REYOLANDO MANOEL LOPES REBELLO DA FONSECA BRASIL

ENGENHEIRO CIVIL, UNIVERSIDADE MACKENZIE (1970)

CONCEITUAÇÃO MATEMÁTICA E FÍSICA DO NÚMERO DE CONDIÇÃO DOS SISTEMAS DE EQUAÇÕES DE EQUILÍBRIO DO MÉTODO DOS DESLOCAMENTOS EM ESTRUTURAS RETICULADAS DE COMPORTAMENTO LINEAR

DISSERTAÇÃO APRESENTADA À
ESCOLA POLITÉCNICA DA UNIVERSIDADE
DE SÃO PAULO PARA OBTENÇÃO DO
TÍTULO DE MESTRE EM ENGENHARIA

ORIENTADOR: PROF. DECIO DE ZAGOTTIS

DEPARTAMENTO DE ENGENHARIA DE
ESTRUTURAS E FUNDAÇÕES

SÃO PAULO

1984

A minha família

Agradeço o apoio persisente e
paciente,
ao longo de seis anos,
do Prof. Dr. Decio Leal de Zagottis

NOTAÇÃO

a	mantissa de um número digital
a_i	(1) dígitos da mantissa de um número (2) constantes de combinação linear de vetores (3) parâmetros iniciais de um problema
a_{ij}	elemento da linha i, coluna j, de matriz $\underline{A}$
$\underline{a}$	vetor de deslocamentos nodais
$\overset{*}{\underline{a}}$	vetor de deslocamentos nodais virtuais
$\underline{a}_i^t$	vetores colunas de uma matriz $\underline{A}$
A	área da seção transversal da barra
$\underline{A}$	matriz de coeficientes de sistema de equações
$\|\| \underline{A} \|\|$	norma da matriz $\underline{A}$
$\|\| \underline{A} \|\|_\infty$	norma linha da matriz $\underline{A}$
$\|\| \underline{A} \|\|_1$	norma coluna da matriz $\underline{A}$
$\|\| \underline{A} \|\|_2$	norma espectral da matriz $\underline{A}$
$\|\| \underline{A} \|\|_E$	norma euclidiana da matriz $\underline{A}$
det A	determinante da matriz $\underline{A}$
b_i	constante da i-ésima equação de um sistema
b_{ij}	elementos de uma matriz $\underline{B}$ qualquer
$\underline{b}$	vetor de constantes de um sistema de equações
B	base de um número digital
$\underline{B}$	(1) matriz qualquer (2) matriz de *Bathe* (3)
c_{ij}	elemento de uma matriz $\underline{C}$ qualquer
cx, cy, cz	cossenos diretores do eixo de uma barra
C_i	amortecimento modal
C_{ij}	elementos da matriz de amortecimento

$\underline{C}$ (1) matriz qualquer
(2) matriz de amortecimento

d_i elemento de matriz diagonal

$\underline{D}$ (1) matriz relações tensão/deformação de um material
(2) matriz diagonal

e, e_i expoente de um número digital

e_{ij} elementos da matriz $\underline{E}$ de erros

E módulo de Young

$\underline{E}$ matriz de erros de arredondamento nos coeficientes de $\underline{A}$

f_{Ii}, f_{Ai}, f_{Ei} componentes de força de inércia, amortecimento, elástica

$\underline{f}_I$, $\underline{f}_A$, $\underline{f}_E$ vetores de força de inércia, amortecimento, elástica

$\underline{f}_m$, $\underline{f}_s$ forças de massa e de superfície

G módulo de elasticidade transversal

$h(\underline{A})$ número de condição espectral da matriz $\underline{A}$

$\underline{I}$ matriz identidade ou unitária

$I_{máx}$, $I_{mín}$, I_{tor} momento de inércia máximo, mínimo e torcional

i, j, k variáveis indiciais inteiras

K constante real ou complexa

$|K|$ módulo de constante

K_i rigidez modal

K_{ij} coeficiente de rigidez da matriz $\underline{K}$

$\underline{K}$ matriz de rigidez

ℓ_{ij} elemento de matriz triangular $\underline{L}$

L vão

$\underline{L}$ (1) operador linear
(2) matriz triangular inferior

$\hat{\underline{L}}$ matriz triangular resultante da decomposição de *Cholesky* (11) (13)

m número de linhas ou colunas de matriz

$m(\underline{A})$ número de condição **M** de $\underline{A}$, segundo *Turing* (19)

$M(\underline{A})$ maior coeficiente em módulo de $\underline{A}$

M_i massa modal

M_{ij} coeficiente da matriz de massa $\underline{M}$

$\underline{M}$ matriz de massa

n (1) número de componentes do vetor
(2) número de linhas ou colunas de matriz

$n(\underline{A})$ número de condição **N** de $\underline{A}$, segundo Turing

$\underline{N}$ funções de interpolação ou de forma

$\underline{p}(t)$ vetor carregamento variável no tempo

P_i carregamento modal

$p(\underline{A})$ número P de condição da matriz $\underline{A}$

$r(\underline{A})$ ranque da matriz $\underline{A}$

$\underline{r}$ vetor de cargas nodais

$\underline{R}$ matriz de Rutishauser

s sinal de número digital

t (1) número de dígitos significativos de número digital
(2) índice indicativo de transposição de vetor ou matriz

u_{ij} elemento de matriz triangular superior

$\underline{u}$ vetor de deslocamentos

$\tilde{\underline{u}}$ vetor de deslocamentos aproximados

$\overset{*}{\underline{u}}$ vetor de deslocamentos virtuais

$\underline{U}$ matriz triangular superior

x, x_i (1) número digital qualquer
(2) componente de vetor $\underline{x}$
(3) resposta(s) de algoritmo

