Álgebra Linear Numérica Aleatória

Thiago Rodrigo Ramos

7 de março de 2025

Sumário

1 Fatos básicos

Vamos seguir principalmente [Trefethen and Bau, 1997].

1.1 Multiplicações

Alguns fatos importantes sobre multiplicações envolvendo uma matriz $A \in \mathbb{R}^{m \times n}$

1.1.1 Matriz-Vetor

Seja A_j a j-ésima coluna de A, um m-vetor. Então, a equação $b = Ax$ pode ser reescrito como:

$$b = Ax = \sum_{j=1}^{n} x_j a_j. \tag{1}$$

Essa equação pode ser representada esquematicamente da seguinte forma:

$$\begin{bmatrix} \\ b \\ \\ \end{bmatrix} = \begin{bmatrix} A_1 & A_2 & \cdots & A_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1 \begin{bmatrix} A_1 \end{bmatrix} + x_2 \begin{bmatrix} A_2 \end{bmatrix} + \cdots + x_n \begin{bmatrix} A_n \end{bmatrix}.$$

Na equação (1.2), b é expresso como uma combinação linear das colunas A_j.

Podemos resumir essas diferentes descrições do produto matriz-vetor da seguinte forma. Como matemáticos, estamos acostumados a interpretar a fórmula $Ax = b$ como uma afirmação de que A age sobre x para produzir b. A forma acima, por outro lado, sugere a interpretação de que x age sobre A para produzir b.

1.1.2 Matriz-Matriz

Para o produto matriz-matriz $B = AC$, *cada coluna de B é uma combinação linear das colunas de A*. Para demonstrar esse fato, começamos com a fórmula usual para produtos de matrizes. Se A é uma matriz de dimensão $\ell \times m$ e C é de dimensão $m \times n$, então B será de dimensão $\ell \times n$, com entradas definidas por

$$B_{ij} = \sum_{k=1}^{m} A_{ik} C_{kj}. \tag{2}$$

Aqui, B_{ij}, A_{ik} e C_{kj} são elementos de B, A e C, respectivamente. Escrito em termos de colunas, o produto é

$$\begin{bmatrix} B_1 & B_2 & \cdots & B_n \end{bmatrix} = A \begin{bmatrix} C_1 & C_2 & \cdots & C_n \end{bmatrix},$$

que implica em:

$$B_j = AC_j = \sum_{k=1}^{m} C_{kj} A_k. \tag{3}$$

Note que isso é só uma generalização da multiplicação anterior, já que $B_j = AC_j$ e podemo utilizar a formulação Matriz-Vetor da seção anterior.

Um exemplo simples de um produto matriz-matriz é o *produto externo*. Este é o produto de um vetor coluna u de dimensão m com um vetor linha v de dimensão n; o resultado é uma matriz $m \times n$ de posto 1. O produto externo pode ser escrito como

$$\begin{bmatrix} u \end{bmatrix} \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix} = \begin{bmatrix} v_1 u & v_2 u & \cdots & v_n u \end{bmatrix} = \begin{bmatrix} v_1 u_1 & \cdots & v_n u_1 \\ \vdots & \ddots & \vdots \\ v_1 u_m & \cdots & v_n u_m \end{bmatrix}.$$

As colunas são todas múltiplos do mesmo vetor u e, da mesma forma, as linhas são todas múltiplos do mesmo vetor v. $\square$

1.2 Núcleo e Imagem

A imagem (range) de uma matriz $A \in \mathbb{R}^{m \times n}$, denotado por $R(A)$, é o conjunto dos vetores z em $\mathbb{R}^m$ que podem ser escritos como $z = Ax$ para algum $xin\mathbb{R}^n$. Note que pela discussão acima, isso significa que

$$z = x_1 A_1 + \ldots z_n A_n,$$

e portanto $R(A)$ é o espaço gerado pelas colunas de A.

O núcleo de A, denotado por $N(A)$ é o conjunto de pontos $x \in \mathbb{R}^n$ tal que $Ax = 0$.

1.3 Posto

O posto de uma matriz A é dado pela dimensão de $R(A)$. Uma matriz $m \times n$ de *posto completo* é aquela que possui o posto máximo possível (o menor valor entre m e n). Isso significa que uma matriz de posto completo com $m \geq n$ deve ter n colunas linearmente independentes. Tal matriz também pode ser caracterizada pela propriedade de que a transformação que ela define é injetiva.

