

Algebra Linear

1. Sistema de Equações Lineares
2. Ortonormalização

Sistema de Equações Lineares

$$\underbrace{\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}}_A \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}}_x = \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}}_y$$

- ↪ $a_{ij}, y_i \in \mathbb{R}$: dados do sistema; $x_i \in \mathbb{R}$: incógnitas
- ↪ Solução para $Ax = y$ quando $m > n$, $m = n$ ou $m < n$
- ↪ Problema: dados A e y
 1. $\exists x : Ax = y?$
 2. Se existe solução, qual o número de soluções LI?

Sistema de Equações Lineares

Definição O espaço **Range** da matriz A é o espaço gerado pelas colunas de A :

$$\mathcal{R}(A) \triangleq \left\{ \mathbf{y} = A\mathbf{x} : \mathbf{x} \in \mathbb{R}^n \right\} \subseteq \mathbb{R}^m$$

Exemplo $\mathbf{y} = \begin{bmatrix} 1 & -1 \end{bmatrix}^T$ está no $\mathcal{R}(A)$, com

$$A = \begin{bmatrix} 1 & -1 & -3 \\ 0 & 10 & 0 \end{bmatrix} ?$$

Em outras palavras, $\mathbf{y} = A\mathbf{x}$ para algum \mathbf{x} ? Resposta positiva, pois as colunas de A geram todo o espaço 2-dimensional

Sistema de Equações Lineares

Rank (ou posto) da matriz $A \in \mathbb{R}^{m \times n}$ é definido como a dimensão do range de A (ou, equivalentemente, é dado pelo número de colunas LI de A , gerando o espaço \mathbb{R}^m). O posto é denotado por $\rho(A)$

- posto de colunas de A é a dimensão do $\mathcal{R}(A)$
- posto de linhas de A é a dimensão do $\mathcal{R}(A^*)$
- $\dim \mathcal{R}(A) = \dim \mathcal{R}(A^*) = r = \text{posto}(A)$ ou $\text{rank}(A)$

Sistema de Equações Lineares

Definição O espaço nulo de A é o espaço gerado por todos os vetores x satisfazendo $Ax = 0$, i.e.:

$$\mathcal{N}(A) \triangleq \{x \in \mathbb{R}^n : Ax = 0\} \subseteq \mathbb{R}^n$$

Exemplo $x = \begin{bmatrix} 1 & 10 & 0 \end{bmatrix}^T$ está no espaço $\mathcal{N}(A)$ com $A = \begin{bmatrix} 1 & -1 & -3 \\ 0 & 10 & 0 \end{bmatrix}$?

Em outras palavras, $Ax = 0$? Note que

$$\begin{bmatrix} 1 & -1 & -3 \\ 0 & 10 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 10 \\ 0 \end{bmatrix} = \begin{bmatrix} -9 \\ 100 \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

e portanto a resposta é negativa

- A dimensão do espaço nulo é chamada de **nulidade** da matriz A e denotada por $\nu(A)$

Sistema de Equações Lineares

Note que $y = Ax$ pode ser escrita como $y = x_1a_1 + x_2a_2 + \cdots + x_na_n$ e, portanto, $x_i \in \mathbb{R}$, $i = 1, \dots, n$ são as ponderações das colunas de A

Propriedade

↪ Para $A \in \mathbb{R}^{m \times n}$:

$$\nu(A) = n - \rho(A)$$

$$\begin{aligned} \text{rank de } A \quad (\rho(A)) &= \text{número de colunas LI de } A \\ &= \text{número de linhas LI de } A \\ &\leq \min(n, m) \end{aligned}$$

Sistema de Equações Lineares

↪ Se $\nu(A) = 0$ então $\mathcal{N}(A) = \{0\}$ e as seguintes afirmações são equivalentes:

- x pode ser determinado de maneira única de $y = Ax$
- colunas de A são LI
- $\det(A'A) \neq 0$

↪ Se $\nu(A) = k$, então $Ax = 0$ possui k soluções LI

Sistema de Equações Lineares

Exemplo: Considere a matriz $A \in \mathbb{R}^{3 \times 5}$ dada por

$$A = \begin{bmatrix} 0 & 1 & 1 & 2 & -1 \\ 1 & 2 & 3 & 4 & -1 \\ 2 & 0 & 2 & 0 & 2 \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & a_3 & a_4 & a_5 \end{bmatrix}$$

$$Ax = x_1 \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} + x_2 \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} + x_4 \begin{bmatrix} 2 \\ 4 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} -1 \\ -1 \\ 2 \end{bmatrix}$$

$$a_3 = a_1 + a_2 ; \quad a_4 = 2a_2 ; \quad a_5 = a_3 - a_4 = a_1 + a_2 - 2a_2 = a_1 - a_2$$

$$Ax = (x_1 + x_3 + x_5)a_1 + (x_2 + x_3 + 2x_4 - x_5)a_2$$

Como a_1 e a_2 são LI $\Rightarrow \rho(A) = 2$

Sistema de Equações Lineares

$$Ax = 0 \iff \begin{cases} x_1 + x_3 + x_5 = 0 \\ x_2 + x_3 + 2x_4 - x_5 = 0 \end{cases}$$

Número de Equações = $\rho(A) = 2$

Número de Incógnitas = 5

Número de Graus de Liberdade = 3

Possíveis soluções LI:

$$v_1 = \begin{bmatrix} -1 \\ -1 \\ 1 \\ 0 \\ 0 \end{bmatrix} ; v_2 = \begin{bmatrix} 0 \\ -2 \\ 0 \\ 1 \\ 0 \end{bmatrix} , v_3 = \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

