Conjunto de exercícios 9.4

1. Um certo computador consegue executar 10 gigaflops por segundo. Em cada caso, use a Fórmula (5) para encontrar o tempo necessário para resolver o sistema usando eliminação de Gauss-Jordan.

(a) Um sistema de 1.000 equações em 1.000 incógnitas.

(b) Um sistema de 10.000 equações em 10.000 incógnitas.

(c) Um sistema de 100.000 equações em 100.000 incógnitas.

2. Um certo computador consegue executar 100 gigaflops por segundo. Em cada caso, use a Fórmula (5) para encontrar o tempo necessário para resolver o sistema usando eliminação de Gauss-Jordan.

(a) Um sistema de 10.000 equações em 10.000 incógnitas.

(b) Um sistema de 100.000 equações em 100.000 incógnitas.

(c) Um sistema de 1.000.000 equações em 1.000.000 incógnitas.

3. Os computadores pessoais de hoje conseguem executar 70 gigaflops por segundo. Em cada caso, use a Tabela 1 para obter uma estimativa do tempo necessário para executar a operação dada com uma matriz invertível A de tamanho 10.000×10.000.

(a) Executar a fase direta da eliminação de Gauss-Jordan.

(b) Executar a fase inversa da eliminação de Gauss-Jordan.

(c) Obter uma decomposição LU de A.

(d) Encontrar A^{-1} reduzindo $[A \mid I]$ a $[I \mid A^{-1}]$.

4. O computador Roadrunner, da IBM, consegue operar a velocidades superiores a um petaflops por segundo (1 petaflop $= 10^{15}$ flops). Em cada caso, use a Tabela 1 para obter uma estimativa do tempo necessário para executar a operação dada com uma matriz invertível A de tamanho 100.000×100.000.

(a) Executar a fase direta da eliminação de Gauss-Jordan.

(b) Executar a fase inversa da eliminação de Gauss-Jordan.

(c) Obter uma decomposição LU de A.

(d) Encontrar A^{-1} reduzindo $[A \mid I]$ a $[I \mid A^{-1}]$.

5. (a) Obtenha uma aproximação do tempo necessário para executar a fase direta da eliminação de Gauss-Jordan num sistema de 100.000 equações em 100.000 incógnitas usando um computador que consiga executar 1 gigaflop por segundo. Faça o mesmo com a fase indireta. (Ver Tabela 1.)

(b) Um computador deve ser capaz de executar quantos gigaflops por segundo para encontrar a decomposição LU de uma matriz de tamanho 10.000×10.000 em menos de 0,5 segundos? (Ver Tabela 1.)

6. Um computador deve ser capaz de executar quantos teraflops por segundo para encontrar a inversa de uma matriz de tamanho 100.000×100.000 em menos de 0,5 segundos? (1 teraflop $= 10^{12}$ flops.)

▶ Nos Exercícios **7–10**, A e B são matrizes $n \times n$ e c é um número real. ◀

7. Quantos flops são necessários para calcular cA?

8. Quantos flops são necessários para calcular $A + B$?

9. Quantos flops são necessários para calcular AB?

10. Supondo que A seja uma matriz diagonal e k um inteiro positivo, quantos flops são necessários para calcular A^k?

9.5 Decomposição em valores singulares

Nesta seção, discutimos uma extensão da teoria de diagonalização de matrizes $n \times n$ simétricas para matrizes $m \times n$ arbitrárias. Os resultados que desenvolvemos nesta seção têm aplicações à compressão, ao armazenamento e à transmissão de informação digitalizada e formam a base de muitos dos melhores algoritmos computacionais que estão atualmente disponíveis para resolução de sistemas lineares.

Decomposições de matrizes quadradas

Vimos na Fórmula (2) da Seção 7.2 que qualquer matriz simétrica A pode ser expressa como

$$A = PDP^T \tag{1}$$

em que P é uma matriz ortogonal $n \times n$ de autovetores de A e D é a matriz diagonal cujas entradas diagonais são os autovalores associados aos vetores coluna de P. Nesta seção, vamos dizer que (1) é uma ***decomposição em autovalores*** de A (abreviada pelas iniciais em inglês EVD de A).

Se uma matriz A de tamanho $n \times n$ não for simétrica, então não existe uma decomposição em autovalores, mas existe uma ***decomposição de Hessenberg***

$$A = PHP^T$$

na qual P é uma matriz ortogonal e H é uma matriz de Hessenberg superior (Teorema 7.2.4). Além disso, se A tiver autovalores reais, existe ainda uma **decomposição de Schur**

$$A = PSP^{\mathrm{T}}$$

em que P é uma matriz ortogonal e S é triangular superior (Teorema 7.2.3).

