Matrizes Hermitianas e Simétricas

Já vimos anteriormente que uma matriz é hermitiana se $A^H = A$, e simétrica se $A \in \mathbb{R}^{n\times n}$ e $A^T = A$. Vimos também que matrizes hermitianas são <u>normais</u>, e portanto, diagonalizáveis por transformações unitárias.

Matrizes hermitianas ocorrem frequentemente na prática, e as propriedades acima são amplamente utilizadas.

Exemplos de matrizes hermitianas:

1) Seja $f : D \to \mathbb{R}^n$ uma função duplamente continuamente derivável no domínio $D \subset \mathbb{R}^n$. A matriz

$$H(x) = [h_{ij}(\underset{\sim}{x})] = \left[\frac{\partial^2 f(\underset{\sim}{x})}{\partial x_i \partial x_j}\right] \in \mathbb{R}^{n\times n}$$

é conhecida como <u>hessiano</u> de $f(x)$. Com as definição de $f(\cdot)$ acima, $\frac{\partial^2 f(\underset{\sim}{x})}{\partial x_i \partial x_j} = \frac{\partial^2 f(\underset{\sim}{x})}{\partial x_j \partial x_i}$, ou seja, $H(x) = H^T(x)$.

O hessiano é útil no estudo de minimização de funções diferenciáveis, tanto teoricaquanto numericamente.

2) Seja $\underset{\sim}{x}_n$ um vetor aleatóreo, e defina

$R = E(\underset{\sim}{x}_n \underset{\sim}{x}_n^H)$, a matriz de autocorrelação de $\underset{\sim}{x}_n$.

(Matriz de autocovariância se $E\underset{\sim}{x}_n = \underset{\sim}{0}$).

R é hermitiana, pois as operações $E(\cdot)$ e $(\cdot)^H$ comutam.

Propriedades de matrizes hermitianas

Algumas propriedades úteis de matrizes hermitianas são

1) Se $A^H = A$, A é diagonalizável por uma transformação unitária: Existe U unitária tal que

$$U^H A U = \Lambda = \text{diag}(\lambda_i).$$

Prova: Teorema 1 sobre matrizes normais.

2) Se $A^H = A$, então

a) $\underset{\sim}{x}^H A \underset{\sim}{x} \in \mathbb{R}$ para todo $\underset{\sim}{x} \in \mathbb{C}^n$;

b) Todos os autovalores de A são reais

c) $S^H A S$ é hermitiana para todo $S \in \mathbb{C}^{n\times n}$.

Prova: (a) $(\underset{\sim}{x}^H A \underset{\sim}{x})^H = \overline{\underset{\sim}{x}^H A \underset{\sim}{x}} = \underset{\sim}{x}^H A^H \underset{\sim}{x} = \underset{\sim}{x}^H A \underset{\sim}{x}$.

Como $\bar{a} = a \Leftrightarrow a \in \mathbb{R}$, o resultado está provado.

(b) Seja λ um autovalor de A, com autovetor $\underset{\sim}{x}$ satisfazendo $\underset{\sim}{x}^H \underset{\sim}{x} = 1$. Então,

$$\lambda = \lambda \underset{\sim}{x}^H \underset{\sim}{x} = \underset{\sim}{x}^H A \underset{\sim}{x} \in \mathbb{R}, \text{ por (a).}$$

(c) $(S^H A S)^H = S^H (S^H A)^H = S^H A^H S = S^H A S.$ ∎

Na verdade, as propriedades acima definem matrizes hermitianas ((b) precisa ser um pouco modificada):

Teorema 1: $A \in \mathbb{C}^{n\times n}$ é hermitiana se e somente se uma das condições abaixo for satisfeita:

(a) $\underset{\sim}{x}^H A \underset{\sim}{x} \in \mathbb{R}$ para todo $\underset{\sim}{x} \in \mathbb{C}^n$,

(b) A é normal e todos os ~~seus~~ autovalores são reais,

(c) $S^H A S$ é hermitiana para todo $S \in \mathbb{C}^{n\times n}$.

