

CAPÍTULO 3

SISTEMAS DE EQUAÇÕES LINEARES

3.1. INTRODUÇÃO

A análise de um sistema físico qualquer se inicia pela observação do espaço físico que o mesmo ocupa, podendo variar dinamicamente de acordo com as leis físicas que norteiam seu funcionamento. Na realidade, o espaço físico é contínuo, sendo constituído de infinitos pontos. A modelagem matemática desse sistema consiste em escrever as equações matemáticas que representam as leis físicas que regem o sistema. Conforme foi discutido nos capítulos anteriores, a grande maioria dos sistemas físicos é modelada por equações matemáticas complexas não lineares e que não apresentam soluções analíticas, i.e., para os infinitos pontos que constituem o domínio em análise.

A questão fundamental de engenharia é a concepção e o projeto de tais sistemas. Para tanto, é necessário resolver as equações matemáticas que modelam o comportamento físico do sistema, para obter sua resposta no tempo e/ou espaço. Como a solução analítica em forma fechada dessas equações não é possível de ser obtida na maior parte dos casos, busca-se uma solução aproximada.

Diante da impossibilidade matemática de se obter soluções numéricas para os infinitos pontos do domínio, aproxima-se o sistema contínuo (infinitos pontos). Por exemplo, uma equação matemática escrita para os infinitos pontos de um domínio pode representar a resposta de temperatura no tempo e/ou espaço de um sistema (aquecimento ou resfriamento) em função das condições do ambiente em que o mesmo é inserido. Esta equação é utilizada para criar um conjunto de equações matemáticas para cada ponto do sistema discreto que por sua vez aproxima o sistema contínuo.

Para ilustrar esse procedimento, considere o problema de análise estrutural de uma placa fina com orifício central que deve ser montada em um suporte mecânico, conforme mostra a Fig. 3.1. A peça deve ser montada por um processo mecânico de ajuste com interferência, o que provoca naturalmente uma deformação no material. A questão de engenharia é saber se cada ponto da peça resistirá ao esforço provocado pela montagem, i. e., se sua estrutura será preservada ou não.

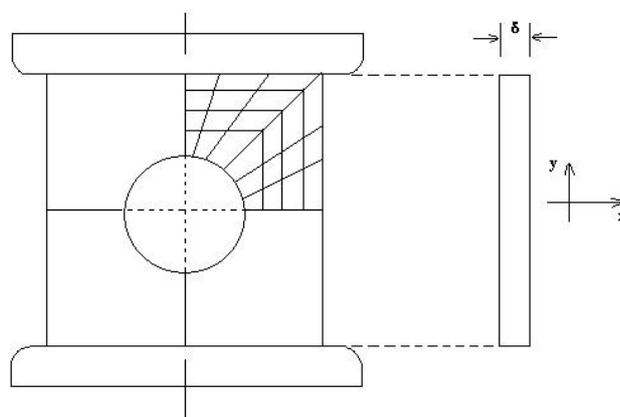


Figura 3.1 – Chapa com orifício montada em suporte via ajuste com interferência.

As tensões, que resultam no material devido ao carregamento imposto pela montagem, são calculadas a partir da teoria da Elasticidade Linear, que tem origem no princípio de conservação da quantidade de movimento [1] (Hughes, 1987). Trata-se de uma equação diferencial parcial linear que relaciona as tensões com a posição espacial (x, y) . Reconhecendo que o problema apresenta dois eixos de simetria, conforme mostra a Fig. 3.1, o domínio de análise se reduz a apenas um quadrante. Verifica-se que o quadrante contínuo (com infinitos pontos) pode ser aproximado por um conjunto discreto de pontos, também representado na Fig. 3.1. Esse processo é chamado de “discretização” e o conjunto de pontos resultante é denominado de “malha”.

A equação diferencial parcial linear escrita para as tensões é aproximada por um conjunto de equações algébricas lineares, uma para cada ponto da malha resultante. Os detalhes desse processo de transformação da equação diferencial parcial em equações algébricas não são objeto do presente tratamento, porém o leitor pode encontrar essas informações na literatura técnica básica [2,3] (Fletcher, 1991, Anderson et al., 1984). O fato importante a destacar neste momento é que o sistema de equações algébricas lineares cuja solução são os deslocamentos resultantes no material da peça, que leva ao cálculo das deformações e tensões no material, permitirá verificar se a estrutura com a configuração especificada (tipo de material, geometria) será preservada ou não, i. e., se a tensão máxima resultante é menor do que a máxima tensão que o material resiste, de acordo com suas propriedades mecânicas pré-conhecidas.

A solução do problema de engenharia assim formulado é obtida, portanto, a partir da solução de um sistema de equações lineares. Desta maneira, é necessária a existência de processos eficazes para resolver numericamente esse sistema de equações. Note que o número de equações resultante pode ser bastante grande, dependendo do número de pontos da malha da Fig. 3.1.

Como o leitor pode perceber, a obtenção da solução de sistemas de equações lineares é vital para uma vasta classe de problemas físicos e de engenharia. Os itens seguintes deste capítulo destinam-se a apresentar e discutir os métodos numéricos de solução de sistemas de equações lineares. Basicamente, esses métodos dividem-se em métodos diretos e indiretos (ou iterativos).

3.2. MÉTODOS DIRETOS

Os métodos diretos são aqueles que permitem obter o vetor solução \vec{x} do sistema

$$A\vec{x} = \vec{b} \quad (3.1)$$

através de procedimentos algébricos diretos. Assim, a solução é obtida diretamente pelo computador, sem a necessidade de realizar nenhum procedimento iterativo.

Um sistema de n equações lineares e n incógnitas é formulado matematicamente da seguinte maneira:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2 \\ \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n = b_n \end{cases} \quad (3.2)$$

com $a_{ij}, b_i \in \mathbb{R}$, para $i, j = 1, \dots, n$. Em forma matricial, o sistema pode ser escrito como

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \quad (3.3)$$

ou ainda na forma geral

$$A\vec{x} = \vec{b} \quad (3.1)$$

onde $A = [a_{ij}]$ e $b = \{b_i\}, i, j = 1, \dots, n$.

3.2.1. Noções de álgebra matricial

Uma matriz quadrada A^T é dita **transposta** de A , se

$$A = [a_{ij}] \text{ e } A^T = [a_{ji}], i, j = 1, \dots, n \quad (3.4)$$

Se $A^T = A$, então A é uma matriz **simétrica** que, por definição, é a matriz quadrada que apresenta os elementos acima da diagonal principal iguais aos elementos abaixo da mesma, i. e., $a_{ij} = a_{ji}$.

Uma operação de multiplicação entre duas matrizes A e B é definida por

$$(AB)_{ij} = \sum_{k=1}^p a_{ik} b_{kj} \quad (3.5)$$

que só é possível de ser realizada se $A(m \times p)$ e $B(p \times n)$, onde m, n, p são números inteiros positivos que definem as dimensões das matrizes. Em outras palavras, na matriz $A(m \times p)$, m define o número de linhas e p o número de colunas. Assim, para realizar uma multiplicação de duas matrizes, o número de colunas da primeira matriz deve ser igual ao número de linhas da segunda matriz. Na Eq. (3.5), a matriz resultante (AB) é, portanto, uma matriz de dimensão $(m \times n)$.

Uma operação de multiplicação entre uma matriz $A(m \times p)$ e um vetor $\vec{x}(p \times 1)$ é também definida pela Eq. (3.5). Note que o vetor \vec{x} é uma matriz de p linhas e 1 coluna. O resultado será, portanto, um vetor $\vec{b}(m \times 1)$.

No caso especial do sistema de equações lineares representado pela Eq. (3.2), a matriz A é quadrada e de dimensão $(n \times n)$. Portanto, o resultado da multiplicação de A por \vec{x} , é um vetor $\vec{b}(n \times 1)$.

Dois sistemas de equações lineares são ditos equivalentes quando têm a mesma solução. Matematicamente escreve-se:

$$A\vec{x} = \vec{b} \quad \text{e} \quad B\vec{x} = \vec{d} \quad (3.6)$$

onde A e B são duas matrizes diferentes de mesma dimensão, e b e d dois vetores diferentes também de mesma dimensão.

Para se obter um sistema equivalente $B\vec{x} = \vec{d}$, a partir de um sistema $A\vec{x} = \vec{b}$, realizam-se procedimentos chamados de **operações elementares com linhas**. São elas:

- i) Troca de duas equações

Esta operação realiza a mudança na ordem das linhas de A , \vec{x} e \vec{b} , o que significa mudar a ordem das equações do sistema original (3.2). Obviamente, este procedimento não alterará a solução. Obtém-se assim um sistema equivalente. Matematicamente, isso é representado pela multiplicação de uma permutação da matriz identidade I por A , \vec{x} e \vec{b} . Por exemplo, para um sistema 3×3 , onde se deseja trocar a linha 2 pela linha 3, faz-se

$$E_1(A, \vec{x}, \vec{b}) \quad (3.7)$$

onde $E_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$ é uma matriz identidade permutada. Assim, no caso deste exemplo, ao

multiplicar a matriz E_1 pelo sistema de equações (A, \vec{x}, \vec{b}) , o resultado é a troca da segunda equação pela terceira.

- ii) Multiplicação por um número diferente de zero

Esta operação elementar é matematicamente representada por

$$E_2(A, \vec{x}, \vec{b}) \quad (3.8)$$

Por exemplo, para um sistema 3×3 , em que se deseja multiplicar a segunda equação por λ , $E_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & 1 \end{bmatrix}$. Analogamente, uma ou mais equações do sistema poderiam ser

multiplicadas por constantes pré-definidas $\lambda_i, i = 1, \dots, n$. É claro que este procedimento não alterará a solução do sistema (A, \vec{x}, \vec{b}) .

- iii) Adição a uma equação de um múltiplo de alguma outra equação:

É possível demonstrar que um sistema equivalente é obtido ao se somar um múltiplo de uma equação a outra equação qualquer do sistema (problema 3.1). Matematicamente, o procedimento é representado por

$$E_3(A, \vec{x}, \vec{b}) \quad (3.9)$$

Por exemplo, para um sistema 3×3 , em que se deseje adicionar a primeira equação multiplicada por uma constante $\lambda \in \mathbb{R}$ à terceira equação, $E_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \lambda & 0 & 1 \end{bmatrix}$. Analogamente, o mesmo procedimento poderia ser aplicado a qualquer equação do sistema $n \times n$.