As decomposições em autovalores, de Hessenberg e de Schur são importantes em algoritmos numéricos não só porque as matrizes D, H e S têm formatos mais simples do que A, mas também porque as matrizes ortogonais que aparecem nessas fatorações não aumentam os erros de arredondamento. Para ver isso, suponha que $\hat{\mathbf{x}}$ seja um vetor coluna cujas entradas são conhecidas exatamente e que

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

seja o vetor que resulta quando ocorrem erros de arredondamento nas entradas de $\hat{\mathbf{x}}$. Se P for uma matriz ortogonal, então a propriedade de preservação de comprimentos de vetores por transformações ortogonais implica

$$\|P\mathbf{x} - P\hat{\mathbf{x}}\| = \|\mathbf{x} - \hat{\mathbf{x}}\| = \|\mathbf{e}\|$$

o que nos diz que o erro em aproximar $P\hat{\mathbf{x}}$ por $P\mathbf{x}$ tem a mesma magnitude que o erro de aproximar $\hat{\mathbf{x}}$ por $\mathbf{x}$.

Existem dois caminhos principais para percorrer na procura de outros tipos de fatoração de uma matriz quadrada A arbitrária. Poderíamos procurar fatorações da forma

$$A = PJP^{-1}$$

em que P é invertível, mas não necessariamente ortogonal, ou poderíamos procurar fatorações da forma

$$A = U\Sigma V^T$$

em que U e V são ortogonais, mas não necessariamente iguais. O primeiro caminho leva a fatorações em que J é ou diagonal ou um certo tipo de matriz diagonal em blocos, denominada ***forma canônica de Jordan***, em homenagem ao matemático francês Camille Jordan (ver página 510). As formas canônicas de Jordan, que não serão consideradas neste texto, têm importância na teoria e em certas aplicações, mas são menos importantes numericamente por causa de dificuldades de arredondamento que decorrem da falta de ortogonalidade de P. Nesta seção, nos concentramos no segundo caminho.

Valores singulares

Como os produtos de matrizes do tipo A^TA desempenham um papel importante no nosso trabalho, começamos com dois teoremas básicos relativos a essas matrizes.

TEOREMA 9.5.1 *Se A for uma matriz $m \times n$, então*

(*a*) *A e A^TA têm o mesmo espaço nulo.*

(*b*) *A e A^TA têm o mesmo espaço linha.*

(*c*) *A^T e A^TA têm o mesmo espaço coluna.*

(*d*) *A e A^TA têm o mesmo posto.*

Provamos a parte (*a*) e deixamos as demais provas como exercícios.

Prova (a) Devemos mostrar que cada solução de $A\mathbf{x} = \mathbf{0}$ é uma solução de $A^TA\mathbf{x} = \mathbf{0}$ e vice-versa. Se $\mathbf{x}_0$ for uma solução qualquer de $A\mathbf{x} = \mathbf{0}$, então $\mathbf{x}_0$ também é solução de $A^TA\mathbf{x} = \mathbf{0}$, pois

$$A^TA\mathbf{x}_0 = A^T(A\mathbf{x}_0) = A^T\mathbf{0} = \mathbf{0}$$

Reciprocamente, se $\mathbf{x}_0$ for uma solução qualquer de $A^TA\mathbf{x} = \mathbf{0}$, então $\mathbf{x}_0$ está no espaço nulo de A^TA e é, portanto, ortogonal a todos os vetores do espaço linha de A^TA, pela parte (q) do

Teorema 4.8.10. No entanto, A^TA é simétrica, de modo que $\mathbf{x}_0$ é ortogonal a cada vetor no espaço coluna de A^TA. Em particular, $\mathbf{x}_0$ deve ser ortogonal ao vetor $(A^TA)\mathbf{x}_0$, ou seja,

$$\mathbf{x}_0 \cdot (A^TA)\mathbf{x}_0 = 0$$

Usando a primeira fórmula da Tabela 1 da Seção 3.2 e as propriedades do operador de transposição, podemos reescrever isso como

$$\mathbf{x}_0^T(A^TA)\mathbf{x}_0 = (A\mathbf{x}_0)^T(A\mathbf{x}_0) = (A\mathbf{x}_0)\cdot(A\mathbf{x}_0) = \|A\mathbf{x}_0\|^2 = 0$$

o que implica $A\mathbf{x}_0 = 0$, provando que $\mathbf{x}_0$ é uma solução de $A\mathbf{x} = \mathbf{0}$. ◄

TEOREMA 9.5.2 *Se A for uma matriz $m \times n$, então*

(*a*) *A^TA é ortogonalmente diagonalizável.*

(*b*) *Os autovalores de A^TA são não negativos.*

Prova (a) A matriz A^TA, por ser simétrica, é ortogonalmente diagonalizável pelo Teorema 7.2.1.

Prova (b) Como A^TA é ortogonalmente diagonalizável, existe alguma base ortonormal de R^n consistindo em autovetores de A^TA, digamos, $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$. Se $\lambda_1, \lambda_2, \ldots, \lambda_n$ forem os autovalores correspondentes, então, dado qualquer $1 \leq i \leq n$, temos

$$\begin{aligned}\|A\mathbf{v}_i\|^2 &= A\mathbf{v}_i \cdot A\mathbf{v}_i = \mathbf{v}_i \cdot A^TA\mathbf{v}_i \qquad \textbf{[Fórmula (26) da Seção 3.2]}\\ &= \mathbf{v}_i \cdot \lambda_i\mathbf{v}_i = \lambda_i(\mathbf{v}_i \cdot \mathbf{v}_i) = \lambda_i\|\mathbf{v}_i\|^2 = \lambda_i\end{aligned}$$

Segue dessa relação que $\lambda_i \geq 0$. ◄

Em toda esta seção, vamos supor que os autovalores de A^TA estão nomeados de tal forma que

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$$

e, portanto, que

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$$

DEFINIÇÃO 1 Se A for uma matriz $m \times n$ e $\lambda_1, \lambda_2, \ldots, \lambda_n$ os autovalores de A^TA, então os números

$$\sigma_1 = \sqrt{\lambda_1}, \quad \sigma_2 = \sqrt{\lambda_2}, \ldots, \quad \sigma_n = \sqrt{\lambda_n}$$

são denominados ***valores singulares*** de A.