(d) Existe uma transformação unitária U e uma matriz real e diagonal Λ tais que $A = U \Lambda U^H$.

Para matrizes hermitianas, pode-se definir autovalores de uma maneira alternativa, através de problemas de minimização e maximização. Esta caracterização levará à definição de valores singulares no futuro.

<u>Teorema 2</u>: Seja $A \in \mathbb{C}^{n\times n}$, e sejam $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ os autovalores de A. Então vale para todo $\underset{\sim}{x} \in \mathbb{C}^n$, com $\|\underset{\sim}{x}\|_2 = 1$,

$$\lambda_1 \leq \underset{\sim}{x}^H A \underset{\sim}{x} \leq \lambda_n, \quad \text{ou} \qquad \lambda_{max} = \lambda_n = \max_{\underset{\sim}{x} \neq \underset{\sim}{0}} \frac{\underset{\sim}{x}^H A \underset{\sim}{x}}{\underset{\sim}{x}^H \underset{\sim}{x}}$$

$$\lambda_{min} = \lambda_1 = \min_{\underset{\sim}{x} \neq \underset{\sim}{0}} \frac{\underset{\sim}{x}^H A \underset{\sim}{x}}{\underset{\sim}{x}^H \underset{\sim}{x}}.$$

<u>Prova</u>: Seja U unitária que diagonaliza A. Então

$$\underset{\sim}{x}^H A \underset{\sim}{x} = \underset{\sim}{x}^H U \Lambda U^H \underset{\sim}{x} \triangleq \underset{\sim}{y}^H \Lambda \underset{\sim}{y} = \sum_{i=1}^{n} \lambda_i |y_i|^2$$

Como $|y_i|^2 \geq 0$, vem

$$(1) \qquad \lambda_{min} \cdot \sum_{i=1}^{n} |y_i|^2 \leq \sum_{i=1}^{n} \lambda_i |y_i|^2 = \underset{\sim}{x}^H A \underset{\sim}{x} \leq \lambda_{max} \sum_{i=1}^{n} |y_i|^2.$$

Como U é unitária, $\sum_{i=1}^{n} |y_i|^2 = \|\underset{\sim}{y}\|_2^2 = \|\underset{\sim}{x}\|_2^2$.

Assuma que $\Lambda = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$ (isto sempre pode ser conseguido reordenando-se as colunas de U). Então, escolhendo $\underset{\sim}{x}$ tal que $U\underset{\sim}{x} = \underset{\sim}{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$, temos

$$\underset{\sim}{x}^H A \underset{\sim}{x} = \lambda_1 \cdot |1|^2 = \lambda_1 \cdot \|\underset{\sim}{x}\|_2^2,$$

ou seja a primeira desigualdade de (1) torna-se uma igualdade para esta particular escolha de $\underset{\sim}{x}$. Portanto,

$$\lambda_1 = \min_{\underset{\sim}{y} \neq \underset{\sim}{0}} \frac{\sum_{i=1}^{n} \lambda_i |y_i|^2}{\|\underset{\sim}{y}\|_2^2} = \min_{\underset{\sim}{x} \neq \underset{\sim}{0}} \frac{\underset{\sim}{x}^H A \underset{\sim}{x}}{\underset{\sim}{x}^H \underset{\sim}{x}}.$$

Usando-se um argumento semelhante para λ_n, conclui-se a demonstração. ■

Matrizes Positivas - Definidas

Vamos começar esta seção definindo e apresentando algumas propriedades de matrizes positivas - definidas. Em seguida apresentaremos alguns exemplos.

Definição 1: Uma matriz hermitiana $A \in \mathbb{C}^{n \times n}$ é positiva - definida se

$$\underset{\sim}{x}^H A \underset{\sim}{x} > 0 \quad \text{para todo} \quad \underset{\sim}{x} \in \mathbb{C}^n.$$

Se valer apenas $\underset{\sim}{x}^H A \underset{\sim}{x} \geq 0$, diz-se que A é positiva - semidefinida. Analogamente, diz-se que A é negativa (semi)definida se $(-A)$ for positiva (semi) definida.

