Ortonormalização

Um vetor x está normalizado se sua norma Euclidiana é igual a 1, ou seja, $x'x = 1$.

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

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

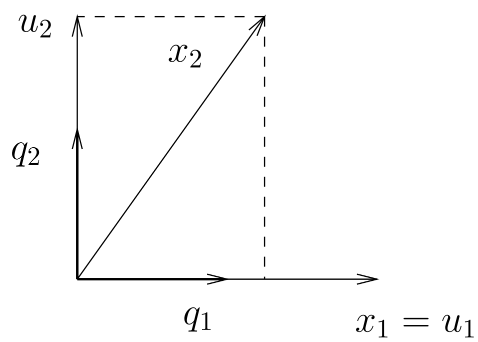
$$x_i'x_j = \begin{cases} 0 & \text{se } i \neq j \\ 1 & \text{se } i = j \end{cases}$$

Dado um conjunto de vetores LI $\{x_1, x_2, \ldots, x_n\}$, pode-se obter um conjunto de vetores ortonormais através do procedimento de ortonormalização de Schmidt:

$$\begin{array}{rclrcl} u_1 & = & x_1 & q_1 & = & u_1/\|u_1\| \\ u_2 & = & x_2 - (q_1'x_2)q_1 & q_2 & = & u_2/\|u_2\| \\ \vdots & & & \vdots & & \\ u_n & = & x_n - \displaystyle\sum_{k=1}^{n-1}(q_k'x_n)q_k & q_n & = & u_n/\|u_n\| \end{array}$$

A primeira equação apenas normaliza o vetor x_1

Na segunda equação, $(q_1'x_2)q_1$ é a projeção do vetor x_2 ao longo de q_1, e $x_2 - (q_1'x_2)q_1$ é necessariamente ortogonal ao vetor q_1.



Se um conjunto de vetores u_1, u_2, $\ldots$, u_n é ortonormal então

$$U \triangleq \left[\, u_1 \;\; u_2 \;\; \cdots \;\; u_n \,\right] \qquad ; \qquad U'U = \mathbf{I}_n$$

Propriedades:

Se $y = Ux$, então $\displaystyle\sum_{i=1}^{n} y_i^2 = \sum_{i=1}^{n} x_i^2$, ou

$$\|y\|^2 = \|Ux\|^2 = (Ux)'(Ux) = x'U'Ux = x'x = \|x\|^2$$

Em outras palavras, $y = Ux$ é um mapeamento isométrico (a multiplicação por U não altera a norma)

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

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

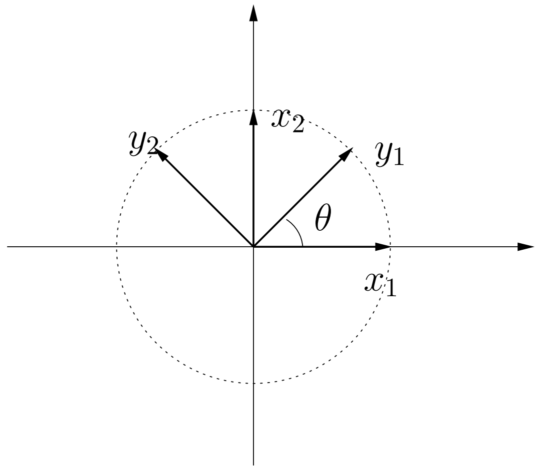
e também não altera o ângulo $\angle\langle y, \tilde{y} \rangle = \angle\langle x, \tilde{x} \rangle$

Se U é ortonormal, a transformação linear $y = Ux$ preserva a norma dos vetores $\|Ux\| = \|x\|$ e preserva o ângulo entre vetores $\angle\langle Ux, U\tilde{x} \rangle = \angle\langle x, \tilde{x} \rangle$

Exemplos: transformações de rotação ou reflexão de vetores (de fato, toda matriz ortogonal descreve ou uma rotação ou uma reflexão).