3.2.2. Matrizes inversas

Seja o produto de duas matrizes $A(n \times p)$ e $B(p \times n)$ dado por

$$AB = I \quad (3.10)$$

onde I é a matriz identidade ($n \times n$). Quando esse resultado ocorre, diz-se que B é a matriz inversa direita de A , e A é a matriz inversa esquerda de B . Note que a seguinte situação pode se verificar:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ \alpha & \beta \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \forall \alpha, \beta \quad (3.11)$$

Portanto, a segunda matriz não é única.

Se as matrizes forem quadradas, a situação é mais simples. As inversas direita e esquerda são coincidentes e únicas. Assim:

$$AA^{-1} = A^{-1}A = I \quad (3.12)$$

Uma matriz elementar é aquela que é obtida através de uma operação elementar sobre a matriz identidade. Uma matriz inversa pode ser obtida através de uma seqüência de operações elementares conforme se segue:

$$\underbrace{E_n \dots E_4 E_3 E_2 E_1}_{A^{-1}} A = I \quad \text{e} \quad E_n \dots E_4 E_3 E_2 E_1 I = A^{-1} \quad (3.13)$$

Exemplo 3.1) Usando operações elementares com linhas, determine a matriz inversa A^{-1} , sabendo que:

$$A = \begin{bmatrix} 2 & 4 & 3 & 2 \\ 3 & 6 & 5 & 2 \\ 2 & 5 & 2 & -3 \\ 4 & 5 & 14 & 14 \end{bmatrix}$$

Solução

Utiliza-se a Eq. (3.13), partindo de:

$$\begin{bmatrix} 2 & 4 & 3 & 2 & : & 1 & 0 & 0 & 0 \\ 3 & 6 & 5 & 2 & : & 0 & 1 & 0 & 0 \\ 2 & 5 & 2 & -3 & : & 0 & 0 & 1 & 0 \\ 4 & 5 & 14 & 14 & : & 0 & 0 & 0 & 1 \end{bmatrix}$$

Divide-se a primeira linha por 2, e a seguir multiplica-se a mesma linha por 3, 2 e 4, subtraindo-a da 2ª, 3ª e 4ª linhas respectivamente:

$$\begin{bmatrix} 1 & 2 & 3/2 & 1 & : & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & -1 & : & -3/2 & 1 & 0 & 0 \\ 0 & 1 & -1 & -5 & : & -1 & 0 & 1 & 0 \\ 0 & -3 & 8 & 10 & : & -2 & 0 & 0 & 1 \end{bmatrix}$$

Troca-se a 3ª linha pela 2ª linha:

$$\begin{bmatrix} 1 & 2 & 3/2 & 1 & : & 1/2 & 0 & 0 & 0 \\ 0 & 1 & -1 & -5 & : & -1 & 0 & 1 & 0 \\ 0 & 0 & 1/2 & -1 & : & -3/2 & 1 & 0 & 0 \\ 0 & -3 & 8 & 10 & : & -2 & 0 & 0 & 1 \end{bmatrix}$$

Multiplica-se por 2 a 3ª linha; soma-se a 4ª linha à 2ª linha multiplicada por 3, e subtrai-se da 1ª linha, a 2ª linha multiplicada por 2:

$$\begin{bmatrix} 1 & 0 & 7/2 & 11 & : & 5/2 & 0 & -2 & 0 \\ 0 & 1 & -1 & -5 & : & -1 & 0 & 1 & 0 \\ 0 & 0 & 1 & -2 & : & -3 & 2 & 0 & 0 \\ 0 & 0 & 5 & -5 & : & -5 & 0 & 3 & 1 \end{bmatrix}$$

Subtrai-se a 3ª linha multiplicada por 7/2 da 1ª linha; soma-se a 3ª linha à 2ª linha, e subtrai-se da 4ª linha a 3ª linha multiplicada por 5:

$$\begin{bmatrix} 1 & 0 & 0 & 18 & : & 13 & -7 & -2 & 0 \\ 0 & 1 & 0 & -7 & : & -4 & 2 & 1 & 0 \\ 0 & 0 & 1 & -2 & : & -3 & 2 & 0 & 0 \\ 0 & 0 & 0 & 5 & : & 10 & -10 & 3 & 1 \end{bmatrix}$$

Divide-se a 4ª linha por 5:

$$\begin{bmatrix} 1 & 0 & 0 & 18 & : & 13 & -7 & -2 & 0 \\ 0 & 1 & 0 & -7 & : & -4 & 2 & 1 & 0 \\ 0 & 0 & 1 & -2 & : & -3 & 2 & 0 & 0 \\ 0 & 0 & 0 & 1 & : & 2 & -2 & 3/5 & 1/5 \end{bmatrix}$$

Multiplica-se a 4ª linha por -2, -7, e 18, e subtrai-se da 3ª, 2ª e 1ª linhas, respectivamente. O resultado final é:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & -23 & 29 & -64/5 & -18/5 \\ 0 & 1 & 0 & 0 & 10 & -12 & 26/5 & 7/5 \\ 0 & 0 & 1 & 0 & 1 & -2 & 6/5 & 2/5 \\ 0 & 0 & 0 & 1 & 2 & -2 & 3/5 & 1/5 \end{bmatrix}$$

Portanto, a matriz inversa é:

$$A^{-1} = \begin{bmatrix} -23 & 29 & -64/5 & -18/5 \\ 10 & -12 & 26/5 & 7/5 \\ 1 & -2 & 6/5 & 2/5 \\ 2 & -2 & 3/5 & 1/5 \end{bmatrix}$$

3.2.3. Sistemas de fácil solução

No problema enunciado por:

$$A\vec{x} = \vec{b} \quad (3.1)$$

quando A é uma matriz do tipo doagonal, D, triangular inferior, L, ou triangular superior, U, diz-se que o sistema de equações lineares é de fácil solução. A razão é que para obter a solução \vec{x} nesses sistemas, não é necessário inverter a matriz A. Observe que a solução é obtida por procedimentos simples:

i) Matriz diagonal, D:

$$D\vec{x} = \vec{b} \quad (3.14)$$

Exemplo 3.2) Dado que:

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

Determine \vec{x} .

Solução

A solução é obtida diretamente como $x_1 = \frac{1}{1} = 1$ e $x_2 = \frac{1}{2}$.

ii) Matriz triangular inferior, L:

$$L\vec{x} = \vec{b} \quad (3.15)$$

Exemplo 3.3) Dado que:

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 2 & 0 \\ 2 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}$$

Determine \bar{x} .

Solução

A solução é obtida calculando a partir da 1ª equação o valor da 1ª incógnita x_1 . A seguir, conhecendo x_1 , calcula-se x_2 com a 2ª equação, e assim sucessivamente até a última equação do sistema. Este procedimento é denominado de **substituição avançada**. No caso deste exemplo, calcula-se:

$$x_1 = 2$$

$$x_1 + 2x_2 = 1 \Rightarrow 2 + 2x_2 = 1 \Rightarrow x_2 = -\frac{1}{2}$$

$$2 \times 2 + 1 \times \left(-\frac{1}{2}\right) + x_3 = 3 \Rightarrow x_3 = -\frac{1}{2}$$

iii) Matriz triangular superior, U:

$$U\bar{x} = \vec{b} \tag{3.16}$$

Exemplo 3.4) Dado que:

$$\begin{bmatrix} 1 & 1 & 2 \\ 0 & 2 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}$$

Determine \bar{x} .

Solução

O procedimento é exatamente o inverso do utilizado para a matriz triangular inferior, no item anterior. O processo é denominado **substituição à retaguarda**. No caso deste exemplo, calcula-se:

$$x_3 = 3$$

$$2x_2 + 1 \times 3 = 3 \Rightarrow x_2 = 0$$

$$1 \times x_1 + 1 \times 0 + 2 \times 3 = 2 \Rightarrow x_1 = -4$$

3.2.4. A fatoração LU

Quando a matriz quadrada associada ao sistema de equações lineares, A , não é diagonal, triangular superior ou inferior, o sistema não é denominado de “fácil solução”. Assim, a obtenção do vetor solução \bar{x} diretamente é calculada invertendo-se a matriz A :

$$\bar{x} = A^{-1}\bar{b} \quad (3.17)$$

ou, de forma mais eficiente por outros métodos que não requeiram a inversão da matriz A que representa alto custo computacional em sistemas com grande número de equações, como pode ser observado no exemplo 3.1.

O método da fatoração LU consiste em fatorar a matriz quadrada associada ao sistema na forma:

$$A = LU \quad (3.18)$$

Quando for possível fatorar a matriz A da forma mostrada na Eq. (3.18), diz-se que a matriz A admite uma decomposição LU.

Admitindo que a matriz quadrada A possui uma decomposição LU, pode-se deduzir o processo de fatoração LU. O processo é desenvolvido como se segue.

Em notação indicial, escreve-se:

$$A = [a_{ij}] \quad (i, j = 1, \dots, n)$$

$$a_{ij} = \sum_{s=1}^n l_{is} u_{sj} = \sum_{s=1}^{\min(i,j)} l_{is} u_{sj} \quad (3.19)$$

Onde foi utilizado o fato de que $l_{is} = 0$ para $s > i$ e $u_{sj} = 0$ para $s > j$.

Fazendo $i = j = k$ e separando a última parcela do somatório, obtém-se:

$$a_{kk} = \sum_{s=1}^{k-1} l_{ks} u_{sk} + l_{kk} u_{kk} \quad (3.20)$$

Se u_{kk} ou l_{kk} (elementos da diagonal principal) for especificado, pode-se determinar o outro. Por exemplo, especificando $u_{kk} \neq 0$, tem-se:

$$l_{kk} = \left(a_{kk} - \sum_{s=1}^{k-1} l_{ks} u_{sk} \right) / u_{kk} \quad (3.21)$$

No início do processo, quando $k = 1$, o segundo termo entre parêntesis é um somatório de $s = 1$ a -1 , portanto, neste caso o somatório é nulo, por definição. Assim, quando $k = 1$, define-se o que se chama de primeiro **menor principal líder** da matriz A , que coincide com o componente a_{11} , um escalar, uma vez que a matriz tem apenas um elemento quando $k = 1$.

