7. DECOMPOSIÇÃO EM VALORES SINGULARES (SVD)

A decomposição em valores singulares é um método muito útil para a análise de sistemas multivariáveis. Em termos da operação de um processo o método SVD facilita a sua avaliação em função da:

- possibilidade de se conseguir os objetivos operacionais desejados;
- determinação dos locais mais apropriados para aumentar a sensibilidade das medidas;
- determinação das variáveis que mais vão influenciar na operação; etc.

O método utiliza a decomposição em valores singulares da matriz de ganhos estáticos do processo:

$$\mathbf{K} = \mathbf{U} \cdot \mathbf{\Sigma} \cdot \mathbf{V}^T$$

Os valores singulares de uma matriz **K** são dados pela seguinte relação:

$$\sigma = +\sqrt{\lambda\left(\mathbf{K}^T\mathbf{K}\right)} \qquad (\text{ou} \qquad \sigma = +\sqrt{\lambda\left(\mathbf{K} \cdot \mathbf{K}^T\right)})$$

onde $\lambda(\mathbf{A})$ representa o valor característico de **A**.

A SVD é usada em álgebra linear para minimizar erros computacionais em operações com matrizes de grande porte.

O interesse da SVD para a área de sistemas multivariáveis é fácil de se justificar. Qualquer estratégia de operação busca, basicamente, determinar a inversa da matriz de funções de transferência:

$$\mathbf{y}(s) = \mathbf{G}(s) \cdot \mathbf{m}(s)$$

Exigindo

$$\mathbf{y}(s) = \mathbf{y}^d(s)$$

resulta

$$\mathbf{m}(s) = \mathbf{G}^{-1}\mathbf{y}^d(s)$$

No estado estacionário

$$\mathbf{m} = \mathbf{K}^{-1}\mathbf{y}^d$$

Os valores singulares medem a inversibilidade da matriz; o menor valor singular é uma medida da distância que a matriz está de se transformar em singular; e a razão entre o maior e o menor valor singular mede a dificuldade de invertê-la.

Os componentes da SVD da matriz **K**, de dimensão n×m, são:

U: matriz **unitária** $\left(\mathbf{U}^{-1} = \mathbf{U}^{T}\right)$ ou **ortonormal**, n×n; suas colunas são os vetores singulares à esquerda da matriz **K**; são os vetores característicos da matriz $\mathbf{K} \cdot \mathbf{K}^{T}$.

V: matriz **unitária** $\left(\mathbf{V}^{-1} = \mathbf{V}^{T}\right)$, m×m; suas colunas são os vetores singulares à direita da matriz **K**; são os vetores característicos da matriz $\mathbf{K}^{T}\mathbf{K}$.

Σ: matriz n×m com a seguinte estrutura:

$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}^{*} & \mathbf{0} \end{bmatrix}$ para n < m

$\mathbf{\Sigma} = \mathbf{\Sigma}^{*}$ para n = m

$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}^{*} \\ \mathbf{0} \end{bmatrix}$ para n > m

onde $\mathbf{\Sigma}^{*} = diag(\sigma_i)$

com $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r \geq 0$ e $\mathrm{r} = \min(\mathrm{n}, \mathrm{m})$.

Define-se o **número de condicionamento** como:

$$CN = \gamma = \frac{\sigma_1}{\sigma_r} = \frac{\overline{\sigma}(\mathbf{K})}{\underline{\sigma}(\mathbf{K})}$$

que mede o "condicionamento" da matriz. Quanto maior esse número pior será o condicionamento da matriz em questão, em termos computacionais. Para a operação de um processo associado a essa matriz, o CN mede a dificuldade para se conseguir boa operabilidade do sistema multivariável.

É possível visualizar o significado dos valores singulares da matriz de ganhos estáticos no caso de um sistema 2×2.

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{bmatrix} \cdot \begin{bmatrix} m_1 \\ m_2 \end{bmatrix} = \begin{bmatrix} K_{11} \cdot m_1 + K_{12} \cdot m_2 \\ K_{21} \cdot m_1 + K_{22} \cdot m_2 \end{bmatrix} = \begin{bmatrix} u_1^1 & u_2^1 \\ u_1^2 & u_2^2 \end{bmatrix} \cdot \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} \cdot \begin{bmatrix} v_1^1 & v_1^2 \\ v_2^1 & v_2^2 \end{bmatrix} \cdot \begin{bmatrix} m_1 \\ m_2 \end{bmatrix}$$

Podemos expressar o vínculo estático da seguinte forma:

$$y_1 = \left(u_1^1 \cdot \sigma_1 \cdot v_1^1 + u_2^1 \cdot \sigma_2 \cdot v_2^1\right) \cdot m_1 + \left(u_1^1 \cdot \sigma_1 \cdot v_1^2 + u_2^1 \cdot \sigma_2 \cdot v_2^2\right) \cdot m_2$$

$$y_2 = \left(u_1^2 \cdot \sigma_1 \cdot v_1^1 + u_2^2 \cdot \sigma_2 \cdot v_2^1\right) \cdot m_1 + \left(u_1^2 \cdot \sigma_1 \cdot v_1^2 + u_2^2 \cdot \sigma_2 \cdot v_2^2\right) \cdot m_2$$

