Definindo $U = U_1 V_2 \cdots V_n$, completamos a prova.

Se $A \in \mathbb{R}^{n\times n}$ e $\lambda_i \in \mathbb{R}$, $i=1,\dots,n$, então todos os autovetores podem ser escolhidos em $\mathbb{R}^n$, e a mesma demonstração acima pode ser usada para se escolher $U \in \mathbb{R}^{n\times n}$, $U^T U = I$, tal que

$$U^T A U = \begin{bmatrix} \lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}.$$

$\blacksquare$

Note que nem U nem R são únicos!

Aplicações:

1) Prove que, se $A \in \mathbb{C}^{n\times n}$ tem autovalores $\lambda_1, \dots, \lambda_n$, então

$$\det(A) = \prod_{i=1}^{n} \lambda_i \quad \text{e} \quad \mathrm{Tr}(A) = \sum_{i=1}^{n} \lambda_i .$$

Solução: Seja U tal que $U^H A U = R$, triangular superior. Então

a) $\det(R) = \det(U^H A U) = \det(U^H)\det(AU) = \det(AU)\cdot\det(U^H) =$
$= \det(A U U^H) = \det(A)$.

Mas $\det(R) = \lambda_1 \cdot \lambda_2 \cdot \dots \cdot \lambda_n = \prod_{i=1}^{n} \lambda_i$.

b) $\mathrm{Tr}(R) = \mathrm{Tr}(U^H A U) = \mathrm{Tr}(A U U^H) = \mathrm{Tr}(A)$. $\blacksquare$

2) Teorema 3 (Cayley-Hamilton). Seja $p_A(s)$ o polinômio característico de $A \in \mathbb{C}^{n\times n}$. Então,

$$p_A(A) = A^n + \alpha_{n-1} A^{n-1} + \dots + \alpha_0 I = 0,$$

onde $p_A(s) = s^n + \alpha_{n-1} \cdot s^{n-1} + \dots + \alpha_0$.

Prova: Sejam $\lambda_1, \dots, \lambda_n$ os autovalores de A, de forma que possamos escrever

$$p_A(s) = (s-\lambda_1)(s-\lambda_2)\cdots(s-\lambda_n).$$

Então $p_A(A) = (A-\lambda_1 I)(A-\lambda_2 I)\cdots(A-\lambda_n I)$.

Introduzindo a decomposição de Schur de A:

$R = U^H A U$, $\quad A = U R U^H$, temos

$$p_A(A) = (URU^H - \lambda_1 I)(URU^H - \lambda_2 I)\cdots(URU^H - \lambda_n I)$$
$$= (URU^H - \lambda_1 UU^H)(URU^H - \lambda_2 UU^H)\cdots(URU^H - \lambda_n UU^H)$$
$$= U\left[(R-\lambda_1 I)(R-\lambda_2 I)\cdots(R-\lambda_n I)\right]U^H.$$

Portanto, $p_A(A) = 0$ se e somente se $p_A(R) = 0$.

Mas $p_A(R)$ pode ser calculado como a seguir:

$$p_A(R) = \begin{bmatrix} 0 & & & * \\ & \lambda_2-\lambda_1 & & \\ & & \ddots & \\ 0 & & & \lambda_n-\lambda_1 \end{bmatrix} \begin{bmatrix} \lambda_1-\lambda_2 & & & * \\ & 0 & & \\ & & \ddots & \\ 0 & & & \lambda_n-\lambda_2 \end{bmatrix} \cdots \begin{bmatrix} \lambda_1-\lambda_n & & & * \\ & \lambda_2-\lambda_n & & \\ & & \ddots & \\ 0 & & & 0 \end{bmatrix} =$$

$$= \left[\begin{array}{cc|c} 0 & 0 & \\ 0 & 0 & * \\ \hline \multicolumn{2}{c|}{0} & \nabla \end{array}\right] \cdot \begin{bmatrix} \lambda_1-\lambda_3 & & & & * \\ & \lambda_2-\lambda_3 & & & \\ & & 0 & & \\ & & & \ddots & \\ 0 & & & & \lambda_n-\lambda_3 \end{bmatrix} \cdots \begin{bmatrix} \lambda_1-\lambda_n & & & * \\ & \lambda_2-\lambda_n & & \\ & & \ddots & \\ 0 & & & 0 \end{bmatrix}$$

$$= \left[\begin{array}{ccc|c} 0 & 0 & 0 & \\ 0 & 0 & 0 & * \\ 0 & 0 & 0 & \\ \hline \multicolumn{3}{c|}{0} & \nabla \end{array}\right] \cdots \begin{bmatrix} \lambda_1-\lambda_n & & & * \\ & \lambda_2-\lambda_n & & \\ & & \ddots & \\ 0 & & & 0 \end{bmatrix} = 0 \qquad \blacksquare$$

Fatoração QR

A fatoração QR é usada em vários aplicações, desde calcular os autovalores de uma matriz até a resolução de problemas de mínimos quadrados e filtros de Kalman. O resultado básico é o seguinte teorema:

Teorema 4 (fatoração QR): Se $A \in \mathbb{C}^{m\times n}$, existe uma matriz $Q \in \mathbb{C}^{m\times m}$ unitária e uma matriz $R \in \mathbb{C}^{m\times n}$ triangular superior e uma matriz de permutação P tais que $A = QRP$.