$|x_i|$ módulo de componente de vetor $\underline{x}$

$\underline{x}$ (1) vetor qualquer
(2) vetor das incógnitas de sistema de equações

$\underline{x}_i$ um vetor de uma coleção de vetores

$||\underline{x}||$ norma de vetor

$||\underline{x}||_\infty$ norma cúbica de vetor

$||\underline{x}||_1$ norma octaédrica de vetor

$||\underline{x}||_2$ norma euclidiana de vetor

$\hat{\underline{x}}_i$ i-ésima forma modal de vibração

y_i (1) componente de vetor $\underline{y}$ qualquer
(2) amplitude do i-ésimo modo de vibração

$\underline{y}$ (1) vetor $\underline{y}$ qualquer
(2) vetor de amplitudes modais

$\underline{W}$ matriz de Wilson

z_i componente do vetor $\underline{z}$ qualquer

$\underline{z}$ vetor $\underline{z}$ qualquer

Δx_i perturbação em respostas de algoritmos

$\Delta \underline{x}$ perturbação em vetor solução de sistema

Δa_i perturbação em parâmetros de algoritmos

$\Delta \underline{b}$ perturbação em vetor constante de sistema

ε, ε_i erros de arredondamento

$|\varepsilon|$ módulo de erro de arredondamento

$\underline{\varepsilon}$ vetor de deformações

$\tilde{\underline{\varepsilon}}$ vetor de deformações aproximadas

$\overset{*}{\underline{\varepsilon}}$ vetor de deformações virtuais

λ, λ_i autovalor ou valor próprio da matriz

λ_1, $\lambda_{mín}$ menor valor próprio da matriz

λ_n, $\lambda_{máx}$ maior valor próprio da matriz

ρ	coeficiente de Rayleigh
$\underline{\sigma}$	vetor de tensões
$\underline{\hat{\sigma}}$	vetor de tensões aproximadas
$\sum$	somatória
$\underline{\phi_i}$	i-ésima forma modal normalizada
$\underline{\Phi}$	matriz modal
ω_i	i-ésima freqüência natural de vibração

INDICE

pág.

pág.

RESUMO

Este trabalho tem por finalidade conceituar matemática e fisicamente o *número de condição* dos sistemas de equações lineares do Método dos Deslocamentos para estruturas reticuladas de comportamento linear. Procura-se, também, indicar sua possível utilização como elemento auxiliar do engenheiro na concepção de modelos numéricos estruturais que melhor se prestem à utilização do Método.

No Primeiro Capítulo, dá-se visão panorâmica dos vários tipos de erros a que a computação numérica está sujeita, limitando o problema a ser estudado aos erros devidos ao truncamento de dados iniciais, e ao que de intrínseco na formulação dos problemas governa a propagação dos mesmos.

O Capítulo 2 recorda brevemente elementos da álgebra linear e do cálculo vetorial e matricial necessários, dando destaque às normas vetoriais e matriciais. Em seguida, dão-se noções do Método dos Elementos Finitos, Processo dos Deslocamentos, para estruturas reticuladas de comportamento linear, exemplificando com um programa para montagem da matriz de rigidez de treliças espaciais.

Os processos numéricos para solução de sistemas de equações lineares simultâneas são expostos no Capítulo 4, centrando a discussão ao algoritmo de Gauss modificado para sistemas simétricos, positivo-definidos, bandeados, para os quais se apresenta um programa.

O Quinto Capítulo lança as bases da análise de erros na computação numérica em geral, em particular visando a utilizar os computadores eletrônicos digitais de programa armazenado com representação de números com quantidade fixa de dígitos significativos. Essa análise é particularizada no capítulo seguinte para os sistemas de equações lineares, para eles derivando rigorosamente os números de condição que detectam suas tendências à amplificação de erros iniciais de truncamento.

O Capítulo mais importante é o Sétimo, em que se conceituam fisicamente os vetores próprios das matrizes de rigidez como modos de deslocamento da estrutura, similares a seus modos naturais de vibração para matriz de massa

unitária. Mostra-se, por meio de exemplos, a influência da rigidez relativa desses modos no condicionamento dos sistemas de equações resultantes.

No Capítulo 8, sugere-se a utilização do número de condição como elemento de avaliação da eficiência do contraventamento de estruturas treliçadas.

O último Capítulo fornece as conclusões e sugestões oriundas do trabalho.

Ressalta-se aqui a influência, ao longo desta Dissertação, entre outros, dos trabalhos fundamentais de *Von Neumann* (13), *Turing* (19), *Shah* (16), *Bathe* (13) e, principalmente, de *Wilkinson* (20).

ABSTRACT

This dissertation intends to assess the mathematical and physical meanings of the *condition number* of the systems of linear equations of the Displacement Method for framed structures of linear behavior. It also endeavors to show its possible application as an auxiliary element for the engineer in the formulation of numerical structural models better suited to the Method.

In the First Chapter, a comprehensive picture is given of the several sources of errors which numerical computation is prone to, limiting the problem to be studied to truncation errors in the initial data, and to what is inherent to the problem formulation which governs their propagation.

Chapter 2 briefly reviews some necessary elements of linear algebra and vector and matrix analyses, in special vector and matrix norms. The next chapter introduces the Finite Element Method, the Displacement Formulation, for framed structures of linear behavior, presenting as an example a computer program to assemble stiffness matrices of space trusses.