Observa-se que $K_{ij,}$ o ganho estático observado, é uma combinação dos valores singulares σ_1 e σ_2, determinada pelos elementos de **U** e **V**:

$$K_{ij} = \sum_{k=1}^{r} u_k^i \cdot v_k^j \cdot \sigma_k$$

Uma outra forma de se representar o vínculo estático é:

$$y_1 = u_1^1 \cdot \sigma_1 \cdot \left(v_1^1 \cdot m_1 + v_1^2 \cdot m_2\right) + u_2^1 \cdot \sigma_2 \cdot \left(v_2^1 \cdot m_1 + v_2^2 \cdot m_2\right)$$

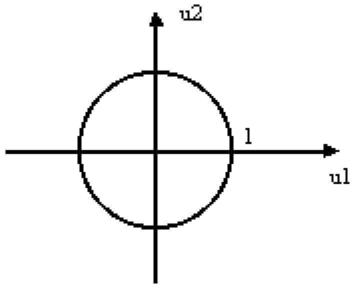
$$y_2 = u_1^2 \cdot \sigma_1 \cdot \left(v_1^1 \cdot m_1 + v_1^2 \cdot m_2\right) + u_2^2 \cdot \sigma_2 \cdot \left(v_2^1 \cdot m_1 + v_2^2 \cdot m_2\right)$$

Neste caso observa-se que a saída "y_i", no estado estacionário, está formada pela soma de termos contendo um valor singular "σ_k". As colunas de **U** indicam como o valor singular se distribui nas diferentes saídas. As colunas de **V** indicam o efeito das entradas sobre cada valor singular. De um modo geral:

$$y_i = \sum_{k=1}^{r} u_k^i \cdot \sigma_k \cdot \left(v_k\right)^T \cdot m$$

É importante ter presente que os valores singulares são ordenados de maior a menor.

Sejam as variáveis de entrada normalizadas tal que $m_1{}^2 + m_2{}^2 = 1$



Para

1) $\mathbf{m} = \begin{bmatrix} m_1 \\ m_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

$$y_1 = u_1^1 \cdot \sigma_1 \cdot v_1^1 + u_2^1 \cdot \sigma_2 \cdot v_2^1$$

$$y_2 = u_1^2 \cdot \sigma_1 \cdot v_1^1 + u_2^2 \cdot \sigma_2 \cdot v_2^1$$

2) $\mathbf{m} = \mathbf{v}_1 = \begin{bmatrix} v_1^1 \\ v_1^2 \end{bmatrix}$

$$y_1 = u_1^1 \cdot \sigma_1$$

$$\Rightarrow \mathbf{y} = \sigma_1 \cdot \mathbf{u}_1$$

$$y_2 = u_1^2 \cdot \sigma_1$$

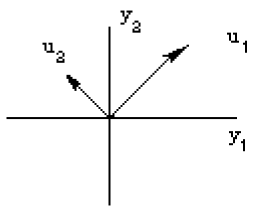
3) $\mathbf{m} = \mathbf{v}_2 = \begin{bmatrix} v_2^1 \\ v_2^2 \end{bmatrix}$

$$y_1 = u_2^1 \cdot \sigma_2$$

$$\Rightarrow \mathbf{y} = \sigma_2 \cdot \mathbf{u}_2$$

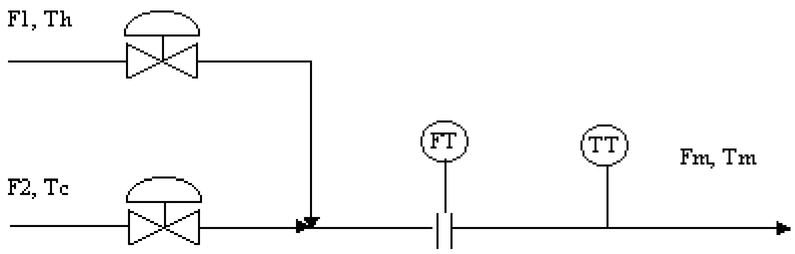
$$y_2 = u_2^2 \cdot \sigma_2$$

Os resultados de (2) e (3) são conseqüência da ortonormalidade dos vetores de **U** e **V**.



A entrada com a direção $\mathbf{v}_1$ sai na direção de $\mathbf{u}_1$ multiplicada por σ_1.

Exemplo



Os dados deste sistema no estado estacionário são:

F1 = 10 gpm

F2 = 20 gpm

Th = 100 ^{o}F

Tc = 65 ^{o}F

Para calcular a matriz de ganhos estáticos

$$\mathbf{K} = \begin{bmatrix} \dfrac{\partial Tm}{\partial F1} & \dfrac{\partial Tm}{\partial F2} \\ \dfrac{\partial Fm}{\partial F1} & \dfrac{\partial Fm}{\partial F2} \end{bmatrix}$$

são estabelecidos os balanços globais de massa e energia no estado estacionário.