Prova: Aplique o algoritmo de Gram-Schmidt às colunas de A:

$A = [a_1, a_2, \dots, a_n]$, $a_i \in \mathbb{C}^m$.

Vamos assumir primeiramente que posto$(A) = n$ (e portanto, $m \geq n$).

Então tomamos

$$q_1 = \frac{a_1}{\|a_1\|_2}$$

$$q_2 = \frac{a_2 - q_1^H a_2 \cdot q_1}{\|a_2 - q_1^H a_2 \cdot q_1\|_2}$$

$$q_3 = \frac{a_3 - q_2^H a_3 \cdot q_2 - q_1^H a_3\, q_1}{\|a_3 - q_2^H a_3 q_2 - q_1^H a_3 q_1\|_2}$$

$$\vdots$$

$$R = \begin{bmatrix} \|a_1\|_2 & q_1^H a_2 & q_1^H a_3 & \cdots \\ 0 & \|a_2 - q_1^H a_2 q_1\|_2 & q_2^H a_3 & \cdots \\ 0 & 0 & \|a_3 - q_2^H a_3 q_2 - q_1^H a_3 q_1\|_2 & \\ 0 & 0 & 0 & \ddots \\ \vdots & \vdots & \vdots & \\ 0 & 0 & 0 & \end{bmatrix}$$

Como posto$(A) = n$, $\{a_1, a_2, \dots, a_n\}$ é linearmente independente, e portanto todos os q_i são todos diferentes de zero, e todas as normas que aparecem nos denominadores das expressões acima são não-nulas.

Com este procedimento, obtemos n vetores $\underset{\sim}{q}_i$ ortonormais. Para completar a decomposição, basta escolher outros $m-n$ vetores $\{\underset{\sim}{q}_{n+1}, \cdots, \underset{\sim}{q}_m\}$ que completem uma base ortonormal $\{\underset{\sim}{q}_1, \cdots, \underset{\sim}{q}_m\}$ do $\mathbb{C}^m$. Definindo então

$$Q = [\underset{\sim}{q}_1, \underset{\sim}{q}_2, \cdots, \underset{\sim}{q}_m], \text{ e } R = \begin{bmatrix} \|\underset{\sim}{a}_1\|_2 & \underset{\sim}{q}_1^H \underset{\sim}{a}_2 & \cdots & \underset{\sim}{q}_1^H \underset{\sim}{a}_n \\ 0 & \|\underset{\sim}{a}_2 - \underset{\sim}{q}_1^H \underset{\sim}{a}_2 \underset{\sim}{q}_1\|_2 & \cdots & \underset{\sim}{q}_2^H \underset{\sim}{a}_n \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \|\underset{\sim}{a}_n - \underset{\sim}{q}_1^H \underset{\sim}{a}_n \underset{\sim}{q}_1 - \cdots - \underset{\sim}{q}_{n-1}^H \underset{\sim}{a}_n \underset{\sim}{q}_{n-1}\|_2 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}.$$

Por construção, $A = QR$.

Caso posto$(A) < n$, haverá um passo k em que

$$\underset{\sim}{a}_k - (\underset{\sim}{q}_1^H \underset{\sim}{a}_k) \cdot \underset{\sim}{q}_1 - \cdots - (\underset{\sim}{q}_{k-1}^H \underset{\sim}{a}_k)\, \underset{\sim}{q}_{k-1} = \underset{\sim}{0}.$$ (ou seja, $\{\underset{\sim}{a}_1, \underset{\sim}{a}_2, \cdots, \underset{\sim}{a}_k\}$ é l.d.)

Nesta situação, podemos permutar a coluna k de A com uma outra coluna de forma a obter $\{\underset{\sim}{a}_1, \underset{\sim}{a}_2, \cdots, \underset{\sim}{a}_{k-1}, \underset{\sim}{a}_k^{(\text{novo})}\}$ linearmente independente. Isto sempre será possível se posto$(A) > k$. Procedendo desta forma, é possível decompor A num produto

$A = QRP$. ∎

Fatoração QR Usando Transformações de Householder

Como dissemos há pouco, o algoritmo de Gram-Schmidt não é numericamente estável. Uma opção melhor é usar as transformações de Householder que definimos há pouco, ou seja, usar matrizes unitárias da forma

$$P = I - \frac{2}{\underset{\sim}{v}^H \underset{\sim}{v}} \cdot \underset{\sim}{v}\,\underset{\sim}{v}^H .$$

Primeiramente, vamos mostrar como, dado um vetor $\underset{\sim}{x} \in \mathbb{C}^n$, escolher uma transformação P tal que

$$P\underset{\sim}{x} = e^{j\theta} \|\underset{\sim}{x}\|_2 \underset{\sim}{e}_1 , \qquad \text{onde } \underset{\sim}{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \text{ e } \theta \text{ é um}$$

ângulo entre 0 e 2π.