$\{v_1, v_2, v_3\}$ formam uma base de $\mathcal{N}(A)$. Nulidade: $\nu(A) = 3$

Sistema de Equações Lineares

Teorema

1. Dada uma matriz $A \in \mathbb{R}^{m \times n}$ e um vetor $y \in \mathbb{R}^{m \times 1}$, existe uma solução $x \in \mathbb{R}^{n \times 1}$ da equação $y = Ax$ se e somente se $y \in \mathcal{R}(A)$ ou, de forma equivalente:

$$\rho(A) = \rho \left(\begin{bmatrix} A & y \end{bmatrix} \right)$$

2. Dada uma matriz $A \in \mathbb{R}^{m \times n}$, uma solução $x \in \mathbb{R}^n$ de $y = Ax$ existe para todo $y \in \mathbb{R}^m$ se e somente se $\rho(A) = m$ (rank completo de linhas)

Sistema de Equações Lineares

Teorema (Parametrização de todas as soluções) Dada uma matriz $A \in \mathbb{R}^{m \times n}$ e um vetor $y \in \mathbb{R}^m$, seja x_p uma solução $x \in \mathbb{R}^n$ da equação $y = Ax$. Seja $k = n - \rho(A) = \nu(A)$ a nulidade de A .

\rightsquigarrow Se $\rho(A) = n$ (rank completo de colunas) então a solução x_p é única

\rightsquigarrow Se $k > 0$, então para todo $\alpha_i \in \mathbb{R}$, $i = 1, \dots, k$ o vetor

$$x = x_p + \alpha_1 n_1 + \alpha_2 n_2 + \dots + \alpha_k n_k = x_p + \sum_{i=1}^k \alpha_i n_i$$

é uma solução de $Ax = y$, sendo $\{n_1, \dots, n_k\}$ uma base de $\mathcal{N}(A)$

Note que da última condição: $Ax = Ax_p + A \sum_{i=1}^k \alpha_i n_i = Ax_p + \mathbf{0} = y$

Sistema de Equações Lineares

Exemplo Considere a descrição:

$$Ax = \begin{bmatrix} 0 & 1 & 1 & 2 \\ 1 & 2 & 3 & 4 \\ 2 & 0 & 2 & 0 \end{bmatrix} x = \begin{bmatrix} a_1 & a_2 & a_3 & a_4 \end{bmatrix} x = \begin{bmatrix} -4 \\ -8 \\ 0 \end{bmatrix} = y$$

Note que como $y \in \mathcal{R}(A)$ então $\exists x$ solução para $Ax = y$ (item 1 do Teorema na pag. 10). Resolvendo para A e y , obtém-se uma solução particular:

$$x_p = \begin{bmatrix} 0 \\ 0 \\ 0 \\ -2 \end{bmatrix}. \quad \text{Outra solução?} \quad x_p = \begin{bmatrix} 0 \\ -4 \\ 0 \\ 0 \end{bmatrix}$$

Matlab:

$$A=[0 \ 1 \ 1 \ 2; 1 \ 2 \ 3 \ 4; 2 \ 0 \ 2 \ 0]; \ y=[-4; -8; 0]; \ xp=linsolve(A,y)$$

Sistema de Equações Lineares

Note que a matriz $A \in \mathbb{R}^{3 \times 4}$ tem posto $\rho(A) = 2$. E.g., a_1 e a_2 são LI. A 3a. coluna é combinação de $a_3 = a_1 + a_2$ e a 4a. coluna é $a_4 = 2a_2$. Desta forma:

$$Ax = (x_1 + x_3)a_1 + (x_2 + x_3 + 2x_4)a_2$$

Note que a nulidade de A é $\nu(A) = n - \rho(A) = 4 - 2 = 2$ e, então, tem-se dois vetores que geram o espaço nulo calculados de:

$$Ax = 0 \iff \begin{cases} x_1 + x_3 = 0 \\ x_2 + x_3 + 2x_4 = 0 \end{cases}$$

Possíveis soluções LI que formam uma base para $\mathcal{N}(A)$ (há diversas soluções!):

$$n_1 = \begin{bmatrix} 1 \\ 1 \\ -1 \\ 0 \end{bmatrix}; n_2 = \begin{bmatrix} 0 \\ 2 \\ 0 \\ -1 \end{bmatrix}. \text{ Usando } \text{null}(A, 'r'): n_1 = \begin{bmatrix} -1 \\ -1 \\ 1 \\ 0 \end{bmatrix}; n_2 = \begin{bmatrix} 0 \\ -2 \\ 0 \\ 0 \end{bmatrix}$$

Sistema de Equações Lineares

Aglutinando estas informações, via Teorema na pág. 11, tem-se uma solução geral para $Ax = y$ da forma:

$$x = x_p + \alpha_1 n_1 + \alpha_2 n_2 = x_p = \begin{bmatrix} 0 \\ 0 \\ 0 \\ -2 \end{bmatrix} + \alpha_1 \begin{bmatrix} 1 \\ 1 \\ -1 \\ 0 \end{bmatrix} + \alpha_2 \begin{bmatrix} 0 \\ 2 \\ 0 \\ -1 \end{bmatrix}$$

para quaisquer valores de $\alpha_1 \in \mathbb{R}$ e $\alpha_2 \in \mathbb{R}$

Notas gerais

↪ Seja $A \in \mathbb{R}^{m \times n}$. Se $C \in \mathbb{R}^{n \times n}$ e $D \in \mathbb{R}^{m \times m}$ são não singulares, então:

$$\rho(AC) = \rho(A) \quad \text{e} \quad \rho(DA) = \rho(A)$$

Isto é, o posto de uma matriz não se altera ao ser pré ou pós-multiplicada por uma matriz não singular

