Solução Numérica de Sistemas Lineares

Muitos problemas provenientes da Física, Matemática, Engenharia, Biologia, Economia, entre outros, envolvem sistemas de equações lineares. Esses sistemas nalguns casos aparecem directamente como parte do modelo matemático do problema real e, noutros casos, surgem em conjunto com a aplicação dum método numérico. Como veremos, a resolução dum sistema de equações não lineares pelo método de Newton conduz a um sistema de equações lineares; também a aplicação de certos métodos numéricos na resolução de equações diferenciais conduz em geral a sistemas lineares da forma:

$$\begin{array}{lccccc} E_1: & a_{11}x_1 & +a_{12}x_2 & +\ldots & +a_{1n}x_n & = b_1 \\ E_2: & a_{21}x_1 & +a_{22}x_2 & +\ldots & +a_{2n}x_n & = b_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ E_n: & a_{n1}x_1 & +a_{n2}x_2 & +\ldots & +a_{nn}x_n & = b_n \end{array} \qquad (I)$$

onde $\begin{cases} x_j, j = 1, 2, \ldots, n \text{ são as incógnitas,} \\ a_{i,j}, b_j, i = 1, 2, \ldots, n,\ j = 1, 2, \ldots, n, \\ \text{são constantes dadas.} \end{cases}$

O sistema linear acima pode ser representado matricialmente por:

$$A\mathbf{x} = \mathbf{b} \qquad (II)$$

onde $\begin{cases} A = (a_{i,j}) & \text{matriz dos coeficientes} \\ \mathbf{b} = (b_j) & \text{vector do lado direito} \\ \mathbf{x} = (x_j) & \text{vector solução.} \end{cases}$

Para resolver (II), existem 2 classes de métodos:

- Métodos Directos: métodos que encontram a solução (exacta) num número finito de operações. Entre os métodos existentes, temos os seguintes: Método de Eliminação de Gauss e os Métodos de factorização LU (Doolittle, Crout e Cholesky).
- Métodos Iterativos: métodos que encontram a solução num número infinito de operações. Mais precisamente, num método iterativo convergente obtém-se uma sucessão de aproximações convergindo para a solução exacta. Aqui estudaremos, em particular, o método de Jacobi e o método de Gauss-Seidel.

Nestas notas começamos com uma revisão do método de Gauss e dos métodos de factorização e descrevemos, em seguida, duas técnicas de pesquisa de pivot. Apresentamos dois tipos de matrizes especiais e completa-se a primeira parte com uma introdução ao condicionamento e análise do erro de sistemas lineares.

Métodos Directos

Método de Eliminação de Gauss

Para obtermos uma solução de (II), assumiremos que a matriz A é invertível ($det(A) \neq 0$).

A idéia básica do método de Eliminação de Gauss é transformar o sistema (II) num sistema equivalente

$$A^{(n)}\mathbf{x} = \mathbf{b}^{(n)}$$

onde $A^{(n)}$ é uma matriz triangular superior. Essa transformação é obtida utilizando as seguintes operações:

$$\begin{cases} i)\ E_i - \lambda E_j \longrightarrow E_i & \{\text{ substituição de } E_i \text{ por a combinação linear de } E_i \text{ e } E_j \quad \} \\ ii)\ \lambda E_i \longrightarrow E_i & \{\text{ substituição de } E_i \text{ por } \lambda E_i,\ \lambda \neq 0 \quad \} \\ iii)\ E_i \longleftrightarrow E_j & \{\text{ substituição de } E_i \text{ por } E_j \quad \} \end{cases}$$

Veremos agora que a aplicação dessas operações ao sistema (II) conduz a um sistema equivalente mais fácil de se resolver.

Exemplo Seja o sistema linear

$$\begin{bmatrix} 4 & -9 & 2 \\ 2 & -4 & 4 \\ -1 & 2 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}.$$

Para resolvê-lo, procedemos como segue:

1. $a_{11} = 4 \neq 0 \Longrightarrow$ podemos colocar zeros abaixo de a_{11}.

   $$\begin{cases} m_{21} = \dfrac{a_{21}}{a_{11}} = \dfrac{2}{4} = \dfrac{1}{2} \\ m_{31} = \dfrac{a_{31}}{a_{11}} = -\dfrac{1}{4} = -\dfrac{1}{4} \end{cases} \text{ e efectuando as operações } \begin{cases} E_2 - m_{21}E_1 \longrightarrow E_2 \\ E_3 - m_{31}E_1 \longrightarrow E_3 \end{cases}$$

   obtém-se o sistema equivalente: $A^{(2)}\mathbf{x} = \mathbf{b}^{(2)}$ onde

   $$\underbrace{\begin{bmatrix} 4 & -9 & 2 \\ 0 & 1/2 & 3 \\ 0 & -1/4 & 5/2 \end{bmatrix}}_{A^{(2)}} \vdots \underbrace{\begin{bmatrix} 2 \\ 2 \\ 3/2 \end{bmatrix}}_{\mathbf{b}^{(2)}}$$

2. $a_{22}^{(2)} = 1/2 \neq 0 \Longrightarrow$ podemos colocar zeros abaixo de $a_{22}^{(2)}$.

$m_{32} = \dfrac{a_{32}^{(2)}}{a_{22}^{(2)}} = -\dfrac{1/4}{1/2} = -\dfrac{2}{4} = -\dfrac{1}{2}$ e efectuando a operação $E_3 - m_{32}E_2 \longrightarrow E_3$ obtem-se os sistema equivalente $A^{(3)}\mathbf{x} = \mathbf{b}^{(3)}$:

$$\underbrace{\left[\begin{array}{cccc} 4 & -9 & 2 & 2 \\ 0 & 1/2 & 3 & 2 \\ 0 & 0 & 4 & 5/2 \end{array}\right]}_{A^{(2)} \quad \mathbf{b}^{(2)}}$$

Portanto, o sistema original for transformado no sistema equivalente:

$$\left[\begin{array}{ccc} 4 & -9 & 2 \\ 0 & 1/2 & 3 \\ 0 & 0 & 4 \end{array}\right] \left[\begin{array}{c} x_1 \\ x_2 \\ x_3 \end{array}\right] = \left[\begin{array}{c} 2 \\ 2 \\ 5/2 \end{array}\right]$$

que tem com solução:

$x_3 = 5/2/4 = 5/8$
$1/2x_2 + 3x_3 = 2 \Longrightarrow x_2 = (2 - 3x_3)/1/2 = (2 - 3 \times 5/8)/1/2 = 1/4$
$4x_1 - 9x_2 + 2x_3 = 2 \Longrightarrow x_1 = 2 + 9x_2 - 2x_3)/4 = (2 + 9 \times 1/4 - 2 \times 5/8)/4 = 3/4$

No caso geral tem-se:

1º Passo: $A\mathbf{x} = \mathbf{b} \longrightarrow A^{(2)}\mathbf{x} = \mathbf{b}^{(2)}$

a_{11} é chamado "elemento pivot". Se $a_{11} \neq 0$ então efectuam-se as seguintes operações:

$$\begin{cases} m_{i1} = a_{i1}/a_{11} \\ E_i - m_{i1}E_1 \longrightarrow E_i, \ i = 2, 3, \ldots, n \end{cases}$$

e obtém-se o sistema equivalente:

$$\left[\begin{array}{cccc:c} a_{11} & a_{12} & \ldots & a_{1n} & b_1 \\ a_{21} & a_{22} & \ldots & a_{2n} & b_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{nn} & b_n \end{array}\right] \longrightarrow \left[\begin{array}{cccc:c} a_{11} & a_{12} & \ldots & a_{1n} & b_1 \\ 0 & a_{22}^{(2)} & \ldots & a_{2n}^{(2)} & b_2^{(2)} \\ \vdots & a_{32}^{(2)} & \ldots & a_{3n}^{(2)} & b_3^{(2)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & a_{n2}^{(2)} & \ldots & a_{nn}^{(2)} & b_n^{(2)} \end{array}\right]$$

2º Passo: $A^{(2)}\mathbf{x} = \mathbf{b}^{(2)} \longrightarrow A^{(3)}\mathbf{x} = \mathbf{b}^{(3)}$

Se $a_{22}^{(2)} \neq 0$ então calculam-se os multiplicadores

$$m_{i1} = a_{i1}^{(2)}/a_{22}^{(2)}, \ i = 3, 4, \ldots, n$$

e efectuam-se as operações:

$$E_i - m_{i1}E_1 \longrightarrow E_i,\ i = 3, 4, \ldots, n$$

Obtém-se o sistema equivalente:

$$[A^{(3)} : \mathbf{b}^{(3)}] = \left[\begin{array}{ccccccc} a_{11} & a_{12} & a_{13} & \ldots & a_{1n} & \vdots & b_1 \\ 0 & a_{22}^{(2)} & a_{23}^{(2)} & \ldots & a_{2n}^{(2)} & \vdots & b_2^{(2)} \\ 0 & 0 & a_{33}^{(3)} & \ldots & a_{3n}^{(3)} & \vdots & b_3^{(3)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\ 0 & 0 & a_{n3}^{(3)} & \ldots & a_{nn}^{(3)} & \vdots & b_n^{(3)} \end{array}\right]$$

Prosseguindo o processo, chegamos, no passo $(n-1)$, ao sistema $A^{(n)}\mathbf{x} = \mathbf{b}^{(n)}$, onde

$$[A^{(n)} : \mathbf{b}^{(n)}] = \left[\begin{array}{ccccccc} a_{11} & a_{12} & a_{13} & \ldots & \ldots & a_{1n} & \vdots & b_1 \\ 0 & a_{22}^{(2)} & a_{23}^{(2)} & \ldots & \ldots & a_{2n}^{(2)} & \vdots & b_2^{(2)} \\ 0 & 0 & a_{33}^{(3)} & \ldots & \ldots & a_{3n}^{(3)} & \vdots & b_3^{(3)} \\ \vdots & \vdots & 0 & \ldots & \vdots & \vdots & \vdots & \\ 0 & 0 & 0 & \ldots & 0 & a_{nn}^{(n)} & \vdots & b_n^{(n)} \end{array}\right]$$

É-se assim conduzido a um sistema linear equivalente ao sistema inicial, cuja matriz é triangular superior. Este sistema pode facilmente ser resolvido, começando com: $x_n = b_n^{(n)}/a_{nn}^{(n)}$ e calculando-se $x_{n-1}, x_{n-2}, \ldots, x_1$, sucessivamente.

Obs: Se no passo k acontecer que o elemento pivot $a_{kk}^{(k)} = 0$, então procura-se um elemento $a_{lk}^{(k)} \neq 0$ na coluna k, onde $l > k$ e trocam-se as linhas l e k, ou seja, efectua-se a operação $E_k \longleftrightarrow E_l$.

Métodos de Factorização: LU

Estes métodos consistem em decompor a matriz A num produto de duas matrizes LU, ou seja, $A = LU$, onde L é uma matriz triangular inferior e U é uma matriz triangular superior.

Suponhamos que existem matrizes L e U tais que $A = LU$, onde L é triangular inferior e U é triangular superior. Então, dado o sistema linear $A\mathbf{x} = \mathbf{b}$, podemos escrever

$$(LU)\mathbf{x} = \mathbf{b}$$

Fazendo $U\mathbf{x} = \mathbf{g}$ tem-se que

$$L\mathbf{g} = \mathbf{b}, \qquad (1)$$

que é um sistema triangular inferior. Resolvendo (1), a solução $\mathbf{x}$ é obtida resolvendo-se a equação

$$U\mathbf{x} = \mathbf{g}, \qquad (2)$$

que é um sistema triangular superior. Portanto, para resolver o sist. $A\mathbf{x} = \mathbf{b}$, basta resolver os dois sistemas triangulares definidos por (1) e (2).

Existem várias maneiras de se encontrar matrizes L e U de modo que $A = LU$, como veremos a seguir.

Cálculo das matrizes L e U.

Considerando a equação matricial

$$\underbrace{\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}}_{A \text{ (matriz dada)}} = \underbrace{\begin{bmatrix} l_{11} & 0 & \dots & \dots & 0 \\ l_{21} & l_{22} & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & 0 \\ \vdots & \vdots & & \ddots & 0 \\ l_{n1} & l_{n2} & \dots & \dots & l_{nn} \end{bmatrix}}_{L} \underbrace{\begin{bmatrix} u_{11} & u_{12} & \dots & \dots & u_{1n} \\ 0 & u_{22} & u_{23} & \dots & u_{2n} \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & u_{nn} \end{bmatrix}}_{U}$$

vemos que temos $(n^2 + n)$–incógnitas: l_{ij} e u_{ij}. Por outro lado, pela regra de produto de matrizes, podemos formar n^2-equações! Logo, precisamos de n–equações adicionais. Vejamos como obtê-las:

Método de Doolittle (ou Factorização de Doolittle): Impor $l_{ii} = 1,\ i = 1, 2, \ldots, n$. Nesse caso temos

$$\begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \ldots & \ldots & 0 \\ l_{21} & 1 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \ddots & 0 \\ \vdots & \vdots & & \ddots & 0 \\ l_{n1} & l_{n2} & \ldots & \ldots & 1 \end{bmatrix} \begin{bmatrix} u_{11} & u_{12} & \ldots & \ldots & u_{1n} \\ 0 & u_{22} & u_{23} & \ldots & u_{2n} \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 & u_{nn} \end{bmatrix}.$$

Efectuando o produto LU e usando a igualdade $A = LU$, elemento a elemento, obtemos as equações:

1a. linha de U : $u_{11} = a_{11}, u_{12} = a_{12}, \ldots, u_{1n} = a_{1n}$.