Note que, como P é unitária, $\|P\underset{\sim}{x}\|_2 = \|\underset{\sim}{x}\|_2$.

Se $P\underset{\sim}{x} = \left(I - \frac{2}{\underset{\sim}{v}^H\underset{\sim}{v}} \underset{\sim}{v}\,\underset{\sim}{v}^H\right)\underset{\sim}{x} = \underset{\sim}{x} - \frac{2\,\underset{\sim}{v}^H\underset{\sim}{x}}{\underset{\sim}{v}^H\underset{\sim}{v}} \cdot \underset{\sim}{v} = e^{j\theta}\|\underset{\sim}{x}\|_2 \cdot \underset{\sim}{e}_1$,

então $\frac{2\,\underset{\sim}{v}^H\underset{\sim}{x}}{\underset{\sim}{v}^H\underset{\sim}{v}} \cdot \underset{\sim}{v} = \underset{\sim}{x} - e^{j\theta}\|\underset{\sim}{x}\|_2 \cdot \underset{\sim}{e}_1$, ou seja,

$$\underset{\sim}{v} \in R\{\underset{\sim}{x}, \underset{\sim}{e}_1\}.$$

Vamos ver se podemos escolher $\underset{\sim}{v}$ da forma

$$\underset{\sim}{v} = \underset{\sim}{x} + \alpha\,\underset{\sim}{e}_1 , \qquad \alpha \in \mathbb{C}.$$

Teremos então

$$P\underset{\sim}{x} = \underset{\sim}{x} - 2\,\frac{(\underset{\sim}{x}^H\underset{\sim}{x} + \bar{\alpha}\,\underset{\sim}{e}_1^H\underset{\sim}{x})}{\underset{\sim}{x}^H\underset{\sim}{x} + \bar{\alpha}x_1 + |\alpha|^2 + \alpha\bar{x}_1} \cdot (\underset{\sim}{x} + \alpha\,\underset{\sim}{e}_1), \text{ onde } \underset{\sim}{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

$$= \left(1 - 2\,\frac{\|\underset{\sim}{x}\|_2^2 + \bar{\alpha}x_1}{\|\underset{\sim}{x}\|_2^2 + \bar{\alpha}x_1 + \alpha\bar{x}_1 + |\alpha|^2}\right) \cdot \underset{\sim}{x} - \frac{2\alpha\,\underset{\sim}{v}^H\underset{\sim}{x}}{\underset{\sim}{v}^H\underset{\sim}{v}} \cdot \underset{\sim}{e}_1 .$$

Para $P\underset{\sim}{x}$ ser igual a $e^{j\theta}\|\underset{\sim}{x}\|_2 \cdot \underset{\sim}{e}_1$, o coeficiente de $\underset{\sim}{x}$ na expressão acima deve ser nulo:

$$\|\underset{\sim}{x}\|_2^2 + \bar{\alpha}x_1 + \alpha\bar{x}_1 + |\alpha|^2 - 2\|\underset{\sim}{x}\|_2^2 - 2\bar{\alpha}x_1 = 0$$

$$\Rightarrow -\|\underset{\sim}{x}\|_2^2 + \alpha\bar{x}_1 - \bar{\alpha}x_1 + |\alpha|^2 = 0$$

Mas $\alpha\bar{x}_1 - \bar{\alpha}x_1 = 2j \cdot \mathrm{Im}\{\alpha\bar{x}_1\}$. Portanto, temos

$$|\alpha|^2 - \|\underset{\sim}{x}\|_2^2 + 2j\,\mathrm{Im}\{\alpha\bar{x}_1\} = 0$$

$$\Rightarrow \begin{cases} |\alpha|^2 = \|\underset{\sim}{x}\|_2^2 \Rightarrow |\alpha| = \|\underset{\sim}{x}\|_2 \\ \text{e } \mathrm{Im}\{\alpha\bar{x}_1\} = 0 \Rightarrow \text{Se } \alpha = r_\alpha e^{j\varphi_\alpha} \text{ e } x_1 = r_{x_1} \cdot e^{j\varphi_{x_1}}, \end{cases}$$

então $r_\alpha = \|\underset{\sim}{x}\|_2$ e $\varphi_\alpha = \varphi_{x_1}$ ou $\varphi_\alpha = \varphi_{x_1} + 180°$.