Numerical procedures to solve systems of linear simultaneous equations are shown in Chapter 4, chiefly discussing Gauss' elimination method for symmetric positive defined banded systems, for which a computer program is presented.

The Fifth Chapter introduces error analysis in numerical computation in general, to be carried out by electronic automatic digital computers with number representation by a fixed number of significant digits. This argument is extended, in the next Chapter, to systems of linear simultaneous equations, where condition numbers are rigorously derived to detect their tendency to amplification of initial truncation errors.

The most important Chapter is number 7, in which it is shown the physical meaning of the eigenvectors of stiffness matrices as being displacement modes of the structure. Through examples, the influence of relative stiffnesss of these modes on the conditioning of resulting systems of equations is also shown.

The last Chapter presents conclusions and suggestions originated from this paper.

The influence along the dissertation from, among others, the fundamental papers by *Von Neumann* (13), *Turing* (19), *Shah* (16), *Bathe* (3) and mainly *Wilkinson* (20), should be stressed.

1. ORIGENS E CLASSIFICAÇÃO DE ERROS EM COMPUTAÇÃO NUMÉRICA

1.1. Histórico

Os processos da análise estrutural foram totalmente reformulados nos últimos 30 anos por duas influências revolucionárias:

a) o computador eletrônico digital de programa armazenado de alta velocidade e precisão;

b) o Método dos Elementos Finitos (MEF), que trata as expressões da mecânica dos contínuos pela análise numérica, aproveitando ao máximo as possibilidades das máquinas.

Uma terceira influência, mais recente, que a curto prazo pode e deve tornar mais revolucionário o efeito das duas primeiras, é o advento do microcomputador pessoal de baixo custo, que tornará acessíveis para o engenheiro, como indivíduo, todos os benefícios citados no primeiro parágrafo.

Uma preocupação constante nesse processo de adaptação tem sido a *precisão dos resultados da análise*.

Este trabalho tem a intenção de dar uma panorâmica do problema, centrando a discussão nos erros cometidos por arredondamento e truncamento na solução das equações de equilíbrio estático de estruturas de barras de comportamento linear pelo MEF, que corresponde ao maior volume e tempo de trabalho nessas análises. Em particular, procurar-se-á interpretar a origem física da propensão a erro de certos problemas.

Erros na solução de sistemas de equações lineares simultâneas têm sido de há muito de interesse dos matemáticos. Restringindo este histórico do advento do computador eletrônico digital de programa armazenado, tem-se: o trabalho de *J. Von Neumann e H. Goldstine* (13) de 1947, deduzindo *delimitações de erros de solução* por processos "exatos" de sistemas simétricos definidos em termos de *normas matriciais*; A. M. *Turing* (19), em 1948, enfatiza que os erros reais nas eliminações de Gauss e Cholesky são usualmente menores que os anteriormente supostos, e introduz o

conceito de *número de condição*; *J. H. Wilkinson* (20) vem, em trabalho de 1960, lançar as bases definitivas das técnicas de delimitação de erros envolvidos na formulação do problema para vários algoritmos.

É claro que também os pesquisadores na área do MEF têm contribuído com outras idéias no assunto. Entre eles, *J. M. Shah* (16) correlaciona os erros com os valores e vetores próprios da matriz de rigidez; em 1968, *B. M. Irons* (7) procura estabelecer critérios de controle; *R. J. Melosh* (11), em 1971, dramatiza as desvantagens da eliminação de Cholesky, bem como os efeitos de discretização e modelagem. Começa a ficar claro que os computadores de grande porte já então utilizados tornavam os erros de arredondamento um problema quase acadêmico.

Um novo interesse na matéria parece necessário e em tempo pela já citada introdução dos microcomputadores pessoais, com toda a sua limitação inerente de memória, implicando uma menor sofisticação de programas e uso de menor precisão para possibilitar espaço ao ataque de problemas do porte do dia a dia do engenheiro.

1.2. Fontes de Erro na Computação Numérica em Geral

1.2.1. Introdução

Quando um problema, em Matemática pura ou aplicada, é resolvido por *computação numérica* (por exemplo, a solução de uma estrutura pelo MEF), desvios da solução numérica em relação à verdadeira, rigorosa, são inevitáveis. Tal solução, assim, não tem sentido sem uma medida da sua *aproximação*, expressa na forma de uma *delimitação do erro*.

A solução de um processo numérico é expressa por um ou mais números obtidos por uma seqüência finita de operações aritméticas elementares chamada *algoritmo*. A delimitação dos erros nessas condições deve levar em conta que eles são um agregado de erros primários de diversas origens.

Começa-se, assim, por uma apresentação rápida dessas fontes. Para ilustração a mais direcionada possível da exposição, procurar-se-á mostrar a importância relativa de cada tipo de erro numa análise estrutural pelo MEF.

As fontes são, em resumo:

- aproximações oriundas do modelo teórico;
- erros devidos à discretização do problema contínuo;
- erros de truncamento em algoritmos que dão aproximações finitas de formulações transcendentais e algébrica;
- erros de arredondamento na representação de números nas operações elementares;
- erros iniciais de observação e arredondamento.

Todo o trabalho nos capítulos posteriores vai-se restringir à influência apenas da última das origens citadas, reputada a que maior importância tem em nossa Área de Concentração. Buscar-se-á saber o que de intrínseco na formulação de um problema torna-o propenso a amplificar os erros iniciais.