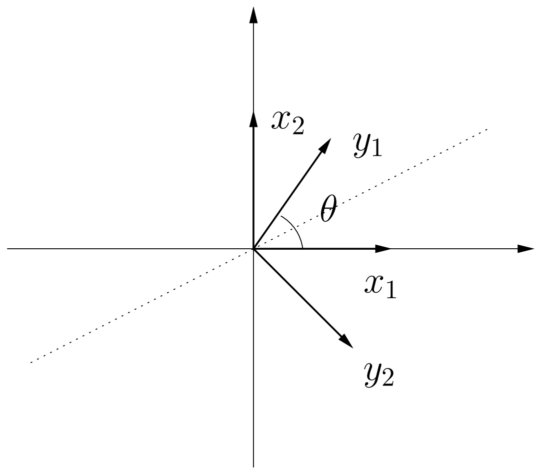
• No $\mathbb{R}^2$, a transformação que roda um vetor (sentido anti-horário) de θ é dada por

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



• A reflexão de um vetor na reta $x_2 = x_1 \tan\left(\frac{\theta}{2}\right)$ é dada por

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



O procedimento de **ortonormalização** pode ser descrito como uma transformação linear. Considere a matriz X formada pelos vetores linearmente independentes $\{x_1, x_2, \ldots, x_n\}$

$$X = \left[\begin{array}{cccc} x_1 & x_2 & \cdots & x_n \end{array} \right]$$

Deseja-se determinar uma matriz F tal que

$$\left[\begin{array}{cccc} q_1 & q_2 & \cdots & q_n \end{array} \right] \triangleq Q = XF$$

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

$$Q'Q = \mathbf{I} \quad \Longrightarrow \quad F'RF = \mathbf{I} \quad ; \quad R \triangleq X'X$$

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

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

$$F'RF = F'L'LF = (LF)'(LF) = \mathbf{I}$$

e $F = L^{-1}$ aparece como solução óbvia.

• 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

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

A escolha da transformação F que ortonormaliza X pode ser ditada por algum critério específico. Por exemplo, o objetivo pode ser encontrar Q solução do problema

$$\begin{array}{l} \min\ \mathbf{Tr}\ (X-Q)'(X-Q) \\ \text{sujeito a}\ \ Q'Q = \mathbf{I} \end{array}$$

que equivale a minimizar a norma (ao quadrado) de $x_i - q_i$, $i = 1, \ldots, n$. Com isso, busca-se uma transformação que tente preservar o máximo possível os vetores originais.

Re-escrevendo em termos de F:

$$\begin{array}{l} \min\ \mathbf{Tr}\ (X-XF)'(X-XF) \\ \text{sujeito a}\ \ F'X'XF = \mathbf{I} \end{array}$$

Desenvolvendo os termos que compõem a função objetivo

$$\begin{array}{l} \min\ \mathbf{Tr}(X'X + F'X'XF - F'X'X - X'XF) \\ \text{sujeito a}\ \ F'X'XF = \mathbf{I} \end{array}$$

Substituindo a restrição e lembrando que os termos constantes na função objetivo não influenciam na solução F:

$$\begin{array}{l} \min\ \mathbf{Tr}\ (-F'X'X - X'XF)\ =\ \max\ \mathbf{Tr}\ (2X'XF) \\ \text{sujeito a}\ \ F'X'XF = \mathbf{I} \end{array}$$

Procedimentos clássicos para resolver problemas de otimização com restrições podem ser usados.

Note que $\mathbf{Tr}(MX)$ é uma função linear dos elementos da matriz X.

Derivadas

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

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

- Algumas derivadas (A e B matrizes complexas):

$$\frac{\partial}{\partial X}\mathbf{Tr}(AXB) = A'B' \quad ; \quad \frac{\partial}{\partial X}\mathbf{Tr}(AX'B) = BA$$

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

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

$$\frac{\partial}{\partial X}\mathbf{Tr}(AX^{-1}B) = -(X^{-1}BAX^{-1})' \quad ; \quad \frac{\partial}{\partial X}\log\det(X) = (X')^{-1}$$

Para $A(\alpha) \in \mathbb{C}^{m\times n}$ e $\alpha \in \mathbb{C}$, $\dfrac{d}{d\alpha}A \triangleq \left[\dfrac{da_{ij}}{d\alpha}\right]$. Então

$$\frac{d}{d\alpha}AB = \frac{dA}{d\alpha}B + A\frac{dB}{d\alpha} \quad ; \quad \frac{d}{d\alpha}A^{-1} = -A^{-1}\frac{dA}{d\alpha}A^{-1}$$

• Escrevendo a função Lagrangeano $\mathbb{L}\,(F, \Lambda)$ (Λ é a variável dual simétrica associada à restrição original do problema)

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

As condições de estacionariedade são dadas por

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

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

Resolvendo em termos de F e de Λ:

$$F = -\Lambda^{-1} \quad \Longrightarrow \quad (-\Lambda^{-1})X'X(-\Lambda^{-1}) = \mathbf{I} \quad \Longrightarrow \quad X'X = \Lambda^2$$

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

A função objetivo original é dada por

$$\min\ \mathbf{Tr}\,(-2X'XF) \;=\; \max\ \mathbf{Tr}\,(2X'XF)$$

e o ótimo é obtido 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. A solução ótima deste problema de otimização é a mesma obtida pela decomposição de Schur.

