y}{\partial \boldsymbol{X}} = \begin{bmatrix} \frac{\partial y}{\partial x_{11}} & \frac{\partial y}{\partial x_{12}} & \cdots & \frac{\partial y}{\partial x_{1M}} \\ \frac{\partial y}{\partial x_{21}} & \frac{\partial y}{\partial x_{22}} & \cdots & \frac{\partial y}{\partial x_{2M}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y}{\partial x_{N1}} & \frac{\partial y}{\partial x_{N2}} & \cdots & \frac{\partial y}{\partial x_{NM}} \end{bmatrix}$$

Conteúdo

Derivada de $a^{\mathrm{H}}x$

- Sejam $\boldsymbol{a}, \boldsymbol{x} \in \mathbb{C}^{N\times 1}$, temos que:

$$\frac{\partial \boldsymbol{a}^{\mathrm{H}}\boldsymbol{x}}{\partial \boldsymbol{x}} = \frac{\partial}{\partial \boldsymbol{x}}\left(\begin{bmatrix} a_1^* & a_2^* & \dots & a_N^* \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}\right) = \frac{\partial}{\partial \boldsymbol{x}}\left(\sum_{i=1}^{N} a_i^* x_i\right)$$

- Recaindo na formulação da derivada escalar-vetor, logo:

$$\frac{\partial \boldsymbol{a}^{\mathrm{H}}\boldsymbol{x}}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial}{\partial x_1}\left(\sum_{i=1}^{N} a_i^* x_i\right) \\ \frac{\partial}{\partial x_2}\left(\sum_{i=1}^{N} a_i^* x_i\right) \\ \vdots \\ \frac{\partial}{\partial x_N}\left(\sum_{i=1}^{N} a_i^* x_i\right) \end{bmatrix} = \begin{bmatrix} a_1^* \\ a_2^* \\ \vdots \\ a_N^* \end{bmatrix}$$

$$\boxed{\frac{\partial \boldsymbol{a}^{\mathrm{H}}\boldsymbol{x}}{\partial \boldsymbol{x}} = \boldsymbol{a}^*}$$

Derivada de $a^{\mathrm{T}}x$

- Considerando os mesmo vetores $\boldsymbol{a}, \boldsymbol{x} \in \mathbb{C}^{N\times 1}$, temos que:

$$\frac{\partial \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}}{\partial \boldsymbol{x}} = \frac{\partial}{\partial \boldsymbol{x}}\left(\begin{bmatrix} a_1 & a_2 & \dots & a_N \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}\right) = \frac{\partial}{\partial \boldsymbol{x}}\left(\sum_{i=1}^{N} a_i x_i\right)$$

- De modo análogo ao caso anterior,

$$\frac{\partial \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial}{\partial x_1}\left(\sum_{i=1}^{N} a_i x_i\right) \\ \frac{\partial}{\partial x_2}\left(\sum_{i=1}^{N} a_i x_i\right) \\ \vdots \\ \frac{\partial}{\partial x_N}\left(\sum_{i=1}^{N} a_i x_i\right) \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{bmatrix}$$

$$\boxed{\frac{\partial \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}}{\partial \boldsymbol{x}} = \boldsymbol{a}}$$

Derivada de $\boldsymbol{x}^{\mathrm{H}}\boldsymbol{a}$

- Considerando os mesmo vetores $\boldsymbol{a}, \boldsymbol{x} \in \mathbb{C}^{N\times 1}$, temos que:

$$\frac{\partial \boldsymbol{x}^{\mathrm{H}}\boldsymbol{a}}{\partial \boldsymbol{x}} = \frac{\partial}{\partial \boldsymbol{x}}\left(\begin{bmatrix} x_1^* & x_2^* & \dots & x_N^* \end{bmatrix}\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{bmatrix}\right) = \frac{\partial}{\partial \boldsymbol{x}}\left(\sum_{i=1}^{N} x_i^* a_i\right)$$