Teorema 1. *Uma matriz $A \in \mathbb{R}^{m \times n}$ com $m \geq n$ tem posto completo se, e somente se, não mapeia dois vetores distintos para o mesmo vetor.*

Demonstração. Se A tem posto completo e $A(x - x') = 0$ com $x - x' \neq 0$, isso significa que existe uma forma de escrever uma das colunas de A em termos da outra já que teríamos algo como $\sum z_i A_i = 0$, o que é contradição já que as colunas são LI.

Para a volta, suponha que as colunas não são LI, então existe $c \in \mathbb{R}^n$ tal que, SPDG, $0 = \sum c_j A_j = Ac$. Mas note que então $Ax = A(x + c)$, contrariando a hipótese. $\square$

1.4 Inversa

Supondo que $A \in R^{m \times m}$, então sua matriz inversa A^{-1} é uma matriz que satisfaz $AA^{-1} = A^{-1}A = I$.

Teorema 2. *Para $A \in \mathbb{R}^{m \times m}$, as seguintes condições são equivalentes:*

(a) A possui uma inversa A^{-1},

(b) $rank(A) = m$,

(c) $range(A) = \mathbb{C}^m$,

(d) $null(A) = \{0\}$,

(e) 0 não é um autovalor de A,

(f) 0 não é um valor singular de A,

(g) $\det(A) \neq 0$.

Com relação ao item (g), mencionamos que o determinante, embora seja uma noção teoricamente conveniente, raramente tem um papel útil em algoritmos numéricos.

1.4.1 Multiplicações

Ao escrever o produto $x = A^{-1}b$, é importante não deixar que a notação de matriz inversa obscureça o que realmente está acontecendo! Em vez de pensar em x como o resultado da aplicação de A^{-1} a b, devemos entendê-lo como o vetor único que satisfaz a equação $Ax = b$.

Uma coisa importante de se notar é que como $AA^{-1}b = b$, se $z = A^{-1}b$ então $b = \sum z_i A_i$, isto é, as coordenadas do vetor $z = A^{-1}b$ indicam os coeficientes necessários para escrever b na base dada pelas colunas de A.

1.5 Aplicações

Com as ideias desenvolvidas nessa seção, somos capazes de desenvolver várias transformações de forma rápida. Por exemplo, suponha que queremos uma matriz C cuja primeira coluna é a primeira coluna de A duplicada, e as outras colunas são iguais as de A. Pela Seção de multiplicação Matriz-Matriz, queremos então que

$$\begin{aligned} C_1 &= 2A_1 + 0A_2 + \ldots 0A_n = A[2,0,\ldots,0]^T \\ &\vdots \\ C_i &= A_i = A[0,0,\ldots,1,\ldots,0]^T, \end{aligned}$$

logo, $C = AB$ onde $B = \text{diag}(2,1,\ldots,1)$.

Suponha agora que D é igual a M, porém com a linha 3 somada com a linha 1. Note que a gente só sabe trabalhar com operações nas colunas, então a primeira coisa é transformas linhas em colunas, fazendo A^T, logo

$$\begin{aligned} D_1 &= A_1^T + A_3^T = A^T[1,0,1,\ldots,0]^T \\ &\vdots \\ D_i &= A_i^T = A^T[0,0,\ldots,1,\ldots,0]^T. \end{aligned}$$

Logo, $D = A^T \begin{pmatrix} 1,0,\ldots,0 \\ 0,1,\ldots,0 \\ 1,0,\ldots,0 \\ \vdots \\ 0,0,\ldots,1. \end{pmatrix} = A^T M$ Como queremos uma expressão em termos de A, podemos fazer $D^T = M^T A$.

Vamos verificar se isso tá certo, considere: $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Note que

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 2a & b \\ 2c & d \end{pmatrix}.$$

E também:

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 2a & b \\ 2c & d \end{pmatrix} = \begin{pmatrix} 2a+2c & b+d \\ 2c & d \end{pmatrix}.$$

Se quiséssemos dobrar a coluna 1 somada com menos a coluna 2 e fazer linha 2 mais o dobro da linha 1, bastaria fazer

$$\begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}^T \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 2 & 0 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 2a-b & b \\ 2c-d & d \end{pmatrix} = \begin{pmatrix} 2a-b & b \\ 4a-2b+2c-d & 2b+d \end{pmatrix}.$$

Ou seja, operações nas colunas de uma matriz são feitas à direita e operações com linhas são feitas à esquerda transposta.

2 Vetores e matrizes ortogonais

Um par de vetores x e y são *ortogonais* se $x^*y = 0$. Se x e y forem reais, isso significa que eles estão em ângulos retos entre si em $\mathbb{R}^m$. Dois conjuntos de vetores X e Y são ortogonais (também expressado como "X é ortogonal a Y") se todo $x \in X$ for ortogonal a todo $y \in Y$.