Exemplo: considere os vetores LI

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

Iterativo:

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

$$u_2 = x_2 - (q_1' x_2) q_1 \ ; \quad q_2 = \frac{u_2}{\|u_2\|} = \begin{bmatrix} -0.7071 \\ 0.0000 \\ 0.7071 \end{bmatrix}$$

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

Usando Cholesky na matriz $R = X'X$:

$$\Longrightarrow \qquad L = \texttt{chol}(R) = \begin{bmatrix} 1.7321 & 3.4641 & 0 \\ 0 & 1.4142 & -0.7071 \\ 0 & 0 & 1.2247 \end{bmatrix} \ , \quad L'L = R$$

$$Q = XL^{-1} = \begin{bmatrix} u_1 & u_2 & u_3 \end{bmatrix} \ , \ Q'Q = \mathbf{I} \quad \left\{ \begin{array}{c} [Q, L] = \texttt{qr}(X) \\ \text{(decomposição triangular} \\ \text{ortonormal)} \end{array} \right.$$

Usando Schur: $\quad [V, \Omega] = \texttt{schur}(R) \ , \quad R^{-0.5} = V \Omega^{-0.5} V'$

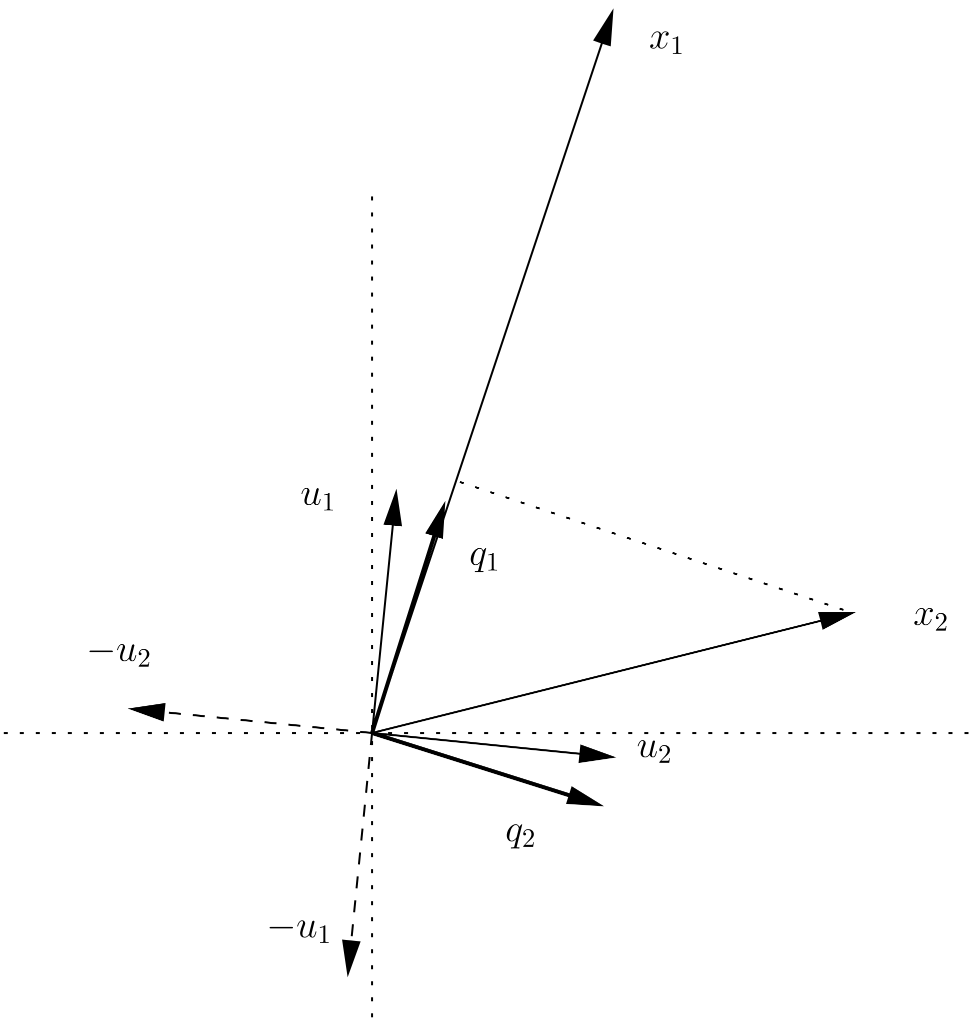
$$Q_2 = XR^{-0.5} = \begin{bmatrix} 0.9908 & -0.0890 & -0.1021 \\ 0.1348 & 0.5758 & 0.8064 \\ 0.0130 & 0.8127 & -0.5825 \end{bmatrix} \ , \quad Q_2' Q_2 = \mathbf{I}$$