1a. coluna de L (supondo que $a_{11} \neq 0$)
$$\begin{cases} l_{21}u_{11} = a_{21} \implies l_{21} = a_{21}/u_{11} \\ l_{31}u_{11} = a_{31} \implies l_{31} = a_{31}/u_{11} \\ \quad \vdots \quad \vdots \qquad\qquad \vdots \\ \quad \vdots \quad \vdots \qquad\qquad \vdots \\ l_{n1}u_{11} = a_{n1} \implies l_{n1} = a_{n1}/u_{11} \end{cases}$$

2a. linha de U
$$\begin{cases} l_{21}u_{12} + u_{22} = a_{22} \implies u_{22} = a_{22} - l_{21}u_{12} \\ l_{21}u_{13} + u_{23} = a_{23} \implies u_{23} = a_{23} - l_{21}u_{13} \\ l_{21}u_{14} + u_{24} = a_{24} \implies u_{24} = a_{24} - l_{21}u_{14} \\ \quad \vdots \quad \vdots \qquad\qquad \vdots \\ \quad \vdots \quad \vdots \qquad\qquad \vdots \\ l_{21}u_{1n} + u_{2n} = a_{2n} \implies u_{2n} = a_{2n} - l_{21}u_{1n} \end{cases}$$

2a. coluna de L
$$\begin{cases} l_{31}u_{12} + l_{32}u_{22} = a_{32} \implies l_{32} = (a_{32} - l_{31}u_{12})/u_{22} \\ l_{41}u_{12} + l_{42}u_{22} = a_{42} \implies l_{42} = (a_{42} - l_{41}u_{12})/u_{22} \\ \quad \vdots \quad \vdots \qquad\qquad \vdots \\ \quad \vdots \quad \vdots \qquad\qquad \vdots \\ l_{n1}u_{12} + l_{n2}u_{22} = a_{n2} \implies l_{n2} = (a_{n2} - l_{n1}u_{12})/u_{22} \end{cases}$$

En geral, tem-se:

$$\text{linha } k \text{ de } U : u_{kj} = a_{kj} - \sum_{r=1}^{k-1} l_{kr}u_{rj},\ j = k, k+1, \ldots, n$$

$$\text{coluna } k \text{ de } L : l_{ik} = \left(a_{ik} - \sum_{r=1}^{k-1} l_{ir}u_{rk} \right) / u_{kk}, i = k+1, k+2, \ldots, n$$

Pode mostrar-se que se o método de Eliminação de Gauss for aplicado ao sistema $A\mathbf{x} = \mathbf{b}$ sem troca de linhas então $A = LU$ onde

$$L = \begin{bmatrix} 1 & 0 & \dots & \dots & \dots & 0 \\ m_{21} & 1 & 0 & \dots & \dots & 0 \\ m_{31} & m_{32} & 1 & \ddots & & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & \vdots & & \ddots & \ddots & 0 \\ m_{n1} & m_{n2} & \dots & \dots & m_{nn-1} & 1 \end{bmatrix}; \; U = A^{(n)} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & \dots & a_{1n} \\ 0 & a_{22}^{(2)} & a_{23}^{(2)} & \dots & \dots & a_{2n}^{(2)} \\ 0 & 0 & a_{33}^{(3)} & \dots & \dots & a_{3n}^{(3)} \\ \vdots & \vdots & \ddots & \dots & \vdots & \vdots \\ \vdots & \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & a_{nn}^{(n)} \end{bmatrix}$$

Exemplo: Determine a solução do sistema linear abaixo pelo método de Doolittle.

$$\begin{bmatrix} 2 & 3 & -2 \\ 2 & -16 & 10 \\ 14 & -7 & -7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ -3 \\ 0 \end{bmatrix}$$

Solução: Pelo método de Doolittle temos:

$$A\mathbf{x} = \mathbf{b} \Longleftrightarrow LU\mathbf{x} = \mathbf{b} \Longleftrightarrow \begin{cases} L\mathbf{g} = \mathbf{b} \\ U\mathbf{x} = \mathbf{g} \end{cases}$$

onde

$$L = \begin{bmatrix} 1 & 0 & 0 \\ l_{21} & 1 & 0 \\ l_{31} & l_{32} & 1 \end{bmatrix} \text{ e } U = \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{bmatrix}$$

As matrizes L e U são obtidas pelas fórmulas:

$$\begin{aligned} u_{kj} &= a_{kj} - \sum_{s=1}^{k-1} l_{ks} u_{sj}, \quad j = k, k+1, \dots, 3 \\ l_{ik} &= \left(a_{ik} - \sum_{s=1}^{k-1} l_{is} u_{sk} \right) / (u_{kk}), \quad i = k+1, k+2, \dots, 3 \end{aligned}$$

- Cálculo de L e U:

$$\left| \begin{array}{c} \text{1a. linha de } U: \\ u_{11} = a_{11} \Longrightarrow u_{11} = 2 \\ u_{12} = a_{12} \Longrightarrow u_{12} = 3 \\ u_{13} = a_{13} \Longrightarrow u_{13} = -2 \end{array} \right| \left| \begin{array}{c} \text{1a. coluna de } L: \\ l_{21} = a_{21}/u_{11} = 2/2 = 1 \\ l_{31} = a_{31}/u_{11} = 14/2 = 7 \end{array} \right.$$

$$\left| \begin{array}{c} \text{2a. linha de } U: \\ u_{22} = a_{22} - l_{21}u_{12} \Longrightarrow u_{11} = -16 - 1 \times 3 = -19 \\ u_{23} = a_{23} - l_{21}u_{13} \Longrightarrow u_{23} = 10 - 1 \times -2 = 12 \end{array} \right| \left| \begin{array}{c} \text{2a. coluna de } L: \\ l_{32} = a_{32} - l_{31}u_{12} = \dfrac{-7 - 7 \times 3}{-19} \\ = 28/19 = 1.47368 \end{array} \right.$$

$$\left|\begin{array}{c} \text{3a. linha de } U: \\ u_{33} = a_{33} - l_{31}u_{13} - -l_{32}u_{23} = -7 - (7 \times -2) - \left(\frac{28}{19} \times 12\right) \\ = -7 + 14 - 17.6842 = -10.68421 \end{array}\right.$$

Portanto,

$$L = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 7 & 1.473684 & 1 \end{bmatrix} \text{ e } U = \begin{bmatrix} 2 & 3 & -2 \\ 0 & -19 & 12 \\ 0 & 0 & -10.68421 \end{bmatrix}$$

Solução dos sistemas lineares $L\mathbf{g} = \mathbf{b}$ e $U\mathbf{x} = \mathbf{g}$

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 7 & 1.473684 & 1 \end{bmatrix} \begin{bmatrix} g_1 \\ g_2 \\ g_3 \end{bmatrix} = \begin{bmatrix} 0 \\ -3 \\ 0 \end{bmatrix} \Longrightarrow \begin{cases} g_1 = 0 \\ g_1 + g_2 = -3 \Longrightarrow g_2 = -3 \\ 7g_1 + 1.473681g_2 + g_3 = 0 \\ \Longrightarrow g_3 = -1.473681g_2 = 4.42105 \end{cases}$$

Portanto, $\mathbf{g} = [0 \ \ -3 \ \ 4.42104]^T$.

$$\begin{bmatrix} 2 & 3 & -2 \\ 0 & -19 & 12 \\ 0 & 0 & -10.68421 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ -3 \\ 4.42105 \end{bmatrix}$$

$$\Longrightarrow \begin{cases} x_3 = \dfrac{4.42104}{-10.68421} = -0.413793 \\ -19x_2 + 12x_3 = -3 \Longrightarrow x_2 = \dfrac{-3 - 12x_3}{-19} \\ = \dfrac{-3 - 12 \times -0.413792}{-19} = -0.103448 \\ 2x_1 + 3x_2 - 2x_3 = 0 \Longrightarrow x_1 = (-3x_2 + 2x_3)/2 \\ \Longrightarrow x_1 = (-3 \times -0.103448 + 2 \times -0.413792)/2 = -0.258621 \end{cases}$$