- De modo análogo aos casos anteriores e sabendo que $\frac{dx^*}{dx} = 0$, temos

$$\frac{\partial \boldsymbol{x}^{\mathrm{H}}\boldsymbol{a}}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial}{\partial x_1}\left(\sum_{i=1}^{N} x_i^* a_i\right) \\ \frac{\partial}{\partial x_2}\left(\sum_{i=1}^{N} x_i^* a_i\right) \\ \vdots \\ \frac{\partial}{\partial x_N}\left(\sum_{i=1}^{N} x_i^* a_i\right) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$\boxed{\frac{\partial \boldsymbol{x}^{\mathrm{H}}\boldsymbol{a}}{\partial \boldsymbol{x}} = \boldsymbol{0}}$$

Derivada de $x^{\mathrm{T}}a$

- Considerando os mesmo vetores $\boldsymbol{a}, \boldsymbol{x} \in \mathbb{C}^{N\times 1}$, temos que:

$$\frac{\partial \boldsymbol{x}^{\mathrm{T}}\boldsymbol{a}}{\partial \boldsymbol{x}} = \frac{\partial}{\partial \boldsymbol{x}}\left(\begin{bmatrix} x_1 & x_2 & \dots & x_N \end{bmatrix}\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{bmatrix}\right) = \frac{\partial}{\partial \boldsymbol{x}}\left(\sum_{i=1}^{N} x_i a_i\right)$$

$$= \begin{bmatrix} \frac{\partial}{\partial x_1}\left(\sum_{i=1}^{N} x_i a_i\right) \\ \frac{\partial}{\partial x_2}\left(\sum_{i=1}^{N} x_i a_i\right) \\ \vdots \\ \frac{\partial}{\partial x_N}\left(\sum_{i=1}^{N} x_i a_i\right) \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{bmatrix}$$

$$\boxed{\frac{\partial \boldsymbol{x}^{\mathrm{T}}\boldsymbol{a}}{\partial \boldsymbol{x}} = \boldsymbol{a}}$$

Derivada de Ax

- Considerando agora uma matriz $\boldsymbol{A} \in \mathbb{C}^{M\times N}$ e um vetor$\boldsymbol{x} \in \mathbb{C}^{N\times 1}$, temos que:

$$\frac{\partial \boldsymbol{Ax}}{\partial \boldsymbol{x}} = \frac{\partial}{\partial \boldsymbol{x}}\left(\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{M1} & a_{M2} & \cdots & a_{MN} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}\right) = \frac{\partial}{\partial \boldsymbol{x}}\left(\begin{bmatrix} \sum_{j=1}^{N} a_{1j}x_j & \sum_{j=1}^{N} a_{2j}x_j & \cdots & \sum_{j=1}^{N} a_{Mj}x_j \end{bmatrix}^{\mathrm{T}}\right)$$

$$= \begin{bmatrix} \frac{\partial}{\partial x_1}\left(\sum_{j=1}^{N} a_{1j}x_j\right) & \frac{\partial}{\partial x_1}\left(\sum_{j=1}^{N} a_{21j}x_j\right) & \cdots & \frac{\partial}{\partial x_1}\left(\sum_{j=1}^{N} a_{Mj}x_j\right) \\ \frac{\partial}{\partial x_2}\left(\sum_{j=1}^{N} a_{1j}x_j\right) & \frac{\partial}{\partial x_2}\left(\sum_{j=1}^{N} a_{21j}x_j\right) & \cdots & \frac{\partial}{\partial x_2}\left(\sum_{j=1}^{N} a_{Mj}x_j\right) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial}{\partial x_N}\left(\sum_{j=1}^{N} a_{1j}x_j\right) & \frac{\partial}{\partial x_N}\left(\sum_{j=1}^{N} a_{21j}x_j\right) & \cdots & \frac{\partial}{\partial x_N}\left(\sum_{j=1}^{N} a_{Mj}x_j\right) \end{bmatrix} = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{N1} \\ a_{12} & a_{22} & \cdots & a_{N2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1M} & a_{2M} & \cdots & a_{NM} \end{bmatrix}$$

$$\boxed{\frac{\partial \boldsymbol{Ax}}{\partial \boldsymbol{x}} = \boldsymbol{A}^{\mathrm{T}}}$$