↪ Seja $A \in \mathbb{C}^{m \times n}$ e A^* sua conjugada transposta. Então

▶ $\rho(A) = n \iff \rho(A^*A) = n \ ; \ \det(A^*A) \neq 0$

▶ $\rho(A) = m \iff \rho(AA^*) = m \ ; \ \det(AA^*) \neq 0$

↪ Observe que $\rho(A) = n$ implica:

▶ $n \leq m$

▶ $A\alpha = 0 \implies \alpha = 0$

Notas gerais

↪ Para uma matriz quadrada $A \in \mathbb{R}^{n \times n}$, A é invertível ou não-singular se $\det(A) \neq 0$ ou, de forma equivalente:

- ▶ as colunas de A formam uma base para o \mathbb{R}^n
- ▶ as linhas de A formam uma base para o \mathbb{R}^n
- ▶ a equação $y = Ax$ tem uma solução única $x = A^{-1}y$ para todo $y \in \mathbb{R}^n$
- ▶ $\mathcal{N}(A) = \{0\}$ (a única solução de $Ax = 0$ é $x = 0$)
- ▶ $\mathcal{R}(A) = \mathbb{R}^n$
- ▶ $\det(A'A) = \det(AA') \neq 0$

Sistema de Equações Lineares

↪ Considere o sistema $y = Ax$ com $A \in \mathbb{R}^{m \times n}$ e $m > n$, i.e., há mais equações do que incógnitas (pode não ter solução)...

Aproximação – define-se o erro: $e = Ax - y$ e busca-se x^* que minimiza $\|e\|$.

A solução x^* gera o ponto no $\mathcal{R}(A)$ que está mais próximo de y , ou seja, Ax_a é a projeção de y no $\mathcal{R}(A)$. Escreva na forma de um problema de otimização:

$$f(x) = \min_x \|e\|^2 = \min_x e'e = \min_x (Ax - y)'(Ax - y)$$

Condição de Otimalidade – derivando em relação a x : $2x'A'A - 2y'A = 0$

► Assumindo que A tem posto completo $\rho(A) = n$ (e portanto $A'A$ é invertível), então uma solução aproximada é:

$$x^* = (A'A)^{-1}A'y$$

Sistema de Equações Lineares

► x^* é a solução exata de $y = Ax$ se $y \in \mathcal{R}(A)$

↪ $A^\dagger \triangleq (A'A)^{-1}A'$ é chamada de pseudo-inversa de A (inversa à esquerda)

↪ A projeção de y no $\mathcal{R}(A)$ é linear e é dada por

$$Ax^* = A(A'A)^{-1}A'y$$

e $A(A'A)^{-1}A'$ é chamada matriz de projeção

↪ Note que o erro:

$$e = Ax^* - y = [A(A'A)^{-1}A' - I]y$$

é ortogonal ao $\mathcal{R}(A)$, i.e.,

$$\langle e, Ax \rangle = y' [A(A'A)^{-1}A' - I]' Ax = 0$$

Estimador Mínimos Quadrados

↪ Vários problemas em engenharia envolvendo reconstrução de informação, inversão ou estimação podem ser colocados na forma $\mathbf{y} = \mathbf{A}\mathbf{x} + \Delta$, sendo \mathbf{x} o vetor a ser estimado ou reconstruído; \mathbf{y} o vetor de informações ou medidas e Δ o vetor de ruídos, erros de medida ou incerteza não modelada

Problema – Obter a estimativa $\hat{\mathbf{x}}$ que minimiza a diferença entre os valores medidos \mathbf{y} e o que deveria ser obtido se $\Delta = \mathbf{0}$

Solução: $\hat{\mathbf{x}} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{y}$

Sistemas de Equações Lineares

↪ Para o sistema $y = Ax$ com $A \in \mathbb{R}^{m \times n}$ e $n > m$, considere o problema de otimização que minimiza a norma-2 de x , i.e.:

$$\begin{cases} \min_x & \|x\|^2 \quad (= x'x) \\ \text{sujeito a} & Ax = y \end{cases}$$

Como resolvê-lo? Introduza o multiplicador de Lagrange λ e o Lagrangeano:

$$\mathbb{L}(x, \lambda) = x'x + \lambda'(Ax - y)$$

As condições de otimalidade são obtidas derivando o Lagrangeano:

$$\frac{\partial \mathbb{L}}{\partial x} = 2x' + \lambda'A = 0 \quad \text{e} \quad \frac{\partial \mathbb{L}}{\partial \lambda} = (Ax - y) = 0$$

Veja que $x = -A'\lambda/2$. Como $(Ax - y)' = 0 \Rightarrow \lambda = -2(AA')^{-1}y$, então:

$$x^* = A'(AA')^{-1}y \quad (\exists x^* \text{ se } \rho(A) = m \text{ implicando } \rho(AA') = m)$$

Ortonormalização

↪ Um vetor x é dito estar normalizado se sua norma Euclidiana é igual a 1, i.e.,
 $x'x = \langle x, x \rangle = 1$

Dois vetores x_1 e x_2 são ortogonais se $x_1'x_2 = x_2'x_1 = 0$

Um conjunto de vetores $\{x_1, x_2, \dots, x_n\}$ é ortonormal se

$$x_i'x_j = \begin{cases} 0 & \text{se } i \neq j \\ 1 & \text{se } i = j \end{cases}$$