► EXEMPLO 1 Valores singulares

Encontre os valores singulares da matriz

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}$$

Solução O primeiro passo é encontrar os autovalores da matriz

$$A^TA = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

O polinômio característico de A^TA é

$$\lambda_2 - 4\lambda + 3 = (\lambda - 3)(\lambda - 1)$$

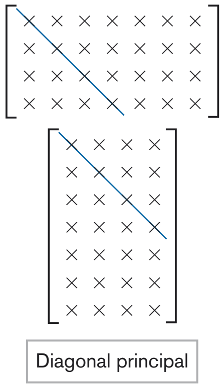
de modo que os autovalores de A^TA são $\lambda_1 = 3$ e $\lambda_2 = 1$ e os valores singulares de A em ordem decrescente de tamanho são

$$\sigma_1 = \sqrt{\lambda_1} = \sqrt{3}, \quad \sigma_2 = \sqrt{\lambda_2} = 1$$ ◄

Antes de passar ao resultado principal desta seção, convém estender a noção de "diagonal principal" para matrizes que não são quadradas. Definimos a ***diagonal principal*** de uma matriz $m \times n$ como a fileira de entradas mostrada na Figura 9.5.1, que começa no canto superior esquerdo e se estende diagonalmente até onde for possível. Dizemos que as entradas nessa diagonal principal são as ***entradas diagonais*** da matriz.

Agora estamos prontos para considerar o resultado principal desta seção, que diz respeito a uma maneira específica de fatorar uma matriz A arbitrária de tamanho $m \times n$. Essa fatoração, denominada ***decomposição em valores singulares*** de A (que abreviamos com as iniciais em inglês SVD de A), será dada em duas versões: uma curta, que captura a ideia principal, e uma expandida, que fornece todos os detalhes. A prova será dada ao final desta seção.

Decomposição em valores singulares



▲ **Figura 9.5.1**

TEOREMA 9.5.3 Decomposição em valores singulares

Se A for uma matriz $m \times n$, então A pode ser expressa como

$$A = U \Sigma V^T$$

em que U e V são matrizes ortogonais e Σ é uma matriz $m \times n$ cujas entradas diagonais são os valores singulares de A e cujas demais entradas são nulas.

TEOREMA 9.5.4 Decomposição em valores singulares (versão expandida)

Se A for uma matriz $m \times n$ de posto k, então A pode ser fatorada como

$$A = U\Sigma V^T = [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \cdots \quad \mathbf{u}_k \mid \mathbf{u}_{k+1} \quad \cdots \quad \mathbf{u}_m] \left[\begin{array}{cccc|c} \sigma_1 & 0 & \cdots & 0 & \\ 0 & \sigma_2 & \cdots & 0 & 0_{k\times(n-k)} \\ \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & \cdots & \sigma_k & \\ \hline & & 0_{(m-k)\times k} & & 0_{(m-k)\times(n-k)} \end{array}\right] \left[\begin{array}{c} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_k^T \\ \hline \mathbf{v}_{k+1}^T \\ \vdots \\ \mathbf{v}_n^T \end{array}\right]$$

em que U, Σ e V têm tamanhos $m \times m$, $m \times n$ e $n \times n$, respectivamente e satisfazem as condições seguintes.

(*a*) $V = [\mathbf{v}_1 \quad \mathbf{v}_2 \quad \cdots \quad \mathbf{v}_n]$ *diagonaliza ortogonalmente* A^TA.

(*b*) *As entradas diagonais não nulas de Σ são* $\sigma_1 = \sqrt{\lambda_1}$, $\sigma_2 = \sqrt{\lambda_2}, \ldots,$ $\sigma_k = \sqrt{\lambda_k}$, *sendo* $\lambda_1, \lambda_2, \ldots, \lambda_k$, *os autovalores não nulos de A^TA associados aos vetores coluna de V.*

(*c*) *Os vetores coluna de V são ordenados de tal modo que* $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k > 0$.

(*d*) $\mathbf{u}_i = \dfrac{A\mathbf{v}_i}{\|A\mathbf{v}_i\|} = \dfrac{1}{\sigma_i}A\mathbf{v}_i \qquad (i = 1, 2, \ldots, k)$

(*e*) $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ *é uma base ortonormal de* $\operatorname{col}(A)$.

(*f*) $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k, \mathbf{u}_{k+1}, \ldots, \mathbf{u}_m\}$ *é uma extensão de* $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ *a uma base ortonormal de* R^m.