Propriedades Básicas (as demonstrações ficam como exercícios)

1) Se $A \in \mathbb{C}^{n \times n}$ for positiva definida, todos os seus autovalores são reais e positivos (reais e não negativos se A for positiva semi definida)

2) Considere as seguintes submatrizes obtidas a partir de $A = [a_{ij}]$ tomando-se apenas algumas das linhas e colunas de A:

Sejam $\alpha = \{ i_1, i_2, \cdots, i_{n_\alpha} \}$, $1 \leq i_1 < i_2 < \cdots < i_{n_\alpha} \leq n$

$\beta = \{ j_1, j_2, \cdots, j_{n_\beta} \}$, $1 \leq j_1 < j_2 < \cdots < j_{n_\beta} < n$

$n_\alpha, n_\beta \leq n$.

A submatriz $A_{\alpha\beta}$ de A é a matriz formada tomando-se apenas os elementos $a_{i_k j_\ell}$, $\begin{cases} i_k \in \alpha \\ j_\ell \in \beta \end{cases}$, ou seja:

$$A_{\alpha\beta} = [a_{i_k j_\ell}] = \begin{bmatrix} a_{i_1 j_1} & a_{i_1 j_2} & \cdots & a_{i_1 j_{n_\beta}} \\ \vdots & & & \\ a_{i_{n_\alpha} j_1} & a_{i_{n_\alpha} j_2} & \cdots & a_{i_{n_\alpha} j_{n_\beta}} \end{bmatrix}$$

Por exemplo, se $A = \begin{bmatrix} \cancel{1} & \cancel{0} & \cancel{2} & \cancel{1} \\ -1 & 1 & \cancel{0} & \cancel{2} \\ -2 & 0 & \cancel{0} & \cancel{3} \\ \cancel{0} & \cancel{-1} & \cancel{1} & \cancel{1} \end{bmatrix}$, e $\alpha = \{2, 3\}$, $\beta = \{1, 2\}$,

então $A_{\alpha\beta} = \begin{bmatrix} -1 & 1 \\ -2 & 0 \end{bmatrix}$.

Caso $\alpha = \beta$, $A_{\alpha\beta}$ é chamada uma <u>submatriz principal</u> de A, e $\det(A_{\alpha\beta})$ é um <u>menor principal</u> de A.

Se A for positiva definida, toda submatriz principal de A também é positiva definida, e todo menor principal é positivo.

Se A for positiva semi definida, as submatrizes principais também o serão, e os menores principais serão não-negativos.

3) Corolário de (1): Se A é positiva definida, A é invertível.

4) Corolário de (2): Se A é positiva definida, os elementos da diagonal principal são positivos. Em particular, $Tr(A) > 0$

5) A soma de duas matrizes positivas-definidas também é positiva-definida.

6) Se $A \in \mathbb{C}^{n \times n}$ é positiva-definida e $C \in \mathbb{C}^{n \times m}$, então $C^H A C$ é positiva definida se e somente se posto$(C) = m$. Se posto$(C) < m$, então $C^H A C$ é positiva semi-definida. Se A é positiva semi-definida, então $C^H A C$ é sempre positiva semi-definida.

7) A é positiva-definida se e somente se A^{-1} for positiva-definida.

8) Seja $A \in \mathbb{C}^{n \times n}$ positiva-definida. Então para qualquer $k \in \mathbb{Z}$, $|k| \geq 1$, existe $B \in \mathbb{C}^{n \times n}$ tal que $B^k = A$.

Prova de 8 ③

Sendo A hermitiana, existem U unitária e Λ diagonal tais que $A = U \Lambda U^H$. Como $\lambda_i > 0$ (λ_i são os elementos da diagonal de Λ), podemos definir

$$B = U \begin{bmatrix} \lambda_1^{1/k} & & & \\ & \lambda_2^{1/k} & & \\ & & \ddots & \\ & & & \lambda_n^{1/k} \end{bmatrix} U^H \triangleq U \Lambda^{1/k} U^H.$$

Confirme que $B^k = A$. Note que sempre é possível escolher B positiva-definida também. ∎

9) Decomposição de Cholesky.