Ortonormalização

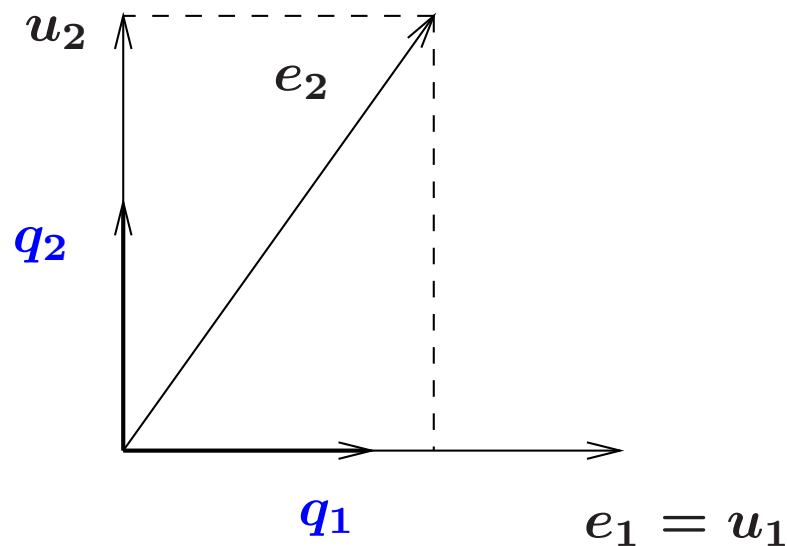
↪ Dado um conjunto de vetores LI $\{e_1, e_2, \dots, e_n\}$, pode-se obter um conjunto de vetores ortonormais usando o procedimento de ortonormalização de Schmidt:

$$\begin{array}{ll} u_1 \triangleq e_1 & q_1 \triangleq u_1 / \|u_1\| \\ u_2 \triangleq e_2 - (q_1' e_2) q_1 & q_2 \triangleq u_2 / \|u_2\| \\ \vdots & \vdots \\ u_n \triangleq e_n - \sum_{k=1}^{n-1} (q_k' e_n) q_k & q_n \triangleq u_n / \|u_n\| \end{array}$$

↪ A primeira equação apenas normaliza o vetor e_1 , tal que $q_1' q_1 = 1$

Ortonormalização

↷ Na segunda equação, $(q'_1 e_2)q_1$ é a projeção do vetor e_2 ao longo de q_1 . E para $u_2 = e_2 - (q'_1 e_2)q_1$, $q_2 = u_2/\|u_2\|$ é ortogonal ao vetor q_1



Ortonormalização - Propriedades

↪ Se um conjunto de vetores u_1, u_2, \dots, u_n é ortonormal então

$$U \triangleq \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix} \Rightarrow U'U = I_n$$

Propriedade: Se $y = Ux$, então $\sum_{i=1}^n y_i^2 = \sum_{i=1}^n x_i^2$ ou, em outras palavras:

$$\|y\|^2 = \|Ux\|^2 = (Ux)'(Ux) = x' \underbrace{U'U}_{=I} x = x'x = \|x\|^2$$

Neste caso, o mapeamento $y = Ux$ é uma isometria (não altera a norma)

Propriedade: Se $y = Ux$ e $\tilde{y} = U\tilde{x}$ então $\langle y, \tilde{y} \rangle = \langle x, \tilde{x} \rangle$ (a multiplicação por U não altera o produto interno) pois

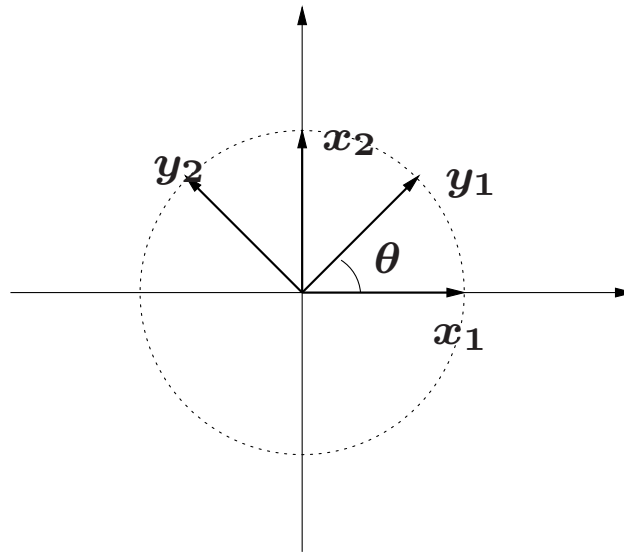
$$\langle y, \tilde{y} \rangle = \langle Ux, U\tilde{x} \rangle = (Ux)'(U\tilde{x}) = x' \underbrace{U'U}_{=I} \tilde{x} = \langle x, \tilde{x} \rangle$$