Harry Bateman (1882–1946)

Nota histórica O termo *valor singular* é aparentemente devido ao matemático Harry Bateman, nascido na Inglaterra, que o utilizou num artigo científico publicado em 1908. Em 1910, Bateman emigrou para os Estados Unidos, lecionando na Escola Bryn Mawr, na Universidade Johns Hopkins e, finalmente, no Instituto Tecnológico da Califórnia. É interessante observar que ele recebeu seu doutorado em Johns Hopkins em 1913, quando já era um matemático eminente, com 60 publicações científicas.

[Imagem: Cortesia dos Arquivos do California Institute of Technology]

Os vetores $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k$ são denominados ***vetores singulares à esquerda*** de A; e $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$, os ***vetores singulares à direita*** de A.

▶ EXEMPLO 2 Decomposição em valores singulares se A não for quadrada

Encontre uma decomposição em valores singulares da matriz

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}$$

Solução Mostramos no Exemplo 1 que os autovalores de A^TA são $\lambda_1 = 3$ e $\lambda_2 = 1$ e que os valores singulares de A são $\sigma_1 = \sqrt{3}$ e $\sigma_2 = 1$. Deixamos para o leitor verificar que

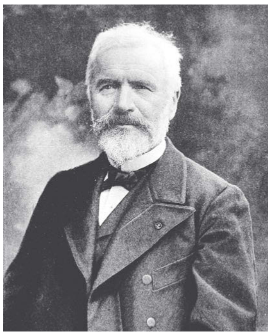
$$\mathbf{v}_1 = \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix} \quad \text{e} \quad \mathbf{v}_2 = \begin{bmatrix} \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \end{bmatrix}$$

são autovetores associados a λ_1 e λ_2, respectivamente, e que $V = [\mathbf{v}_1 \mid \mathbf{v}_2]$ diagonaliza A^TA ortogonalmente. Pela parte (*d*) do Teorema 9.5.4, os vetores

$$\mathbf{u}_1 = \frac{1}{\sigma_1}A\mathbf{v}_1 = \frac{\sqrt{3}}{3}\begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{6}}{3} \\ \frac{\sqrt{6}}{6} \\ \frac{\sqrt{6}}{6} \end{bmatrix}$$

$$\mathbf{u}_2 = \frac{1}{\sigma_2}A\mathbf{v}_2 = (1)\begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \end{bmatrix} = \begin{bmatrix} 0 \\ -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix}$$

Eugenio Beltrami
(1835–1900)



Camille Jordan
(1838–1922)

Herman Klaus Weyl
(1885–1955)

Gene H. Golub
(1932–)

Nota histórica A teoria da decomposição em valores singulares pode ser traçada até o trabalho de cinco pessoas: o matemático italiano Eugenio Beltrami, o matemático francês Camille Jordan, o matemático inglês James Sylvester (ver página 34), e os matemáticos alemães Erhard Schmidt (ver página 360) e Herman Weyl. Mais recentemente, os esforços pioneiros do matemático norte-americano Gene Golub produziram um algoritmo estável e eficaz para calculá-la. Beltrami e Jordan foram os pais da decomposição, sendo que, em 1873, Beltrami deu uma prova para o caso de matrizes reais invertíveis com valores singulares distintos. Subsequentemente, Jordan refinou a teoria e eliminou as restrições desnecessárias impostas por Beltrami. Sylvester, aparentemente desconhecendo o trabalho de Beltrami e Jordan, redescobriu o resultado em 1889 e indicou sua importância. Schmidt foi a primeira pessoa a mostrar que a decomposição em valores singulares poderia ser usada para aproximar uma matriz por outra de posto menor e, ao fazer isso, ele transformou-a de uma curiosidade matemática numa importante ferramenta prática. Weyl mostrou como encontrar a aproximação de posto menor na presença de erro.

[Imagens: Wikipedia (Beltrami); The Granger Collection, New York (Jordan); Cortesia de Electronic Publishing Services, Inc., New York (Weyl); Wikipedia (Golub)]

são dois dos três vetores coluna de U. Observe que $\mathbf{u}_1$ e $\mathbf{u}_2$ são ortonormais, como esperávamos. Poderíamos estender o conjunto $\{\mathbf{u}_1, \mathbf{u}_2\}$ a uma base ortonormal de R^3. Contudo, os cálculos simplificam se primeiro removermos as incômodas raízes multiplicando $\mathbf{u}_1$ e $\mathbf{u}_2$ por escalares apropriados. Assim, procuramos um vetor unitário $\mathbf{u}_3$ ortogonal a

$$\sqrt{6}\,\mathbf{u}_1 = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} \quad \text{e} \quad \sqrt{2}\,\mathbf{u}_2 = \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}$$

Para satisfazer essas duas condições de ortogonalidade, o vetor $\mathbf{u}_3$ deve ser uma solução do sistema linear homogêneo

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

Deixamos para o leitor mostrar que uma solução geral desse sistema é

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = t \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$$

Normalizando o vetor do lado direito, obtemos

$$\mathbf{u}_3 = \begin{bmatrix} -\frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{bmatrix}$$

Assim, a decomposição em valores singulares de A é

$$\underset{A}{\begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}} = \underset{U}{\begin{bmatrix} \frac{\sqrt{6}}{3} & 0 & -\frac{1}{\sqrt{3}} \\ \frac{\sqrt{6}}{6} & -\frac{\sqrt{2}}{2} & \frac{1}{\sqrt{3}} \\ \frac{\sqrt{6}}{6} & \frac{\sqrt{2}}{2} & \frac{1}{\sqrt{3}} \end{bmatrix}} \underset{\Sigma}{\begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}} \underset{V^T}{\begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{bmatrix}}$$

O leitor pode querer confirmar a validade dessa equação multiplicando as matrizes do lado direito. ◀

Concluímos esta seção com uma prova opcional do Teorema 9.5.4.