Método de Crout (ou Factorização de Crout)

Neste método, a matriz L é triangular inferior e U é triangular superior com $u_{ii} = 1$, ou seja, a decomposição é da forma:

$$\underbrace{\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}}_{A \text{ (matriz dada)}} = \underbrace{\begin{bmatrix} l_{11} & 0 & \dots & \dots & 0 \\ l_{21} & l_{22} & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & 0 \\ \vdots & \vdots & & \ddots & 0 \\ l_{n1} & l_{n2} & \dots & \dots & l_{nn} \end{bmatrix}}_{L} \underbrace{\begin{bmatrix} 1 & u_{12} & \dots & \dots & u_{1n} \\ 0 & 1 & u_{23} & \dots & u_{2n} \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix}}_{U}$$

A obtenção das matrizes L e U segue o mesmo processo descrito na factorização de Doolittle, só que agora determina-se primeiro a coluna k de L seguido pelo cálculo da linha k de U. As equações para a obtenção de L e U são:

- Coluna k de L: $l_{ik} = a_{ik} - \sum_{r=1}^{k-1} l_{ir} u_{rk},\ i = k, k+1, \dots\ n$
- Linha k de U $u_{kj} = \left(a_{kj} - \sum_{r=1}^{k-1} l_{kr} u_{rj} \right) / l_{kk},\ j = k+1, \dots\ n$

Método de Cholesky (ou Factorização de Cholesky)

No caso especial em que A é uma matriz real e simétrica, a decomposição LU pode ser modificada de maneira que $L = U^T$, ou seja, $A = LL^T$. As equações para obter a matriz L seguem o mesmo procedimento do método LU, ou seja,

$$\underbrace{\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}}_{A \text{ (matriz dada)}} = \underbrace{\begin{bmatrix} l_{11} & 0 & \dots & \dots & 0 \\ l_{21} & l_{22} & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & 0 \\ \vdots & \vdots & & \ddots & 0 \\ l_{n1} & l_{n2} & \dots & \dots & l_{nn} \end{bmatrix}}_{L} \underbrace{\begin{bmatrix} l_{11} & l_{21} & \dots & \dots & l_{n1} \\ 0 & l_{22} & \dots & \dots & l_{n2} \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & l_{nn} \end{bmatrix}}_{L^T}$$

Efectuando o produto LL^T e usando a igualdade $A = LU$, elemento a elemento, obtém-se as seguintes equações para a determinação das colunas de L:

$$\begin{aligned} l_{kk} &= \sqrt{a_{kk} - \sum_{r=1}^{k-1} l_{kr}^2} \\ l_{ik} &= \left[a_{ik} - \sum_{r=1}^{k-1} l_{ir} l_{kr} \right] / l_{kk}\ , \quad i = k+1, k+2, \dots, n \end{aligned}$$

Cálculo da Matriz Inversa

Seja A uma matriz não-singular e A^{-1} sua inversa. Então,

$$AA^{-1} = I$$

Seja $\mathbf{x_j} = [x_{1j} \;\; x_{2j} \;\; \ldots \;\; x_{nj}]$, a coluna j $(j = 1, ...n)$ de A^{-1}. Então temos:

$$\begin{bmatrix} a_{11} & a_{12} & \ldots & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & \ldots & a_{2n} \\ \vdots & \vdots & \ldots & \ldots & \vdots \\ \vdots & \vdots & \ldots & \ldots & \vdots \\ a_{n1} & a_{n2} & \ldots & \ldots & a_{nn} \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} & \ldots & \ldots & x_{1n} \\ x_{21} & x_{22} & \ldots & \ldots & x_{2n} \\ \vdots & \vdots & \ldots & \ldots & \vdots \\ \vdots & \vdots & \ldots & \ldots & \vdots \\ x_{n1} & x_{n2} & \ldots & \ldots & x_{nn} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \ldots & \ldots & 0 \\ 0 & 1 & \ddots & & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \ldots & \ldots & 0 & 1 \end{bmatrix}$$

Desta igualdade, vemos que o cálculo de A^{-1} equivale a resolver n sistemas lineares

$$\begin{cases} A\mathbf{x_1} = \mathbf{e_1},\; A\mathbf{x_2} = \mathbf{e_2},\; \ldots,\; A\mathbf{x_n} = \mathbf{e_n} \\ \text{onde } \mathbf{e_1} = [1\;0\;\ldots\;0]^T,\; \mathbf{e_2} = [0\;1\;\ldots\;0]^T,\; \mathbf{e_n} = [0\;\ldots\;1]^T \end{cases}$$

para a determinação das colunas de A^{-1}. Se a matriz A for factorizada na forma $A = LU$, então faz-se

$(LU)\mathbf{x_1} = \mathbf{e_1},\; (LU)\mathbf{x_2} = \mathbf{e_2},\; \ldots, (LU)\mathbf{x_n} = \mathbf{e_n}$. Ou seja, faz-se a factorização LU uma única vez e resolvem-se os n sistemas lineares cujas soluções são as colunas de A^{-1}.

Técnicas de Pesquisa de Pivot

Devido aos erros de arredondamento, o método de Eliminação de Gauss pode conduzir a soluções erróneas. Ou seja, podemos ter problemas de instabilidade numérica. As técnicas de pesquisa de pivot surgem numa tentativa de minorar o efeito da propagação dos erros de arredondamento.

Começamos por apresentar o seguinte exemplo.

Exemplo

$$\begin{array}{lrcrcl} E_1: & 0.003000x_1 & + & 59.14x_2 & = & 59.17 \\ E_2: & 5.291x_1 & - & 6.130x_2 & = & 47.78 \end{array} \qquad (1)$$

que tem como solução exacta $x_1 = 10.0$ e $x_2 = 1.0$. Suponhamos um computador que usa um sistema $VF(10, 4, -10, 10)$ e arredondamento simétrico. Então, aplicando o método de Eliminação de Gauss ao sistema linear (1), temos:

$$m_{21} = \frac{5.291}{0.003000} = 1763.66 \approx 0.1764 \times 10^4$$

e, efectuando a operação $E_2 - m_{21}E_1 \longrightarrow E_2$, obtém-se o sistema equivalente:

$$\begin{cases} 0.003000x_1 + 59.14x_2 = 59.17 \\ \qquad\qquad - 104300x_2 = -104400 \end{cases}$$

que tem como solução:

$$x_2 = \frac{-0.1044 \times 10^6}{-0.1043 \times 10^6} = 1.00096 \approx 1.001 \quad \text{e} \quad x_1 = \frac{59.17 - (59.14 \times 1.001)}{0.003000} = -10.0$$

Vemos que um pequeno erro em x_2 : $|\delta x_2| = \dfrac{|1.0 - 1.001|}{1.0} = 0.001$(0.1 % erro) resultou num grande erro em x_1 : $|\delta x_1| = \dfrac{|10.0 - (-10.0)|}{10.0} = 2$(ou seja, 200 % de erro)

Vejamos que não é indiferente a ordem das equações !