Exemplo: considere os vetores LI no plano x_1, x_2

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

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

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



Sistema de Equações Algébricas Lineares

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} \qquad (Ax = y)$$

$A\ (m \times n)$, $x\ (n \times 1)$, $y\ (m \times 1)$
$a_{ij}, y_i \in \mathbb{R}$: dados do sistema; $x_i \in \mathbb{R}$: incógnitas

Três situações: $m > n$, $m = n$ ou $m < n$

- Problema: dados A e y
  1. $\exists x : \ Ax = y$?
  2. Se existe solução, qual o número de soluções LI?

Range da matriz $A \in \mathbb{R}^{m \times n}$ é definido como o conjunto de todas as possíveis combinações lineares das colunas de A

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

Rank da matriz $A \in \mathbb{R}^{m \times n}$ é definido como a dimensão do range de A (ou, equivalentemente, como o número de colunas LI em A) e denotado $\rho(A)$.

Espaço nulo da matriz A consiste no conjunto de vetores $x \in \mathbb{R}^n$ tais que $Ax = 0$. A dimensão do espaço nulo é chamada de **nulidade** da matriz A e denotada $\nu(A)$.

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

Note que $y = Ax$ pode ser escrito

$$y = x_1 a_1 + x_2 a_2 + \cdots + x_n a_n$$

e portanto $x_i \in \mathbb{R}$, $i = 1, \ldots, n$ são as ponderações das colunas de $A = \left[\begin{array}{cccc} a_1 & a_2 & \cdots & a_n \end{array}\right]$

Propriedades

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

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

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

• $\mathcal{N}(A)$ e $\mathcal{R}(A)$ são espaços lineares; ($\mathcal{N}(A)$ é um subespaço do $\mathbb{R}^n$ e $\mathcal{R}(A)$ é um subespaço do $\mathbb{R}^m$)

• Se $\nu(A) = 0$, $\mathcal{N}(A) = \{0\}$ 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 $\nu(A) = k$, $Ax = 0$ possui k soluções LI

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

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

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

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

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

Como a_1, a_2 são LI $\implies$ $\rho(A) = 2$

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

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 $\mathcal{N}(A)$; $\nu(A) = 3$

Teorema

• 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, equivalentemente,

$$\rho(A) = \rho(\left[\begin{array}{cc} A & y \end{array}\right])$$

• Dada uma matriz A, uma solução x de $y = Ax$ existe para todo y se e somente se $\rho(A) = m$ (rank completo de linhas).

Teorema (parametrização de todas as soluções)

• Dada uma matriz $A \in \mathbb{R}^{m\times n}$ e um vetor $y \in \mathbb{R}^{m\times 1}$, seja x_p uma solução $x \in \mathbb{R}^{n\times 1}$ da equação $y = Ax$ e seja $k = n - \rho(A) = \nu(A)$ a nulidade de A. Se $\rho(A) = n$ (rank completo de colunas) a solução x_p é única. Se $k > 0$, então para todo $\alpha_i \in \mathbb{R}$, $i = 1, \ldots, k$ o vetor