1.2.2. Aproximações Oriundas do Modelo Teórico

A formulação matemática que se escolhe do problema real que se quer resolver envolve, via de regra, idealizações, hipóteses e simplificações. O modelo necessariamente representa uma mais ou menos explícita teoria sobre parte da realidade e não dela toda.

Tais aproximações teóricas envolvem, na Mecânica das Estruturas, as próprias hipóteses básicas e simplificações dos estudos: das tensões, deformações e do equilíbrio dinâmico; da Reologia; dos teoremas de energia; das Teorias Gerais (da Elasticidade, da Plasticidade, da Viscoelasticidade, da Resistência dos Materiais, da Estática das Estruturas, Placas, Cascas, Dinâmica, Rótulas e Charneiras Plásticas); das Teorias Específicas (Concreto Simples, Armado e Protendido; Estruturas Metálicas e Mistas, Obras de Terra etc); da Teoria da Estabilidade do Equilíbrio, e assim por diante.

Tal estudo foge completamente ao escopo deste trabalho.

É lógico que aquele que escrever e/ou utilizar um programa de cálculo estrutural deverá estar seguro da aderência de seu modelo teórico à

estrutura real, sem o que qualquer análise dos erros introduzidos ao longo do processo subseqüente não terá qualquer sentido. Tal segurança vem da sólida formação profissional adquirida na escola e na experiência prática.

1.2.3. Erros Devidos à Discretização do Modelo Contínuo

No Método dos Elementos Finitos aplicado à Mecânica das Estruturas, o contínuo constituído de barras ou superfícies ou volumes é discretizado em elementos de tamanho e forma convenientes, que se conectam em "nós" cujo conjunto de deslocamentos (translações e/ou rotações) se constitui, no "*approach*" geralmente adotado do Processo dos Deslocamentos, nas incógnitas do problema, sendo os deslocamentos internos nos elementos relacionados aos nodais por funções de interpolação apropriadas.

O campo real de deslocamentos da estrutura é, pois, aproximado por uma combinação linear das funções escolhidas, sendo os coeficientes dessa combinação o conjunto discreto dos deslocamentos nodais, a determinar pela minimização da energia potencial elástica do sistema. Reconhece-se no MEF uma versão do método mais geral de Ritz-Galerkin.

Torna-se evidente que a aproximação do resultado que se obtém, e mesmo a convergência para ele, depende em alto grau da forma como se realiza essa discretização no que respeita a:

a) número de elementos que, em princípio, quanto maior for melhor representará a estrutura;

b) tamanho e rigidez relativa dos elementos;

c) natureza das funções de interpolação utilizadas no que respeita a continuidade e conformidade.

Aqui ainda não se tentará abordar o problema em nenhuma profundidade.

1.2.4. Erros de Truncamento em Algoritmos que dão Aproximações Finitas de Formulações Transcendentais e Algébricas

A formulação matemática de um problema pode, em geral, envolver operações transcendentais (seno, logaritmo, integração, diferenciação) e definições implícitas (soluções de operações algébricas ou transcendentais, valores próprios, etc).

Para aproximá-las por cálculo numérico, essas operações devem ser substituídas por operações aritméticas elementares que o computador pode realizar diretamente, em uma seqüência linear de passos, i.e., um algoritmo (programa).

Da mesma forma, todo processo de levar a um limite, que, estritamente, é infinito, tem que ser interrompido em um estágio considerado satisfatório.

A diferença entre a solução do problema dado e a do processo numérico que o substitui, supondo-se que o computador utilizado não cometa erros de arredondamento em operações elementares, chama-se *erro de truncamento do algoritmo*. Sua delimitação é dificultada pelo fato mesmo de não ser conhecida, em geral, a solução exata.

O problema básico da solução de sistemas de equações lineares simultâneas por processos "exatos" não está sujeito a este tipo particular de erro.

1.2.5. Erros de Arredondamento na Representação de Números nas Operações Elementares

Nenhum procedimento ou aparelho de caráter digital pode realizar suas operações "elementares" (ou, pelo menos, *todas* elas) rigorosamente.

Isso decorre de que a representação de um número *real* contínuo qualquer, num sistema de posição que utiliza um número *finito* de algarismos significativos, é, em geral, impossível.

Nos aparelhos digitais, os números são representados por um agregado de dígitos da forma

$$x = s\,(0, a_1, a_2, \ldots a_t) \cdot B^e \qquad (1.1)$$

onde B é a base (em geral 2 ou 10), e, o expoente ou característica, s é o sinal, e a fração é a mantissa com t dígitos, variável de máquina para máquina como

ter-se-á oportunidade de ver com detalhe. O processo de decidir qual o último dígito da mantissa que melhor aproxima o número real é chamado *arredondamento*.

É denominada *aritmética racional* ou de *precisão infinita* aquela em que todos os resultados são expressos por frações ordinárias, o que não é prático nem factível no caso geral. Os computadores utilizam a chamada *aritmética arredondada*.

Percebe-se que a soma de dois números de t dígitos é de novo um número de t dígitos, mas o produto terá 2t dígitos e o quociente, em geral, infinitos dígitos. Caso se queira evitar erro de arredondamento, deve-se manter, em cada novo produto, o dobro dos dígitos anteriores, o que é impraticável. Em cada caso estar-se-ia praticando não uma operação elementar de fato, mas uma *pseudo-operação* que já apresenta o resultado com um número dado de dígitos, pela inclusão de um procedimento qualquer de arredondamento.