Derivada de $x^{\mathrm{T}}Ax$

- Considerando a mesma matriz $\boldsymbol{A} \in \mathbb{C}^{M\times N}$ e um vetor$\boldsymbol{x} \in \mathbb{C}^{N\times 1}$ do caso anterior, temos que:

$$
\begin{aligned}
\frac{\partial \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}} &= \frac{\partial}{\partial \boldsymbol{x}}\left(\begin{bmatrix} x_1 & x_2 & \dots & x_N \end{bmatrix}\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1N} \\ a_{21} & a_{22} & \dots & a_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{N1} & a_{N2} & \dots & a_{NN} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}\right) \\
&= \frac{\partial}{\partial \boldsymbol{x}}\left(\begin{bmatrix} \sum_{i=1}^{N} x_i a_{i1} & \sum_{i=1}^{N} x_i a_{i2} & \dots & \sum_{i=1}^{N} x_i a_{iN} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}\right) \\
&= \frac{\partial}{\partial \boldsymbol{x}}\left(\sum_{i=1}^{N}\sum_{j=1}^{N} x_i a_{ij} x_j\right)
\end{aligned}
$$

Derivada de $x^{\mathrm{T}}Ax$

$$\frac{\partial \mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial}{\partial x_1}\left(\sum_{i=1}^{N}\sum_{j=1}^{N} x_i a_{ij} x_j\right) \\ \frac{\partial}{\partial x_2}\left(\sum_{i=1}^{N}\sum_{j=1}^{N} x_i a_{ij} x_j\right) \\ \vdots \\ \frac{\partial}{\partial x_N}\left(\sum_{i=1}^{N}\sum_{j=1}^{N} x_i a_{ij} x_j\right) \end{bmatrix} = \begin{bmatrix} 2x_1 a_{11} + \sum\limits_{\substack{j=1 \\ j\neq 1}}^{N} a_{1j} x_j + \sum\limits_{\substack{i=1 \\ i\neq 1}}^{N} a_{i1} x_i \\ 2x_2 a_{22} + \sum\limits_{\substack{j=1 \\ j\neq 2}}^{N} a_{2j} x_j + \sum\limits_{\substack{i=1 \\ i\neq 2}}^{N} a_{i2} x_i \\ \vdots \\ 2x_N a_{NN} + \sum\limits_{\substack{j=1 \\ j\neq N}}^{N} a_{Nj} x_j + \sum\limits_{\substack{i=1 \\ i\neq N}}^{N} a_{iN} x_i \end{bmatrix}$$

$$= \begin{bmatrix} \sum\limits_{j=1}^{N} a_{1j} x_j + \sum\limits_{i=1}^{N} a_{i1} x_i \\ \sum\limits_{j=1}^{N} a_{2j} x_j + \sum\limits_{i=1}^{N} a_{i2} x_i \\ \vdots \\ \sum\limits_{j=1}^{N} a_{Nj} x_j + \sum\limits_{i=1}^{N} a_{iN} x_i \end{bmatrix}$$

Derivada de $x^{\mathrm{T}}Ax$

- Que pode ser escrito matricialmente como:

$$\frac{\partial \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{x} + \boldsymbol{A}\boldsymbol{x}$$

ou seja,

$$\boxed{\frac{\partial \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}} = (\boldsymbol{A}^{\mathrm{T}} + \boldsymbol{A})\boldsymbol{x}}$$

- No caso de $\boldsymbol{A}$ ser uma matriz simétrica, então $\boldsymbol{A} = \boldsymbol{A}^{\mathrm{T}}$, logo:

$$\boxed{\frac{\partial \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}} = 2\boldsymbol{A}\boldsymbol{x}}$$

Derivada do Traço de uma matriz X

- Seja uma matriz $\boldsymbol{A} \in \mathbb{C}^{N\times M}$ e uma matriz $\boldsymbol{x} \in \mathbb{C}^{M\times N}$, temos que a derivada do traço de $\boldsymbol{A}\boldsymbol{X}$ é dada por:

$$
\begin{aligned}
\frac{\partial \operatorname{tr}(\boldsymbol{A}\boldsymbol{X})}{\partial \boldsymbol{X}} &= \frac{\partial}{\partial \boldsymbol{X}}\left(\operatorname{tr}\left(\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1M} \\ a_{21} & a_{22} & \cdots & a_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ a_{N1} & a_{N2} & \cdots & a_{NM} \end{bmatrix}\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1N} \\ x_{21} & x_{22} & \cdots & x_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ x_{M1} & x_{M2} & \cdots & x_{MN} \end{bmatrix}\right)\right) \\
&= \frac{\partial}{\partial \boldsymbol{X}}\left(\sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij}x_{ji}\right) \\
&= \begin{bmatrix}
\frac{\partial}{\partial x_{11}}\left(\sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij}x_{ji}\right) & \frac{\partial}{\partial x_{12}}\left(\sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij}x_{ji}\right) & \cdots & \frac{\partial}{\partial x_{1N}}\left(\sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij}x_{ji}\right) \\
\frac{\partial}{\partial x_{21}}\left(\sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij}x_{ji}\right) & \frac{\partial}{\partial x_{22}}\left(\sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij}x_{ji}\right) & \cdots & \frac{\partial}{\partial x_{2N}}\left(\sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij}x_{ji}\right) \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial}{\partial x_{M1}}\left(\sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij}x_{ji}\right) & \frac{\partial}{\partial x_{M2}}\left(\sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij}x_{ji}\right) & \cdots & \frac{\partial}{\partial x_{MN}}\left(\sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij}x_{ji}\right)
\end{bmatrix}
\end{aligned}
$$

Derivada do Traço de uma matriz X

- Resultando em

$$\frac{\partial\, \mathrm{tr}\,(\boldsymbol{A}\boldsymbol{X})}{\partial \boldsymbol{X}} = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{N1} \\ a_{12} & a_{22} & \cdots & a_{N2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1M} & a_{2M} & \cdots & a_{NM} \end{bmatrix}$$

$$\boxed{\frac{\partial\, \mathrm{tr}\,(\boldsymbol{A}\boldsymbol{X})}{\partial \boldsymbol{X}} = \boldsymbol{A}^{\mathrm{T}}}$$

Derivada do Determinante de uma matriz X

- Seja uma matriz $\boldsymbol{X} \in \mathbb{C}^{N\times N}$ quadrada.
- Através da expansão de Laplace (expansão em cofatores), podemos reescrever o determinante de $\boldsymbol{X}$ como a soma dos cofatores de uma linha ou coluna qualquer, multiplicada pelo seu elemento gerador, ou seja,

$$\det\left(\boldsymbol{X}\right) = \sum_{i=1}^{N} x_{ki} \det\left(\boldsymbol{C}_{ki}\right) = \sum_{i=1}^{N} x_{ik} \det\left(\boldsymbol{C}_{ik}\right) \quad \forall k \in (1, N),$$

  em que $\boldsymbol{C}_{ij}$ representa o cofator da matriz $\boldsymbol{X}$ gerado a partir do elemento x_{ij}.
- Vale salientar que o cofator $\det\left(\boldsymbol{C}_{ij}\right)$ independe do valor de qualquer elemento da linha i ou da coluna j de $\boldsymbol{X}$.

Derivada do Determinante de uma matriz X

- Assim, temos que:

$$\frac{\partial \det(\boldsymbol{X})}{\partial \boldsymbol{X}} = \frac{\partial}{\partial \boldsymbol{X}} \left(\sum_{i=1}^{N} x_{ki} \det(\boldsymbol{C}_{ki}) \right) \quad \forall k \in (1, N)$$

$$= \frac{\partial}{\partial \boldsymbol{X}} \left(\begin{bmatrix} \sum_{i=1}^{N} x_{1i} \det(\boldsymbol{C}_{1i}) & \sum_{i=1}^{N} x_{1i} \det(\boldsymbol{C}_{1i}) & \dots & \sum_{i=1}^{N} x_{1i} \det(\boldsymbol{C}_{1i}) \\ \sum_{i=1}^{N} x_{2i} \det(\boldsymbol{C}_{2i}) & \sum_{i=1}^{N} x_{2i} \det(\boldsymbol{C}_{2i}) & \dots & \sum_{i=1}^{N} x_{2i} \det(\boldsymbol{C}_{2i}) \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{N} x_{Ni} \det(\boldsymbol{C}_{Ni}) & \sum_{i=1}^{N} x_{Ni} \det(\boldsymbol{C}_{Ni}) & \dots & \sum_{i=1}^{N} x_{Ni} \det(\boldsymbol{C}_{Ni}) \end{bmatrix} \right)$$