Ortonormalização

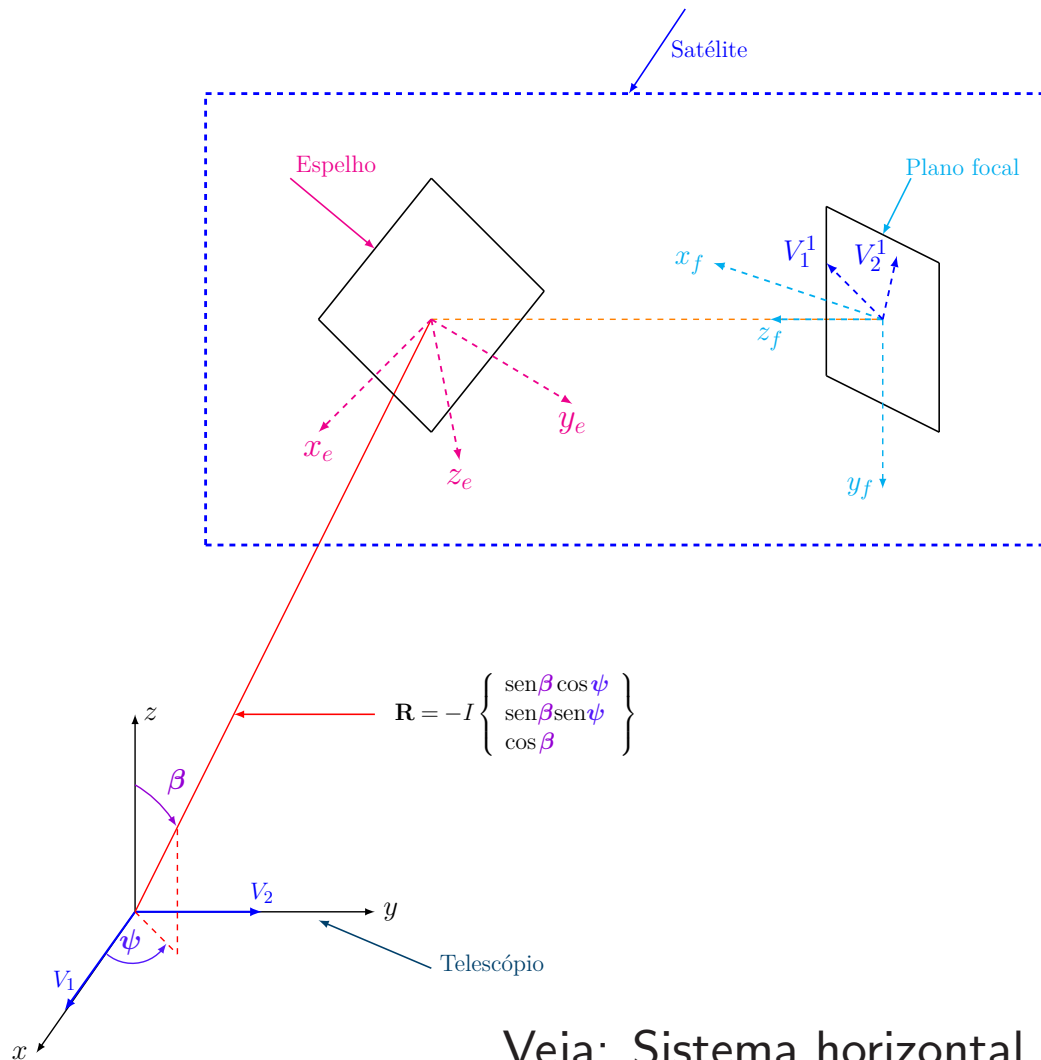
Exemplo: Transformações de rotação, reflexão e translação de vetores via matriz ortogonal

↪ A transformação que rotaciona um vetor no sentido anti-horário de θ no \mathbb{R}^2 é:

$$y = U_{\theta}x \quad ; \quad U_{\theta} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \quad ; \quad U_{\theta}'U_{\theta} = I$$



Rastreamento de satélite – Telescópio terrestre



Quatro sistemas de coordenadas

- 1) Coordenadas na Terra (x, y, z)
- 2) Coordenadas do Satélite
- 3) Espelho (x_e, y_e, z_e)
- 4) Plano focal (x_f, y_f, z_f)

Posição do Satélite

definida na figura em relação a:

- azimute ψ
- zênite β
- inclinação I

Projeção, Rotação e Reflexão

Veja: Sistema horizontal de coordenadas

PCA – Análise de Componentes Principais

Outro Exemplo – Encontrar projeções que minimizem o erro de reconstrução com base ortonormal

▷ A ideia é projetar um conjunto de dados no espaço n-dimensional em outro de dimensão menor, preservando o máximo de informação possível. Obter a melhor aproximação, por exemplo, de dados em 3D no 2D

▷ Para quê?

↪ Comprimir, redução de ruído

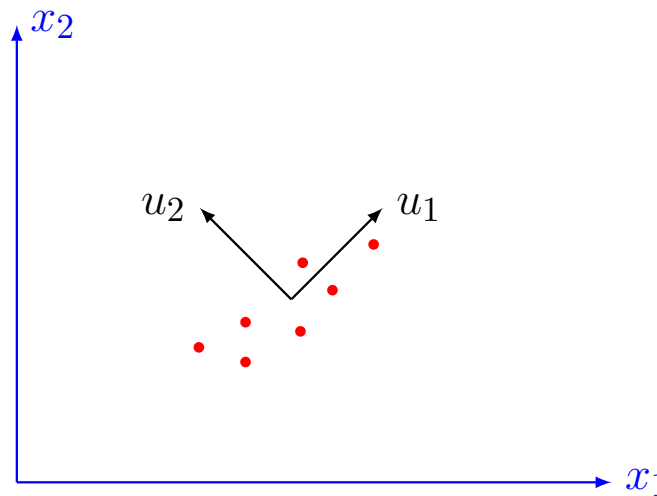
↪ Reduzir a dimensão do espaço de características para fins de classificação

↪ Pode-se usar uma projeção que minimiza o erro quadrático em reconstruir os dados originais