Se os dados do problema forem considerados exatos e exatamente representáveis no sistema de posição utilizado, a diferença entre a solução obtida pela seqüência de pseudo-operações e a que se obteria com o mesmo algoritmo utilizando aritmética de precisão infinita será o chamado *erro de arredondamento*. Essa comparação não é, em geral, factível.

O erro de arredondamento pode ser tornado, pelo menos teoricamente, desprezível, pelo simples expediente, quando possível, de utilizar um número suficientemente elevado de dígitos (na prática, a chamada *precisão dupla*, definida de forma diferente nos diversos computadores, como será visto). Considerar-se-á, assim, essa fonte de erro como controlável nos programas de EF.

1.2.6. Erros Iniciais de Observação e Arredondamento

Admitindo desde já que não é questionada a validade do modelo teórico e da discretização feita, a coleta dos parâmetros iniciais da estrutura (geometria, materiais, condições de contorno), sua transcrição, transmissão e truncamento ou arredondamento dos valores para representação dentro da base e formato numérico do equipamento e algoritmo (programa) utilizados, já levam à solução de uma estrutura que na realidade não é mais a mesma inicial.

Mesmo que os dados sejam exatamente conhecidos, podem não ser exatamente representáveis com o número de dígitos disponível, restringindo assim a precisão mesmo antes de iniciado o cálculo.

Se os dados forem tirados da observação física, serão inevitavelmente sujeitos aos erros inerentes a essas medidas, e serão, em geral, de ordem muito grosseira frente à precisão de representação propiciada pela maioria das máquinas disponíveis; o que é mais grave, não existe, nem mesmo teoricamente, a possibilidade de reduzir seus efeitos por procedimentos posteriores durante o cálculo. Na execução das operações, os erros iniciais se propagam e se interam, podendo até levar a situações críticas que o problema inicial poderia não ter. Em geral isso ocorre com rapidez, quando informações importantes estão embutidas em dígitos de baixa ordem.

A título de exemplo, considere-se a solução de um sistema de equações lineares simultâneas que, como se sabe, só existe se o determinante da matriz de coeficiente é não-nulo. Observa-se que a computação do determinante a partir dos valores aproximados da matriz, considerados como exatos, pode resultar diferente de zero, enquanto que mudando os elementos dentro dos limites de precisão pode-se chegar à singularidade. Um sistema assim não pode ser resolvido com confiança: num caso existe uma solução única, no outro, infinitas soluções ou nenhuma.

É essa fonte de erros que será a preocupação predominante neste trabalho, bem como o que de intrínseco na formulação de um problema o torna propenso a eles.

2. VETORES, MATRIZES E SUAS NORMAS

2.1. Introdução

Neste trabalho, serão utilizados os conceitos elementares da álgebra linear, vetores e matrizes, de conhecimento geral, não merecendo, assim, uma apresentação completa e rigorosa. Nesta introdução, são citadas apenas algumas definições e resultados que serão usados, remetendo aos trabalhos padrões nesse campo, como (7), (5) e (8), para provas e conceituação rigorosa. Na seção 2.2. seguinte, será examinado com detalhe o assunto de normas vetoriais e matriciais.

2.1.1 Vetores e suas Operações Elementares

Um vetor $\underline{x}$, na geometria de três dimensões, é dito uma entidade, com magnitude, direção e sentido, e é descrito por suas três coordenadas (x_1, x_2, x_3) em uma base de referência composta por quaisquer três vetores não-coplanares.

Estende-se essa idéia para conceituar um vetor $\underline{x}$ em n dimensões como um ente definido por componentes $x_i (i = 1, \ldots, n)$ em uma base de referência composta de n vetores linearmente independentes.

Uma coleção de vetores $\underline{x}_1, \underline{x}_2, \ldots$, é dita *linearmente dependente* se existem números $a_1, a_2, \ldots$, não todos nulos, tais que

$$a_1\underline{x}_1 + a_2\underline{x}_2 + \ldots = 0 \qquad (2.1)$$

caso contrário são *linearmente independentes*. Para um espaço de vetores definidos em n dimensões, uma coleção de n vetores linearmente independentes constituem uma *base*.

Entidades aqui tratadas como vetores na análise estrutural serão, por exemplo, o conjunto das componentes dos deslocamentos de um número de pontos de uma estrutura segundo um sistema de referência ou o conjunto de componentes das cargas aplicadas nesses mesmos pontos de acordo com essa mesma referência.

Serão usadas as seguintes operações com vetores:

- Produto de escalar por vetor: $\underline{z} = k\underline{x}$ como $z_i = kx_i$ (2.2)

- Soma de vetores: $\underline{z} = \underline{x} + \underline{y}$ como $z_i = x_i + y_i$ (2.3)

- Produto escalar de vetores: $(\underline{x}, \underline{y}) = \sum_{i=1}^{n} x_i \cdot y_i$ (2.4)