Por outro lado, se trocarmos as equações em (1) ($E_1 \longleftrightarrow E_2$), tem-se:

$$\begin{array}{lrcrcl} & 5.291x_1 & - & 6.130x_2 & = & 46.78 \\ & 0.00300x_1 & + & 59.14x_2 & = & 59.17 \end{array} \qquad (2)$$

e aplicarmos o método de Eliminação de Gauss, obtemos:

$$m_{21} = \frac{0.003000}{5.291} = 0.0005670 \approx 0.0006$$

e a operação $E_2 - m_{21}E_1 \longrightarrow E_2$ conduz agora ao sistema equivalente:

$$\begin{cases} 5.291x_1 - 6.130x_2 = 46.78 \\ \qquad\qquad 59.14x_2 = 59.14 \end{cases}$$

que tem como solução $x_1 = 10.0$ e $x_2 = 1.0$!

Pesquisa Parcial de Pivot

Esta técnica consiste em escolher, para elemento pivot, no k-passo do método de Eliminação de Gauss, o elemento de maior valor absoluto na coluna k, ou seja,

$$\begin{cases} \text{Seja } a_{rk}^k \text{ tal que } |a_{rk}^k| = max\{|a_{ik}^k|, i = k, k+1, \ldots, n\} \\ \text{Então, se } r \neq k \text{ troca-se as linhas } E_r^k \text{ e } E_k^k. \end{cases}$$

Notemos que a aplicação desta técnica, no exemplo anterior, ao sistema (1) conduz ao sistema (2).

obs: Estudos feitos sobre o mét. de Elim. de Gauss mostram que se os multiplicadores m_{ik} são tais que $|m_{ik}| \leq 1$ então o efeito da propagação dos erros de arredondamento na solução é de algum modo reduzido. Esse é o objectivo da técnica de pesquisa parcial de pivot.

Pesquisa Total de Pivot

Aqui o elemento pivot é tal que:

$$|a_{rk}^k| = max\{|a_{ik}^k|, i = j = k, k+1, \ldots, n\}$$

No exemplo anterior, conduziria ao sistema

$$\begin{array}{rcrl} 59.14x_1 & + & 0.00300x_2 & = 59.17 \\ -6.130x_1 & + & 5.291x_2 & = 46.78 \end{array}$$

A pesquisa total de pivot é a técnica que permite maior redução dos erros de arredondamento. Contudo ela requer um maior tempo computacional, sendo, por isso, mais dispendiosa.

Matrizes Especiais

No caso de certo tipo de matrizes, que consideraremos nesta secção, o método de eliminação de Gauss tem um comportamento estável, não sendo necessário recorrer ao uso de pesquisa de pivot.

Definição (Matriz de Diagonal Dominante): Uma matriz $A_{n\times n}$ é chamada "matriz de diagonal dominante" se:

$$|a_{ii}| \geq \sum_{\substack{j=1 \\ j\neq i}}^{n} |a_{ij}| \quad i = 1, 2, \ldots, n \text{ diagonal dominante por linhas}$$

$$|a_{jj}| \geq \sum_{\substack{i=1 \\ i\neq j}}^{n} |a_{ij}| \quad j = 1, 2, \ldots, n \text{ diagonal dominante por colunas}$$

com desigualdade estrita $''>''$ válida para pelo menos um índice.

Definição (Matriz de Diagonal Estritamente Dominante): Se nas desigualdades acima o sinal $\geq$ for substituído por $''>''$ (ou seja, desigualdade estrita sempre) então dizemos que a matriz é de diagonal estritamente dominante por linhas (ou por colunas).

Exemplo Considere a matriz

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

Tem-se: $|4| > |-1|+|0|$, $|4| > |-1|+|1|$ e $|4| = |2|+|2|$, donde vemos que esta matriz é de diagonal dominante por linhas, mas não é de diagonal estritamente dominante por linhas. Por outro lado, A é de diagonal estritamente dominante por colunas: $|4| > |-1|+|2|$, $|4| > |-1|+|-2|$ e $|4| > |1|+|0|$

Definição (Matriz Definida Positiva): Seja $A_{n\times n}$ uma matriz simétrica. Se $\mathbf{x}^T A\mathbf{x} > 0$ qualquer que seja o vector não nulo $\mathbf{x} = [x_1, x_2, ..., x_n]^T \in \mathbf{R^n}$, dizemos que A é uma matriz definida positiva.

Na prática é mais fácil verificar o seguinte.

Teorema: Uma matriz $A_{n\times n}$ simétrica é definida positiva se e só se a submatriz A_k, constituída pelas k primeiras linhas e k primeiras colunas de A, verifica:

$$det(A_k) > 0, \quad k = 1, 2, ..., n.$$

Exemplo Para a matriz

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

tem-se

$$A_1 = [2], \quad A_2 = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}, \quad A_3 = A$$

e, sendo A simétrica, com $det(A_1) = 2, \quad det(A_2) = 3 > 0, \quad det(A_3) = det(A) = 4 > 0$, então A é definida positiva.

Observações

obs1: Se A é simétrica e os valores próprios de A são positivos então A é definida positiva.

obs2: Se A é definida positiva então os elementos da diagonal são positivos.

Teorema: Seja A uma matriz de um dos dois tipos seguintes: i) simétrica definida positiva ou ii) de diagonal estritamente dominante por linhas ou por colunas. Então A é não singular e, além disso, o método de Eliminação de Gauss (também o método LU) pode ser aplicado ao sistema linear $A\mathbf{x} = \mathbf{b}$ sem troca de linhas. Ou seja, o processo é estável em relação à propagação dos erros de arredondamento, não sendo preciso usar nenhuma técnica de pesquisa de pivot.

Tem-se ainda

Teorema: Se A é uma matriz simétrica definida positiva então o método de Cholesky pode ser aplicado ao sistema linear $A\mathbf{x} = \mathbf{b}$. (existe uma única matriz triangular inferior L tal que $A = LL^T$).

Normas Vectoriais e Matriciais

Normas de Vectores

Uma norma em $\mathbf{R^n}$ é uma função denotada por $\|\cdot\|$ com valores em $\mathbf{R}$, satisfazendo:

N1. $\|\mathbf{x}\| \geq 0,\ \forall\ \mathbf{x}\ \in\ \mathbf{R^n}$

N2. $\|\mathbf{x}\| = 0 \Longleftrightarrow \mathbf{x} = 0$

N3. $\|\alpha\mathbf{x}\| = |\alpha|\|\mathbf{x}\|,\ \forall\ \alpha\ \in\ \mathbf{R},\ \forall\ \mathbf{x}\ \in\ \mathbf{R^n}$

N4. $\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{x}\|,\ \forall\ \mathbf{x},\mathbf{y}\ \in\ \mathbf{R^n}$

Em $\mathbf{R^n}$ usaremos as seguints normas:

I. $\|\mathbf{x}\|_\infty = \max\limits_{1\leq i\leq n}\{|x_i|\}$

II. $\|\mathbf{x}\|_1 = \sum\limits_{i=1}^{n}|x_i|$

III. $\|\mathbf{x}\|_2 = \left(\sum\limits_{i=1}^{n}|x_i|^2\right)^{1/2}$

Exemplo: seja $\mathbf{x} = [-1\ 2\ 4]^T$. Então:

$\|\mathbf{x}\|_\infty = max\ \{|-1|, |2|, |4|\} = 4$

$\|\mathbf{x}\|_1 = \sum\limits_{i=1}^{3}|x_i| = |-1| + |2| + |4| = 7$

$\|\mathbf{x}\|_2 = \left(\sum\limits_{i=1}^{3}|x_i|^2\right)^{1/2} = (|-1|^2 + |2|^2 + |4|^2)^{1/2} = \sqrt{21}$

Normas de Matrizes

Uma norma de matriz é uma função definida no conjunto das matrizes quadradas reais com valores em $\mathbf{R}$ ($\|\cdot\| : \mathbf{R^n} \times \mathbf{R^n} \longrightarrow \mathbf{R}$) satisfazendo:

M1. $\|A\| \geq 0, \ \forall \ A$

M2. $\|A\| = 0 \Longleftrightarrow A = 0$

M3. $\|\alpha A\| = |\alpha|\|A\|, \ \forall \ \alpha \ \in \ \mathbf{R}, \ \forall \ A$

M4. $\|A + B\| \leq \|A\| + \|B\|, \ \forall \ A$ e B

M5. $\|AB\| \leq \|A\|\|B\|, \ \forall \ A$ e B

Embora uma norma de matriz possa ser definida de várias maneiras, vamos considerar somente normas provenientes de normas de vectores.