OPCIONAL

Prova do Teorema 9.5.4 Para simplificar a notação, vamos provar o teorema no caso em que A é uma matriz $n \times n$. Para estender o argumento para uma matriz $m \times n$, basta ajustar a notação usada para levar em conta as possibilidades $m > n$ ou $n > m$.

A matriz A^TA é simétrica, portanto, possui uma decomposição em autovalores

$$A^TA = VDV^T$$

em que os vetores coluna de

$$V = [\mathbf{v}_1 \mid \mathbf{v}_2 \mid \cdots \mid \mathbf{v}_n]$$

são autovetores unitários de A^TA e D é a matriz diagonal cujas entradas diagonais sucessivas $\lambda_1, \lambda_2, \ldots, \lambda_n$ são os autovalores de A^TA que correspondem em sucessão aos vetores coluna de V. Como A tem posto k por hipótese, segue do Teorema 9.5.1 que A^TA também tem posto k. Segue disso que também D tem posto k, por ser semelhante a A^TA e o posto ser um invariante de semelhança. Assim, podemos escrever D como

$$D = \begin{bmatrix} \lambda_1 & & & & & & 0 \\ & \lambda_2 & & & & & \\ & & \ddots & & & & \\ & & & \lambda_k & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ 0 & & & & & & 0 \end{bmatrix} \tag{2}$$

em que $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k > 0$. Agora considere o conjunto de imagens

$$\{A\mathbf{v}_1, A\mathbf{v}_2, \ldots, A\mathbf{v}_n\} \tag{3}$$

Esse conjunto é ortogonal, pois, se $i \neq j$, então a ortogonalidade de $\mathbf{v}_i$ e $\mathbf{v}_j$ implica

$$A\mathbf{v}_i \cdot A\mathbf{v}_j = \mathbf{v}_i \cdot A^TA\mathbf{v}_j = \mathbf{v}_i \cdot \lambda j\, \mathbf{v}_j = \lambda_j\, (\mathbf{v}_i \cdot \mathbf{v}_j\,) = 0$$

Além disso, os k primeiros vetores em (3) são não nulos, pois mostramos na prova do Teorema 9.5.2b que $\|A\mathbf{v}_i\|^2 = \lambda_i$ com $i = 1, 2, \ldots, n$ e estamos supondo que as k primeiras entradas diagonais em (2) são positivas. Assim

$$S = \{A\mathbf{v}_1, A\mathbf{v}_2, \ldots, A\mathbf{v}_k\}$$

é um conjunto ortogonal de vetores *não nulos* no espaço coluna de A. Mas o espaço coluna de A tem dimensão k, pois

$$\text{pos}(A) = \text{pos}(A^TA) = k$$

e, portanto, sendo um conjunto linearmente independente de k vetores, S necessariamente é uma base ortogonal de col(A). Normalizando esses vetores em S, obtemos uma base ortonormal $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ de col(A) em que

$$= \frac{A\mathbf{v}_i}{\|A\mathbf{v}_i\|} = \frac{1}{\sqrt{\lambda_i}} A\mathbf{v}_i \qquad (1 \leq i \leq k)$$

ou, equivalentemente,

$$A\mathbf{v}_1 = \sqrt{\lambda_1}\mathbf{u}_1 = \sigma_1\mathbf{u}_1, \quad A\mathbf{v}_2 = \sqrt{\lambda_2}\mathbf{u}_2 = \sigma_2\mathbf{u}_2, \ldots, \quad A\mathbf{v}_k = \sqrt{\lambda_k}\mathbf{u}_k = \sigma_k\mathbf{u}_k \tag{4}$$

Segue do Teorema 6.3.6 que podemos estender isso a uma base ortonormal

$$\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k, \mathbf{u}_{k+1}, \ldots, \mathbf{u}_n\}$$

de R^n. Sejam, agora, U a matriz ortogonal

$$U = [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \cdots \quad \mathbf{u}_k \quad \mathbf{u}_{k+1} \quad \cdots \quad \mathbf{u}_n]$$

e Σ a matriz diagonal

$$\Sigma = \begin{bmatrix} \sigma_1 & & & & & & 0 \\ & \sigma_2 & & & & & \\ & & \ddots & & & & \\ & & & \sigma_k & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ 0 & & & & & & 0 \end{bmatrix}$$