Conseqüentemente, define-se o k -ésimo menor principal líder da matriz $A(n \times n)$, para $1 \leq k \leq n$, como sendo a matriz a seguir:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kk} \end{bmatrix} \quad (3.22)$$

Na seqüência, a partir de $k = 1$, e do conhecimento de u_{kk} e l_{kk} , é possível determinar os outros elementos da coluna k e da matriz L . Isso é feito de forma progressiva a partir de $k = 1$ até $k = n$, conforme mostra a Fig. 3.2:

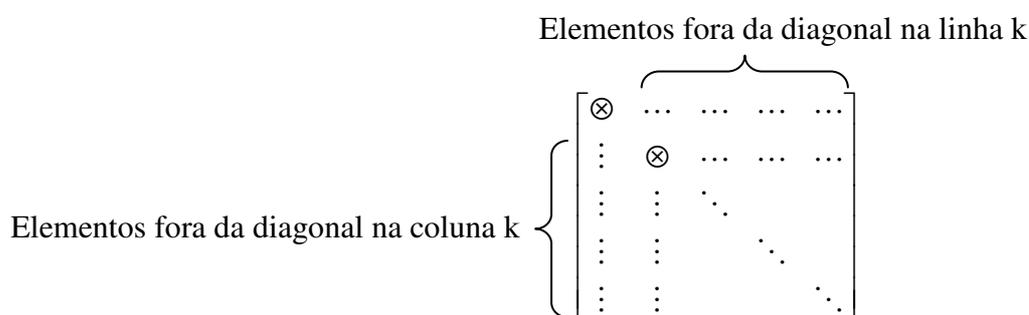


Figura 3.2 – Determinação de elementos fora da diagonal na fatoração $A = LU$.

Para determinar os elementos fora da diagonal na linha k , identificados na Fig. 3.2, utiliza-se a fórmula de multiplicação de matrizes rearranjada para obter u_{kj} , conforme se segue:

$$u_{kj} = \left(a_{kj} - \sum_{s=1}^{k-1} l_{ks} u_{sj} \right) / l_{kk} \quad (k+1 \leq j \leq n) \quad (3.23)$$

Para determinar os elementos fora da diagonal na coluna k , identificados na Fig. 3.2, utiliza-se a fórmula de multiplicação de matrizes rearranjada para obter l_{ik} , conforme se segue:

$$l_{ik} = \left(a_{ik} - \sum_{s=1}^{k-1} l_{is} u_{sk} \right) / u_{kk} \quad (k+1 \leq i \leq n) \quad (3.24)$$

A pergunta ainda a ser respondida é: que condição ou condições a matriz quadrada A deve satisfazer para possuir uma fatoração do tipo $A = LU$? A condição necessária e suficiente é que A possua uma inversa, i.e., seu determinante seja diferente de zero, o que estabelece que A seja não singular e que a Eq. (3.17) seja verdadeira. No entanto, por vezes um sistema é construído de forma que a ordem das equações leva ao aparecimento de zeros na diagonal principal, o que não significa necessariamente que a matriz não possui uma inversa. O Teorema a seguir fornece uma condição suficiente para verificar se a matriz quadrada A admite uma fatoração do tipo $A = LU$.

Teorema 3.1

Se todos os n **menores principais líderes** da matriz $A(n \times n)$ forem não singulares, então A tem infinitas decomposições do tipo $A = LU$.

Prova :

Sejam A_k , L_k e U_k os k -ésimos menores principais líderes das matrizes A , L e U , respectivamente. A hipótese é que A_1, A_2, \dots, A_n sejam não singulares.

Para realizar uma prova por indução, suponha que L_{k-1} e U_{k-1} tenham sido obtidos. Na Equação (3.19), se i e j estiverem na faixa $1, 2, \dots, k-1$, então s também está. Assim, a Eq. (3.19) estabelece que

$$A_{k-1} = L_{k-1} U_{k-1}$$

Como A_{k-1} é não singular por hipótese, L_{k-1} e U_{k-1} são também não singulares. Como L_{k-1} é não singular, pode-se resolver o sistema

$$\sum_{s=1}^{k-1} l_{is} u_{sk} = a_{ik} \quad (1 \leq i \leq k-1)$$

para u_{sk} com $1 \leq s \leq k-1$. Esses estão na k -ésima coluna de U . Como U_{k-1} é não singular, pode-se resolver o sistema

$$\sum_{s=1}^{k-1} l_{ks} u_{sj} = a_{kj} \quad (1 \leq j \leq k-1)$$

para l_{ks} com $1 \leq s \leq k-1$. Esses estão na k -ésima linha de L . A partir de

$$a_{kk} = \sum_{s=1}^k l_{ks} u_{sk} = \sum_{s=1}^{k-1} l_{ks} u_{sk} + l_{kk} u_{kk}$$

pode-se obter u_{kk} ou l_{kk} , uma vez que um deles, u_{kk} ou l_{kk} , for especificado como um valor diferente de zero. Conseqüentemente, todos os novos elementos necessários para a formação de L_k e U_k estão determinados. A indução é finalizada observando que $l_{11} u_{11} = a_{11}$ e, portanto, a partir de u_{11} ou l_{11} especificado como um valor diferente de zero, pode-se determinar $l_{11} = a_{11}/u_{11}$ ou $u_{11} = a_{11}/l_{11}$, respectivamente. Fica claro, portanto que A admite infinitas decomposições do tipo $A = LU$.

Com base na teoria apresentada, é possível escrever um algoritmo para implementação computacional do método de fatoração LU de uma matriz quadrada $A(n \times n)$.

Algoritmo computacional para a fatoração $A = LU$

Leia $n, [a_{ij}], \{b_i\}$

para $k = 1, 2, \dots, n$ **faça**

Especifique um valor diferente de zero para u_{kk} ou l_{kk} e compute o outro usando

$$l_{kk} u_{kk} = a_{kk} - \sum_{s=1}^{k-1} l_{ks} u_{sk}$$

para $j = k+1, k+2, \dots, n$ **faça**

$$u_{kj} = \left(a_{kj} - \sum_{s=1}^{k-1} l_{ks} u_{sj} \right) / l_{kk}$$

fim

para $i = k+1, k+2, \dots, n$ **faça**

$$l_{ik} = \left(a_{ik} - \sum_{s=1}^{k-1} l_{is} u_{sk} \right) / u_{kk}$$

fim

fim do laço k

imprima $[l_{ij}], [u_{ij}]$

para $i = 1, 2, \dots, n$ **faça**

$$z_i = \left(b_i - \sum_{j=1}^{i-1} l_{ij} z_j \right) / l_{ii}$$

fim

para $k = n, n-1, \dots, 1$ **faça**

$$x_k = \left(z_k - \sum_{j=k+1}^n u_{kj} x_j \right) / u_{kk}$$

fim

imprima $\{x_i\}$

Na análise apresentada, para o cálculo de $\sum_{j=a}^b F_j$, onde F_j representa o cálculo de cada parcela do somatório, é necessário adotar uma convenção. Quando $a = 1$ e $b < 1$, bem como quando $b = n$, $a > n$, $\sum_{j=a}^b F_j = 0$, por definição.

Existindo uma fatoração do tipo $A = LU$, alguns casos particulares importantes são:

a) Fatoração Doolittle's:

L é calculada tal que $l_{ii} = 1$, para $1 \leq i \leq n$.

Exemplo 3.5) Obtenha a fatoração Doolittle's $A = LU$ para $A = \begin{vmatrix} 1 & 2 \\ 2 & 1 \end{vmatrix}$.

Solução

$$A = \begin{vmatrix} 1 & 2 \\ 2 & 1 \end{vmatrix} = \overbrace{\begin{vmatrix} 1 & 0 \\ 2 & 1 \end{vmatrix}}^L \overbrace{\begin{vmatrix} 1 & 2 \\ 0 & -3 \end{vmatrix}}^U$$

b) Fatoração Crout's:

U é calculada tal que $u_{ii} = 1$, para $i \leq i \leq n$.

Exemplo 3.6) Obtenha a fatoração Crout's $A = LU$ para $A = \begin{vmatrix} 5 & 10 \\ 3 & 8 \end{vmatrix}$.

Solução

$$A = \begin{vmatrix} 5 & 10 \\ 3 & 8 \end{vmatrix} = \overbrace{\begin{vmatrix} 5 & 0 \\ 3 & 2 \end{vmatrix}}^L \overbrace{\begin{vmatrix} 1 & 2 \\ 0 & 1 \end{vmatrix}}^U$$

c) Fatoração Cholesky's:

Observe que para $A = LU$, dividindo U em duas partes pela divisão de cada linha pelo elemento "pivot" (por definição, o elemento da diagonal principal em cada linha), para matrizes simétricas obtém-se um resultado muito importante, mostrado a seguir:

$$A = LDL^T \quad (3.25)$$

Definição: uma matriz simétrica com pivots positivos é uma matriz **positivamente definida**.

Teorema 3.2

Uma matriz A positivamente definida admite uma fatoração do tipo $A = LL^T$, que é denominada fatoração Cholesky's, i.e., $U = L^T$, tal que $l_{ii} = u_{ii}$, para $1 \leq i \leq n$.

Prova :

Usando a Eq. (3.25) para uma matriz A positivamente definida: $A = LDL^T$ e todos os elementos de D são positivos. Assim:

$$A = LD^{1/2}D^{1/2}L^T = \bar{L}\bar{L}^T, \text{ onde } \bar{L} = LD^{1/2} \text{ e } \bar{L}^T = D^{1/2}L^T.$$

Exemplo 3.7) Obtenha a fatoração Cholesky's $A = LL^T$ para $A = \begin{vmatrix} 4 & 3 \\ 3 & 16 \end{vmatrix}$.

Solução

$$A = \begin{vmatrix} 4 & 3 \\ 3 & 16 \end{vmatrix} = \overbrace{\begin{vmatrix} 1 & 0 \\ 3/4 & 1 \end{vmatrix}}^L \overbrace{\begin{vmatrix} 4 & 3 \\ 0 & 55/4 \end{vmatrix}}^U = \overbrace{\begin{vmatrix} 1 & 0 \\ 3/4 & 1 \end{vmatrix}}^L \overbrace{\begin{vmatrix} 4 & 0 \\ 0 & 55/4 \end{vmatrix}}^D \overbrace{\begin{vmatrix} 1 & 3/4 \\ 0 & 1 \end{vmatrix}}^{L^T}$$

$$\begin{aligned}
 &= \begin{array}{c|c|c|c} \overbrace{}^L & \overbrace{}^{D^{1/2}} & \overbrace{}^{D^{1/2}} & \overbrace{}^{L^T} \\ \hline 1 & 0 & 2 & 0 \\ \hline 3/4 & 1 & 0 & \sqrt{55}/2 \\ \hline 0 & \sqrt{55}/2 & 0 & 1 \end{array} \\
 &= \begin{array}{c|c} \overline{L} & \overline{L^T} \\ \hline 2 & 0 \\ \hline 3/2 & \sqrt{55}/2 \end{array} \begin{array}{c|c} \overline{L} & \overline{L^T} \\ \hline 2 & 3/2 \\ \hline 0 & \sqrt{55}/2 \end{array}
 \end{aligned}$$

3.2.5. Eliminação Gaussiana

Um procedimento estruturado de se obter a fatoração Doolittle's é denominado eliminação gaussiana. O procedimento consiste em escalonar a matriz A por operações elementares sucessivas, obtendo sistemas equivalentes até que todos os elementos abaixo da diagonal principal tenham sido anulados, i.e., sejam iguais a zero.