$$x = x_p + \alpha_1 n_1 + \alpha_2 n_2 + \cdots + \alpha_k n_k$$

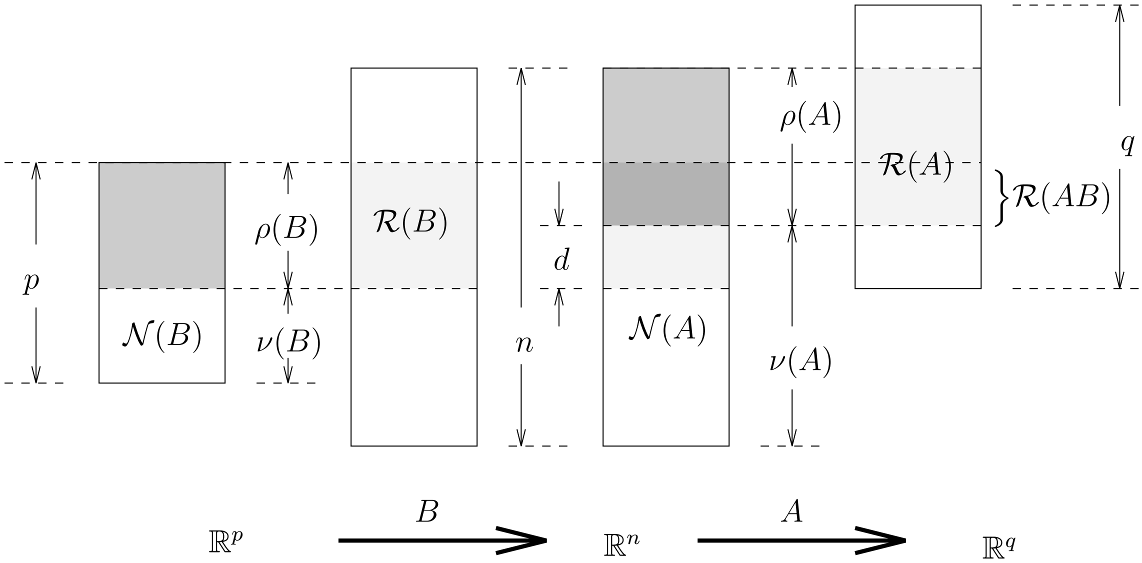
sendo $\{n_1, \ldots, n_k\}$ uma base de $\mathcal{N}(A)$ é uma solução de $Ax = y$.

De fato,

$$Ax_p + \sum_{i=1}^{k} \alpha_i A n_i = Ax_p + 0 = y$$

Desigualdade de Sylvester: para $A \in \mathbb{R}^{q\times n}$ e $B \in \mathbb{R}^{n\times p}$

$$\rho(A) + \rho(B) - n \leq \rho(AB) \leq \min\left(\rho(A), \rho(B)\right)$$



- domínio de $AB = \mathcal{R}(B)$;
- $\mathcal{R}(AB)$ = subespaço de $\mathcal{R}(A)$;
- $\rho(AB) \leq \min\left(\rho(A), \rho(B)\right)$;
- $\rho(AB) = \rho(B) - d \ \ ; \ \ \nu(A) = n - \rho(A) \quad \Rightarrow \quad d \leq n - \rho(A)$;
- $\rho(AB) \geq \rho(A) + \rho(B) - n$

Se B é uma matriz $n \times n$ não singular

$$\rho(A) + \rho(B) - n = \rho(A) \leq \rho(AB) \leq \min\left(\rho(A), n\right) \leq \rho(A)$$

Seja $A \in \mathbb{R}^{m\times n}$. Então

$$\rho(AC) = \rho(A) \quad ; \quad \rho(DA) = \rho(A)$$

para quaisquer matrizes $C \in \mathbb{R}^{n\times n}$ e $D \in \mathbb{R}^{m\times m}$ não singulares.

• O rank 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,

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

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

$A^*A \; n\times n \quad ; \quad AA^* \; m\times m$

Observe que $\rho(A) = n$ implica

• $n \leq m$
• $A\alpha = 0 \quad \Rightarrow \quad \alpha = 0$

Inversa

Se $A \in \mathbb{R}^{n \times n}$ (matriz quadrada), A é invertível ou não-singular se $\det(A) \neq 0$.

Condições equivalentes:

- 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 $y \in \mathbb{R}^n$. Em particular, a única solução de $Ax = 0$ é $x = 0$
- $AA^{-1} = A^{-1}A = \mathbf{I}$
- $\mathcal{N}(A) = \{0\}$
- $\mathcal{R}(A) = \mathbb{R}^n$
- $\det(A'A) = \det(AA') \neq 0$

$$A^{-1} = \frac{1}{\det(A)} \text{Adj } (A)$$

Adj (A): matriz **adjunta** da matriz A

$$\text{Adj } (A) = [\text{Co } (A)]'$$

Co (A): matriz **cofatora** de A, composta pelos cofatores C_{ij} da matriz A.