Segue de (4) e do fato de que $A\mathbf{v}_i = 0$ com $i > k$, que

$$\begin{aligned} U\Sigma &= [\sigma_1\mathbf{u}_1 \quad \sigma_2\mathbf{u}_2 \quad \cdots \quad \sigma_k\mathbf{u}_k \quad 0 \quad \cdots \quad 0] \\ &= [A\mathbf{v}_1 \quad A\mathbf{v}_2 \quad \cdots \quad A\mathbf{v}_k \quad A\mathbf{v}_{k+1} \quad \cdots \quad A\mathbf{v}_n] \\ &= AV \end{aligned}$$

que, usando a ortogonalidade de V, pode ser reescrito como $A = U\Sigma V^T$. ◀

Revisão de conceitos

- Decomposição em autovalores
- Decomposição de Hessenberg
- Decomposição de Schur
- Magnitude do erro de arredondamento
- Propriedades comuns a A e A^TA
- A^TA é ortogonalmente diagonalizável
- Os autovalores de A^TA são não negativos
- Valores singulares
- Entradas diagonais de uma matriz que não é quadrada
- Decomposição em valores singulares

Aptidões desenvolvidas

- Encontrar os valores singulares de uma matriz $m \times n$.
- Encontrar uma decomposição em valores singulares de uma matriz $m \times n$.

Conjunto de exercícios 9.5

▶ Nos Exercícios **1–4**, encontre os valores singulares distintos de A. ◀

1. $A = [1 \quad 2 \quad 0]$

2. $A = \begin{bmatrix} 3 & 0 \\ 0 & 4 \end{bmatrix}$

3. $A = \begin{bmatrix} 1 & -2 \\ 2 & 1 \end{bmatrix}$

4. $A = \begin{bmatrix} \sqrt{2} & 0 \\ 1 & \sqrt{2} \end{bmatrix}$

▶ Nos Exercícios **5–12**, encontre uma decomposição em valores singulares de A. ◀

5. $A = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$

6. $A = \begin{bmatrix} -3 & 0 \\ 0 & -4 \end{bmatrix}$

7. $A = \begin{bmatrix} 4 & 6 \\ 0 & 4 \end{bmatrix}$

8. $A = \begin{bmatrix} 3 & 3 \\ 3 & 3 \end{bmatrix}$

9. $A = \begin{bmatrix} -2 & 2 \\ -1 & 1 \\ 2 & -2 \end{bmatrix}$

10. $A = \begin{bmatrix} -2 & -1 & 2 \\ 2 & 1 & -2 \end{bmatrix}$

11. $A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ -1 & 1 \end{bmatrix}$

12. $A = \begin{bmatrix} 6 & 4 \\ 0 & 0 \\ 4 & 0 \end{bmatrix}$

13. Prove: se A for uma matriz $m \times n$, então A^TA e AA^T têm o mesmo posto.

14. Prove a parte (d) do Teorema 9.5.1 usando a parte (a) desse teorema e o fato de que A e A^TA têm n colunas.

15. (a) Prove a parte (b) do Teorema 9.5.1 mostrando primeiro que $\text{lin}(A^TA)$ é um subespaço de $\text{lin}(A)$.

(b) Prove a parte (c) do Teorema 9.5.1 usando a parte (b).

16. Sejam $T: R^n \to R^m$ uma transformação linear cuja matriz canônica A tem uma decomposição em valores singulares $A = U\Sigma V^T$ e $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ e $B' = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$ os vetores coluna de V e U, respectivamente. Mostre que $\Sigma = [T]_{B',B}$.

17. Mostre que os valores singulares de A^TA são os quadrados dos valores singulares de A.

18. Mostre que se $A = U\Sigma V^T$ for uma decomposição em valores singulares de A, então U diagonaliza AA^T ortogonalmente.

Exercícios verdadeiro/falso

Nas partes (a)-(g), determine se a afirmação é verdadeira ou falsa, justificando sua resposta.

(a) Se A for uma matriz $m \times n$, então A^TA é uma matriz $m \times n$.

(b) Se A for uma matriz $m \times n$, então A^TA é uma matriz simétrica.

(c) Se A for uma matriz $m \times n$, então os autovalores de A^TA são números reais positivos.

(d) Se A for uma matriz $n \times n$, então A é ortogonalmente diagonalizável.

(e) Se A for uma matriz $m \times n$, então A^TA é ortogonalmente diagonalizável.

(f) Os autovalores de A^TA são valores singulares de A.

(g) Qualquer matriz $m \times n$ tem uma decomposição em valores singulares.

CAPÍTULO 18

Aplicações lineares e valores singulares

1. A decomposição em valores singulares

Seja $A : \mathbb{R}^n \to \mathbb{R}^m$ uma aplicação linear. Os espaços R^n e R^m são munidos do produto interno canônico.

Vamos mostrar que existem uma base ortonormal $(\mathbf{v}_1, \ldots, \mathbf{v}_n)$ de $\mathbb{R}^n$, e uma base ortonormal $(\mathbf{u}_1, \ldots, \mathbf{u}_m)$ de $\mathbb{R}^m$, tais que a matriz associada $\Sigma = A_{\mathbf{u},\mathbf{v}}$ a A relativa a essas duas bases é diagonal positiva.