Fm = F1 + F2

F1 Th + F2 Tc = Fm Tm

Então

$$\frac{\partial \mathrm{Tm}}{\partial \mathrm{F1}} = \frac{\partial}{\partial \mathrm{F1}}\left[\frac{\mathrm{F1}\cdot\mathrm{Th}+\mathrm{F2}\cdot\mathrm{Tc}}{\mathrm{F1}+\mathrm{F2}}\right] = \frac{\mathrm{Th}\cdot(\mathrm{F1}+\mathrm{F2})-\mathrm{F1}\cdot\mathrm{Th}-\mathrm{F2}\cdot\mathrm{Tc}}{(\mathrm{F1}+\mathrm{F2})^2}$$

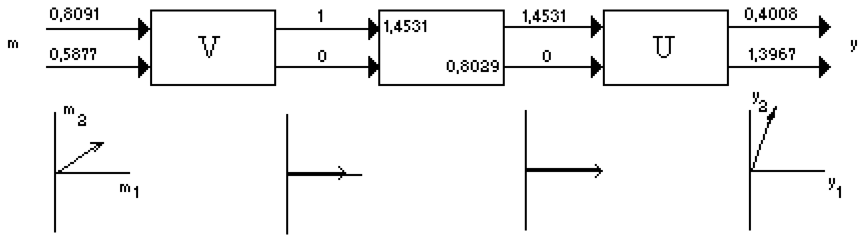
$$= \frac{\mathrm{F2}(\mathrm{Th}-\mathrm{Tc})}{(\mathrm{F1}+\mathrm{F2})^2} = \frac{20(100-65)}{(10+20)^2} = 0{,}7778$$

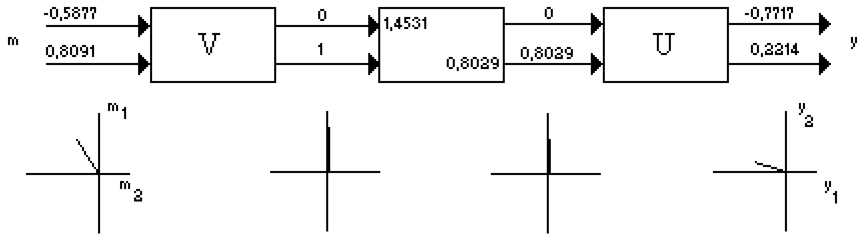
Seguindo o mesmo procedimento chega-se a

$$\mathbf{K} = \begin{bmatrix} 0,7778 & -0,3889 \\ 1 & 1 \end{bmatrix} = \mathbf{U} \cdot \mathbf{\Sigma} \cdot \mathbf{V}^T =$$

$$= \begin{bmatrix} 0,2758 & -0,9612 \\ 0,9612 & 0,2758 \end{bmatrix} \cdot \begin{bmatrix} 1,4531 & 0 \\ 0 & 0,8029 \end{bmatrix} \cdot \begin{bmatrix} 0,8091 & 0,5877 \\ -0,5877 & 0,8091 \end{bmatrix}$$

CN=1,7





Interpretação física dos elementos na decomposição em valores singulares

K: é matriz de ganhos estáticos do processo

$$K_{ij} = \left. \frac{\partial y_i}{\partial m_j} \right|_{m_k}$$

Os elementos desta matriz refletem a sensibilidade da variável de saída a uma variação na variável de entrada. Os ganhos estáticos dependem das unidades utilizadas.

U: é uma matriz cujas colunas, $\mathbf{u}^i$, são os vetores singulares à esquerda da matriz **K**.

$\mathbf{u}_1$: indica a direção em que as variáveis de saída mudam mais, frente a variações nas variáveis de entrada.

$\mathbf{u}_2, \mathbf{u}_3, \cdots, \mathbf{u}_n$: indicam, nessa ordem, direções cada vez menos sensíveis.

V: é uma matriz cujas colunas, $\mathbf{v}_i$, são os vetores singulares à direita da matriz **K**.

$\mathbf{v}_1$: indica a combinação de variáveis de entrada cuja variação afeta as saídas de forma mais significativa.

Σ: é uma matriz diagonal cujos elementos, os valores singulares, fornecem os ganhos estáticos do sistema "desacoplado" (em termos de variáveis estruturadas).

$$\begin{aligned}
&\mathbf{y} = \mathbf{K} \cdot \mathbf{u} \\
&\mathbf{y} = \mathbf{U} \cdot \boldsymbol{\Sigma} \cdot \mathbf{V}^{\mathbf{T}} \cdot \mathbf{u} \\
&\mathbf{U}^{\mathbf{T}} \cdot \mathbf{y} = \mathbf{U}^{\mathbf{T}} \cdot \mathbf{U} \cdot \boldsymbol{\Sigma} \cdot \mathbf{V}^{\mathbf{T}} \cdot \mathbf{u} = \boldsymbol{\Sigma} \cdot \mathbf{V}^{\mathbf{T}} \cdot \mathbf{u} \\
&\mathbf{y}' = \boldsymbol{\Sigma} \cdot \mathbf{u}'
\end{aligned}$$

$CN = \gamma = \dfrac{\sigma_1}{\sigma_r}$: é o número de condicionamento; mede o grau de dificuldade para se operar o sistema multivariável; CN grande indica que há direções de variação muito altas e muito baixas; dependendo da qualidade da instrumentação utilizada deverá se abrir mão de certos objetivos; há menos graus de liberdade para resolver problemas operacionais; não podem ser satisfeitos simultaneamente todos os objetivos almejados.

Exemplo

Se no exemplo anterior forem mudadas as condições de estado estacionário:

F1 = 100 gpm

F2 = 150 gpm

Th = 100 ^{o}F

Tc = 65 ^{o}F

resulta

$$\mathbf{K} = \begin{bmatrix} 0{,}084 & -0{,}056 \\ 1{,}000 & 1{,}000 \end{bmatrix}$$

Os valores singulares são σ_1 = 1,414 e σ_2 = 0,0990, gerando um número de condicionamento:

CN = 14,28

O sistema responde 14 vezes mais na direção mais forte. Apesar de haver dois graus de liberdade, um deles é pouco útil.