Portanto, $\underset{\sim}{v} = \underset{\sim}{x} \pm \|\underset{\sim}{x}\|_2 \cdot e^{j\varphi_{x_1}} \cdot \underset{\sim}{e}_1$, onde φ_{x_1} é tal que

$$x_1 = r_{x_1} \cdot e^{j\varphi_{x_1}}.$$

Com esta escolha de v, temos

$$P\underset{\sim}{x} = \mp \|\underset{\sim}{x}\|_2 \cdot e_1 \cdot e^{j\varphi_{x_1}}.$$ (note a inversão da ordem dos sinais).

Usando transformações como acima, podemos facilmente obter uma fatoração QR de uma matriz A qualquer. Por exemplo, seja

$$A = \begin{bmatrix} \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \end{bmatrix} \triangleq [\underset{\sim}{x}_1 \quad \underset{\sim}{x}_2 \quad \underset{\sim}{x}_3] \in \mathbb{R}^{4\times 3}$$

Defina P_1 tal que $P_1 \underset{\sim}{x}_1 = \begin{bmatrix} \|\underset{\sim}{x}_1\|_2 \\ 0 \\ 0 \\ 0 \end{bmatrix}$, e aplique P_1 a A:

$$P_1 A = \begin{bmatrix} \|\underset{\sim}{x}_1\|_2 & \times & \times \\ 0 & \times & \times \\ 0 & \times & \times \\ 0 & \times & \times \end{bmatrix} \triangleq \left[\begin{array}{c|cc} \|\underset{\sim}{x}_1\|_2 & \times & \times \\ \hline \underset{\sim}{0} & \underset{\sim}{x}_2^{(1)} & \underset{\sim}{x}_3^{(1)} \end{array}\right].$$

Defina agora P_2 tal que $P_2 \underset{\sim}{x}_2^{(1)} = \begin{bmatrix} \|\underset{\sim}{x}_2^{(1)}\|_2 \\ 0 \\ 0 \end{bmatrix}$, e

defina também $H_2 = \left[\begin{array}{c|c} 1 & \underset{\sim}{0}^T \\ \hline \underset{\sim}{0} & P_2 \end{array}\right]$. Então

$$H_2 P_1 A = \begin{bmatrix} \|\underset{\sim}{x}_1\|_2 & \times & \times \\ 0 & \|\underset{\sim}{x}_2^{(1)}\|_2 & \times \\ 0 & 0 & \times \\ 0 & 0 & \times \end{bmatrix} = \left[\begin{array}{cc|c} \|\underset{\sim}{x}_1\|_2 & \times & \times \\ 0 & \|\underset{\sim}{x}_2^{(1)}\|_2 & \times \\ \hline \underset{\sim}{0} & \underset{\sim}{0} & \underset{\sim}{x}_3^{(2)} \end{array}\right].$$

Finalmente, defina P_3 e H_3 tais que

$$P_3 \underset{\sim}{x}_3^{(2)} = \begin{bmatrix} \|\underset{\sim}{x}_3^{(2)}\|_2 \\ 0 \end{bmatrix}, \quad H_3 = \left[\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & \\ \hline \multicolumn{2}{c|}{0} & P_3 \end{array}\right].$$

Então, $$H_3 H_2 P_1 A = \begin{bmatrix} \|\underset{\sim}{x}_1\|_2 & \times & \times \\ 0 & \|\underset{\sim}{x}_2^{(1)}\|_2 & \times \\ 0 & 0 & \|\underset{\sim}{x}_3^{(2)}\|_2 \\ 0 & 0 & 0 \end{bmatrix} \triangleq R.$$

A nossa fatoração será então

$$A = \underbrace{P_1^H H_2^H H_3^H}_{\triangleq Q} . R .$$

Note que este algoritmo é geral. Se em algum passo tivermos $\underset{\sim}{x}_k^{(k-1)} = \underset{\sim}{0}$, basta tomar $H_k = I$ (ou permutar as colunas de A para poder prosseguir)

Neste caso, a matriz resultante já fica exatamente na forma desejada.

Estabilidade Numérica do Algoritmo QR para solução de Sistemas Lineares

Se o algoritmo acima, usando transformações de Householder, for usado para resolver o problema $A\underset{\sim}{x} = \underset{\sim}{b}$, pode-se demonstrar (ver Higham, pg. 369) que a solução calculada numericamente, $\underset{\sim}{y}$, satisfaz (se $A \in \mathbb{R}^{n\times n}$, $\underset{\sim}{b} \in \mathbb{R}^n$)

$$(A + \Delta A)\underset{\sim}{y} = \underset{\sim}{b} + \underset{\sim}{\Delta b},$$

onde $$|||\Delta A|||_F \leq n^2 \cdot \frac{c\,n.eps}{1 - c\,n.eps} \cdot |||A|||_F,$$

$$\|\underset{\sim}{\Delta b}\|_2 \leq n^2 \cdot \frac{c\,n\,eps}{1 - c\,n\,eps} \cdot \|b\|_2 ,$$

e c é uma constante pequena.

Portanto, se $n^3 eps \ll 1$, podemos concluir que o algoritmo acima para o cálculo da solução de um sistema linear é numericamente estável.