Vamos escrever esse resultado em forma de fatoração matricial:

Teorema 18.1. *Seja A uma matriz real de tamanho $m \times n$. Então existem $U \in O(m)$, $V \in O(n)$ e Σ diagonal positiva de tamanho $m \times n$, tais que*

$$A = U \Sigma V^T.$$

Além disso, podemos escolher Σ de maneira a que

$$\Sigma_{ij} = \begin{cases} 0 & \text{se } i \neq j \\ \sigma_i & \text{se } i = j, \end{cases}$$

com $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$.

Os números $\sigma_1, \ldots, \sigma_n$ são chamados de *valores singulares* da aplicação A. Dependem dos produtos internos utilizados nos espaços $\mathbb{R}^m$ e $\mathbb{R}^n$. Os vetores $\mathbf{u}_i$, ou colunas de U, são chamados de *vetores singulares* à esquerda e os vetores $\mathbf{v}_j$ (colunas de V) de *vetores singulares* à direita.

Demonstração. Vamos assumir sem perda de generalidade que $m \geq n$. (Para o caso $m < n$, basta substituir A por A^T).

A matriz $A^T A$ é real e simétrica. Pelo Teorema 17.1, ela admite uma base ortonormal $(\mathbf{v}_1, \ldots, \mathbf{v}_n)$ de autovetores. Denotamos por λ_i os autovalores de $A^T A$. Para todo $i = 1, \ldots, n$, definimos

$$\sigma_i = \|A\mathbf{v}_i\|$$

e assumimos que a base $\mathbf{v}_i$ está ordenada de maneira que $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$. Seja r o posto de A, então $\sigma_1, \ldots, \sigma_r \neq 0$ e para todo $i \in \{1, 2, \ldots, r\}$, podemos definir

$$\mathbf{u}_i = \sigma_i^{-1} A \mathbf{v}_i$$

Por construção, $\|\mathbf{u}_i\| = 1$. (Note que isso implica que $\sigma_i^2 = \lambda_i$). Como para todo $i \neq j$, $\mathbf{v}_i (A^T A) \mathbf{v}_j = \lambda_j \mathbf{v}_i^T \mathbf{v}_j = 0$, teremos que $\langle \mathbf{u}_i, \mathbf{u}_j \rangle = 0$ e os $\mathbf{u}_i$ formam um conjunto ortonormal.

Existe uma base $(\mathbf{u}_1, \dots, \mathbf{u}_r, \mathbf{w}_{r+1}, \dots, \mathbf{w}_m)$ de $\mathbb{R}^n$. Aplicando Gram-Schmidt a essa base, obtemos um base ortonormal $(\mathbf{u}_1, \dots, \mathbf{u}_m)$. Como todo vetor ortogonal à imagem de A pertence ao núcleo de A^T,

$$A = U \begin{bmatrix} \sigma_1 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & 0 & \cdots & 0 \\ 0 & 0 & \sigma_3 & & \\ \vdots & \vdots & & \ddots & \\ 0 & 0 & & & \sigma_r \\ 0 & 0 & & & 0 \\ \vdots & \vdots & & & \vdots \\ 0 & 0 & & & 0 \end{bmatrix} V^T$$

□

Uma interpretação geométrica é a seguinte. Assuma que A é de tamanho $m \times n$, com $m \geq n$ e posto n). Seja

$$\mathcal{E} = \{A\mathbf{x} : \|\mathbf{x}\| \leq 1\}$$

O conjunto $\mathcal{E}$ é o elipsoide de centro zero, e semieixos $\sigma_i \mathbf{u}_i$.

2. Aplicações à mineração de dados

A Internet é acessada por aproximadamente 1,2 bilhões de pessoas, que frequentemente precisam procurar páginas contendo uma ou mais palavras chave.

Em janeiro de 2005, foi estimado[1] que existiriam 11,5 bilhões de páginas na *World Wide Web* indexadas pelos principais sistemas de busca.

Para evitar a necessidade de armazenar uma cópia da World Wide Web inteira no site de busca, o seguinte algoritmo foi sugerido para fazer a associação entre páginas e palavras:

Define-se a matriz A, de tamanho $m \times n$, onde m é o número de páginas (11,5 bilhões) e n o número de palavras indexáveis (dicionário, procuras frequentes, etc...). (No início da *Google*, em torno de 24 milhões). (Retira-se artigos como "o", "e", etc...).

A coordenada A_{ij} é o número de ocorrências da j-ésima palavra na i-ésima página. A rede é constantemente vasculhada por programas *rastejadores* (em inglês, *crawlers* ou *bots*, contração de 'robôs') que fazem a atualização da matriz.

Google[2] afirma manter cópia comprimida de todos os documentos na internet. Mas com o aumento vertiginoso do conteúdo visível na internet, pode não ser viável (e talvez não seja conveniente hoje) armazenar a matriz A inteira, mesmo de modo esparso. Uma possibilidade é armazenar uma aproximação de A. Por exemplo, a melhor aproximação de posto k para a matriz A.