Exemplo 3.8) Ache o vetor solução \bar{x} para o sistema de 4 equações e 4 incógnitas ($n = 4$) dado a seguir:

$$\begin{array}{c|c|c} \overbrace{}^A & \overbrace{}^{\bar{x}} & \overbrace{}^{\bar{b}} \\ \hline 6 & -2 & 2 & 4 & x_1 & 12 \\ 12 & -8 & 6 & 10 & x_2 & 34 \\ 3 & -13 & 9 & 3 & x_3 & 27 \\ -6 & 4 & 1 & -18 & x_4 & -38 \end{array} =$$

Solução

1º passo: a 1ª linha é a linha pivot. Calculam-se os novos elementos da matriz por operações elementares com linhas, fazendo

$$L_i = L_i - M_i L_p \quad (3.26)$$

onde $i = 2, \dots, n$, L a linha da matriz, M o multiplicador correspondente à linha i, e p o número da linha pivot. O resultado é

$$\begin{array}{c|c|c} \overbrace{}^{\text{Multiplicadores } M_i} & \overbrace{}^A & \overbrace{}^{\bar{x}} & \overbrace{}^{\bar{b}} \\ \hline 2 & 0 & -4 & 2 & 2 & x_1 & 10 \\ 1/2 & 0 & -12 & 8 & 1 & x_2 & 21 \\ -1 & 0 & 2 & 3 & -14 & x_3 & -26 \end{array} =$$

2º passo: a linha 2 é a linha pivot. Calculam-se os novos elementos da matriz por operações elementares com linhas, usando a Eq. (3.26), onde $i = 3, \dots, n$. O resultado é

$$\underbrace{\text{Multiplicadores } M_i}_{\substack{3 \\ -1/2}} \left| \begin{array}{cccc|c} 6 & -2 & 2 & 4 & x_1 \\ 0 & -4 & 2 & 2 & x_2 \\ 0 & 0 & 2 & -5 & x_3 \\ 0 & 0 & 4 & -13 & x_4 \end{array} \right| = \left| \begin{array}{c} 12 \\ 10 \\ -9 \\ -3 \end{array} \right|$$

3º passo: a linha 3 é a linha pivot. Calculam-se os novos elementos da matriz por operações elementares com linhas, usando a Eq. (3.26), onde $i = 4, \dots, n$. O resultado é

$$\underbrace{\text{Multiplicadores } M_i}_{2} \left| \begin{array}{cccc|c} 6 & -2 & 2 & 4 & x_1 \\ 0 & -4 & 2 & 2 & x_2 \\ 0 & 0 & 2 & -5 & x_3 \\ 0 & 0 & 0 & -3 & x_4 \end{array} \right| = \left| \begin{array}{c} 12 \\ 10 \\ -9 \\ -3 \end{array} \right|$$

Portanto, com o novo sistema equivalente, resolve-se por substituição à retaguarda. O resultado é

$$\bar{x}^T = [1, -3, -2, 1]$$

Observe que a fatoração do tipo Doolittle é obtida com os multiplicadores para a matriz L, e com a matriz U acima. O resultado é

$$\underbrace{\begin{bmatrix} 6 & -2 & 2 & 4 \\ 12 & -8 & 6 & 10 \\ 3 & -13 & 9 & 3 \\ -6 & 4 & 1 & -18 \end{bmatrix}}_A = \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 1/2 & 3 & 1 & 0 \\ -1 & -1/2 & 2 & 1 \end{bmatrix}}_L \underbrace{\begin{bmatrix} 6 & -2 & 2 & 4 \\ 0 & -4 & 2 & 2 \\ 0 & 0 & 2 & -5 \\ 0 & 0 & 0 & -3 \end{bmatrix}}_U$$

O algoritmo computacional para realizar o procedimento de eliminação gaussiana é dado por

Leia $n, [a_{ij}], \{b_i\}$

para $k = 1, 2, \dots, n-1$ **faça**

para $i = k+1, k+2, \dots, n$ **faça**

$$z = a_{ik} / a_{kk}$$

$$a_{ik} = 0$$

$$b_i = b_i - z b_k$$

para $j = k+1, k+2, \dots, n$ **faça**

$$a_{ij} = a_{ij} - z a_{kj}$$

fim

fim

fim

imprima $\{a_{ij}\}, \{b_i\}$ e calcule $\{x_i\}$ por substituição à retaguarda.

Um cuidado adicional deve ser tomado ao realizar o procedimento de eliminação gaussiana. Deve-se buscar eliminar os zeros da diagonal principal, caso existam, criando um sistema equivalente, com uma nova matriz A e um novo vetor \bar{b} , utilizando operações elementares de troca de linhas.

3.2.6. Pivoteamento

O método de eliminação gaussiana falha quando:

$$i) \begin{array}{c} \overbrace{\begin{vmatrix} 0 & 1 \\ 1 & 1 \end{vmatrix}}^A \quad \overbrace{\begin{vmatrix} x_1 \\ x_2 \end{vmatrix}}^{\bar{x}} = \overbrace{\begin{vmatrix} 1 \\ 2 \end{vmatrix}}^{\bar{b}} \end{array}$$

ou

$$ii) \begin{array}{c} \overbrace{\begin{vmatrix} \varepsilon & 1 \\ 1 & 1 \end{vmatrix}}^A \quad \overbrace{\begin{vmatrix} x_1 \\ x_2 \end{vmatrix}}^{\bar{x}} = \overbrace{\begin{vmatrix} 1 \\ 2 \end{vmatrix}}^{\bar{b}}, \text{ onde } \varepsilon \rightarrow 0. \end{array}$$

Assim, usando o algoritmo de eliminação gaussiana, obtém-se:

$$\begin{vmatrix} \varepsilon & 1 \\ 1 & 1 \end{vmatrix} \begin{vmatrix} x_1 \\ x_2 \end{vmatrix} = \begin{vmatrix} 1 \\ 2 - \varepsilon^{-1} \end{vmatrix}$$

$$x_2 = \frac{2 - \varepsilon^{-1}}{1 - \varepsilon^{-1}} \cong 1$$

$$x_1 = (1 - x_2)\varepsilon^{-1} \cong 0$$

A solução correta obtida por substituição algébrica seria:

$$x_1 = \frac{1}{1 - \varepsilon} \cong 1$$

$$x_2 = \frac{1 - 2\varepsilon}{1 - \varepsilon} \cong 1$$

Verifica-se que, com o método de eliminação gaussiana, a solução é correta para x_2 , mas é extremamente imprecisa para x_1 .

Observe que dividindo a primeira equação por ε , obtém-se:

$$\left[\begin{array}{cc|c} \overbrace{1}^A & \overbrace{\epsilon^{-1}}^{\bar{x}} & \overbrace{1}^{\bar{b}} \\ 1 & 1 & 2 \end{array} \right]$$

Então, $a_{11} < a_{ij}$, $j \neq 1$, que é na realidade a origem do problema. Se a ordem das equações for trocada, o problema é resolvido e obtém-se a solução correta por eliminação gaussiana pelo sistema a seguir:

$$\left[\begin{array}{cc|c} \overbrace{1}^A & \overbrace{1}^{\bar{x}} & \overbrace{2}^{\bar{b}} \\ \epsilon & 1 & 1 \end{array} \right]$$

A solução para problemas desse tipo, portanto, é mudar a ordem das linhas (ou equações). Escolhem-se linhas **pivot** (p_1, p_2, \dots, p_n) ao invés de $(1, 2, \dots, n)$. Assim, ao aplicar o método de eliminação gaussiana, os múltiplos das linha p_1 são subtraídos das linhas p_2, p_3, \dots, p_n , e assim por diante, com p_2, p_3, \dots, p_{n-1} atuando como linhas pivot.

3.2.7. Pivoteamento com escala

O procedimento para mudar a ordem das linhas é denominado “pivoteamento com escala”. A lógica obedece a seguinte sequência:

- 1) Primeiramente computa-se uma “escala”:

$$S_i = \max_{1 \leq j \leq n} |a_{ij}| = \max\{|a_{i1}|, |a_{i2}|, \dots, |a_{in}|\}$$

para $1 \leq i \leq n$.

- 2) A seguir seleciona-se a linha pivot, onde $\frac{|a_{i1}|}{S_i}$ é o maior valor dentre todos calculados no passo 1. O índice encontrado “i” é chamado de p_1 . Portanto, inicia-se a eliminação gaussiana com $(p_1, p_2, \dots, p_n) = (1, 2, \dots, n)$, na ordem decrescente dos valores calculados.

Algoritmo computacional para eliminação gaussiana com pivoteamento com escala:

Leia $n, \{a_{ij}\}, \{b_i\}$

para $i = 1, 2, \dots, n$ **faça**

$p_i = i$

$s_i = \max_{1 \leq j \leq n} |a_{ij}|$

fim

para $k = 1, 2, \dots, n-1$ **faça**

selecione $j \geq k$ **tal que**

$|a_{p_j,k}|/s_{p_j} \geq |a_{p_i,k}|/s_{p_i}$ **para** $i = k, k + 1, \dots, n$
 $p_k = p_j$
para $i = k + 1, k + 2, \dots, n$ **faça**
 $z = a_{p_i,k} / a_{p_k,k}$
 $a_{p_i,k} = z$
 $b_{p_i} = b_{p_i} - z b_{p_k}$
para $j = k + 1, k + 2, \dots, n$ **faça**
 $a_{p_i,j} = a_{p_i,j} - z a_{p_k,j}$
fim
fim
fim
imprima $\{a_{ij}\}, \{b_i\} \text{ e } \{p_i\}$
 Calcule $\{x_i\}$ por substituição à retaguarda, conforme se segue:
para $i = n, n - 1, \dots, 1$ **faça**

$$x_{p_i} = \left(b_{p_i} - \sum_{j=i+1}^n a_{p_i,j} x_j \right) / a_{p_i,i}$$

fim
imprima $\{x_{p_i}\}$

Contagem de operações

Um importante conceito em programação para o analista numérico é saber qual o esforço e tempo gasto na obtenção da solução de um problema numericamente, através do uso do computador. O custo computacional está associado a operações de multiplicação e divisão, chamadas de operações longas.