Encontrar Projeções que minimizem o erro de reconstrução

Por exemplo, suponha que os dados são representados por um conjunto de ϕ vetores com dimensão n cada um deles: $\chi = \{x'_1, x'_2, \dots, x'_\phi\}$

É possível escrever o k -ésimo vetor da forma: $x_k = \sum_{i=1}^{\text{posto}(\chi)} z_i^k u_i$, sendo que os vetores u_i formam um conjunto de vetores ortonormais

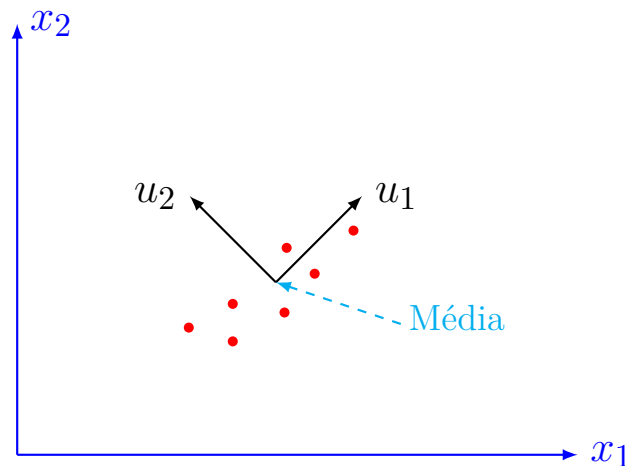


Encontrar Projeções que minimizem o erro de reconstrução

PCA: Dado $\zeta < \phi$ encontre o conjunto de vetores ortonormais $\{u_1, u_2, \dots, u_\zeta\}$ tal que minimize o erro quadrático:

$$J_\zeta = \sum_{k=1}^{\phi} \|x_k - \hat{x}_k\|^2$$

sendo que $\hat{x}_k = \bar{x} + \sum_{i=1}^{\zeta} z_i^k u_i$ e \bar{x} é a média dada por $\bar{x} = \frac{1}{\phi} \sum_{i=1}^{\phi} x_i$



Ortonormalização

↪ O procedimento de ortonormalização pode ser descrito como uma transformação linear. Considere a matriz X formada por n vetores LI, $\{x_1, x_2, \dots, x_n\}$, da forma: $X = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}$

Deseja-se determinar uma matriz F tal que

$$XF = Q \triangleq \begin{bmatrix} q_1 & q_2 & \cdots & q_n \end{bmatrix}$$

A condição para que os vetores $\{q_1, q_2, \dots, q_n\}$ sejam ortonormais pode ser escrita na forma

$$Q'Q = I \implies F'RF = I \ ; \ R \triangleq X'X$$

Sistema com $n(n+1)/2$ restrições e n^2 incógnitas ($n(n-1)/2$ elementos de F são arbitrários). Solução?

Ortonormalização

↪ Uma solução particular pode ser obtida pela fatorização de Cholesky da matriz R , que produz uma matriz triangular superior L tal que $R = L'L$:

$$F'RF = I = F'L'LF = (LF)'(LF)$$

e por inspeção uma solução é dada por $F = L^{-1}$

↪ A fatorização de Schur aplicada à matriz R produz uma matriz unitária V e uma matriz diagonal Ω composta pelos autovalores de R tal que $R = V\Omega V'$, sugerindo como solução

$$F'RF = I = F'V\Omega V'F = (V\Omega^{0.5}V'F)'(V\Omega^{0.5}V'F)$$

$$\therefore F = V\Omega^{-0.5}V'$$

Ortonormalização

↪ A escolha da transformação F que ortonormaliza X pode ser orientada também por algum critério específico, descrito por problemas de otimização restritos. Por exemplo, considere encontrar Q (que guarda a base ortonormal) e seja a solução para:

$$\begin{cases} \min_Q & \text{Tr} \{ (X - Q)'(X - Q) \} \\ \text{sujeito a} & Q'Q = I \end{cases}$$

i.e., Q minimiza a norma ao quadrado de $x_i - q_i$, $i = 1, \dots, n$ (denotado pelo traço). Esta transformação procura preservar o máximo possível os vetores originais. Note que $\text{Tr}(X - Q)$ é uma função linear dos elementos da matriz Q

Re-escrevendo o problema de otimização em termos de F tem-se

Ortonormalização

Re-escrevendo em termos de F

$$\begin{cases} \min_F & \text{Tr} \{ (X - XF)'(X - XF) \} \\ \text{sujeito a} & F'X'XF = I \end{cases}$$



$$\begin{cases} \min_F & \text{Tr}(X'X + F'X'XF - F'X'X - X'XF) \\ \text{sujeito a} & F'X'XF = I \end{cases}$$

Substituindo a restrição ($F'X'XF = I$) na função objetivo tem-se o termo constante ($X'X + I$). Note que termos constantes na função objetivo não influenciam na solução F , então pode-se escrever:

$$\begin{cases} \min_F & \text{Tr} (-F'X'X - X'XF) = \max \text{Tr} (2X'XF) \\ \text{sujeito a} & F'X'XF = I \end{cases} \quad \dots \text{ e a solução?}$$

Derivadas de funções matriciais

↪ Considere $X \triangleq [x_{ij}]$, $X \in \mathbb{C}^{m \times n}$ e $F(X)$ uma função escalar (real ou complexa) de X . Então

$$\frac{\partial}{\partial X} F(X) \triangleq \left[\frac{\partial}{\partial x_{ij}} F(X) \right]$$