[1] Antonio Gulli e Alessio Signorini, *The Indexable Web is more than 11.5 billion pages*, Preprint, http://www.cs.uiowa.edu/~asignori/web-size/

[2] http://www.google.com.br/intl/pt-BR/features.html

Assuma que

$$A = U \begin{bmatrix} \sigma_1 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & 0 & \cdots & 0 \\ 0 & 0 & \sigma_3 & & \\ \vdots & \vdots & & \ddots & \\ 0 & 0 & & & \sigma_r \\ 0 & 0 & & & 0 \\ \vdots & \vdots & & & \vdots \\ 0 & 0 & & & 0 \end{bmatrix} V^T$$

Então prova-se que a melhor aproximação de posto k é

$$A = \begin{bmatrix} \mathbf{u}_1 & \dots & \mathbf{u}_k \end{bmatrix} \begin{bmatrix} \sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_k \end{bmatrix} \begin{bmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_k \end{bmatrix}^T$$

que é um objeto de tamanho $(m + n + 1)k$, bem menor do que mn.

Um algoritmo de compressão similar foi proposto para imagens (A_{ij} seria a intensidade do pixel (i, j). Mas não parece existir evidência de alguma vantagem em relação aos formatos de compressão usuais, como jpg.

3. A pseudo-inversa

Nem toda matriz é inversível. Mas toda matriz (mesmo sendo quadrada) tem uma pseudo-inversa.

Definição 18.2. Seja A uma matriz real de tamanho $m \times n$. A *pseudo-inversa* de A, denotada por $A^\dagger$, é a matriz de tamanho $n \times m$ dada por:

$$A^\dagger \mathbf{y} = \mathbf{x}$$

onde $A\mathbf{x} = \pi \mathbf{y}$, π denota a projeção ortogonal de $\mathbf{y}$ em $\mathrm{Im} A$, e $\|\mathbf{x}\|$ é minimal sob essas condições. Em particular, $\mathbf{x} \perp \ker A$.

A pseudo-inversa de uma matriz diagonal (por exemplo, $m > n$) é portanto:

$$\begin{bmatrix} \sigma_1 & \cdots & 0 \\ & \ddots & \\ 0 & \cdots & \sigma_n \\ 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{bmatrix}^\dagger = \begin{bmatrix} \rho_1 & \cdots & 0 & 0 & \cdots & 0 \\ & \ddots & & \vdots & & \vdots \\ 0 & \cdots & \rho_n & 0 & \cdots & 0 \end{bmatrix}$$

onde $\rho_j = \sigma_j^{-1}$ quando $\sigma_j \neq 0$, e $\rho_j = 0$ se $\sigma_j = 0$.

A pseudo-inversa de uma matriz quadrada inversível A é sempre $A^\dagger = A^{-1}$.

Outro caso importante é o de matrizes ortogonais. Se Q for ortogonal, da definição de pseudo-inversa concluimos que:

$$Q^\dagger = Q^T \qquad \text{e} \qquad (Q^T)^\dagger = Q$$

Em geral, se $A = U\Sigma V^T$ é a decomoposição em valores singulares de A, $A^\dagger = V\Sigma^\dagger U$ (Exercício 18.6).

Finalmente, podemos escrever a solução geral do problema de mínimos quadrados utilizando a pseudo-inversa: O valor de $\mathbf{x}$ que minimiza $\|A\mathbf{x} - \mathbf{b}\|^2$ com $\|\mathbf{x}\|$ minimal é $\mathbf{x} = A^\dagger \mathbf{b}$.

4. Exercícios

Exercício 18.1. Compare os autovalores e autovetores de $\begin{bmatrix} 0 & A^T \\ A & 0 \end{bmatrix}$ com os valores singulares e vetores singulares de A.

Exercício 18.2. Mostre que para todo vetor $\mathbf{x}$, $\|A\mathbf{x}\| \leq \sigma_1 \|\mathbf{x}\|$, onde σ_1 é o maior valor singular de A.

Exercício 18.3. Se A for sobrejetiva, mostre que $AA^\dagger$ é a identidade. Mostre um exemplo de matriz A não sobrejetiva com $AA^\dagger \neq I$.

Exercício 18.4. Se A for injetiva, mostre que $A^\dagger A$ é a identidade. Mostre um exemplo de matriz A não injetiva com $A^\dagger A \neq I$.

Exercício 18.5. Mostre que se A é uma matriz qualquer, então $A^\dagger A$ é a identidade, ou é uma projeção.

Exercício 18.6. Mostre que se $A = U\Sigma V^T$ é a decomoposição em valores singulares de A, $A^\dagger = V\Sigma^\dagger U^T$.

Exercício 18.7. Ache um exemplo onde $(AB)^\dagger \neq B^\dagger A^\dagger$.'

Exercício 18.8. Seja A uma matriz $m \times n$, com $m > n$ e valores singulares $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$. Seja W um subespaço r-dimensional de $\mathbb{R}^n$. Mostre que $\|A_{|W}\| \geq \sigma_{n-r+1}$.

Exercício 18.9. Descreva o conjunto de todas as triplas (U, Σ, V) tal que $A = U\Sigma V^T$ seja a decomposição em valores singulares de A, para $A = \begin{bmatrix} 2 & 3 \\ 1 & 1 \end{bmatrix}$.

Exercício 18.10. Mesmo problema, para $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 3 \\ 0 & 1 & 1 \end{bmatrix}$.