Uma matriz $A \in \mathbb{C}^{n\times n}$ é positiva-definida se e somente se existir $L \in \mathbb{C}^{n\times n}$ (triangular inferior com elementos positivos na diagonal) tal que

$$A = L L^H.$$

Se A for real, L pode ser escolhida real também.

Prova: Por (8), existe B tal que $B^2 = A$, e podemos escolher B positiva definida. Sejam Q e R as matrizes unitária e triangular superior tais que $B = QR$.

Então

$$A = B^2 = B^H B = R^H Q^H Q R = R^H R.$$

Como A é positiva-definida e R é triangular, temos $0 < \det(A) = |\det(R)|^2 = \prod_{i=1}^{n} |r_{ii}|^2$, mas em geral $r_{ii} \notin \mathbb{R}$ (e portanto, não podemos dizer que $r_{ii} > 0$). No entanto, podemos definir

$$L^H = \begin{bmatrix} e^{j\theta_1} & & & \\ & e^{j\theta_2} & & \\ & & \ddots & \\ & & & e^{j\theta_n} \end{bmatrix} \cdot R = \begin{bmatrix} e^{j\theta_1} r_{11} & e^{j\theta_1} r_{12} & \cdots & e^{j\theta_1} r_{1n} \\ 0 & e^{j\theta_2} r_{22} & \cdots & e^{j\theta_2} r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e^{j\theta_n} r_{nn} \end{bmatrix}$$

Se escolhermos θ_i de forma que

$e^{j\theta_i} \cdot r_{ii} = |r_{ii}|$, a matriz L satisfaz as condições do teorema. ∎

Exemplo: Análise de Filtro Adaptativo LMS.

Considere duas seqüências de variáveis aleatóreas $\{\underset{\sim}{x}_k\}_{k=0}^{\infty}$, $\underset{\sim}{x}_k \in \mathbb{R}^n$, e $\{y(k)\}_{k=0}^{\infty}$, $y(k) \in \mathbb{R}$. Pode-se demonstrar que, se $\{y(k), \underset{\sim}{x}_k\}$ for uma seqüência estacionária ($E\,y(k)=0$, $E\,\underset{\sim}{x}_k=\underset{\sim}{0}$), então existe um vetor de parâmetros $\underset{\sim}{w}_*$ que relaciona $\underset{\sim}{x}_k$ e $y(k)$ linearmente como abaixo:

$$y(k) = \underset{\sim}{w}_*^T \underset{\sim}{x}_k + v(k),$$

onde $v(k)$ é o erro de estimação (uma outra variável aleatórea), e $v(k)$ é <u>não correlacionado</u> com $\underset{\sim}{x}_k$, isto é, $E\,v(k)\,\underset{\sim}{x}_k = \underset{\sim}{0}$.

O filtro adaptativo LMS procura calcular uma estimativa para $\underset{\sim}{w}_*$ atravéz da recursão

$$\underset{\sim}{w}_{k+1} = \underset{\sim}{w}_k + \mu \underset{\sim}{x}_k \left(y(k) - \underset{\sim}{w}_k^T \underset{\sim}{x}_k\right). \quad (1)$$

A análise estocástica do filtro acima procura calcular para que valores a média e a matriz de autocorrelação de $\underset{\sim}{w}_k$ convergem, isto é,

$$\lim_{k\to\infty} E\,\tilde{\underset{\sim}{w}}_k \quad , \quad \text{e} \quad \lim_{k\to\infty} E\,\tilde{\underset{\sim}{w}}_k \tilde{\underset{\sim}{w}}_k^T, \text{ onde } \tilde{\underset{\sim}{w}}_k = \underset{\sim}{w}_* - \underset{\sim}{w}_k.$$