Considere o sistema $y = Ax$ com $A \in \mathbb{R}^{m\times n}$ e $m > n$ (matriz retangular em pé), isto é, o sistema tem mais equações do que incógnitas (pode não ter solução).

Solução aproximada: define-se um erro

$$e = Ax - y$$

e busca-se x_a que minimiza $\|e\|$ (solução aproximada de **mínimos quadrados** de $Ax = y$). A solução x_a gera o ponto no $\mathcal{R}(A)$ que está mais próximo de y, ou seja, Ax_a é a projeção de y no $\mathcal{R}(A)$.

$$\min \quad \|e\|^2 \;=\; \min\,(Ax - y)'(Ax - y)$$

Derivando a função objetivo (em relação a x) e igualando a 0

$$2x'A'A - 2y'A = 0$$

Assumindo que A tem rank completo (e portanto $A'A$ é invertível)

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

Note que x_a é uma função linear de y, e $x_a = A^{-1}y$ se A é quadrada.

x_a 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

A projeção de y no $\mathcal{R}(A)$ é linear, dada por $Ax_a = A(A'A)^{-1}A'y$ e $A(A'A)^{-1}A'$ é chamada matriz de projeção.

O erro $e = Ax_a - y = [A(A'A)^{-1}A' - \mathbf{I}]y$ é ortogonal ao $\mathcal{R}(A)$:

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

Estimador Mínimos Quadrados

Muitos problemas de reconstrução, inversão ou estimação podem ser colocados na forma $y = Ax + \Delta$, sendo que x é o vetor a ser estimado ou reconstruído; y é o vetor de medidas e Δ é o vetor de ruídos ou erros de medida.

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

Solução: $\hat{x} = (A'A)^{-1}A'y$

Considere o sistema $y = Ax$ com $A \in \mathbb{R}^{m \times n}$ e $m < n$ (matriz retangular deitada), isto é, o sistema tem mais variáveis do que equações (várias escolhas de x podem levar ao mesmo y).

Assumindo que A tem rank completo de linhas m, o conjunto de todas as soluções tem a forma

$$\{x \ : \ Ax = y\} = \{x_p + z \ : \ z \in \mathcal{N}(A)\}$$

sendo x_p uma solução particular qualquer. Os vetores z do espaço nulo de A caracterizam as várias soluções para o problema ($\mathcal{N}(A) = n - m$ graus de liberdade).

Uma solução (de mínima norma) é dada por $x_m = A'(AA')^{-1}y$ (como A tem rank completo de linhas, AA' é invertível). De fato, para qualquer outra solução x tem-se $A(x - x_m) = 0$ e

$$(x - x_m)'x_m = (x - x_m)'A'(AA')^{-1}y =$$
$$= [A(x - x_m)]'(AA')^{-1}y = 0 \quad \Longrightarrow \quad (x - x_m) \perp x_m$$

$$\|x\|^2 = \|x_m + x - x_m\|^2 = \|x_m\|^2 + \|x - x_m\|^2 \geq \|x_m\|^2$$

A matriz $A'(AA')^{-1}$ é chamada de pseudo-inversa de A (rank completo de linhas). Note que x_m é ortogonal a $\mathcal{N}(A)$ (x_m é a projeção de 0 no conjunto de soluções $\{x \ : \ Ax = y\}$).

O mesmo resultado poderia ser obtido com multiplicadores de Lagrange.

$$\begin{array}{l} \min \ x'x \\ \text{sujeito a} \ \ Ax = y \end{array}$$

Introduzindo o multiplicador de Lagrange λ e escrevendo o Lagrangeano

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

As condições de otimalidade são

$$\frac{\partial \mathbb{L}}{\partial x} = 2x' + \lambda' A = 0$$

$$\frac{\partial \mathbb{L}}{\partial \lambda} = (Ax - y)' = 0$$

$$\Longrightarrow \quad x = -A'\lambda/2 \quad ; \quad \lambda = -2(AA')^{-1}y$$

$$\Longrightarrow \quad x = A'(AA')^{-1}y$$