Como exemplo de como avaliar o custo computacional a partir do número total de operações longas de um algoritmo computacional, toma-se a solução de sistemas de equações lineares por um método direto. É claro que, com base nesse exemplo, um procedimento análogo poderá ser utilizado em qualquer algoritmo pelo analista numérico posteriormente.

Considerando uma matriz $A(n \times n)$ e o algoritmo de eliminação gaussiana com pivoteamento com escala, pode-se avaliar o número de operações longas realizadas na solução computacional do problema, conforme o raciocínio a seguir:

- i) A computação necessária para definir p_1 requer n operações de divisão;
- ii) Para cada $n - 1$ linhas p_1, p_2, \dots, p_n , 1 multiplicador é calculado, então um múltiplo da linha p_1 é subtraído da linha p_i , para $2 \leq i \leq n$. O multiplicador e o processo de eliminação consomem n operações por linha. Como $n - 1$ linhas são processadas dessa maneira, então são realizadas $n(n - 1)$ operações longas. Assim, o total resulta em $n(n - 1) + n = n^2$ operações longas;
- iii) O processo de fatoração consiste da repetição dos passos anteriores em matrizes seqüencialmente menores. Assim, no segundo passo, a linha p_1 e a coluna 1 não são utilizadas, i.e., o cálculo do número de operações agora envolve uma matriz

$(n-1) \times (n-1)$, e assim sucessivamente, de forma que o número total de operações longas é obtido por uma relação de recorrência representada pela série:

$$n^2 + (n-1)^2 + \dots + 3^2 + 2^2 = \frac{1}{3}n^3 + \frac{1}{2}n^2 + \frac{1}{6}n - 1 \cong \frac{1}{3}n^3 + \frac{1}{2}n^2 \quad (3.26)$$

iv) Para a atualização do vetor do lado direito \vec{b} , há $n - 1$ passos. No primeiro há $n - 1$ operações longas, no segundo $n - 2$ operações longas, e assim sucessivamente. Portanto, o total de operações longas é representado pela série:

$$(n-1) + (n-2) + \dots + 1 = \frac{1}{2}n^2 - \frac{1}{2}n \quad (3.27)$$

v) Para a substituição a retaguarda para obtenção do vetor solução \vec{x} o total de operações longas é representado pela série:

$$1 + 2 + 3 + \dots + n = \frac{1}{2}n^2 + \frac{1}{2}n \quad (3.28)$$

vi) Portanto, somando as Eqs. (3.26) a (3.28), obtém-se o número de operações total aproximado para a solução do sistema $A\vec{x} = \vec{b}$, como $\frac{1}{3}n^3 + \frac{3}{2}n^2$.

3.2.8. Sistemas com matrizes diagonalmente dominantes

O processo de eliminação gaussiana para solução do sistema $A\vec{x} = \vec{b}$ pode ser usado sem pivoteamento quando:

$$|a_{ii}| > \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}| \quad (1 \leq i \leq n) \quad (3.29)$$

Note que neste caso, a primeira parte do algoritmo de eliminação gaussiana com pivoteamento com escala não é necessária, conforme se constata analisando a Eq. (3.29) e a parte do algoritmo reproduzida a seguir:

...

para $i = 1, 2, \dots, n$ **faça**

$$p_i = i$$

$$s_i = \max_{1 \leq j \leq n} |a_{ij}|$$

fim

para $k = 1, 2, \dots, n-1$ **faça**

selecione $j \geq k$ **tal que**

$$|a_{p_j,k}| / s_{p_j} \geq |a_{p_i,k}| / s_{p_i} \quad \text{para } i = k, k+1, \dots, n$$

$$p_k = p_j$$

...

Neste caso, diz-se que a matriz A , associada do sistema $A\vec{x} = \vec{b}$, é diagonalmente dominante. Mais do que isso, é possível demonstrar que toda matriz diagonalmente dominante é não singular e tem uma fatoração LU (Problema 3.1).

3.2.9. Sistemas do tipo banda

É comum encontrar problemas aplicados em que a formulação do equacionamento do modelo matemático do sistema físico em análise leva a matrizes do tipo banda. Essas matrizes são definidas dessa forma, por apresentarem elementos não nulos apenas na diagonal principal e em algumas outras diagonais. Nesses casos, pode ser conveniente utilizar algoritmos simplificados, em relação ao algoritmo de eliminação gaussiana com pivoteamento com escala, a fim de reduzir o número total de operações longas e ganhar tempo computacional na solução do problema. Esses sistemas podem ser do tipo tri, penta, heptadiagonal e assim por diante, conforme o número de diagonais que são não nulas, com a denominação genérica de n -diagonal, sendo n o número de diagonais não nulas, incluindo a diagonal principal.

Como exemplo de sistema n -diagonal, esta seção toma os sistemas ditos **tridiagonais**, que são aqueles em que a matriz associada A , do sistema $A\vec{x} = \vec{b}$ é dada por:

$$A = [a_{ij}], \quad a_{ij} = 0 \quad \text{para } |i - j| > 1 \quad (3.30)$$

Assim, na linha i da matriz A , somente $a_{i,i-1}$, $a_{i,i}$ e $a_{i,i+1}$ são diferentes de zero.

Como três diagonais apenas são diferentes de zero, utiliza-se um vetor para cada diagonal, ao invés de utilizar uma matriz para armazenar os coeficientes das equações do sistema $A\vec{x} = \vec{b}$.

Por simplicidade e para evitar pivoteamento com escala, assume-se neste exemplo que a matriz é diagonalmente dominante. Portanto, o sistema $A\vec{x} = \vec{b}$ é representado por:

$$\begin{bmatrix} d_1 & e_1 & \cdots & 0 \\ f_1 & d_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & e_n \\ 0 & \cdots & f_n & d_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \quad (3.31)$$

onde \vec{d} é o vetor que contém os elementos da diagonal principal, \vec{e} da diagonal superior e \vec{f} da diagonal inferior.

Ao usar eliminação gaussiana simples, processa-se o vetor \vec{b} simultaneamente. Assim, um múltiplo da linha 1 é subtraído da linha 2 a fim de anular o elemento abaixo da diagonal principal. Neste caso, d_2 e b_2 são alterados, mas e_2 não é alterado. O multiplicador adequado é f_1 / d_1 . Os cálculos para substituição são dados por:

$$d_2 = d_2 - \frac{f_1}{d_1} e_1 \quad (3.32)$$

$$b_2 = b_2 - \frac{f_1}{d_1} b_1 \quad (3.33)$$

Os passos subseqüentes são análogos até a última linha da matriz. Para a obtenção do vetor solução é realizada a substituição à retaguarda simplificada por:

$$x_n = \frac{b_n}{d_n} \quad (3.34)$$

$$x_{n-1} = (b_{n-1} - e_{n-1}x_n) / d_{n-1} \quad (3.35)$$

Algoritmo computacional para solução de sistemas tridiagonais:

Leia $n, \{b_i\}, \{d_i\}, \{e_i\}, \{f_i\}$

para $i = 2, 3, \dots, n$ **faça**

$$d_i = d_i - \frac{f_{i-1}}{d_{i-1}} e_{i-1}$$

$$b_i = b_i - \frac{f_{i-1}}{d_{i-1}} b_{i-1}$$

fim

$$x_n = \frac{b_n}{d_n}$$

para $i = n-1, n-2, \dots, 1$ **faça**

$$x_i = (b_i - e_i x_{i+1}) / d_i$$

fim

imprima $\{x_i\}$

3.3. MÉTODOS INDIRETOS OU ITERATIVOS

Um método é dito **direto** para resolver o sistema de equações lineares $A\vec{x} = \vec{b}$ quando produz uma solução \vec{x} que seria completamente precisa se não fossem os erros de arredondamento dos cálculos no computador, em um número finito de passos.

Um método **indireto** produz uma seqüência de vetores que idealmente converge para a solução, dentro de uma tolerância pré-especificada.

Neste ponto, o leitor poderia questionar: porque buscar métodos iterativos para resolver sistemas lineares, se já existem os métodos diretos que resolvem o problema? Uma resposta para esta pergunta tem seu fundamento na questão da eficiência computacional. Em problemas de grandes simulações numéricas em Engenharia, em que os sistemas de equações a resolver têm uma dimensão grande, i.e., n é grande e há um grande número de equações a resolver, é necessário que o número de operações longas seja o menor possível, a fim de que o tempo total de processamento computacional seja minimizado. Na seção anterior deste capítulo, foi visto que o número de operações total aproximado para a solução do sistema

$A\vec{x} = b$ por um método **direto** é estimado aproximadamente como $\frac{1}{3}n^3 + \frac{3}{2}n^2$, i.e., $O(n^3)$.

Assim, quanto maior o valor de n , mais o problema se agrava. Além disso, em um método **direto**, para n grande exige-se grande alocação de memória, o que não ocorre com os métodos **indiretos**, uma vez que as equações são processadas seqüencialmente em cada passo do processo iterativo para obtenção da solução.

Os métodos **indiretos**, portanto, passam a ser vantajosos quando:

- a) Há um grande número de equações a resolver, e

b) A matriz A é esparsa, i.e., contém muitos elementos nulos.

Além desses aspectos, os métodos **indiretos** são usualmente estáveis, porque reduzem os erros de arredondamento devido ao processo iterativo.

3.3.1. Fórmula geral de iteração

Um método **indireto (ou iterativo)** pode ser deduzido matematicamente de uma forma genérica, a partir da equação vetorial formuladora do problema:

$$A\bar{x} = \bar{b} \quad (3.1)$$

Para obter uma fórmula geral de iteração, define-se uma matriz auxiliar do processo $Q(n \times n)$, que deve ser prescrita, sendo inerente a cada método. Assim, parte-se da Eq. (3.1) como se segue:

$$Q\bar{x} = Q\bar{x} - A\bar{x} + \bar{b} \quad (3.36)$$

$$Q\bar{x} = (Q - A)\bar{x} + \bar{b} \quad (3.37)$$

A Equação (3.37) sugere, portanto, o seguinte processo iterativo:

$$Q\bar{x}_k = Q\bar{x}_{k-1} - A\bar{x} + \bar{b} \quad (k \geq 1) \quad (3.38)$$

A próxima questão a ser resolvida é a escolha da matriz Q . Como se deve escolhê-la? A resposta é baseada nos seguintes requisitos:

- a) A seqüência $[\bar{x}_k]$ deve ser facilmente computada, e
- b) A seqüência $[\bar{x}_k]$ deve ser convergente.