↪ Considere A e B matrizes complexas:

$$\frac{\partial}{\partial X} \text{Tr}(AXB) = A'B'$$

$$\frac{\partial}{\partial X} \text{Tr}(AX'B) = BA$$

$$\frac{\partial}{\partial X} \text{Tr}(AXBX) = A'X'B' + B'X'A'$$

$$\frac{\partial}{\partial X} \text{Tr}(AXBX') = A'XB' + AXB$$

Ortonormalização

↪ Para obter uma solução, escreva o Lagrangeano $\mathbb{L}(\mathbf{F}, \Lambda)$ (Λ é a variável dual simétrica associada à restrição original do problema: $\mathbf{F}'\mathbf{X}'\mathbf{X}\mathbf{F} = \mathbf{I}$):

$$\mathbb{L}(\mathbf{F}, \Lambda) = \text{Tr} [2\mathbf{X}'\mathbf{X}\mathbf{F} + \Lambda (\mathbf{F}'\mathbf{X}'\mathbf{X}\mathbf{F} - \mathbf{I})]$$

As condições necessárias de otimalidade são dadas por:

$$\frac{\partial \mathbb{L}}{\partial \mathbf{F}} = 2(\mathbf{X}'\mathbf{X} + \mathbf{X}'\mathbf{X}\mathbf{F}\Lambda) = 2\mathbf{X}'\mathbf{X}(\mathbf{I} + \mathbf{F}\Lambda) = 0$$

$$\frac{\partial \mathbb{L}}{\partial \Lambda} = (\mathbf{F}'\mathbf{X}'\mathbf{X}\mathbf{F} - \mathbf{I}) = 0$$

Ortonormalização

↪ Note que X é não singular. Portanto resolvendo em termos de F e de Λ em $2X'X(I + F\Lambda) = 0$, necessariamente $2X'X \neq 0$ e então $(I + F\Lambda) = 0$, tal que:

$$F = -\Lambda^{-1} \implies FX'X'F = I \implies (-\Lambda^{-1})X'X(-\Lambda^{-1}) = I \implies X'X = \Lambda^2$$

$$\Lambda = \begin{cases} +(X'X)^{0.5} \\ -(X'X)^{0.5} \end{cases} \implies F = \begin{cases} -(X'X)^{-0.5} \\ +(X'X)^{-0.5} \end{cases}$$

Veja que o custo inicial é dado por:

$$\min \text{Tr} (-2X'XF) = \max \text{Tr} (2X'XF)$$

e a solução ótima (que minimiza) é obtida para $F = (X'X)^{-0.5}$

↪ Note que a solução $F = -(X'X)^{-0.5}$ fornece a mesma base ortonormalizada Q (apenas trocando o sinal dos vetores) e pode ser interpretada como a transformação F que maximiza a diferença entre as normas (ao quadrado) dos vetores de cada base

Ortonormalização

Exemplo Considere $X \triangleq \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}$, formada pelos vetores LI

$$x_1 = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}', \quad x_2 = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}', \quad x_3 = \begin{bmatrix} 0 & 1 & -1 \end{bmatrix}'$$

Algoritmo: $u_1 = x_1; q_1 = \frac{u_1}{\|u_1\|} = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.5774 \\ 0.5774 \\ 0.5774 \end{bmatrix}$

$$u_2 = x_2 - (q_1' x_2) q_1; \quad q_2 = \frac{u_2}{\|u_2\|} = \begin{bmatrix} -0.7071 & 0 & 0.7071 \end{bmatrix}'$$

$$u_3 = x_3 - (q_2' x_3) q_2 - (q_1' x_3) q_1; \quad q_3 = \frac{u_3}{\|u_3\|} = \begin{bmatrix} -0.4082 & 0.8165 & -0.4082 \end{bmatrix}'$$

Ortonormalização

↪ Aplicando Cholesky na matriz $R = X'X$ (MATLAB[©]):

$$L = \text{chol}(R) = \begin{bmatrix} 1.7321 & 3.4641 & 0 \\ 0 & 1.4142 & -0.7071 \\ 0 & 0 & 1.2247 \end{bmatrix} \Rightarrow L'L = R$$

$$Q = XL^{-1} = \underbrace{\begin{bmatrix} u_1 & u_2 & u_3 \end{bmatrix}}_U, Q'Q = I \quad \left\{ \begin{array}{l} \text{Uma outra alternativa.} \\ \text{Decomposição Triangular} \\ \text{Ortonormal – QR:} \\ [Q,L]=\text{qr}(X) \end{array} \right.$$

Ortonormalização

Veja que de Cholesky:

$$Q = X * \text{inv}(L) = \begin{bmatrix} 0.5774 & -0.7071 & -0.4082 \\ 0.5774 & -0.0000 & 0.8165 \\ 0.5774 & 0.7071 & -0.4082 \end{bmatrix}$$

E usando QR?

$[Q1, L1] = \text{qr}(X) \Rightarrow$ observe o sinal...

$$Q1 = \begin{bmatrix} -0.5774 & 0.7071 & 0.4082 \\ -0.5774 & 0.0000 & -0.8165 \\ -0.5774 & -0.7071 & 0.4082 \end{bmatrix}; L1 = \begin{bmatrix} -1.7321 & -3.4641 & 0 \\ 0 & -1.4142 & 0.7071 \\ 0 & 0 & -1.2247 \end{bmatrix}$$