Cuidados com a escolha do vetor de Householder (Golub, pg. 210)

O algoritmo acima é estável, mas devem ser tomados alguns cuidados na implementação. Por exemplo, para vetores reais, é possível escolher

$\underset{\sim}{v} = \underset{\sim}{x} - \|\underset{\sim}{x}\|_2 \underset{\sim}{e}_1$ para termos $\left(I - \frac{2}{\underset{\sim}{v}^T\underset{\sim}{v}} \underset{\sim}{v}\underset{\sim}{v}^T\right)\underset{\sim}{x} = \|\underset{\sim}{x}\|_2 \cdot \underset{\sim}{e}_1$.

No entanto, se o vetor $\underset{\sim}{x}$ for quase paralelo a $\underset{\sim}{e}_1$, teremos

$$\underset{\sim}{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad \text{com} \quad |x_1| \simeq \|\underset{\sim}{x}\|_2 .$$

Se $x_1 > 0$, então poderá haver um cancelamento grave no cálculo de $x_1 - \|\underset{\sim}{x}\|_2 = v_1$ $\left(\underset{\sim}{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}\right)$. Uma fórmula alternativa de cálculo é fazer

$$v_1 = x_1 - \|\underset{\sim}{x}\|_2 = \frac{x_1^2 - \|\underset{\sim}{x}\|_2^2}{x_1 + \|\underset{\sim}{x}\|_2} = -\frac{(x_2^2 + \cdots + x_n^2)}{x_1 + \|\underset{\sim}{x}\|_2} .$$

$$\times \frac{x_1 + \|\underset{\sim}{x}\|_2}{x_1 + \|\underset{\sim}{x}\|_2}$$

Rotações de Givens (Golub, pp. 215–221)

As rotações de Givens permitem zerar um elemento qualquer de um vetor, e podem ser usadas também para resolver problemas de mínimos quadrados. Rotações de Givens podem ser mais úteis que transformações de Householder em problemas estruturados (onde se sabe de antemão que alguns elementos da matriz inicial são zero) e em computadores paralelos.

Exemplo: Vamos achar uma rotação de Givens que transforme o vetor

$$\underset{\sim}{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_i \\ \vdots \\ x_k \\ \vdots \\ x_n \end{bmatrix} \quad \text{em} \quad G(i,k,\theta)\cdot \underset{\sim}{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \pm(x_i^2 + x_k^2)^{1/2} \\ \vdots \\ 0 \\ \vdots \\ x_n \end{bmatrix} \begin{matrix} \\ \\ \\ \leftarrow \text{posição } i \\ \\ \leftarrow \text{posição } k \\ \\ \\ \end{matrix}$$

Note que

$$\begin{bmatrix} c & s \\ -s & c \end{bmatrix} \begin{bmatrix} x_i \\ x_k \end{bmatrix} = \begin{bmatrix} c x_i + s x_k \\ -s x_i + c x_k \end{bmatrix},$$

onde vamos escolher c e s tais que $c^2 + s^2 = 1$. Então

$$-s x_i + c x_k = 0 \quad \text{se escolhermos} \quad c = \frac{x_i}{\sqrt{x_i^2 + x_k^2}}, \quad s = \frac{+x_k}{\sqrt{x_i^2 + x_k^2}}.$$

Definindo

$$G(i,k,\theta) = \begin{bmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots & & \vdots \\ 0 & \cdots & c & \cdots & s & \cdots & 0 \\ \vdots & & \vdots & \ddots & \vdots & & \vdots \\ 0 & \cdots & -s & \cdots & c & \cdots & 0 \\ \vdots & & \vdots & & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{bmatrix},$$

obtemos o resultado desejado.

Novamente, é importante calcular c e s de forma a não se introduzir erros numéricos desnecessários. Ver Golub, pg. 216.

As propriedades numéricas da fatoração QR usando rotações de Givens são semelhantes às propriedades da fatoração com transformações de Householder.

$\mathcal{R}(A)$ via fatoração QR

Vimos que sempre é possível fatorar $A \in \mathbb{C}^{m \times n}$ como

$$A = Q \begin{bmatrix} R_{11} & R_{12} \\ 0 & 0 \end{bmatrix} P,$$

onde $Q \in \mathbb{C}^{m \times m}$ é unitária, $P \in \mathbb{C}^{n \times n}$ é uma matriz de permutação, e $R_{11} \in \mathbb{C}^{r \times r}$ é triangular superior e invertível.

Vamos dividir as colunas de Q como abaixo:

$$Q = \left[\underset{\sim}{q}_1 \; \underset{\sim}{q}_2 \cdots \underset{\sim}{q}_r \,\vdots\, \underset{\sim}{q}_{r+1} \cdots \underset{\sim}{q}_m \right] = \left[Q_1 \,\vdots\, Q_2 \right]$$

$$Q_1 \in \mathbb{C}^{m \times r}$$
$$Q_2 \in \mathbb{C}^{m \times (m-r)}$$

As colunas de Q_1 e Q_2 são ortonormais, e $Q_1^H Q_2 = 0$.