2.1.2. Matrizes e suas Operações Elementares

Na análise estrutural, em muitas ocasiões, um vetor será expresso como função linear de outro, e.g., cargas nodais em função de deslocamentos nodais. Em geral, as coordenadas x_i ($i = 1, \ldots, m$) de um vetor $\underline{x}$ seriam relacionadas às coordenadas b_i ($i = 1, \ldots, n$) de outro vetor $\underline{b}$ por um conjunto de n relações lineares, do tipo:

$$\begin{aligned} &a_{11}\, x_1 + a_{12}\, x_2 + \ldots + a_{1m}\, x_m = b_1 \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ &a_{n1}\, x_1 + a_{n2}\, x_2 + \ldots + a_{nm}\, x_m = b_n \end{aligned} \qquad (2.5)$$

que se poderia escrever como:

$$\begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1m} \\ \ldots & \ldots & \ldots & \ldots \\ a_{n1} & a_{n2} & \ldots & a_{nm} \end{pmatrix} \cdot \begin{Bmatrix} x_1 \\ \vdots \\ x_m \end{Bmatrix} = \begin{Bmatrix} b_1 \\ \vdots \\ b_n \end{Bmatrix} \text{ ou } \underline{A}\ \underline{x} = \underline{b} \qquad (2.6)$$

onde o conjunto $\underline{A}$ de $n \cdot m$ coeficientes a_{ij} pode ser denominado *matriz*, onde se utilizou a operação:

$$\sum_{j=1}^{m} a_{ij} \cdot x_j = b_i \qquad (2.7)$$

que é o produto escalar de cada linha de $\underline{A}$ pelo vetor $\underline{x}$, só possível se o número m de colunas da matriz é igual ao número m de componentes do vetor $\underline{x}$.

Usar-se-ão as seguintes operações com matrizes:

- Produto de escalar por matriz: $\underline{C} = k\ \underline{A}$ como $c_{ij} = k\ a_{ij}$ (2.8)
- Soma de matrizes: $\underline{C} = \underline{A} + \underline{B}$ como $c_{ij} = a_{ij} + b_{ij}$ (2.9)
  (onde $\underline{A}$ e $\underline{B}$ devem ter a mesma ordem $n \cdot m$)
- Produto de matrizes: $\underline{C} = \underline{A} \cdot \underline{B}$ como $c_{ij} = \sum_{k=1}^{p} a_{ik}\, b_{kj}$ (2.10)

(em que: se $\underline{A}$ for n · p, $\underline{B}$ deverá ser p · m, e C resultará n · m; a operação não é comutativa, importando a ordem)

Chama-se *matriz transposta* a matriz $\underline{A}^t$ obtida escrevendo-se ordenadamente as linhas de $\underline{A}$ como colunas.

2.1.3. Matrizes Quadradas, Simétricas, Definidas, Autovalores e Autovetores

Trabalhar-se-á com matrizes quadradas, onde n = m, em que os elementos a_{ij} para i = j pertencem à *diagonal principal*. Se os elementos da diagonal principal são unitários e os demais elementos da matriz são nulos, tem-se a *matriz identidade* $\underline{I}$ de ordem n, para a qual, para qualquer $\underline{x}$, tem-se:

$$\underline{x} = \underline{I}\,\underline{x} \tag{2.11}$$

A *matriz inversa* $\underline{A}^{-1}$, se existir, será tal que:

$$\underline{A}\,\underline{A}^{-1} = \underline{I} \tag{2.12}$$

Se $\underline{A}\,\underline{x} = \underline{b}$ é um mapeamento um a um de todos os vetores $\underline{x}$ em todos os vetores $\underline{b}$, então $\underline{A}^{-1}$ existe e é equivalente a dizer que o determinante de $\underline{A}$ é não-nulo ou que $\underline{A}$ é *não-singular*.

O *determinante* de $\underline{A}$ de ordem n é um escalar, calculado pela expressão de recorrência:

$$\det \underline{A} = \sum_{j=1}^{n} (-1)^{1+j} \cdot \det \underline{A}_{1j} \tag{2.13}$$

onde $\underline{A}_{1j}$ é a matriz (n - 1) · (n - 1) obtida eliminando-se a primeira linha e a j-ésima coluna de $\underline{A}$, e definindo o determinante de uma matriz de ordem 1 como sendo igual a seu único elemento.

Uma matriz é dita simétrica se $\underline{A} = \underline{A}^t$, isto é, se $a_{ij} = a_{ji}$.

$\underline{A}$ é *definida* se, além de simétrica, tem-se o produto escalar $(\underline{A} \cdot \underline{x}, \underline{x})$ sempre não-negativo para qualquer $\underline{x}$.

Nota-se que $\underline{A}\,\underline{A}^t$ é sempre definida.

Definem-se *valores próprios* ou *autovalores* $\lambda_1, \ldots, \lambda_n$ de uma matriz quadrada n • n como as raízes do polinômio de ordem n obtido pelo cálculo do determinante de $(\lambda \underline{I} - \underline{A})$, chamado *polinômio característico*. Neste trabalho, interessará o caso de $\underline{A}$ simétrica quando *os valores próprios são todos reais*. Se $\underline{A}$ é ainda definida, eles são ainda todos não-negativos e, por convenção, escritos na ordem decrescente monotônica:

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq 0 \qquad (2.14)$$

com

$$\det \underline{A} = \text{produto dos } \lambda_i \qquad (i = 1, \ldots, n) \qquad (2.15)$$

A cada valor λ_i $(i = 1, \ldots, n)$ corresponde um vetor $\underline{x}_i$ $(i = 1, \ldots, n)$ não-nulo, chamado *vetor próprio* ou *autovetor*, tal que:

$$\underline{A}\, x_i = \lambda_i\, \underline{x}_i \quad , \text{ ou, no geral, } \quad \underline{A}\, \underline{x} = \lambda\, \underline{x} \qquad (2.16)$$

Na prática, a obtenção desses valores e vetores próprios para matrizes de ordem n relativamente elevada seria impossível de calcular pelo polinômio característico. Utilizam-se processos iterativos com vetores tentativos e o chamado *coeficiente de Rayleigh*:

$$\rho(\underline{x}) = \underline{x}^t\, \underline{A}\, \underline{x} / \underline{x}^t\, \underline{x} \text{ , que obedece a } \lambda_1 \leq \rho(\underline{x}) \leq \lambda_n \qquad (2.17)$$

2.2. Normas

2.2.1. Normas Vetoriais

A magnitude relativa de números reais é um conceito primitivo geral. O comprimento de um vetor da geometria de três dimensões também é dominado com facilidade.