Para tanto, usam-se muitas vezes as hipóteses de independência, ou seja, assume-se que $\{y(k), \underset{\sim}{x}_k\}$ são brancos, isto é, que

$$\begin{cases} y(k) \text{ é independente de } \underset{\sim}{x}_j \text{ se } k \neq j \\ y(k) \text{ é independente de } y(j) \text{ se } k \neq j \\ \underset{\sim}{x}_k \text{ é independente de } \underset{\sim}{x}_j \text{ se } k \neq j \\ v(k) \text{ é independente de todas as outras variáveis.} \end{cases}$$

Neste caso, podemos calcular $E\tilde{\underset{\sim}{w}}_k$ e $E\tilde{\underset{\sim}{w}}_k\tilde{\underset{\sim}{w}}_k^T$ como a seguir. Primeiramente, vamos reescrever a recursão (1) em termos de $\tilde{w}_k$:

$$
\begin{aligned}
\underset{\sim}{w}_* - \underset{\sim}{w}_{k+1} &= \underset{\sim}{w}_* - \underset{\sim}{w}_k - \mu \underset{\sim}{x}_k\left(y(k) - \underset{\sim}{w}_k^T \underset{\sim}{x}_k\right) \\
&= \tilde{\underset{\sim}{w}}_k - \mu \underset{\sim}{x}_k\left(\underset{\sim}{w}_*^T \underset{\sim}{x}_k - \underset{\sim}{w}_k^T \underset{\sim}{x}_k\right) \\
&= \tilde{\underset{\sim}{w}}_k - \mu \underset{\sim}{x}_k \underset{\sim}{x}_k^T \tilde{\underset{\sim}{w}}_k = \left(I - \mu \underset{\sim}{x}_k \underset{\sim}{x}_k^T\right)\tilde{\underset{\sim}{w}}_k. \quad (2)
\end{aligned}
$$

Portanto,

$$E\tilde{\underset{\sim}{w}}_{k+1} = E\left[\left(I - \mu \underset{\sim}{x}_k \underset{\sim}{x}_k^T\right)\tilde{\underset{\sim}{w}}_k\right].$$

Como $\underset{\sim}{x}_k$ é independente de $\underset{\sim}{x}_j$ para $j<k$, e de $v(j)$ para $j\leq k$, e como $\tilde{\underset{\sim}{w}}_k$ depende apenas de $\underset{\sim}{x}_j, v(j)$ para $j \leq k-1$, concluímos que $\tilde{\underset{\sim}{w}}_k$ é independente de $\underset{\sim}{x}_k$. Portanto,

$$E\tilde{\underset{\sim}{w}}_{k+1} = E\left(I - \mu \underset{\sim}{x}_k \underset{\sim}{x}_k^T\right) E\tilde{\underset{\sim}{w}}_k.$$

Definindo $R = E\underset{\sim}{x}_k \underset{\sim}{x}_k^T$, temos

$$E\tilde{\underset{\sim}{w}}_{k+1} = (I - \mu R)\, E\tilde{\underset{\sim}{w}}_k,$$

e $$E\tilde{\underset{\sim}{w}}_{k+1} = (I-\mu R)^{k+1} \cdot E\tilde{\underset{\sim}{w}}_0.$$

Portanto, $E\tilde{\underset{\sim}{w}}_k \to \underset{\sim}{0}$ se e somente se $(I-\mu R)^k \to 0$.

Vimos anteriormente que isto ocorre se e somente se $\rho(I-\mu R) < 1$. Vamos então calcular $\rho(I-\mu R)$ (ou seja, o módulo maximo dos autovalores de $I-\mu R$).