Dada uma norma vectorial $\|\cdot\|$, a aplicação

$$\|A\| = \max_{\mathbf{x}\neq\mathbf{0}} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|} \qquad \{ \text{ norma natural ou induzida } \} \qquad (1)$$

satisfaz as condições **M1**-**M5** para normas de matrizes. Como vemos, a norma de matriz definida por (1) depende da norma de vector adoptada. Vejamos alguns casos particulares:

1. Consideremos a norma vectorial $\|\mathbf{x}\|_\infty = \max_{1\leq i\leq n}\{|x_i|\}$. Então tem-se:

   $\|A\|_\infty = \max_{\mathbf{x}\neq\mathbf{0}} \frac{\|A\mathbf{x}\|_\infty}{\|\mathbf{x}\|_\infty}$. Nesse caso, pode-se provar que:

$$\|A\|_\infty = \max_{1\leq i\leq n} \left\{ \sum_{j=1}^{n} |a_{ij}| \right\} \qquad \{ \text{ conhecida como "norma das linhas" } \}$$

Exemplo Se $A = \begin{bmatrix} 1 & 3 & -5 \\ 1 & 4 & 2 \\ -6 & 5 & 7 \end{bmatrix}$, $\|A\|_\infty = max \left\{ \begin{array}{l} |1| + |3| + |-5| = 9, \\ |1| + |4| + |2| = 7, \\ |-6| + |5| + |7| = 18 \end{array} \right\} = 18$

2. Para a norma vectorial $\|\mathbf{x}\|_1 = \sum_{i=1}^{n}\{|x_i|\}$, tem-se:

$$\|A\|_1 = \max_{\mathbf{x}\neq\mathbf{0}} \frac{\|A\mathbf{x}\|_1}{\|\mathbf{x}\|_1}. \text{ e pode-se provar que:}$$

$$\|A\|_1 = \max_{1\leq j\leq n}\left\{\sum_{i=1}^{n}|a_{ij}|\right\} \qquad \{ \text{ conhecida como "norma das colunas" } \}$$

Exemplo

Para a matriz acima, tem-se: $\|A\|_1 = max\left\{\begin{array}{l} |1|+|1|+|-6| = 8, \\ |3|+|4|+|5| = 12, \\ |-5|+|2|+|7| = 14 \end{array}\right\} = 14$

3. Para definirmos a norma $\|A\|_2$ precisamos de introduzir o conceito de raio espectral. Começamos por rever o seguinte.

Definição: Os valores próprios de uma matriz $A_{n\times n}$ são as raízes da equação

$$det(A-\lambda I) = 0$$

Exemplo Determine os valores próprios da matriz $A = \begin{bmatrix} 1 & 0 & 1 \\ 2 & 2 & 1 \\ -1 & 0 & 0 \end{bmatrix}$

Solução:

$$A-\lambda I = \begin{bmatrix} 1-\lambda & 0 & 1 \\ 2 & 2-\lambda & 1 \\ -1 & 0 & -\lambda \end{bmatrix}$$

Então $det(A-\lambda I) = 0 \iff (1-\lambda)\begin{vmatrix} 2-\lambda & 1 \\ 0 & -\lambda \end{vmatrix} + \begin{vmatrix} 2 & 2-\lambda \\ -1 & 0 \end{vmatrix} = 0 \implies$

$(1-\lambda)\left[(2-\lambda)(-\lambda)\right] + (2-\lambda) = 0$

$\implies (2-\lambda)\left[-\lambda(1-\lambda)+1\right] = 0 \implies (2-\lambda)\left[\lambda^2-\lambda+1\right] = 0$

$\implies \lambda_1 = 2,\ \lambda_2 = \dfrac{1+\sqrt{3}i}{2},\ \lambda_3 = \dfrac{1-\sqrt{3}i}{2}$

Definição: Seja A uma matriz quadrada. O raio espectral de A, denotado por $\rho(A)$, é definido por:

$$\rho(A) = \max_{1\leq i\leq m}|\lambda_i| \text{ onde } \lambda_i \text{ é valor próprio de } A$$

Exemplo Para a matriz acima tem-se:

$$\rho(A) = \max_{1\leq i\leq 3}|\lambda_i| = max\{2,\ 1,\ 1\} = 2$$

À norma vectorial $\|\cdot\|_2$ está associada a norma matricial

$$\|A\|_2 = \left[\rho(AA^T)\right]^{1/2}$$

onde $\rho(AA^T)$ é o raio espectral da matriz produto de A por A^T.

Exemplo No caso da matriz anterior, pode verificar-se que os valores próprios de AA^T são: $1, (11-\sqrt{105})/2, (11+\sqrt{105})/2$. Então $\rho(AA^T) = (11+\sqrt{105})/2 \simeq 10.62348$ e $\|A\|_2 \simeq 3.25937$. Note que também se tem $\|A\|_\infty = 5$ e $\|A\|_1 = 4$. Por outro lado, obtivémos $\rho(A) = 2$, que é um valor inferior a qualquer das três normas obtidas.

Tem-se a seguinte propriedade do raio espectral.

Teorema: i) Qualquer que seja a norma matricial $\|\cdot\|$, induzida por uma norma vectorial, tem-se, para qualquer matriz A:

$$\rho(A) \leq \|A\|$$

ii) Dada uma matriz A qualquer, para todo o $\epsilon > 0$, existe uma norma induzida $\|\cdot\|$, tal que:

$$\|A\| \leq \rho(A) + \epsilon$$

Ou seja, o raio espectral é o ínfimo de todas as normas induzidas (entre os valores $\rho(A)$ e $\rho(A) + \epsilon$ existe sempre uma norma de A)

Condicionamento de Sistemas Lineares

Notação: Seja $\mathbf{x}$ um vector e $\bar{\mathbf{x}}$ uma aproximação para $\mathbf{x}$:

$\mathbf{e_x} = \mathbf{x} - \bar{\mathbf{x}}$ é chamado erro de $\bar{\mathbf{x}}$ (um vector)

$\|\mathbf{e}\| = \|\mathbf{x} - \bar{\mathbf{x}}\|$ é o erro absoluto de $\bar{\mathbf{x}}$

$\|\delta_{\mathbf{x}}\| = \dfrac{\|\mathbf{e_x}\|}{\|\mathbf{x}\|} = \dfrac{\|\mathbf{x} - \bar{\mathbf{x}}\|}{\|\mathbf{x}\|}$, se $\mathbf{x} \neq 0$, é o erro relativo de $\bar{\mathbf{x}}$

$100\|\delta_{\mathbf{x}}\|\%$ dá a percentagem de erro em $\bar{\mathbf{x}}$

Quando se pretende resolver um sistema

$$A\mathbf{x} = \mathbf{b}, \tag{1}$$

por vezes os elementos de A ou $\mathbf{b}$ (os dados) podem ter de ser arredondados. Também ao utilizarmos o método de eliminação de Gauss, os erros resultantes dos arredondamentos efectuados conduzem a erros na solução final.