Para um vetor de n dimensões geral, a avaliação de seu "tamanho" vai depender de todos os seus elementos, e será feita por certas funções denominadas *normas*.

A *norma de um vetor* $\underline{x}$ é uma função de seus elementos que associa a $\underline{x}$ um número real não-negativo, representado por $\| \underline{x} \|$, satisfazendo:

1. $\| \underline{x} \| > 0$ para $\underline{x} \neq 0$, e $\| \underline{x} \| = 0$ se e somente se $\underline{x} = 0$ (2.18)

2. $\|k\underline{x}\| = |k| \, \|\underline{x}\|$ para qualquer multiplicador real ou complexo k (2.19)

3. $\|\underline{x} + \underline{y}\| \leq \|\underline{x}\| + \|\underline{y}\|$ (desigualdade triangular) (2.20)

Dessa última condição, têm-se:

$$\|\underline{x} - \underline{y}\| \geq \|\underline{x}\| - \|\underline{y}\| \quad \text{e} \quad \|\underline{x} - \underline{y}\| \geq \|\underline{y}\| - \|\underline{x}\| \tag{2.21}$$

uma delas sempre trivial.

Toda norma determina uma "esfera unitária" (5): um conjunto de vetores cujas normas não excedem a 1. A "esfera unitária" é um corpo convexo simétrico em relação ao centro, isto é, um conjunto que contém, para cada $\underline{x}$, um vetor $-\underline{x}$ (simetria em relação ao eixo), e para quaisquer dois vetores $\underline{x}_1$ e $\underline{x}_2$ contém um vetor $t\underline{x}_1 + (1 - t)\,\underline{x}_2$, $0 \leq t \leq 1$, sobre o segmento que une as extremidades dos vetores $\underline{x}_1$ e $\underline{x}_2$ (convexidade).

Há três normas vetoriais em uso corrente. São definidas, para um vetor $\underline{x} = (x_1, x_2, \ldots, x_n)$, real ou complexo, como:

$$\| \underline{x} \|_p = (|x_1|^p + |x_2|^p + \ldots + |x|^p)^{1/p} \quad (p = \infty, 1, 2) \tag{2.22}$$

1a. norma: Norma infinita, cúbica ou de Chebyshev.

$$\|\underline{x}\|_\infty = \max_i |x_i| \tag{2.23}$$

O conjunto de vetores do espaço real com norma que não excede a 1, preenche o cubo unitário:

$$-1 \leq x_1 \leq 1, \ldots, -1 \leq x_n \leq 1$$

2a. norma: Norma octaédrica ou de Manhattan.

$$\|\underline{x}\|_1 = |x_1| + |x_2| + \ldots + |x_n| \tag{2.24}$$

O conjunto dos vetores reais para os quais $\|\underline{x}\|_1 \leq 1$, preenche o equivalente n-dimensional de um octaedro.

3a. norma: Norma euclidiana ou esférica.

$$\|\underline{x}\|_2 = \sqrt{|x_1|^2 + |x_2|^2 + \ldots + |x_n|^2} \qquad (2.25)$$

Esta norma nada mais é que o *comprimento* do vetor como normalmente entendido. O conjunto de vetores para os quais a norma euclidiana não excede à unidade preenche uma esfera de raio unitário.

As normas acima introduzidas relacionam-se através das desigualdades seguintes:

$$\|\underline{x}\|_\infty \leq \|\underline{x}\|_1 \leq n \cdot \|\underline{x}\|_\infty \qquad (2.26)$$

$$\|\underline{x}\|_\infty \leq \|\underline{x}\|_2 \leq \sqrt{n}\, \|\underline{x}\|_\infty \qquad (2.27)$$

$$1/\sqrt{n}\ \|\underline{x}\|_1 \leq \|\underline{x}\|_2 \leq \|\underline{x}\|_1 \qquad (2.28)$$

Uma aplicação usual do conceito de norma vetorial é o estabelecimento da condição necessária e suficiente para a convergência de uma seqüência de vetores $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_k$ para um vetor $\underline{x}$. A condição é que, para cada uma das três normas, tenha-se:

$$\lim_{k \to \infty} \|\underline{x}_k - \underline{x}\| = 0 \qquad (2.29)$$

O cálculo da ordem de convergência p e da razão de convergência c é feito (3) por:

$$\lim_{k \to \infty} \frac{\|\underline{x}_{k+1} - \underline{x}\|}{\|\underline{x}_k - \underline{x}\|^p} = c \qquad (2.30)$$

2.2.2. Normas Matriciais

Em analogia às normas vetoriais, define-se *norma de uma matriz* $\underline{A}$ quadrada de ordem n qualquer função de seus elementos que associa à matriz um número não-negativo $\|\underline{A}\|$ tal que:

1. $\|\underline{A}\| > 0$, e $\|\underline{A}\| = 0$ se e somente se $\underline{A} = 0$ (2.31)

2. $\|k \cdot \underline{A}\| = |k| \cdot \|\underline{A}\|$ para qualquer número real ou complexo k (2.32)