Para que $A\bar{x} = \bar{b}$ tenha solução, deve-se assumir que a matriz A é não singular. Assim, escolhendo a matriz Q , como não singular, pode-se escrever:

$$\bar{x}_k = (I - Q^{-1}A)\bar{x}_{k-1} + Q^{-1}\bar{b} = F(\bar{x}_{k-1}) \quad (3.39)$$

A solução \bar{x} satisfaz a equação a seguir:

$$\bar{x} = (I - Q^{-1}A)\bar{x} + Q^{-1}\bar{b} = F(\bar{x}) \quad (3.40)$$

Portanto, verifica-se que a Eq. (3.39) representa uma iteração de ponto fixo.

Condições de convergência

- I) Veja-se primeiro uma condição de convergência baseada na norma matricial subordinada:
 - i) Subtraindo a Eq. (3.40) da Eq. (3.39)

$$\bar{x}_k - \bar{x} = (\mathbf{I} - \mathbf{Q}^{-1}\mathbf{A})(\bar{x}_{k-1} - \bar{x}) \quad (3.41)$$

Com qualquer norma vetorial e matricial, obtém-se:

$$\begin{aligned} \|\bar{x}_k - \bar{x}\| &\leq \|\mathbf{I} - \mathbf{Q}^{-1}\mathbf{A}\| \|\bar{x}_{k-1} - \bar{x}\| \\ &\vdots \quad \quad \quad \vdots \\ \|\bar{x}_k - \bar{x}\| &\leq \|\mathbf{I} - \mathbf{Q}^{-1}\mathbf{A}\|^k \|\bar{x}_0 - \bar{x}\| \end{aligned}$$

ii) Então, se $\|\mathbf{I} - \mathbf{Q}^{-1}\mathbf{A}\| < 1$, conclui-se que:

$$\lim_{k \rightarrow \infty} \|\bar{x}_k - \bar{x}\| = 0 \quad (3.42)$$

Assim, pela Eq. (3.42), a seqüência $[\bar{x}_k]$ converge $\forall \bar{x}_0$, e um critério de convergência aceitável seria a verificação de $\|\bar{x}_k - \bar{x}_{k-1}\| < \varepsilon$, onde ε representa uma tolerância pré-especificada, i.e., um número pequeno próximo de zero.

II) Veja-se agora uma condição de convergência alternativa:

i) Parte-se de:

$$\bar{x}_k = (\mathbf{I} - \mathbf{Q}^{-1}\mathbf{A})\bar{x}_{k-1} + \mathbf{Q}^{-1}\bar{\mathbf{b}} \quad (3.39)$$

ii) A condição de convergência para esta iteração de ponto fixo, pelo teorema do mapeamento contrativo, é que:

$$|\mathbf{F}'(\bar{x})| = \left\| \underbrace{\mathbf{I} - \mathbf{Q}^{-1}\mathbf{A}}_{\mathbf{T}} \right\| < 1 \quad (3.43)$$

Note que $\bar{x} = \sum_i \alpha_i \bar{v}_i$, onde $\alpha_i \in \Re$ e \bar{v}_i é um autovetor de T. Assim

$$\mathbf{T}^k \bar{x} = \mathbf{T}^k \sum_i \alpha_i \bar{v}_i = \sum_i \alpha_i \lambda_i^k \bar{v}_i \quad (3.44)$$

onde $\lambda_i \in \Re$ são os autovalores de T. Portanto, se:

$$|\lambda_i| < 1 \Rightarrow \lim_{k \rightarrow \infty} \mathbf{T}^k \bar{x} = 0 \quad (3.45)$$

Desta maneira, a condição para convergência é que:

$$\max_{1 \leq i \leq n} |\lambda_i| = \rho(\mathbf{I} - \mathbf{Q}^{-1}\mathbf{A}) < 1 \quad (3.46)$$

onde a terminologia $\rho(\cdot)$ indica o raio espectral da matriz $I - Q^{-1}A$, aqui definido como o maior autovalor em módulo da matriz $I - Q^{-1}A$.

O raio espectral de uma matriz B qualquer pode ser calculado por:

$$\rho(B) = \max\{|\lambda| : \det(B - \lambda I) = 0\} \quad (3.47)$$

A partir da Eq. (3.47), entende-se que $\rho(B)$ é o menor número tal que um círculo com raio $\rho(B)$ centrado na origem do plano complexo, contém todos os autovalores de B . No caso de uma matriz triangular, os autovalores são os elementos de sua diagonal, fato este que pode ser demonstrado a partir da definição estabelecida pela Eq. (3.47) e que é deixado como exercício para o leitor (Problema 3.1).

Definição

A é similar a B , se $\exists S$ não singular, tal que:

$$S^{-1}AS = B \quad (3.48)$$

Teorema 3.3

Se A for similar a B , então os autovalores de A são os mesmos de B . A demonstração deste teorema é deixada como exercício para o leitor (Problema 3.2).

Teorema 3.4

O raio espectral de uma matriz A é menor ou igual que $\|A\|$.

Prova :

$$\rho(A)\|\vec{x}\| = \|\rho(A)\vec{x}\| = \|\lambda \vec{x}\| = \|A \vec{x}\| \leq \|A\|\|\vec{x}\|$$

Portanto:

$\rho(A)\|\vec{x}\| \leq \|A\|\|\vec{x}\|$, e dividindo lado a lado por $\|\vec{x}\|$, obtém-se finalmente

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

3.3.2. Métodos clássicos

a) Método de Richardson

Talvez o método iterativo mais simples que possa ser concebido seja o método de Richardson, que utiliza a matriz identidade como matriz auxiliar do processo, $Q = I$. Portanto, a fórmula de geral de iteração dada pela Eq. (3.39) toma a forma seguinte:

$$\vec{x}_k = (I - A)\vec{x}_{k-1} + \vec{b} = \vec{x}_{k-1} + \vec{r}_{k-1} \quad (3.49)$$

onde $\vec{r}_{k-1} = \vec{b} - A\vec{x}_{k-1}$, i.e., o vetor residual obtido na iteração $k-1$.

Condição de convergência

A condição de convergência para o método de Richardson é que $\|I - A\| < 1$. Duas classes de matrizes que permitem que a iteração de Richardson seja bem sucedida são as matrizes com diagonais unitárias, diagonalmente dominantes em relação a linha ou coluna, definidas por:

$$a_{ii} = 1 > \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}| \quad (1 \leq i \leq n) \quad \text{ou} \quad a_{ii} = 1 > \sum_{\substack{i=1 \\ i \neq j}}^n |a_{ij}| \quad (1 \leq j \leq n) \quad (3.50)$$

A prova da Equação (3.50) é deixada como exercício para o leitor (Problema 3.3).

Em conclusão, o método de Richardson é algebricamente simples, mas é mais restritivo quanto às condições requeridas para convergência para a solução de $A\vec{x} = \vec{b}$ do que os métodos que utilizam matrizes auxiliares do processo $Q(n \times n)$ diferentes da matriz identidade.

Algoritmo computacional do método de Richardson:

```

Leia  $n, \{b_i\}, [a_{ij}], \{x_i\}, M$ 
para  $k = 1, \dots, M$  faça
  para  $i = 1, \dots, n$  faça
     $r_i = b_i - \sum_{j=1}^n a_{ij}x_j$ 
  fim
   $\|\vec{r}\| < \varepsilon \Rightarrow$  imprima  $\{x_i\}$ 
  para  $i = 1, \dots, n$  faça
     $x_i = x_i + r_i$ 
  fim
fim

```

b) Método de Jacobi

O método de Jacobi utiliza uma matriz diagonal como matriz auxiliar do processo $Q(n \times n)$. Os elementos dessa matriz diagonal são os mesmos da matriz $A = [a_{ij}]$. Portanto, divide-se a matriz A em partições, tal que $A = D - L - U$, com D sendo uma matriz diagonal, $-U$ estritamente triangular superior e $-L$ estritamente triangular inferior. Essa divisão é graficamente representada por:

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & & & -U \\ \vdots & & D & \vdots \\ & -L & & a_{n-1,n} \\ a_{n1} & \cdots & a_{n,n-1} & a_{n,n} \end{pmatrix} \quad (3.51)$$

Portanto, a fórmula geral de iteração dada pela Eq. (3.39) toma a forma seguinte:

$$\bar{x}_k = (I - D^{-1}A)\bar{x}_{k-1} + D^{-1}\bar{b} = D^{-1}(L + U)\bar{x}_{k-1} + D^{-1}\bar{b} \quad (3.52)$$

Condição de convergência

Na iteração de Jacobi, os componentes genéricos da matriz $Q^{-1}A$ são dados por a_{ij}/a_{ii} , cujos elementos da diagonal principal são todos iguais a 1, portanto:

$$\|I - Q^{-1}A\|_{\infty} = \max_{1 \leq i \leq n} \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}/a_{ii}| \quad (3.53)$$

Assim, verifica-se que se A for diagonalmente dominante, a iteração de Jacobi converge para a solução de $A\bar{x} = \bar{b}$ para qualquer vetor estimativo inicial. O fato de uma matriz ser diagonalmente dominante significa que:

$$|a_{ii}| > \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}| \quad (i \leq i \leq n) \quad (3.54)$$

Por inspeção direta da Eq. (3.53), conclui-se que $\|I - Q^{-1}A\|_{\infty} < 1$.

A convergência do método de Jacobi também pode ser verificada a partir do raio espectral, utilizando a Eq. (3.46). Assim, deve-se verificar se:

$$\max_{1 \leq i \leq n} |\lambda_i| = \rho(I - Q^{-1}A) = \rho(D^{-1}(L + U)) < 1 \quad (3.55)$$

Algoritmo computacional do método de Jacobi:

Leia $n, \{b_i\}, \{a_{ij}\}, \{x_i\}, M$

para $k = 1, \dots, M$ **faça**

para $i = 1, \dots, n$ **faça**

$$r_i = b_i - \sum_{j=1}^n a_{ij}x_j$$

fim

$$\|\bar{r}\| < \varepsilon \Rightarrow \text{imprima } \{x_i\}$$

para $i = 1, \dots, n$ **faça**

$$u_i = \frac{\left(b_i - \sum_{\substack{j=1 \\ j \neq i}}^n a_{ij} x_j \right)}{a_{ii}}$$

fim

para $i = 1, \dots, n$ **faça**

$$x_i = u_i$$

fim

fim

Exemplo 3.9) Obtenha a fórmula de iteração de Jacobi para o sistema de equações dado a seguir:

$$\begin{cases} x + 2y = 3 \\ 2x - 3y = -5 \end{cases}$$

Solução

Para obter a fórmula de iteração de Jacobi para o sistema dado, basta que seja explicitada uma incógnita por cada equação dada, conforme se segue:

$$\begin{cases} x_{k+1} = 3 - 2y_k \\ y_{k+1} = (5 + 2x_k)/3 \end{cases} \quad \text{para } k \geq 0$$

c) Método de Gauss-Seidl

O método de Gauss-Seidl utiliza uma matriz tridiagonal inferior como matriz auxiliar do processo $Q(n \times n)$. Os elementos dessa matriz tridiagonal inferior são os mesmos da matriz $A = [a_{ij}]$, usando a divisão da matriz A em partições apresentada na Eq. (3.51), tal que $Q = D - L$, com D sendo uma matriz diagonal, e $-L$ estritamente triangular inferior.