Para avaliar até que ponto a solução $\mathbf{x}$ é sensível à propagação dos erros de arredondamento, podemos pensar que o sistema original é substituído por um outro sistema $\bar{A}\bar{\mathbf{x}} = \bar{\mathbf{b}}$, onde $\bar{A}$ e $\bar{\mathbf{b}}$ são aproximações de A e $\mathbf{b}$, resultantes de pequenas "perturbações" ou "mudanças" nos elementos de A e b.

Interessa-nos saber se o sistema (1) é sensível a pequenas mudanças (perturbações) nos dados.

Dizemos, informalmente, que um sistema linear é *bem condicionado* se a pequenas "perturbações" nos dados correspondem pequenas "perturbações" na solução $\mathbf{x}$. Caso contrário, dizemos que é *mal condicionado*.

No seguinte exemplo estuda-se o efeito de uma pequena perturbação no vector $\mathbf{b}$.

Exemplo Consideremos o sistema $A\mathbf{x} = \mathbf{b}$ dado por

$$\left.\begin{array}{rcrl} 1.0001x_1 & + & 2x_2 & = 3.0001 \\ x_1 & + & 2x_2 & = 3.0 \end{array}\right\} \begin{array}{l} \text{cuja solução exacta é} \\ x_1 = 1.0 \text{ e } x_2 = 1.0 \end{array}$$

Agora, se "perturbarmos" o lado direito:

$$\begin{array}{rcrl} 1.0001x_1 & + & 2x_2 & = 3.0003 \\ x_1 & + & 2x_2 & = 3.0 \end{array}$$

este sistema tem por solução: $x_1 = 3.0$ e $x_2 = 0.0$.

Quais foram as mudanças (relativas) em b e em x ?

Calculemos o erro de $\bar{\mathbf{b}}$ e o consequente erro na solução, usando, por exemplo a norma $\|\cdot\|_\infty$.

Tem-se, para o vector $\bar{\mathbf{b}}$,

$$\mathbf{e_b} = \mathbf{b} - \bar{\mathbf{b}} = [3.0001\ 3.0]^T - [3.0003\ 3.0]^T = [-0.0002\ 0.0]^T$$

o que levou a um erro na solução

$$\mathbf{e_x} = \mathbf{x} - \bar{\mathbf{x}} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 3 \\ 0 \end{bmatrix} = \begin{bmatrix} -2 \\ 1 \end{bmatrix}.$$

Calculando os respectivos erros relativos na norma $\|\cdot\|_\infty$ temos:

$$\|\delta_{\mathbf{b}}\| = \frac{\|\mathbf{e_b}\|_\infty}{\|\mathbf{b}\|_\infty} = \frac{0.0002}{3.0001} \approx 0.000067 \approx 0.007\%$$

e

$$\|\delta_{\mathbf{x}}\| = \frac{\|\mathbf{e_x}\|_\infty}{\|\mathbf{x}\|_\infty} = \frac{2}{1} = 200\%$$

Assim vemos que uma pequena perturbação relativa em $\mathbf{b}$ conduziu a uma grande perturbação relativa na solução $\mathbf{x}$. O sistema não é bem condicionado. Trata-se dum sistema mal condicionado.

Como identificar um sistema mal condicionado ?

Definição: Seja A uma matriz não-singular. O número de condição de A é definido por:

$$cond(A) = \|A\|\ \|A^{-1}\|, \text{ onde } \|\cdot\| \text{ é uma norma natural de matrizes.}$$

Note-se que o número de condição $cond(A)$ depende da norma utilizada, mas, para as normas induzidas, é sempre maior ou igual a um.

Previsão do erro na solução do sistema

Teorema: Seja $\mathbf{x}$ a solução exacta do sistema $A\mathbf{x} = \mathbf{b}$ e $\bar{\mathbf{x}}$ a solução obtida do sistema pertubado $A\bar{\mathbf{x}} = \bar{\mathbf{b}}$. Então,

$$\frac{1}{cond(A)} \frac{\|\mathbf{b} - \bar{\mathbf{b}}\|}{\|\mathbf{b}\|} \leq \frac{\|\mathbf{x} - \bar{\mathbf{x}}\|}{\|\mathbf{x}\|} \leq cond(A) \frac{\|\mathbf{b} - \bar{\mathbf{b}}\|}{\|\mathbf{b}\|}$$

ou

$$\frac{1}{cond(A)}\|\delta_{\mathbf{b}}\| \leq \|\delta_{\mathbf{x}}\| \leq cond(A)\ \|\delta_{\mathbf{b}}\|$$

Prova: de $A\mathbf{x} = \mathbf{b}$ e $A\bar{\mathbf{x}} = \bar{\mathbf{b}}$ temos:

$$A\mathbf{x} - A\bar{\mathbf{x}} = \mathbf{b} - \bar{\mathbf{b}} \implies A(\mathbf{x} - \bar{\mathbf{x}}) = \mathbf{b} - \bar{\mathbf{b}} \implies \mathbf{x} - \bar{\mathbf{x}} = A^{-1}(\mathbf{b} - \bar{\mathbf{b}}) \quad (1)$$

$$\implies \|\mathbf{x} - \bar{\mathbf{x}}\| = \|A^{-1}(\mathbf{b} - \bar{\mathbf{b}})\| \leq \|A^{-1}\|\ \|\mathbf{b} - \bar{\mathbf{b}}\|$$

$$\implies \|\mathbf{x} - \bar{\mathbf{x}}\| \leq \|A^{-1}\|\ \|\mathbf{b} - \bar{\mathbf{b}}\| \quad (2)$$

Por outro lado, $A\mathbf{x} = \mathbf{b} \implies \|\mathbf{b}\| = \|A\mathbf{x}\| \leq \|A\|\ \|\mathbf{x}\|$ donde obtém-se

$$\frac{1}{\|\mathbf{x}\|} \leq \frac{\|A\|}{\|\mathbf{b}\|} \quad (3)$$

Multiplicando (2) e (3) tem-se:

$$\frac{\|\mathbf{x} - \bar{\mathbf{x}}\|}{\|\mathbf{x}\|} \leq \|A^{-1}\|\ \|A\|\frac{\|\mathbf{b} - \bar{\mathbf{b}}\|}{\|\mathbf{b}\|} \quad (4)$$

Agora, de (1) tem-se que:

$$\|\mathbf{b} - \bar{\mathbf{b}}\| = \|A(\mathbf{x} - \bar{\mathbf{x}})\| \leq \|A\|\ \|\mathbf{x} - \bar{\mathbf{x}}\| \implies \|\mathbf{x} - \bar{\mathbf{x}}\| \geq \frac{\|\mathbf{b} - \bar{\mathbf{b}}\|}{\|A\|} \quad (5)$$

e de $A\mathbf{x} = \mathbf{b}$ vem:

$$\mathbf{x} = A^{-1}\mathbf{b} \implies \|\mathbf{x}\| = \|A^{-1}\mathbf{b}\| \leq \|A^{-1}\|\ \|\mathbf{b}\|$$