Vamos achar $\mathcal{R}(A)$. Seja $\underset{\sim}{x} \in \mathbb{C}^n$ um vetor qualquer. Então,

$$\mathcal{R}(A) = \{ \underset{\sim}{y} \in \mathbb{C}^m : \underset{\sim}{y} = A\underset{\sim}{x} \text{ para algum } \underset{\sim}{x} \in \mathbb{C}^n \}.$$

Mas

$$A\underset{\sim}{x} = Q \begin{bmatrix} R_{11} & R_{12} \\ 0 & 0 \end{bmatrix} P \underset{\sim}{x} = \left[Q_1 \,\vdots\, Q_2 \right] \begin{bmatrix} R_{11} & R_{12} \\ 0 & 0 \end{bmatrix} P \underset{\sim}{x} =$$

$$= \left[Q_1 R_{11} \;\; Q_1 R_{12} \right] \cdot P \underset{\sim}{x} = Q_1 \left[R_{11} \;\; R_{12} \right] P \underset{\sim}{x}.$$

Como posto$(R_{11}) = r$, posto$([R_{11} \;\; R_{12}]) = r$, e portanto

$$\mathcal{R}(A) = \mathcal{R}(Q_1) = \text{Span}\{ \underset{\sim}{q}_1, \underset{\sim}{q}_2, \ldots, \underset{\sim}{q}_r \}.$$

Ou seja, as r primeiras colunas de Q formam uma base ortonormal para $\mathcal{R}(A)$.

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

Exemplos de Aplicações

1) Matrizes unitárias.

Matrizes unitárias são bastante comuns em Engenharia Elétrica. O exemplo mais comum é a transformada discreta de Fourier: Seja $\{s(0), \dots, s(N-1)\}$ uma seqüência (um sinal). Sua TDF é

$$S(k) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} s(n) \cdot e^{-j 2\pi \frac{nk}{N}}.$$

A seqüência original pode ser recuperada a partir da seqüência transformada $\{S(0), S(1), \dots, S(N-1)\}$ pela relação

$$s(n) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} S(k) \cdot e^{+j 2\pi \cdot \frac{nk}{N}}.$$

Definindo os vetores $\underline{s} = \begin{bmatrix} s(0) \\ s(1) \\ \vdots \\ s(N-1) \end{bmatrix}$ e $\underline{S} = \begin{bmatrix} S(0) \\ S(1) \\ \vdots \\ S(N-1) \end{bmatrix}$,

podemos escrever

$$\underline{S} = \frac{1}{\sqrt{N}} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & e^{-j\frac{2\pi}{N}} & \cdots & e^{-j 2\pi \frac{(N-1)}{N}} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & e^{-j 2\pi \frac{(N-1)}{N}} & \cdots & e^{-j 2\pi \frac{(N-1)^2}{N}} \end{bmatrix} \cdot \underline{s} \triangleq F \cdot \underline{s}$$

e

$$\underline{s} = \frac{1}{\sqrt{N}} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & e^{j\frac{2\pi}{N}} & \cdots & e^{j 2\pi \frac{(N-1)}{N}} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & e^{+j 2\pi \frac{(N-1)}{N}} & \cdots & e^{j 2\pi \frac{(N-1)^2}{N}} \end{bmatrix} \cdot \underline{S} = F^H \cdot \underline{S}.$$

Note que se for escolhida outra constante para definir F (em vez de $1/\sqrt{N}$), a matriz não será mais unitária, mas ainda terá número condicional igual a 1.

Como as relações acima valem para qualquer seqüência $\{s(0), \dots, s(N-1)\}$, segue que

$$F^H \cdot F = I \Rightarrow F \text{ é unitária.}$$

Outras transformações que envolvem matrizes unitárias são a DCT (discrete cosine transform), DST (discrete sine transform), KLT (Karhunen-Loève transform).

2) Teorema da Diagonalização Unitária

Seja $\{\underset{\sim}{x}_n\}_{n=0}^{\infty}$ um processo estocástico estacionário no sentido amplo (WSS, ou seja: a média $E(\underset{\sim}{x}_n)$ é constante, a autocorrelação $E\,\underset{\sim}{x}_n \underset{\sim}{x}_{n+t}^H$ depende apenas da diferença t).

Assuma que $E\,\underset{\sim}{x}_n = 0$ e $E\,\underset{\sim}{x}_n \underset{\sim}{x}_n^H = R$, $\underset{\sim}{x}_n = \begin{bmatrix} \alpha_n(0) \\ \vdots \\ \alpha_n(N-1) \end{bmatrix}$.