Um conjunto de vetores não nulos S é *ortogonal* se seus elementos forem ortogonais entre si, ou seja, se para $x, y \in S$, $x \neq y \Rightarrow x^*y = 0$. Um conjunto de vetores é *ortonormal* se for ortogonal e, além disso, todo $x \in S$ satisfaz $\|x\| = 1$.

Teorema 3. *Os vetores em um conjunto ortogonal S são linearmente independentes.*

Demonstração. Se os vetores em S não forem independentes, então algum $v_k \in S$ pode ser expresso como uma combinação linear de outros membros $v_1, \ldots, v_n \in S$,

$$v_k = \sum_{i=1, i\neq k}^{n} c_i v_i.$$

Como $v_k \neq 0$, temos $v_k^* v_k = \|v_k\|^2 > 0$. Utilizando a bilinearidade dos produtos internos e a ortogonalidade de S, calculamos

$$v_k^* v_k = \sum_{i=1, i\neq k}^{n} c_i v_k^* v_i = 0,$$

o que contradiz a suposição de que os vetores em S são não nulos. □ □

A ideia mais importante a se extrair dos conceitos de produtos internos e ortogonalidade é a seguinte: produtos internos podem ser usados para decompor vetores arbitrários em componentes ortogonais. Por exemplo, suponha que $\{q_1, q_2, \ldots, q_n\}$ seja um conjunto ortonormal e seja v um vetor arbitrário. A quantidade $q_i^* v$ é um escalar. Utilizando esses escalares como coordenadas em uma expansão, encontramos que o vetor

$$r = v - (q_1^* v)q_1 - (q_2^* v)q_2 - \cdots - (q_n^* v)q_n$$

é ortogonal a $\{q_1, q_2, \ldots, q_n\}$. Isso pode ser verificado computando $q_i^* r$:

$$q_i^* r = q_i^* v - (q_1^* v)(q_i^* q_1) - \cdots - (q_n^* v)(q_i^* q_n).$$

Essa soma se anula, pois $q_i^* q_j = 0$ para $i \neq j$:

$$q_i^* r = q_i^* v - (q_i^* v)(q_i^* q_i) = 0.$$

Assim, vemos que v pode ser decomposto em $n+1$ componentes ortogonais:

$$v = r + \sum_{i=1}^{n} (q_i^* v) q_i = r + \sum_{i=1}^{n} (q_i q_i^*) v.$$

Nessa decomposição, r é a parte de v ortogonal ao conjunto de vetores $\{q_1, q_2, \dots, q_n\}$, ou, equivalentemente, ao subespaço gerado por esse conjunto de vetores, e $(q_i^* v)q_i$ é a parte de v na direção de q_i.

Se $\{q_i\}$ for uma base de $\mathbb{C}^m$, então n deve ser igual a m e r deve ser o vetor nulo, de modo que v é completamente decomposto em m componentes ortogonais nas direções dos q_i:

$$v = \sum_{i=1}^{m} (q_i^* v) q_i = \sum_{i=1}^{m} (q_i q_i^*) v.$$

Nas equações acima, escrevemos a fórmula de duas maneiras diferentes: uma vez com $(q_i^* v)q_i$ e outra vez com $(q_i q_i^*)v$. Essas expressões são iguais, mas possuem interpretações diferentes. No primeiro caso, vemos v como uma soma de coeficientes $q_i^* v$ multiplicados pelos vetores q_i. No segundo, vemos v como uma soma de projeções ortogonais de v sobre as várias direções q_i. A i-ésima projeção é obtida pelo operador especial de matriz de posto um $q_i q_i^*$. Discutiremos essa operação mais adiante.

2.1 Aplicações

Alguns resultados importante para lembrarmos são:

- Se uma matriz A é simultaneamente triangular e unitária, então ela é diagonal. Para ver isso, veja o que acontece com $\langle A_1, A_j \rangle$. Veja que isso zera todos os elementos da primeira linha fora da diagonal, o resto sai por indução.

- O teorema de Pitágoras afirma que, para um conjunto de n vetores ortogonais $\{x_i\}$, vale a identidade:

  $$\left\| \sum_{i=1}^{n} x_i \right\|^2 = \sum_{i=1}^{n} \|x_i\|^2.$$

  Primeiramente, verifica-se essa igualdade no caso $n = 2$ por meio do cálculo explícito de $\|x_1 + x_2\|^2$. Em seguida, observa-se que esse cálculo se estende ao caso geral por indução.

- Seja $A \in \mathbb{C}^{m \times m}$ uma matriz hermitiana. Sabemos que todos os autovalores de A são reais, para isso basta fazer $\bar{\lambda} \langle A^* x, x \rangle = \langle Ax, x \rangle = \lambda \langle x, x \rangle$.