Portanto, a fórmula geral de iteração dada pela Eq. (3.39) toma a forma seguinte:

$$\begin{aligned} \bar{x}_k &= \left(I - (D - L)^{-1} A \right) \bar{x}_{k-1} + (D - L)^{-1} \bar{b} \\ &= \left((D - L)^{-1} (D - L) - (D - L)^{-1} A \right) \bar{x}_{k-1} + (D - L)^{-1} \bar{b} \\ &= (D - L)^{-1} U \bar{x}_{k-1} + (D - L)^{-1} \bar{b} \end{aligned} \quad (3.56)$$

Condição de convergência

Fazendo $M = (D - L)^{-1}$ e $N = U$, uma condição de convergência baseada no raio espectral é verificada por:

$$\max_{1 \leq i \leq n} |\lambda_i| = \rho(I - Q^{-1}A) = \rho(M^{-1}N) < 1 \quad (3.57)$$

Teorema 3.5

Se a matriz A é diagonalmente dominante, então o método de Gauss-Seidl converge.

Prova :

Seja um autovetor \bar{x} normalizado, tal que $\|\bar{x}\|_\infty = 1$, $\max_{1 \leq i \leq n} |x_i| = 1$.

$$(M^{-1}N) \bar{x} = \lambda \bar{x} \Rightarrow U \bar{x} = \lambda(D-L) \bar{x}$$

Rearranjando:

$$\lambda D \bar{x} = \lambda L \bar{x} + U \bar{x}$$

Assim:

$$\lambda d_{ii} x_i = \lambda \sum_{j=1}^{i-1} a_{ij} x_j + \sum_{j=i+1}^n a_{ij} x_j$$

Aplicando $\|\cdot\|_\infty$ em ambos os lados da equação e a desigualdade do triângulo, obtém-se:

$$|\lambda| |d_{ii}| |x_i| \leq |\lambda| \sum_{j=1}^{i-1} |a_{ij}| |x_j| + \sum_{j=i+1}^n |a_{ij}| |x_j|$$

Como $\max_{1 \leq i \leq n} |x_i| = 1$:

$$|\lambda| |d_{ii}| \leq |\lambda| \sum_{j=1}^{i-1} |a_{ij}| + \sum_{j=i+1}^n |a_{ij}|$$

$$|\lambda| \left(|d_{ii}| - \sum_{j=1}^{i-1} |a_{ij}| \right) \leq \sum_{j=i+1}^n |a_{ij}|$$

Como A é diagonalmente dominante:

$$|d_{ii}| = |a_{ii}| > \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}| \Rightarrow |d_{ii}| - \sum_{j=1}^{i-1} |a_{ij}| > \sum_{j=i+1}^n |a_{ij}|$$

$$|\lambda| \leq \frac{\sum_{j=i+1}^n |a_{ij}|}{|d_{ii}| - \sum_{j=1}^{i-1} |a_{ij}|} \leq 1, \quad \forall \lambda \in \mathfrak{R}$$

Assim, $\rho(M^{-1}N) < 1$.

Algoritmo computacional do método de Gauss-Seidl:

Leia $n, \{b_i\}, [a_{ij}], \{x_i\}, M$
para $k = 1, \dots, M$ **faça**
 para $i = 1, \dots, n$ **faça**

$$r_i = b_i - \sum_{j=1}^n a_{ij} x_j$$

 fim
 $\|\vec{r}\| < \varepsilon \Rightarrow$ **imprima** $\{x_i\}$
 para $i = 1, \dots, n$ **faça**

$$x_i = \frac{\left(b_i - \sum_{j=1}^{i-1} a_{ij} x_j - \sum_{j=i+1}^n a_{ij} x_j \right)}{a_{ii}}$$

 fim
fim

Exemplo 3.10) Obtenha a fórmula de iteração de Gauss-Seidl para o sistema de equações dado a seguir:

$$\begin{cases} x + 2y = 3 \\ 2x - 3y = -5 \end{cases}$$

Solução

Para obter a fórmula de iteração de Gauss-Seidl para o sistema dado, basta que seja explicitada uma incógnita por cada equação dada, utilizando-se as incógnitas calculadas pelas equações anteriores na iteração atual, conforme se segue:

$$\begin{cases} x_{k+1} = 3 - 2y_k \\ y_{k+1} = (5 + 2x_{k+1})/3 \end{cases} \quad \text{para } k \geq 0$$

d) Método de sobre relaxação sucessiva (SOR)

Como o método de Gauss-Seidl, o método SOR utiliza uma matriz tridiagonal inferior como matriz auxiliar do processo $Q(n \times n)$. A diferença é que o método SOR acrescenta um novo parâmetro ao processo, que atua como um peso, aumentando ou diminuindo a contribuição dos elementos abaixo da diagonal de Q . A construção dessa matriz tridiagonal inferior utiliza a matriz $A = [a_{ij}]$, usando a divisão da matriz A em partições apresentada na Eq. (3.51), tal que $Q = \omega^{-1}(D - \omega L)$, com D sendo uma matriz diagonal, e $-L$ estritamente triangular inferior e ω , um parâmetro a ser escolhido para o processo.

Portanto, usando a Eq. (3.51), a fórmula geral de iteração dada pela Eq. (3.39) toma a forma seguinte:

$$\begin{aligned}
\bar{x}_k &= \left(I - (\omega^{-1}(D - \omega L))^{-1} A \right) \bar{x}_{k-1} + (\omega^{-1}(D - \omega L))^{-1} \bar{b} \\
&= \left((D - \omega L)^{-1} (D - \omega L) - \omega (D - \omega L)^{-1} A \right) \bar{x}_{k-1} + \omega (D - \omega L)^{-1} \bar{b} \\
&= \left((D - \omega L - \omega(D - L - U)) \bar{x}_{k-1} + \omega \bar{b} \right) \\
&= (D - \omega L)^{-1} (\omega(U \bar{x}_{k-1} + \bar{b}) + (1 - \omega)D \bar{x}_{k-1})
\end{aligned} \tag{3.58}$$

Condição de convergência

Fazendo $M = (D - \omega L)^{-1}$ e $N = (1 - \omega)D + \omega U$, uma condição de convergência baseada no raio espectral é verificada por:

$$\max_{1 \leq i \leq n} |\lambda_i| = \rho(I - Q^{-1}A) = \rho(M^{-1}N) < 1 \tag{3.59}$$

É importante ressaltar que a escolha do parâmetro ω ainda é intrigante. Em geral, quando a iteração é convergente, observa-se que $0 < \omega < 2$ [4] (Kincaid e Cheney, 1991). No entanto, a questão de escolher um valor ótimo de ω para convergência mais rápida da iteração SOR é discutida por muitos autores [5-7] (e.g., Young, 1971; Wachpress, 1966; Golub e van Loan, 1989).

Algoritmo computacional do método SOR:

Leia $n, \{b_i\}, \{a_{ij}\}, \{x_i\}, M$

para $k = 1, \dots, M$ **faça**

para $i = 1, \dots, n$ **faça**

$$r_i = b_i - \sum_{j=1}^n a_{ij}x_j$$

fim

$$\|\bar{r}\| < \varepsilon \Rightarrow \text{imprima } \{x_i\}$$

para $i = 1, \dots, n$ **faça**

$$x_i = \frac{\omega \left(b_i - \sum_{j=1}^{i-1} a_{ij}x_j - \sum_{j=i+1}^n a_{ij}x_j \right) + (1 - \omega)a_{ii}x_i}{a_{ii}}$$

fim

fim

Exemplo 3.11) Obtenha a fórmula de iteração SOR para o sistema de equações dado a seguir:

$$\begin{cases} x + 2y = 3 \\ 2x - 3y = -5 \end{cases}$$

Solução

Para obter a fórmula de iteração SOR para o sistema dado, basta que seja explicitada uma incógnita por cada equação dada, utilizando-se as incógnitas calculadas pelas equações anteriores na iteração atual, conforme se segue:

$$\begin{cases} x_{k+1} = \omega(3 - 2y_k) + (1 - \omega)x_k \\ y_{k+1} = (\omega(5 + 2x_{k+1}) + (1 - \omega)3y_k)/3 \end{cases} \quad \text{para } k \geq 0$$

3.4. EXERCÍCIOS

3.1. Solucione os sistemas a seguir empregando-se a Eliminação de Gauss.

$$\text{a) } \begin{cases} 2x + 3y + 4z = 20 \\ x + y + z = 6 \\ 4x - y - 2z = -4 \end{cases}$$

$$\text{b) } \begin{cases} 2x + y + z = 3 \\ -x - y - z = -1 \\ 6x - 2y + 4z = 14 \end{cases}$$

$$\text{c) } \begin{cases} x + y + 2z = 27 \\ 3x - 5y + z = -9 \\ 2x + y - z = -1 \end{cases}$$

$$\text{d) } \begin{cases} 4x + 3y + z = 4 \\ 2x - 6y - z = -2 \\ x + y - z = -1/6 \end{cases}$$

3.2. Encontre a inversa das seguintes matrizes, através da decomposição LU:

$$\text{a) } A = \begin{bmatrix} 2 & 3 & 4 \\ 1 & 1 & 1 \\ 4 & -1 & -2 \end{bmatrix}$$

$$\text{b) } B = \begin{bmatrix} 4 & 8 & 3 \\ 1 & 1 & 1 \\ 2 & -1 & -1 \end{bmatrix}$$

$$\text{c) } C = \begin{bmatrix} 1 & 1 & 1 \\ 3 & -1 & 1 \\ -1 & -1 & 1 \end{bmatrix}$$

3.3. Demonstre que os autovalores de uma matriz triangular são os elementos da diagonal principal.

3.4. Demonstre que os autovalores de B são os mesmos de A, se S não singular, tal que $S^{-1}AS = B$.

3.5. Resolva o seguinte sistema de equações

$$\begin{cases} 1,05x_1 + 2,05x_2 = 5,15 \\ 1,1x_1 + 2x_2 = 5,1 \end{cases}$$

a) Gráficamente;

- b) Por Eliminação de Gauss, de modo exato;
- c) Por Eliminação de Gauss, mas com 3 algarismos significativos. Utilize aritmética de arredondamento.
- d) Substitua os valores obtidos no item (c) nas equações originais e faça a verificação dos resultados.
- e) Compare os resultados obtidos nos itens (a), (b) e (c). O que pode ser dito sobre o sistema fornecido?