Como R é normal, existe Q unitária tal que

$$Q^H R Q = \Lambda = \mathrm{diag}(\lambda_i).$$

Portanto,

$$Q^H(I-\mu R)Q = I - \mu Q^H R Q = I - \mu\Lambda.$$

Concluímos que os autovalores de $I-\mu R$ são $1-\mu\lambda_i$, $1 \leq i \leq n$, e que $E\,\tilde{\underset{\sim}{w}}_k \to \underset{\sim}{0}$ se e somente se

$$|1-\mu\lambda_i| < 1 \quad \text{para todo } \lambda_i.$$

Como R é positiva-semidefinida (por ser uma matriz de autocorrelação), então $\lambda_i \geq 0$ para $1 \leq i \leq n$, e a condição acima fica

$$-1 < 1-\mu\lambda_i < 1 \iff \begin{cases} 1-\mu\lambda_i < 1 \Rightarrow \mu\lambda_i > 0 \quad (A) \\ -1 < 1-\mu\lambda_i \Rightarrow \mu\lambda_i < 2 \quad (B) \end{cases}$$

A condição (A) é satisfeita se R for positiva-definida (ou seja, se R for invertível). (B) é satisfeita se μ for escolhido suficientemente pequeno.

Exemplo: Suponha que a e b são duas variáveis aleatórias reais com $Ea = Eb = 0$,

$$E a^2 = \sigma_a^2, \quad E b^2 = \sigma_b^2, \quad E ab = \sigma_{ab} \neq 0.$$

Suponha também que $\{a, b\}$ são conjuntamente Gaussianas.

Como calcular $E(a^2 b^2)$?

Solução: Defina $R = \begin{bmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{bmatrix}$. R é positiva-semidefinida e normal, portanto existe Q ortogonal tal que

$$Q^T R Q = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}. \qquad (3)$$

Defina $x = q_1^T \begin{bmatrix} a \\ b \end{bmatrix}$, e $y = q_2^T \begin{bmatrix} a \\ b \end{bmatrix}$, onde

$[q_1 \quad q_2] = Q$ (as colunas de Q).

Como a e b são conjuntamente Gaussianas, x e y também serão conjuntamente Gaussianas, e (3) implica que x e y são <u>independentes</u>.

Por outro lado,

$$\begin{bmatrix} x \\ y \end{bmatrix} = Q^T \begin{bmatrix} a \\ b \end{bmatrix} \Rightarrow \begin{bmatrix} a \\ b \end{bmatrix} = Q \begin{bmatrix} x \\ y \end{bmatrix}, \text{ e}$$

$$a = q_{11} x + q_{12} y \qquad b = q_{21} x + q_{22} y$$

$$a^2 = q_{11}^2 x^2 + 2 q_{11} q_{12} xy + q_{12}^2 y^2$$

$$b^2 = q_{21}^2 x^2 + 2 q_{21} q_{22} xy + q_{22}^2 y^2$$

e Portanto,

$$E a^2 b^2 = (E x^4) q_{11}^2 q_{21}^2 + 2(q_{11} q_{12} + q_{21} q_{22})$$

Com isto podemos agora calcular $E a^2 b^2$: ⑧

Como x e y têm média zero (verifique!), e são independentes,

$$E(x \cdot y^3) = \underbrace{\cancel{E x}}_{=0} \cdot E y^3 = 0$$

$$E(x^3 \cdot y) = E x^3 \cdot \underbrace{\cancel{E y}}_{=0} = 0$$

$$E(x^2 \cdot y^2) = (E x^2)(E y^2) = \lambda_1 \cdot \lambda_2$$

Portanto, temos

$$E(a^2 \cdot b^2) = q_{11}^2 \cdot q_{21}^2 \cdot E x^4 + (q_{11}^2 \cdot q_{22}^2 + q_{21}^2 \cdot q_{12}^2 + 4 q_{11} q_{12} q_{21} q_{22}) \lambda_1 \lambda_2$$
$$+ q_{12}^2 \cdot q_{22}^2 \cdot E y^4 .$$

Falta apenas calcular $E x^4$ e $E y^4$, mas isto consegue-se de um livro qualquer sobre probabilidades (por exemplo,)): Como x e y são Gaussianos com médias iguais a 0 e variâncias λ_1 e λ_2 respectivamente, segue que

$$E x^4 = 3 \lambda_1^2 \qquad E y^4 = 3 \lambda_2^2 .$$