Apêndice (Notas do Nivelamento)

Transformação linear: Sejam U e V espaços vetoriais sobre $\Re$. Uma aplicação F: $U \to V$ é chamada transformação linear de U em V se, e somente se,

(a) $F(u_1 + u_2) = F(u_1) + F(u_2)$, $\forall\ u_1, u_2 \in U$, e

(b) $F(\alpha\, u) = \alpha\, F(u)$, $\forall \alpha \in \Re$ e $\forall\ u \in U$.

No caso em que $U = V$, uma transformação linear F: $U \to U$ é chamada também de **operador linear**.

<u>Exemplo</u>: Para verificar se a aplicação F: $\Re^3 \to \Re^2$ definida por $F(x, y, z) = (x, 2x - z)$, $\forall\ (x, y, z) \in \Re^3$, é uma transformação linear, sejam $\mathbf{u}_1 = (x_1, y_1, z_1)$ e $\mathbf{u}_2 = (x_2, y_2, z_2)$ em $\Re^3$ e:

(a) $F(\mathbf{u}_1 + \mathbf{u}_2) = F(x_1 + x_2, y_1 + y_2, z_1 + z_2) = (x_1 + x_2, 2x_1 + 2\,x_2 - z_1 - z_2) = (x_1, 2x_1 - z_1) + (x_2, 2x_2 - z_2) = F(\mathbf{u}_1) + F\,(\mathbf{u}_2)$

(b) $F(\alpha\, \mathbf{u}_1) = F(\alpha\, x_1, \alpha\, y_1, \alpha\, z_1) = (\alpha\, x_1, 2\alpha\, x_1 - \alpha\, z_1) = \alpha\,(x_1, 2x_1 - z_1) = \alpha\, F(\mathbf{u}_1)$, $\forall \alpha \in \Re$.

logo, a aplicação é uma transformação linear.

Verifica-se prontamente que a multiplicação de uma matriz $\mathbf{A} \in \Re^{m \times n}$ por um vetor $\mathbf{x} \in \Re^n$ é uma transformação linear, pois $\mathbf{A} \cdot (\mathbf{x}_1 + \mathbf{x}_2) = \mathbf{A} \cdot \mathbf{x}_1 + \mathbf{A} \cdot \mathbf{x}_2$ e $\mathbf{A} \cdot \alpha\, \mathbf{x} = \alpha\, \mathbf{A} \cdot \mathbf{x}$.

Núcleo ou espaço nulo: Sejam U e V espaços vetoriais sobre $\Re$ e F: $U \to V$ uma transformação linear. Indica-se por Ker(F) ou null(F) e denomina-se núcleo ou espaço nulo de F o seguinte subconjunto de U:

$$\text{Ker}(F) = \text{null}(F) = \{\text{u} \in U \mid F(u) = 0\} \subseteq U$$

<u>Exemplo</u>: Seja F: $\Re^2 \to \Re^3$ a transformação linear dada por $F(x, y) = (0, x + y, 0)$, $\forall\ (x, y) \in \Re^2$, o núcleo de F é obtido fazendo $F(x, y) = (0, 0, 0) = (0, x + y, 0)$. Logo, $x = -y$ e $\text{Ker}(F) = \{(x, -x) \mid x \in \Re\}$, como ilustra a figura abaixo:

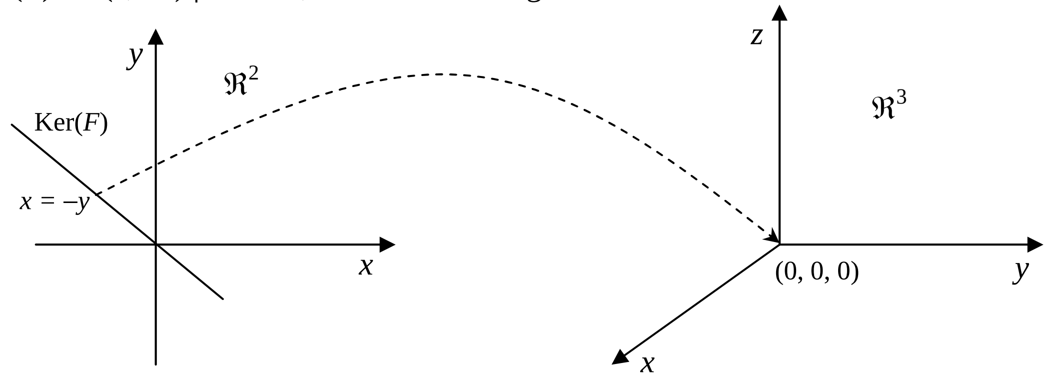


Figura 1: Núcleo de uma transformação linear

O espaço nulo de uma matriz $\mathbf{A} \in \Re^{m\times n}$ é o subespaço Ker($\mathbf{A}$) = null($\mathbf{A}$) = $\{\mathbf{x} \in \Re^n \mid \mathbf{A}\cdot\mathbf{x} = \mathbf{0}\} \subseteq \Re^n$, ou seja, é o conjunto de todas as soluções do sistema homogêneo $\mathbf{A}\cdot\mathbf{x} = \mathbf{0}$. No caso de uma matriz regular (inversível), Ker($\mathbf{A}$) = $\{\mathbf{0}\}$.