A transformada de Karhunen-Loève é uma transformação T tal que

$$\underset{\sim}{v}_n = T\,\underset{\sim}{x}_n, \qquad E\,\underset{\sim}{v}_n \underset{\sim}{v}_n^H = \mathrm{diag}(\lambda_i),\ 0 \le i \le N-1.$$

Ou seja, a KLT corresponde a uma mudança de base na qual os elementos do vetor transformado $\underset{\sim}{v}_n$ <u>não são correlacionados</u>. A existência da KLT segue do Teorema da Diagonalização Unitária, já que

$$R^H = (E\,\underset{\sim}{x}_n \underset{\sim}{x}_n^H)^H = E(\underset{\sim}{x}_n \underset{\sim}{x}_n^H)^H = R \rightarrow R \text{ é hermitiana.}$$

Veremos mais sobre esta transformada após falarmos sobre decomposição em valores singulares (SVD).

3) Mínimos quadrados.

Considere o modelo estatístico linear

$$A\underset{\sim}{x} = \underset{\sim}{b} + \varepsilon,$$

onde $A \in \mathbb{C}^{m \times n}$ e $\underset{\sim}{b} \in \mathbb{C}^m$ são conhecidos, e ε

é um vetor aleatório de incertezas com média zero e variância $E(\varepsilon\varepsilon^H) = \sigma^2 . I$, e $\underset{\sim}{x}$ é um vetor de parâmetros desconhecido. Então pode-se demonstrar que o melhor estimador não viezado de $\underset{\sim}{x}$ é a solução de $\min_{\hat{\underset{\sim}{x}}} \|A\hat{\underset{\sim}{x}} - \underset{\sim}{b}\|_2$.

(não viezado ⇒ média do estimador é igual a $\underset{\sim}{x}$).

A resposta à nossa pergunta anterior é o seguinte teorema.

Teorema 1: Se $A \in \mathbb{C}^{n\times n}$ tem autovalores $\lambda_1, \lambda_2, \ldots, \lambda_n$, as seguintes afirmações são equivalentes:

(a) A é normal;

(b) A é unitariamente diagonalizável;

(c) $\sum_{i=1}^{n}\sum_{j=1}^{n} |a_{ij}|^2 = \sum_{i=1}^{n} |\lambda_i|^2$;

(d) Há um conjunto ortonormal de n autovetores de A.

Prova: Seja T uma matriz triangular superior tal que $UTU^H = A$ com U unitária (T e U existem pelo teorema de Schur). Como A é normal, temos

$$TT^H = (U^HAU)(U^HA^HU) = U^HAA^HU = U^HA^HAU =$$
$$= (U^HA^HU)(U^HAU) = T^HT$$

$\Rightarrow$ T é normal se A for normal.

Como T é triangular superior, temos

$$T^HT = \begin{bmatrix} \bar{t}_{11} & 0 & \cdots & 0 \\ \bar{t}_{12} & \bar{t}_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \bar{t}_{1n} & \bar{t}_{2n} & \cdots & \bar{t}_{nn} \end{bmatrix} \begin{bmatrix} t_{11} & t_{12} & \cdots & t_{1n} \\ 0 & t_{22} & \cdots & t_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & t_{nn} \end{bmatrix} =$$

$$= \begin{bmatrix} \bar{t}_{11}t_{11} & \cdots & \\ \vdots & \ddots & \\ & & \end{bmatrix}$$

Mas $T^HT = TT^H = \begin{bmatrix} \bar{t}_{11}t_{11} + |t_{12}|^2 + \cdots + |t_{1n}|^2 & \cdots \\ \vdots & \ddots \\ & \end{bmatrix}$.

Como a igualdade acima implica

$$|t_{11}|^2 = |t_{11}|^2 + |t_{12}|^2 + \cdots + |t_{1n}|^2,$$

concluímos que $t_{12} = t_{13} = \cdots = t_{1n} = 0$.

Analogamente, o elemento $(2,2)$ de $T^H T$ é

$\cancel{|t_{12}|^2} + |t_{22}|^2$, e o elemento $(2,2)$ de TT^H é

$=0$

$|t_{22}|^2 + |t_{23}|^2 + \dots + |t_{2n}|^2$.

Novamente, concluímos que $t_{23} = t_{24} = \dots = t_{2n} = 0$.

Continuando desta forma, mostramos que todos os elementos t_{ij} de T com $j > i$ são nulos, ou seja, T é

<u>diagonal</u>.

Com isto, provamos que (a) implica (b). Vamos continuar, provando que (b) implica (c). De fato,

$$\sum_{i,j=1}^{n} |a_{ij}|^2 = \mathrm{Tr}\,(A^H A) = \mathrm{Tr}\,(U^H T^H \underbrace{U U^H}_{=I} T U) = \mathrm{Tr}\,(T^H T \underbrace{U U^H}_{=I}) =$$

$$= \mathrm{Tr}\,(T^H T) = \sum_{i=1}^{n} |\lambda_i|^2.$$

Agora vamos provar que (c) implica (b):

$$\sum_{i=1}^{n}\sum_{j=1}^{n} |a_{ij}|^2 = \underbrace{|||A|||_F^2 = |||UAU^H|||_F^2}_{\text{propriedade de } |||\cdot|||_F} = |||T|||_F^2 \overset{\text{hipótese}}{=} \sum_{i=1}^{n} |\lambda_i|^2 \Rightarrow T \text{ é diagonal}$$

U é unitária

T é triangular superior (Teorema de Schur) $\rightarrow T = \begin{bmatrix} \lambda_1 & x & \cdots & x \\ 0 & \lambda_2 & \cdots & x \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$

A equivalência de (b) e (d) fica como exercício.