3.6. O processo de discretização da equação diferencial parcial que rege a transferência de calor, em regime permanente, para uma geometria bidimensional com uma taxa constante de geração de calor resultou no seguinte sistema de equações lineares:

$$\begin{cases} 4T_1 - T_2 - T_3 = 350 \\ -T_1 + 4T_2 - T_4 = 550 \\ -T_1 + 4T_3 - T_4 = 150 \\ -T_2 - T_3 + 4T_4 = 350 \end{cases}$$

Verifique se, para a matriz apresentada, os critérios de convergência para os métodos de Jacobi, Gauss-Seidl e SOR, com $\omega = 0,5$, são verificados. Caso seja aplicar tais métodos, utilize como estimativa inicial $T_1 = T_2 = T_3 = T_4 = 250$ e realize quatro iterações com cada um dos métodos possíveis, sendo utilizados cinco algarismos significativos. Calcule, também, a norma euclidiana (norma l_2) do resíduo, ao longo das iterações.

3.7. Faça as quatro primeiras iterações dos métodos de Jacobi e de Gauss-Seidel para os seguintes sistemas lineares. Compare os resultados obtidos àqueles provenientes do método de eliminação de Gauss. Empregue em seus cálculos 4 algarismos significativos. Como estimativa inicial, empregue o vetor $x = (0 \ 0 \ 0)^T$. A cada iteração, calcule também a norma euclidiana do resíduo para o sistema.

$$\begin{cases} 2x_1 + x_2 + x_3 = 8 \\ x_1 + 3x_2 - x_3 = -6 \\ 2x_2 + 4x_3 = 18 \end{cases}$$

3.8. Deseja-se obter o polinômio interpolador de mais elevado grau que passe através de um dado conjunto de pontos. Uma possibilidade é baseada na equação geral do polinômio (outras técnicas mais eficientes de interpolação serão vistas posteriormente). Neste caso, caso se tenha três pares de pontos, o polinômio de mais elevado grau obtido é uma parábola, cuja equação geral é $y = ax^2 + bx + c$. Considere, então, os seguintes conjuntos de pares ordenados (x, y) :

- a) (-2;25); (0;5); (1;13)
- b) (-3;35); (1;3); (2;0)
- c) (-1;11); (2;-5); (5;-35)
- d) (-2;-11); (1;-5); (4;10)

Obtenha o sistema linear a ser resolvido para se determinar os coeficientes a , b e c referentes ao polinômio interpolador quadrático para cada caso. Na sequência, verifique se o método de Gauss-Seidl podem ser empregado, de modo a solucionar cada um dos sistemas originados.

Caso seja possível, empregue-o e, em sua impossibilidade, utilize um método direto, para obtenção dos coeficientes do polinômio interpolador.

Projetos:

1. Implemente um código computacional para solucionar o problema de condução de calor em uma aleta de seção reta uniforme, cujo fenômeno é governado pela seguinte equação diferencial:

$$\frac{d^2T}{dx^2} - \frac{hP}{kA_{tr}}(T - T_{\infty}) = 0$$

onde T é a temperatura local, x é posição axial, h é o coeficiente de transferência de calor por convecção, P é o perímetro da aleta, k é a condutividade térmica do material da aleta, A_{tr} é a área da seção transversal da aleta e T_{∞} é a temperatura do fluido em contato com a aleta.

Com a finalidade de solucionar a equação diferencial, emprega-se uma metodologia chamada de discretização, na qual a equação original é aproximada por meio de um conjunto de equações algébricas lineares. Uma possibilidade de discretização é feita empregando-se o método de diferenças finitas, associado a um balanço de energia ao redor de cada ponto da malha. Nesse caso, o sistema de equações algébricas apresenta o seguinte aspecto:

$$T_1 = T_b$$

$$(2kA_{tr} + hP\Delta x)T_i - kA_{tr}T_{i-1} - kA_{tr}T_{i+1} = hP\Delta xT_{\infty}, i = 2, N-1$$

$$\left(kA_{tr} + hP\frac{\Delta x}{2}\right)T_N - kA_{tr}T_{N-1} = hP\frac{\Delta x}{2}T_{\infty}$$

sendo N o número total de pontos em que a temperatura será calculada e Δx o espaçamento entre dois pontos consecutivos (considerando-se um espaçamento constante), podendo ser expresso como

$$\Delta x = \frac{L}{N-1}$$

onde L é o comprimento da aleta.

Considere o caso em que o comprimento da aleta (L) é de 0,1 m, a temperatura da base (T_b) é de 100°C, a temperatura do fluido (T_{∞}) é de 20°C, o coeficiente de convecção é de 50 W/m²K, a condutividade térmica do material da aleta (k) é de 400 W/mK, o perímetro da aleta (P) é 0,20 m, a área da seção transversal da aleta (A_{tr}) é de $2,5 \times 10^{-3}$ m² e o número de pontos (N) a ser considerado é igual a 11.

Compare as soluções numéricas da temperatura com a solução analítica:

$$\frac{T - T_{\infty}}{T_b - T} = \frac{\cosh[m(L-x)]}{\cosh(mL)}$$

sendo

$$m = \left(\frac{hP}{kA_{tr}}\right)^{1/2}$$

e x a posição do ponto, que pode ser avaliada como

$$x = (i-1)\Delta x, \quad i = 1, N$$

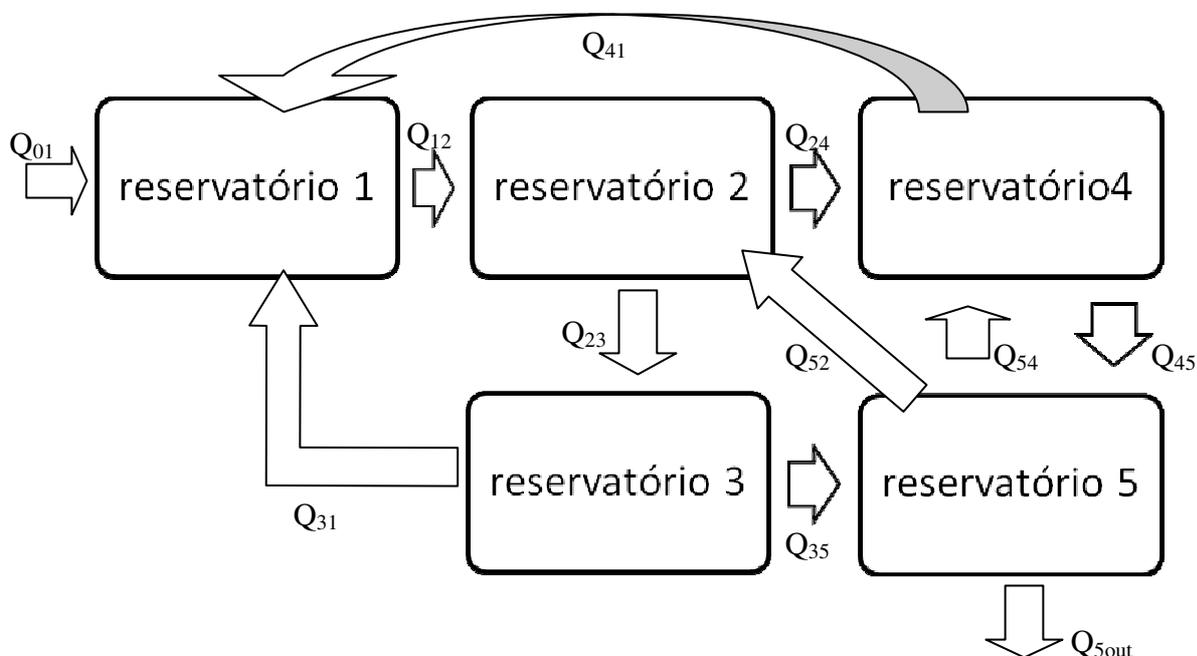
Sugestão: na solução do sistema linear de equações, empregue o algoritmo para sistemas tridiagonais.

2. Em uma treliça estaticamente determinada apenas forças axiais (tração ou compressão) estão presentes em seus membros. Para a determinação de tais forças, pode-se empregar um diagrama de corpo livre para cada um dos nós (pontos de encontro) entre os membros.

Considerando-se os eixos coordenados x e y , para cada nó existem duas direções componentes das forças e o somatório das forças atuantes em cada direção deve ser nulo. Pode-se, assim, gerar um sistema de equações lineares, que ao ser resolvido, apresenta todas as forças atuantes sobre os elementos da treliça.

Obtenha o sistema de equações lineares e implemente um código computacional para solucioná-lo, empregando um método direto e um método iterativo. Compare os resultados obtidos bem como o tempo necessário à obtenção da solução. Para o método iterativo, empregue como critério de parada uma tolerância de 10^{-9} para a norma l_2 , baseada no resíduo da estimativa inicial (pode-se empregar para tal estimativa que todos os valores sejam nulos).

3. Considere um sistema de reatores acoplados, cuja configuração é representada na figura a seguir. Deve-se obter as concentrações de um dado produto químico A, referentes a cada um dos reservatórios, sabendo-se o sistema opera em regime permanente.



São fornecidos os seguintes dados: $Q_{01} = 5$ kg/s; $Q_{31} = 0,5$ kg/s; $Q_{41} = 1$ kg/s; $Q_{12} = 6,5$ kg/s; $Q_{23} = 4$ kg/s; $Q_{52} = 1,5$ kg/s; $Q_{24} = 4$ kg/s; $Q_{54} = 2$ kg/s; $Q_{45} = 5$ kg/s; $Q_{35} = 3,5$ kg/s; $Q_{5out} = 5$ kg/s. Sabe-se, também, que a concentração do produto químico A no fluxo de entrada no reservatório 1 é de 0,10 e que na saída, a concentração do reservatório 5, a concentração observada é de 0,05.