Imagem ou espaço gerado: Sejam U e V espaços vetoriais sobre $\Re$ e F: $U \to V$ uma transformação linear. Indica-se por Im(F) ou range(F) e denomina-se imagem ou espaço gerado de F o seguinte subconjunto de V:

$$\mathrm{Im}(F) = \mathrm{range}(F) = \{F(u) \mid u \in U\} \subseteq V$$

A imagem ou espaço gerado por uma matriz $\mathbf{A} \in \Re^{m\times n}$ é o subespaço Im($\mathbf{A}$) $\subseteq \Re^m$ gerado pela transformação linear $F(\mathbf{x}) = \mathbf{A}\cdot\mathbf{x}$, isto é,

$$\mathrm{Im}(\mathbf{A}) = \mathrm{range}(\mathbf{A}) = \{\mathbf{A}\cdot\mathbf{x} \mid \mathbf{x} \in \Re^n\} \subseteq \Re^m$$

Da mesma forma, a imagem ou espaço gerado pela transposta de uma matriz $\mathbf{A} \in \Re^{m\times n}$ é o subespaço Im($\mathbf{A}^T$) $\subseteq \Re^n$ gerado pela transformação linear $F(\mathbf{y}) = \mathbf{A}^T\cdot\mathbf{y}$, isto é,

$$\mathrm{Im}(\mathbf{A}^T) = \mathrm{range}(\mathbf{A}^T) = \{\mathbf{A}^T\cdot\mathbf{y} \mid \mathbf{y} \in \Re^m\} \subseteq \Re^n$$

Um resultado importante é sobre a dimensão destes subespaços. Sejam U e V espaços vetoriais de dimensão finita sobre $\Re$. Dada uma transformação linear F: $U \to V$, então

$$\dim U = \dim \mathrm{Ker}(F) + \dim \mathrm{Im}(F)$$

Conjunto gerador: para um conjunto de vetores $S = \{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_r\}$ o subespaço:

$$\mathrm{span}(S) = \alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_r\mathbf{v}_r$$

gerado por todas as combinações lineares dos vetores de S é chamado espaço gerado por S, denominado conjunto gerador.

Com o uso deste conceito, pode-se dizer que range($\mathbf{A}$) é o espaço gerado pelos vetores coluna de $\mathbf{A}$ e range($\mathbf{A}^T$) é o espaço gerado pelos vetores linha de $\mathbf{A}$.

Posto ou *rank*: o posto ou rank de uma matriz $\mathbf{A} \in \Re^{m\times n}$ é o número máximo de linhas ou colunas linearmente independentes. Observe que dim Im($\mathbf{A}$) = rank($\mathbf{A}$) e, portanto, se Ker($\mathbf{A}$) = $\{\mathbf{0}\}$, então rank($\mathbf{A}$) = n, pois $\dim \Re^n = n = \dim \mathrm{Ker}(\mathbf{A}) + \dim \mathrm{Im}(\mathbf{A})$. Da mesma forma, se Ker($\mathbf{A}^T$) = $\{\mathbf{0}\}$, então rank($\mathbf{A}$) = m.

Matriz simétrica: $\mathbf{A} = \mathbf{A}^T$

Matriz hermitiana (ou auto-adjunta): $\mathbf{A} = \mathbf{A}^H$, ou seja, matriz complexa igual a sua transposta conjugada ($a_{ij} = \overline{a}_{ji}$).

Exemplo de matriz hermitiana: $A = \begin{bmatrix} 1 & 3+i \\ 3-i & 4 \end{bmatrix}$

Matriz ortogonal (**Q**): $\mathbf{Q}^{-1} = \mathbf{Q}^T$

Matriz unitária (**U**): $\mathbf{U}^{-1} = \mathbf{U}^H$

As matrizes hermitiana e unitária em $\mathbb{C}$ (campo complexo) estão, respectivamente, para as matrizes simétrica e ortogonal em $\Re$. Por isso, as descrições a seguir serão limitadas ao campo dos números reais, podendo ser estendidas para $\mathbb{C}$ substituindo as matrizes ortogonais por matrizes unitárias e as matrizes simétricas por matrizes hermitianas.

Dada uma matriz $\mathbf{A} \in \Re^{m \times n}$ e lembrando dos conceitos básicos que o range(**A**) é o espaço gerado pelos vetores coluna de **A** e range($\mathbf{A}^T$) é o espaço gerado pelos vetores linha de **A**, a decomposição SVD é capaz de obter simultaneamente as bases ortonormais destes subespaços.