$$= \begin{bmatrix} \frac{\partial}{\partial x_{11}} \left(\sum_{i=1}^{N} x_{1i} \det(\boldsymbol{C}_{1i}) \right) & \frac{\partial}{\partial x_{12}} \left(\sum_{i=1}^{N} x_{1i} \det(\boldsymbol{C}_{1i}) \right) & \dots & \frac{\partial}{\partial x_{13}} \left(\sum_{i=1}^{N} x_{1i} \det(\boldsymbol{C}_{1i}) \right) \\ \frac{\partial}{\partial x_{21}} \left(\sum_{i=1}^{N} x_{2i} \det(\boldsymbol{C}_{2i}) \right) & \frac{\partial}{\partial x_{22}} \left(\sum_{i=1}^{N} x_{2i} \det(\boldsymbol{C}_{2i}) \right) & \dots & \frac{\partial}{\partial x_{33}} \left(\sum_{i=1}^{N} x_{2i} \det(\boldsymbol{C}_{2i}) \right) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial}{\partial x_{N1}} \left(\sum_{i=1}^{N} x_{Ni} \det(\boldsymbol{C}_{Ni}) \right) & \frac{\partial}{\partial x_{N2}} \left(\sum_{i=1}^{N} x_{Ni} \det(\boldsymbol{C}_{Ni}) \right) & \dots & \frac{\partial}{\partial x_{N3}} \left(\sum_{i=1}^{N} x_{Ni} \det(\boldsymbol{C}_{Ni}) \right) \end{bmatrix}$$

Derivada do Determinante de uma matriz X

- Resultando em:

$$\frac{\partial \det(\boldsymbol{X})}{\partial \boldsymbol{X}} = \begin{bmatrix} \det(\boldsymbol{C}_{11}) & \det(\boldsymbol{C}_{12}) & \dots & \det(\boldsymbol{C}_{1N}) \\ \det(\boldsymbol{C}_{21}) & \det(\boldsymbol{C}_{22}) & \dots & \det(\boldsymbol{C}_{2N}) \\ \vdots & \vdots & \ddots & \vdots \\ \det(\boldsymbol{C}_{N1}) & \det(\boldsymbol{C}_{N2}) & \dots & \det(\boldsymbol{C}_{NN}) \end{bmatrix}$$

- A matriz $N \times N$ composta pelo determinante dos cofatores de $\boldsymbol{X}$ é denominada matriz adjunta de $\boldsymbol{X}$, ou simplesmente $\mathrm{adj}\,(\boldsymbol{X})$.

$$\boxed{\frac{\partial \det(\boldsymbol{X})}{\partial \boldsymbol{X}} = \mathrm{adj}\,(\boldsymbol{X})}$$

Conteúdo

Bibliografia

[BV04] S. Boyd and L. Vandenberghe, *Convex optimization*, 1st ed. Cambridge University Press, 2004.

[Che99] C.-T. Chen, *Linear system theory and design*, 3rd ed. Oxford Univeristy Press, 1999.

[dF10] A. B. H. de Ferreira, *Dicionário do Aurélio*, 5th ed. Positivo Editora, 2010.

[GS08] A. Gilat and V. Subramanian, *Métodos numéricos para engenheiros e cientistas: uma introdução com aplicações usando o MATLAB*, 1st ed. Bookman, 2008.

[HJ12] R. A. Horn and C. R. Johnson, *Matrix Analysis*, 2nd ed. Cambridge University Press, 2012.

[MK06] T. F. Maciel and A. Klein, "A low-complexity SDMA grouping strategy for the downlink of Multi-User MIMO systems," in *Proceedings of the IEEE Personal, Indoor and Mobile Radio Communications (PIMRC)*, sep 2006.

[NW06] J. Nocedal and S. J. Wright, *Numerical optimization*, 2nd ed., T. V. Mikosh, S. M. Robinson, and S. I. Resnick, Eds. Springer, 2006.

[PP08] K. B. Petersen and M. S. Pedersen, "The matrix cookbook," Technical University of Denmark, oct 2008, version 20081110. [Online]. Available: http://www2.imm.dtu.dk/pubdb/p.php?3274

[Rug96] W. J. Rugh, *Linear system theory*, 2nd ed., ser. Information and systems sciences, T. Kailath, Ed. Prentice Hall, 1996.

[Str88] G. Strang, *Linear algebra and its applications*, 3rd ed. Harcourt College Publishers, 1988.