Ortonormalização

Usando Schur: $[V, \Omega] = \text{schur}(R)$

$$V = \begin{bmatrix} 0.8850 & 0.2376 & -0.4005 \\ -0.4035 & -0.0381 & -0.9142 \\ -0.2325 & 0.9706 & 0.0622 \end{bmatrix}; \Omega = \begin{bmatrix} 0.2643 & 0 & 0 \\ 0 & 2.0392 & 0 \\ 0 & 0 & 16.6965 \end{bmatrix}$$

$$\rightsquigarrow R^{-0.5} = V\Omega^{-0.5}V' \Rightarrow RR = V * \text{sqrt}(\text{inv}(\Omega)) * V'$$

$$Q_2 = XR^{-0.5} = X * RR = \begin{bmatrix} 0.9908 & -0.0890 & -0.1021 \\ 0.1348 & 0.5758 & 0.8064 \\ 0.0130 & 0.8127 & -0.5825 \end{bmatrix}, Q_2'Q_2 = I$$

Ortonormalização

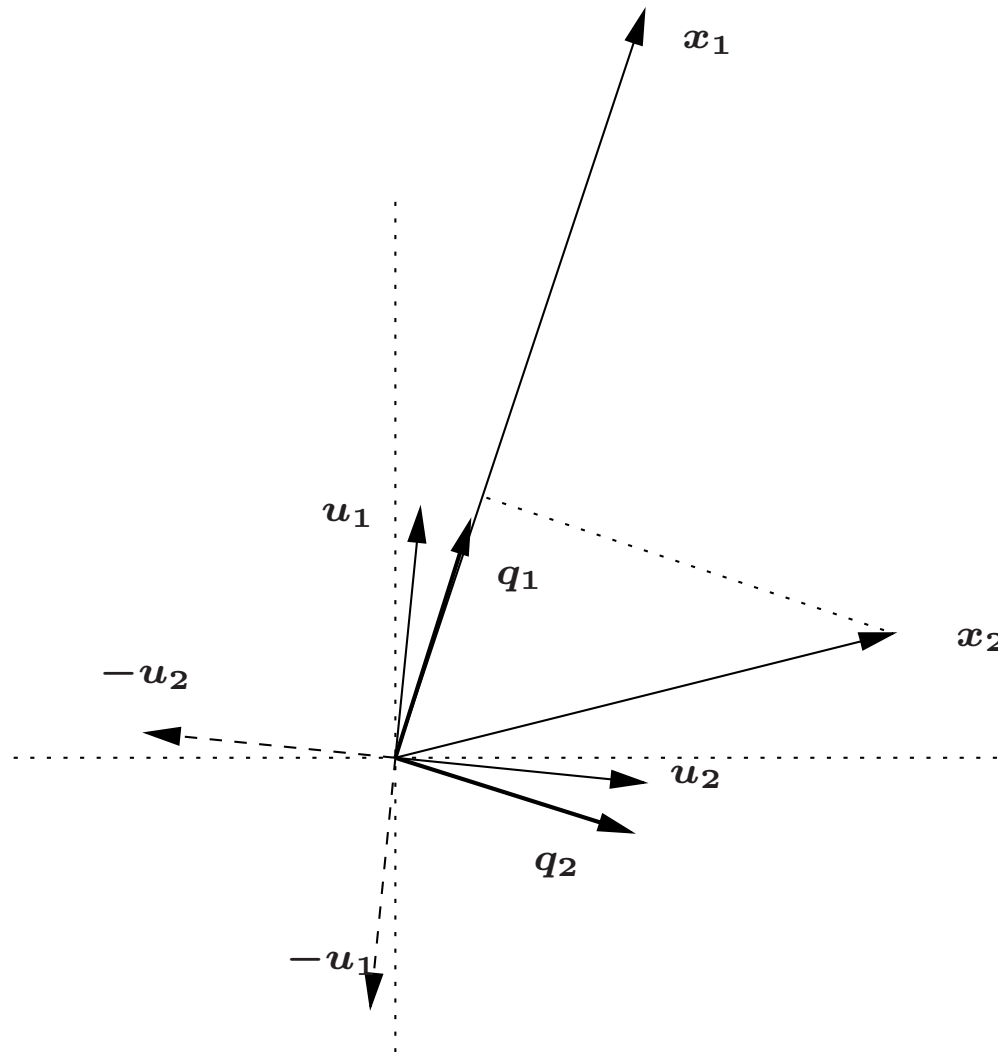
Exemplo Considere dois vetores LI no plano \mathbb{R}^2 :

$$X = \left[\begin{array}{c|c} x_1 & x_2 \end{array} \right] = \left[\begin{array}{c|c} 1 & 2 \\ 3 & 0.5 \end{array} \right]$$

Schmidt: $Q = \left[\begin{array}{c|c} q_1 & q_2 \end{array} \right] = \left[\begin{array}{c|c} 0.32 & 0.95 \\ 0.95 & -0.32 \end{array} \right]$

Schur: $U = \left[\begin{array}{c|c} u_1 & u_2 \end{array} \right] = \left[\begin{array}{c|c} 0.1 & 1 \\ 1 & -0.1 \end{array} \right]$

Ortonormalização



MATLAB

- ↪ `trace(A)` – traço de uma matriz quadrada A
- ↪ `rank(A)` – posto de uma matriz A
- ↪ `chol(A)` – fatoração de Cholesky de uma matriz A
- ↪ `qr(A)` – fatoração QR de uma matriz A
- ↪ `null(A)` – retorna uma base ortonormal para o espaço nulo de A
- ↪ `orth(A)` – retorna uma base ortonormal para o range de A
- ↪ `pinv(A)` – retorna uma matriz X com mesma dimensões de A' tal que

$$A^* X^* A = A \ ; \ X^* A^* X = X$$

e $A^* X$ e $X^* A$ são matrizes Hermitianas