Qualquer matriz $\mathbf{A} \in \Re^{m \times n}$ pode ser decomposta na forma:

$$\mathbf{A} = \mathbf{U} \cdot \Sigma \cdot \mathbf{V}^T$$

onde $\mathbf{U} \in \Re^{m \times m}$ é uma matriz ortogonal cujas colunas são os vetores característicos de $\mathbf{A} \cdot \mathbf{A}^T$, $\mathbf{V} \in \Re^{n \times n}$ é uma matriz ortogonal cujas colunas são os vetores característicos de $\mathbf{A}^T \cdot \mathbf{A}$, $\Sigma \in \Re^{m \times n}$ é uma matriz diagonal contento a raiz quadrada dos valores característicos de $\mathbf{A} \cdot \mathbf{A}^T$ (que são equivalentes aos valores característicos de $\mathbf{A}^T \cdot \mathbf{A}$), arranjados em ordem decrescente. Os vetores característicos de $\mathbf{A} \cdot \mathbf{A}^T$ e $\mathbf{A}^T \cdot \mathbf{A}$ estão arranjados nas colunas de **U** e **V**, respectivamente, na ordem de seus valores característicos na matriz Σ. Os elementos, σ_i, da diagonal de Σ são denominados de valores singulares de **A**, sendo todos não-negativos. Além disso, o número de valores singulares positivos é igual ao rank(**A**). Os vetores coluna de **U** são denominados de vetores singulares à esquerda de **A** e os vetores coluna de **V** são denominados de vetores singulares à direita de **A**, e as relações entre estes vetores são:

$$\mathbf{A} \cdot \mathbf{v}_i = \sigma_i \, \mathbf{u}_i \quad \text{e} \quad \mathbf{A}^T \cdot \mathbf{u}_i = \sigma_i \, \mathbf{v}_i$$

A decomposição SVD revela várias propriedades intrínsecas da matriz **A** e é numericamente estável para os cálculos. Algumas propriedades são listadas abaixo para uma matriz com *r* valores singulares positivos:

1) rank(**A**) = r

2) null(**A**) = span($\mathbf{v}_{r+1}$, $\mathbf{v}_{r+2}$, ..., $\mathbf{v}_n$)

3) range(**A**) = span($\mathbf{u}_1$, $\mathbf{u}_2$, ..., $\mathbf{u}_r$)

4) range($\mathbf{A}^{\mathrm{T}}$) = span($\mathbf{v}_1$, $\mathbf{v}_2$, ..., $\mathbf{v}_r$)

5) $\mathbf{A} = \sum_{i=1}^{r} \sigma_i \, \mathbf{u}_i \cdot \mathbf{v}_i^T$

6) $\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^{r} \sigma_i^2}$ (norma de Frobenius)

7) $\|\mathbf{A}\|_2 = \sigma_1$

A matriz:

$$\mathbf{A}^{\dagger} = \mathbf{V} \cdot \Sigma^{\dagger} \cdot \mathbf{U}^T = (\mathbf{A}^T \cdot \mathbf{A})^{-1} \cdot \mathbf{A}^T$$

é chamada de pseudo-inversa de **A**, onde os elementos da diagonal de $\Sigma^{\dagger}$ consistem no recíproco dos valores singulares positivos de Σ, na mesma ordem. A pseudo-inversa tem a propriedade $\mathbf{A}^{\dagger} \cdot \mathbf{A} = \mathbf{A} \cdot \mathbf{A}^{\dagger} = \mathbf{I}$.

A solução do problema de valores singulares para o sistema linear, $\mathbf{y} = \mathbf{A} \cdot \mathbf{x}$, corresponde resolver o seguinte problema de otimização:

$$\max_{\mathbf{x}} \| \mathbf{y} \|_2^2$$

$$\text{sujeito a } \| \mathbf{x} \|_2^2 = 1$$

onde $\| \mathbf{y} \|_2^2 = \mathbf{y}^T \cdot \mathbf{y}$. Usando o conceito dos multiplicadores de Lagrange, o problema acima pode ser reescrito como:

$$\max_{\mathbf{x},\lambda} \left\{ S(\mathbf{x}) = \mathbf{x}^T \cdot \mathbf{A}^T \cdot \mathbf{A} \cdot \mathbf{x} - \lambda \, (\mathbf{x}^T \cdot \mathbf{x} - 1) \right\}$$

cuja primeira condição de otimalidade é $\nabla S(\mathbf{x}, \lambda) = 2 (\mathbf{A}^T \cdot \mathbf{A} \cdot \mathbf{x} - \lambda \, \mathbf{x}) = 0$ e $\| \mathbf{x} \|_2 = 1$, ou seja, a solução é equivalente ao problema de valor característico $\mathbf{A}^T \cdot \mathbf{A} \cdot \mathbf{x} = \lambda \, \mathbf{x}$, cujos **x** ótimos locais correspondem aos vetores característicos de $\mathbf{A}^T \cdot \mathbf{A}$ ou os vetores singulares de **A** (também chamados de **componentes principais** de variação, pois indicam as direções de máxima variação de **y** em função das variações em **x** com mesma energia, isto é, $\| \mathbf{x} \|_2 = 1$) e os respectivos multiplicadores de Lagrange são os valores característicos de $\mathbf{A}^T \cdot \mathbf{A}$ ou o quadrado dos valores singulares de **A**. A matriz Hessiana em relação a **x** deste problema de otimização, $H(\mathbf{x}, \lambda) = 2 (\mathbf{A}^T \cdot \mathbf{A} - \lambda \mathbf{I})$, é negativa semi-definida (máximo) para o maior valor singular e positiva semi-definida (mínimo) para o menor valor singular, sendo indefinida para os demais valores intermediários.