Também fica como exercício provar que (b) implica (a). ∎

Matrizes Normais

Vimos até agora que matrizes unitárias têm várias propriedades interessantes: seu número condicional é baixo, sua inversa é fácil de calcular, e a transformação de um vetor por uma matriz unitária não altera o comprimento euclideano ($\|\cdot\|_2$) do vetor. De modo geral, tanto em teoria quanto em algoritmos práticos, é interessante usar matrizes unitárias sempre que tiver que ser formado um produto de matrizes.

Vimos também que é possível achar uma matriz triangular superior similar (via similaridade unitária) a qualquer matriz $A \in \mathbb{C}^{n\times n}$. A próxima pergunta seria, que classe de matrizes pode ser <u>diagonalizada</u> a partir de transformações unitárias? A resposta a esta pergunta involve matrizes normais:

<u>Definição 1</u>: A Matriz $A \in \mathbb{C}^{n \times n}$ é dita <u>normal</u> se

$$A^H A = A A^H.$$

<u>Exemplos</u>:

1) $U^H U = I = U U^H$ para toda matriz unitária. Portanto, matrizes unitárias são normais.
2) Se $A^H = A$, A é dita <u>hermitiana</u> (<u>simétrica</u> se $A \in \mathbb{R}^{n\times n}$). Portanto, $A^H A = A^2 = AA^H \Rightarrow$ matrizes hermitianas são normais.
3) Se $A^H = -A$, A é dita <u>skew-hermitian</u>. Neste caso, $A^H A = -A^2 = AA^H$; portanto estas matrizes também são normais.
4) Existem matrizes normais que não caem em nenhum dos casos acima, como $A = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$ por exemplo.

<u>Definição 2</u>: Uma matriz $A \in \mathbb{C}^{n\times n}$ é unitariamente diagonalizável se existir $U \in \mathbb{C}^{n\times n}$ unitária tal que

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

$R(A)$ via fatoração QR

Vimos que sempre é possível fatorar $A \in \mathbb{C}^{m\times n}$ como

$$A = Q\begin{bmatrix} R_{11} & R_{12} \\ 0 & 0 \end{bmatrix} P,$$

onde $Q \in \mathbb{C}^{m\times m}$ é unitária, $P \in \mathbb{C}^{n\times n}$ é uma matriz de permutação, e $R_{11} \in \mathbb{C}^{r\times r}$ é triangular superior e invertível.

Vamos dividir as colunas de Q como abaixo:

$$Q = [\,\underset{\sim}{q}_1\ \underset{\sim}{q}_2 \cdots \underset{\sim}{q}_r \mid \underset{\sim}{q}_{r+1} \cdots \underset{\sim}{q}_m\,] = [\,Q_1 \mid Q_2\,]$$

$$Q_1 \in \mathbb{C}^{m\times r}$$
$$Q_2 \in \mathbb{C}^{m\times(m-r)}$$

As colunas de Q_1 e Q_2 são ortonormais, e $Q_1^H Q_2 = 0$.

Vamos achar $R(A)$. Seja $\underset{\sim}{x} \in \mathbb{C}^n$ um vetor qualquer. Então,

$$R(A) = \{\underset{\sim}{y} \in \mathbb{C}^m : \underset{\sim}{y} = A\underset{\sim}{x} \text{ para algum } \underset{\sim}{x} \in \mathbb{C}^n\}.$$

Mas $A\underset{\sim}{x} = Q\begin{bmatrix} R_{11} & R_{12} \\ 0 & 0 \end{bmatrix} P\underset{\sim}{x} = [\,Q_1 \mid Q_2\,]\begin{bmatrix} R_{11} & R_{12} \\ 0 & 0 \end{bmatrix} P\underset{\sim}{x} =$

$$= [\,Q_1 R_{11} \quad Q_1 R_{12}\,] \cdot P\underset{\sim}{x} = Q_1 [\,R_{11} \quad R_{12}\,] P\underset{\sim}{x}.$$

Como posto$(R_{11}) = r$, posto$([R_{11}\ R_{12}]) = r$, e portanto

$$R(A) = R(Q_1) = \text{Span}\{\underset{\sim}{q}_1, \underset{\sim}{q}_2, \dots, \underset{\sim}{q}_r\}.$$

Ou seja, as r primeiras colunas de Q formam uma base ortonormal para $R(A)$.