Obtém-se: $\frac{1}{\|\mathbf{x}\|} \geq \frac{1}{\|A^{-1}\|\ \|\mathbf{b}\|}$ (6)

e, multiplicando (5) por (6), vem:

$$\frac{1}{\|A\|\ \|A^{-1}\|}\frac{\|\mathbf{b} - \bar{\mathbf{b}}\|}{\|\mathbf{b}\|} \leq \frac{\|\mathbf{x} - \bar{\mathbf{x}}\|}{\|\mathbf{x}\|} \quad (7)$$

Finalmente, das equações (4) e (7) resulta:

$$\frac{1}{\|A\|\ \|A^{-1}\|}\frac{\|\mathbf{b} - \bar{\mathbf{b}}\|}{\|\mathbf{b}\|} \leq \frac{\|\mathbf{x} - \bar{\mathbf{x}}\|}{\|\mathbf{x}\|} \leq \|A^{-1}\|\ \|A\|\frac{\|\mathbf{b} - \bar{\mathbf{b}}\|}{\|\mathbf{b}\|}$$

ou

$$\frac{1}{cond(A)}\|\delta_{\mathbf{b}}\| \leq \|\delta_{\mathbf{x}}\| \leq cond(A)\|\delta_{\mathbf{b}}\|$$

Observações: Comecemos por notar que o resultado do teorema anterior compara o erro relativo de $\bar{\mathbf{b}}$ com o erro relativo de $\bar{\mathbf{x}}$, sendo independente do método utilizado para resolver o sistema.

Examinemos a desigualdade que nos dá uma majoração para o erro:

$$\|\delta_{\mathbf{x}}\| \leq cond(A)\ \|\delta_{\mathbf{b}}\|$$

1. Se $cond(A) \approx 1$ (*pequeno*) então, se $\|\delta_{\mathbf{b}}\|$ for pequeno, vem que $\|\delta_{\mathbf{x}}\|$ tambem é pequeno. Neste caso, o sistema é um *sistema bem condicionado*!

2. Se $cond(A) >> 1$ (*grande*) então $\|\delta_{\mathbf{b}}\|$ pequeno $\not\Longrightarrow \|\delta_{\mathbf{x}}\|$ pequeno. Ou seja, pode haver situações em que resulte um $\delta_{\mathbf{x}}$ pequeno (ver exemplo abaixo) e outras em que o erro relativo de $\bar{\mathbf{x}}$ possa ser muito maior do que o de $\bar{\mathbf{b}}$. Na verdade é garantido isso acontecer para certas escolhas de $\mathbf{b}$ e $\bar{\mathbf{b}}$, só que na prática não se sabe quais são essas escolhas. Um número $cond(A)$ muito grande traduz-se numa grande sensibilidade do sistema a pequenas mudanças relativas em $\mathbf{b}$, ou, por outras palavras, o sistema é *mal condicionado*.

Exemplo Para o exemplo anterior temos:

$$A = \begin{bmatrix} 1.0001 & 2 \\ 1 & 2 \end{bmatrix}, \quad A^{-1} = \begin{bmatrix} 10000 & -10000 \\ -5000 & 5000.5 \end{bmatrix}$$

donde vemos que: $\|A\|_\infty = 3.0001$ e $\|A^{-1}\|_\infty = 20000$. Portando,

$$cond(A)_\infty = 60002 \ >> 1$$

Previsão:

Usemos a fórmula de majoração, concretizada para a norma $\|\cdot\|_\infty$:

$$\|\delta_{\mathbf{x}}\|_\infty \leq cond_\infty(A) \ \|\delta_{\mathbf{b}}\|_\infty$$

$$\frac{\|\mathbf{x} - \bar{\mathbf{x}}\|_\infty}{\|\mathbf{x}\|_\infty} \leq 60002 \frac{\|\mathbf{b} - \bar{\mathbf{b}}\|}{\|\mathbf{b}\|} = 4, \quad \text{ou seja} \ \ 400\%$$

Pode concluir-se para este exemplo: quando o erro relativo em $\bar{\mathbf{b}}$ é 0.0067, o erro relativo na solução pode ir até 400%. Na verdade, já tínhamos calculado $\delta_{\mathbf{x}} = 2$ (200%), o que está de acordo com a previsão (que dá um majorante para o erro).

Para melhor ilustrar a incerteza inerente ao mau condicionamento de sistemas, incluímos ainda o seguinte exemplo.

Consideremos de novo o mesmo sistema, mas sujeito a uma perturbação em $\mathbf{b}$ diferente da considerada. Assim, seja o sistema $A\bar{\mathbf{x}} = \bar{\mathbf{b}}$ dado por:

$$\begin{array}{rcll} 1.0001\bar{x}_1 & + & 2\bar{x}_2 & = 3.0011 \\ \bar{x}_1 & + & 2\bar{x}_2 & = 3.001 \end{array}$$

Este sistema tem por solução: $\bar{x}_1 = 1.$ e $\bar{x}_2 = 1.0005$.

É fácil de verificar que $\|\delta_{\mathbf{b}}\| \simeq 0.33\ 10^{-5}\%$ e $\|\delta_{\mathbf{x}}\| = 0.5\ 10^{-5}\%$. Ou seja, agora a uma pequena perturbação relativa em $\mathbf{b}$ correspondeu também uma pequena perturbação relativa em $\mathbf{x}$.

As observações acima sobre mudanças no vector $\mathbf{b}$ também são válidas quando há mudanças nos elementos da matriz A. Em particular, se compararmos as soluções de $A\mathbf{x} = \mathbf{b}$ e $\bar{A}\bar{\mathbf{x}} = \mathbf{b}$, tem-se, no caso em que o erro relativo de $\bar{A}$ é suficientemente pequeno:

$$\frac{\|\mathbf{x} - \bar{\mathbf{x}}\|}{\|\mathbf{x}\|} \leq \frac{cond(A)}{1 - cond(A)\frac{\|A-\bar{A}\|}{\|A\|}} \frac{\|A - \bar{\mathbf{A}}\|}{\|\mathbf{A}\|} \tag{2}$$

desde que $\|A - \bar{\mathbf{A}}\| < 1/\|A^{-1}\|$

De novo pequenas mudanças em A podem levar a grandes mudanças em $\mathbf{x}$, se $cond(A)$ for grande. Em geral, o software para resolver sistemas lineares tem incorporados métodos para se estimar $cond(A)$, sem ter de se calcular A^{-1}. Quando se suspeita que a matriz é mal condicionada, pode ser usada uma técnica (método da correcção residual) para melhorar a solução aproximada obtida.

Concluímos esta secção com a seguinte observação. Como vimos, as técnicas de pesquisa de pivot são utilizadas para minorar as dificuldades resultantes da propagação dos erros de arredondamento. Contudo, no caso de sistemas mal condicionados, a sua utilização não vai em princípio melhorar grandemente os resultados, já que o mau condicionamento resulta sempre em instabilidade numérica.