Observe que para uma matriz de rank(**A**) = r, $\mathbf{A} = \mathbf{U} \cdot \Sigma \cdot \mathbf{V}^{\mathrm{T}} = \sum_{i=1}^{r} \sigma_i \, \mathbf{u}_i \cdot \mathbf{v}_i^T$ e, portanto, $\mathbf{y} = \mathbf{A} \cdot \mathbf{x} = \mathbf{U} \cdot \Sigma \cdot \mathbf{V}^{\mathrm{T}} \cdot \mathbf{x} = \sum_{i=1}^{r} \sigma_i \, \mathbf{u}_i \cdot \mathbf{v}_i^T \cdot \mathbf{x}$, indicando que a projeção do vetor **x** na direção do vetor $\mathbf{v}_i$ (ou seja, $\mathbf{v}_i^T \cdot \mathbf{x}$) é amplificada por σ_i na direção $\mathbf{u}_i$ do vetor **y**, sendo i = 1 a direção de maior amplificação e $i = r$ a direção de menor amplificação. Dependendo do valor de σ_i, uma pequena mudança em **x** pode causar uma grande mudança em **y**, mas isto vai depender do ângulo entre os vetores **x** e $\mathbf{v}_i$.

Exemplo: usando a função demonstrativa **eigshow(A)** do MATLAB, que apresenta de forma gráfica um vetor unitário **x** e sua transformação linear **A·x** para uma matriz de dimensão 2 × 2 e valores de **x**, com ||**x**|| = 1, é possível pela movimentação do mouse sobre a figura localizar as direções características da matriz. As Figuras 2a, 2b e 2c ilustram esta movimentação para a matriz

$$\boldsymbol{A} = \begin{bmatrix} 1/4 & 3/4 \\ 1 & 1/2 \end{bmatrix}$$

onde os valores característicos, λ, e os vetores característicos, **x**, resultantes das soluções não-triviais do sistema de equações lineares

$$\mathbf{A}\cdot\mathbf{x} = \lambda\,\mathbf{x}$$

$$\lambda_1 = 5/4 \quad \text{e} \quad \mathbf{x}_1 = [3/5 \quad 4/5]^T \qquad ; \qquad \lambda_2 = -1/2 \quad \text{e} \quad \mathbf{x}_2 = [\sqrt{2}/2 \quad -\sqrt{2}/2\,]^T$$

são visualizados quando os dois vetores (**x** e **A·x**) estão na mesma direção (Figuras 2b e 2c). Mostrando que o operador **A**, na direção de **x**, corresponde a uma redução ou ampliação por um fator λ. Quando os sentidos dos dois vetores são opostos tem-se um valor característico negativo.

Pode-se observar que os dois vetores característicos não são os eixos maior e menor da elipse formada pelas transformações lineares. Seriam para o caso particular de matrizes simétricas. Matrizes 2 × 2 com um par de valores característicos complexo não possuem vetores característicos reais.

Movendo dois vetores unitários, **x** e **y**, perpendiculares e suas correspondentes transformações lineares, **A·x** e **A·y**, pode-se visualizar os valores e vetores singulares, resultantes das soluções não-triviais dos sistemas de equações lineares

$$\mathbf{A}^T\cdot\mathbf{A}\cdot\mathbf{v} = \sigma^2\,\mathbf{v} \qquad\qquad \mathbf{A}\cdot\mathbf{A}^T\cdot\mathbf{u} = \sigma^2\,\mathbf{u}$$

$$\sigma_1 = 1{,}2792, \quad \mathbf{x} = \mathbf{v}_1 = [0{,}7678 \quad 0{,}6407]^T \quad \text{e} \quad \mathbf{A}\cdot\mathbf{x} = \sigma_1\,\mathbf{u}_1 = [0{,}6725 \quad 1{,}0881]^T$$

$$\sigma_2 = 0{,}4886, \quad \mathbf{y} = \mathbf{v}_2 = [-0{,}6407 \quad 0{,}7678]^T \quad \text{e} \quad \mathbf{A}\cdot\mathbf{y} = \sigma_2\,\mathbf{u}_2 = [0{,}4156 \quad -0{,}2569]^T$$

onde σ são os valores singulares, **v** e **u** são os vetores singulares à direita e à esquerda, respectivamente. Eles surgem no momento em que as transformações são perpendiculares entre si, conforme mostra a Figura 2d. Observa-se que isto acontece quando os vetores das transformações são os eixos maior e menor da elipse, mostrando, por exemplo, as direções de máxima e mínima amplificação de sinais, respectivamente. Para o caso particular de uma matriz quadrada, simétrica e positiva definida, as decomposições em valores característicos e em valores singulares são equivalentes.