- Se x e y são autovetores correspondentes a autovalores distintos, então x e y são ortogonais. Basta ver que $\lambda_x \langle x, z \rangle = \langle Ax, z \rangle = \langle x, Az \rangle = \lambda_z \langle x, z \rangle$

- Os autovalores de uma matriz unitária pertencem ao círculo unitário no plano complexo, ou seja, possuem módulo igual a 1. Basta ver o que acontece com $\langle Ax, Ax \rangle$

  Seja $A = I + uv^T$, dizemos que A é uma perturbação de posto 1 da identidade. Note que tomando $A' = I + \alpha uv^T$, com $\alpha = \frac{-1}{1+\langle u,v \rangle}$ temos

$$\begin{aligned} AA' &= I + \alpha uv^T + uv^T + \alpha uv^T uv^T \\ &= I + uv^T(\alpha(1 + \langle u, v \rangle) + 1) \\ &= 0. \end{aligned}$$

  Isso implica que $A' = A^{-1}$, contanto que $\langle u, v \rangle \neq -1$. Se u, v são normalizados, então isso é equivalente a dizer que u, v nao podem ser diametralmente opostos.

3 Decomposição em valores singulares

A utilidade da Decomposição em Valores Singulares (SVD) torna-se evidente ao catalogar suas conexões com outros tópicos fundamentais da álgebra linear. Nos teoremas a seguir, assume-se que A tem dimensões $m \times n$. Seja p o mínimo entre m e n, e seja $r \leq p$ o número de valores singulares não nulos de A. Além disso, $\langle x, y, \ldots, z \rangle$ denota o espaço gerado pelos vetores $x, y, \ldots, z$.

Teorema 4. *A matriz A possui posto r, que corresponde ao número de valores singulares não nulos.*

Demonstração. O posto de uma matriz diagonal é igual ao número de suas entradas não nulas. Na decomposição $A = U\Sigma V^*$, as matrizes U e V são de posto completo. Assim, tem-se que

$$\text{rank}(A) = \text{rank}(\Sigma) = r.$$

□

Teorema 5. *O espaço imagem de A é dado por* $\text{range}(A) = \langle u_1, \ldots, u_r \rangle$ *e o espaço nulo de A é dado por* $\text{null}(A) = \langle v_{r+1}, \ldots, v_n \rangle$.

Demonstração. Essa propriedade decorre do fato de que o espaço imagem de Σ é

$$\text{range}(\Sigma) = \langle e_1, \ldots, e_r \rangle \subseteq \mathbb{C}^m$$

e que o espaço nulo de Σ é

$$\text{null}(\Sigma) = \langle e_{r+1}, \ldots, e_n \rangle \subseteq \mathbb{C}^n.$$

□

Teorema 6. *A norma espectral de A é dada por* $\|A\|_2 = \sigma_1$ *e a norma de Frobenius por*

$$\|A\|_F = \sqrt{\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_r^2}.$$

Demonstração. O primeiro resultado já foi estabelecido na demonstração do Teorema 4.1. Como $A = U\Sigma V^*$ com U e V unitários, tem-se que

$$\|A\|_2 = \|\Sigma\|_2 = \max\{\sigma_j\} = \sigma_1,$$

de acordo com o Teorema 3.1. Para o segundo resultado, pelo Teorema 3.1 e pela observação a seguir, a norma de Frobenius é invariante por multiplicação unitária. Assim, conclui-se que

$$\|A\|_F = \|\Sigma\|_F,$$

o que, pela equação (3.16), resulta na fórmula apresentada. □

Teorema 7. *Os valores singulares não nulos de A são as raízes quadradas dos autovalores não nulos de* A^*A *ou* AA^*. *Essas matrizes possuem os mesmos autovalores não nulos.*

Demonstração. A partir do cálculo

$$A^*A = (U\Sigma V^*)^*(U\Sigma V^*) = V\Sigma^*U^*U\Sigma V^* = V(\Sigma^*\Sigma)V^*,$$

observa-se que A^*A é semelhante a $\Sigma^*\Sigma$ e, portanto, tem os mesmos n autovalores. Os autovalores da matriz diagonal $\Sigma^*\Sigma$ são $\sigma_1^2, \sigma_2^2, \ldots, \sigma_p^2$, com $n - p$ autovalores adicionais iguais a zero se $n > p$. Um cálculo análogo se aplica aos m autovalores de AA^*. □

Referências

[Trefethen and Bau, 1997] Trefethen, L. N. and Bau, D. (1997). *Numerical Linear Algebra*. SIAM. 2