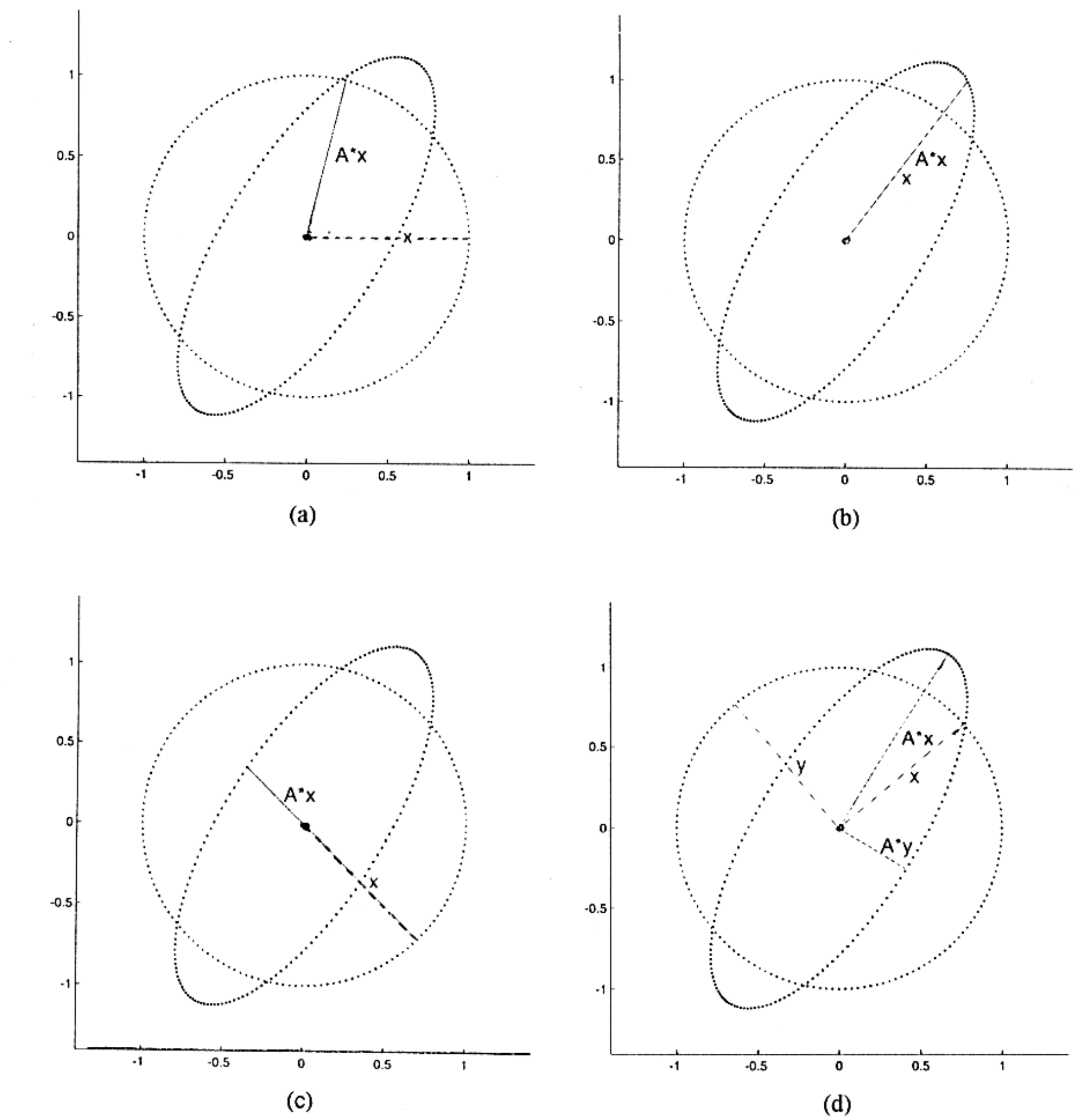


Figura 2: Direções características (b, c) e singulares (d) de uma matriz.

Para determinar se o sistema linear, **y** = **A·x**, está bem escalonado deve-se verificar o condicionamento da matriz **A**, que na norma 2 é dado por:

$$\gamma(\mathbf{A}) = \overline{\sigma}(\mathbf{A}) / \underline{\sigma}(\mathbf{A})$$

onde $\overline{\sigma}(\mathbf{A})$ é o maior valor singular de **A** e $\underline{\sigma}(\mathbf{A})$ é o menor valor singular não-nulo de **A**.

A melhor maneira de escalonar um sistema é atacando a origem do problema, ou seja, um apropriado adimensionamento das variáveis dependentes e das equações do problema. Uma maneira de determinar quais as variáveis devem ser re-escalonadas é através do cálculo do condicionamento mínimo, isto é, determinar as matrizes que pré- e pós-multiplicadas pela matriz **A** resultam em um γ mínimo (γ*), isto é:

$$\gamma^*(\mathbf{A}) = \min_{\mathbf{L},\mathbf{R}} \gamma(\mathbf{L} \cdot \mathbf{A} \cdot \mathbf{R})$$

Considerando **L** e **R** matrizes diagonais, tem-se como resultado do problema de otimização acima quais as saídas e entradas que devem ser re-escalonadas, respectivamente, pois:

$$\mathbf{y}_e = \mathbf{L} \cdot \mathbf{y} \qquad \text{e} \qquad \mathbf{x}_e = \mathbf{R}^{-1} \cdot \mathbf{x}$$

No MATLAB, o condicionamento de uma matriz **A** pode ser obtido através da função *cond*(**A**), e as matrizes de escalonamento podem ser calculadas pela função:

$$[ub, \mathbf{D}] = mu\ (\mathbf{H}, ones\ (2*n, 2), \text{'C'})$$

onde n é a ordem da matriz **A**, $\mathbf{H} = \begin{bmatrix} 0 & \mathbf{A}^{-1} \\ \mathbf{A} & 0 \end{bmatrix}$, $\gamma^*(\mathbf{A}) = \left(\dfrac{ub(1)}{ub(2)} \right)^2$,

L = *diag*(**D**(n+1:2*n)) e **R** = *inv*(*diag*(**D**(1:n))). Para verificar se os cálculos estão corretos, basta calcular o *cond*(**L*****A*****R**), que deve ser igual a $\gamma^*(\mathbf{A})$.