3. $\|\underline{A} + \underline{B}\| \leq \|\underline{A}\| + \|\underline{B}\|$ para matrizes $\underline{A}$ e $\underline{B}$ (desigualdade triangular) (2.33)

4. $\|\underline{A} \cdot \underline{B}\| \leq \|\underline{A}\| \cdot \|\underline{B}\|$ para matrizes $\underline{A}$ e $\underline{B}$ (2.34)

Desde que a maioria dos problemas envolve, simultaneamente, vetor e matriz, é recomendável introduzirem-se normas matriciais que se relacionem às vetoriais. Diz-se que uma norma matricial é *compatível* ou *consistente* com uma dada norma vetorial se:

$$\|\underline{A}\ \underline{x}\| \leq \|\underline{A}\| \cdot \|\underline{x}\| \text{ para todo } \underline{A} \text{ e } \underline{x} \tag{2.35}$$

Diz-se ainda que uma norma matricial é subordinada a uma norma vetorial se existe pelo menos um vetor $\underline{x}$ tal que:

$$\|\underline{A}\ \underline{x}\| = \|\underline{A}\| \cdot \|\underline{x}\| \tag{2.36}$$

de onde se conclui pela necessidade de que $\|\underline{I}\| = 1$ para que a norma seja subordinada.

Em correspondência às três normas vetoriais já vistas, têm-se as seguintes normas matriciais:

1a. norma: Norma linha ou infinita

$$\|\underline{A}\|_\infty = \max_{1 \leq i \leq n} \sum_{j=1}^{n} |a_{ij}| \tag{2.37}$$

que é a máxima soma por linha dos módulos dos elementos da matriz. Esta norma é subordinada à norma $\|\underline{x}\|_\infty$ (de Chebyshev).

2a. norma: Norma coluna

$$\|\underline{A}\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^{n} |a_{ij}| \tag{2.38}$$

que é a máxima soma por coluna dos módulos dos elementos da matriz. Esta norma é subordinada à norma $\|\underline{x}\|_1$ (de Manhattan).

Para uma matriz $\underline{A}$ simétrica, é claro que:

$$\|\underline{A}\|_\infty = \|A\|_1 \tag{2.39}$$

3a. norma: Norma espectral

$$\|\underline{A}\|_2 = (\text{máximo valor próprio de } \underline{A}^t \cdot \underline{A})^{1/2} \tag{2.40}$$

e é subordinada à norma euclidiana ou esférica do vetor:

$$\|\underline{x}\|_2 = \left(\sum_{i=1}^{n} |x_i|^2\right)^{1/2} \tag{2.41}$$

Essa norma é chamada, às vezes, de delimitação superior da matriz. Em correspondência, o menor valor próprio da matriz $\underline{A}^t \cdot \underline{A}$ é chamado de delimitação inferior da matriz $\underline{A}$. É óbvio que a delimitação inferior de $\underline{A}$ é um número inverso da delimitação superior da matriz inversa.

Duas outras normas são também utilizadas nos diversos estudos sobre matrizes, e são:

- Norma euclidiana ou de Schur

$$\|\underline{A}\|_E = \left(\sum_{i,j} |a_{ij}|^2\right)^{1/2} \tag{2.42}$$

Verifica-se que ela é consistente com $\|\underline{x}\|_2$, porém não é a ela subordinada, já que:

$$\|\underline{I}\|_E = n^{1/2} \tag{2.43}$$

- Maior coeficiente em módulo

$$M(\underline{A}) = \max_{i,j} |a_{ij}| \tag{2.44}$$

Essa norma, utilizada por *Turing* (19), é consistente com todas as três normas vetoriais vistas.

A seguir, transcreve-se uma série de desigualdades a que obedecem as normas matriciais vistas, segundo os trabalhos de *Turing* (19) e *Faddeev* (5):

$$M(\underline{A}) \leq \| \underline{A} \|_\infty \leq n \cdot M(\underline{A}) \tag{2.45}$$

$$M(\underline{A}) \leq \| \underline{A} \|_1 \leq n \cdot M(\underline{A}) \tag{2.46}$$

$$M(\underline{A}) \leq \| \underline{A} \|_2 \leq n \cdot M(\underline{A}) \tag{2.47}$$

$$M(\underline{A}) \leq \| \underline{A} \|_E \leq n \cdot M(A) \tag{2.48}$$

$$1/\sqrt{n} \quad \|\underline{A}\|_E \leq \|\underline{A}\|_2 \leq \|\underline{A}\|_E \tag{2.49}$$

$$1/\sqrt{n} \quad \|\underline{A}\|_E \leq \|\underline{A}\|_\infty \leq \sqrt{n} \ \|\underline{A}\|_E \tag{2.50}$$

$$1/\sqrt{n} \quad \|\underline{A}\|_E \leq \|\underline{A}\|_1 \leq \sqrt{n} \ \|\underline{A}\|_E \tag{2.51}$$

$$1/\sqrt{n} \quad \|\underline{A}\|_2 \leq \|\underline{A}\|_\infty \leq \sqrt{n} \ \|\underline{A}\|_2 \tag{2.52}$$

$$1/\sqrt{n} \quad \|\underline{A}\|_2 \leq \|\underline{A}\|_1 \leq \sqrt{n} \ \|\underline{A}\|_2 \tag{2.53}$$

$$1/\sqrt{n} \quad \|\underline{A}\|_\infty \leq \|\underline{A}\|_1 \leq n \cdot \|\underline{A}\|_\infty \tag{2.54}$$

Na aplicação de normas à análise de erros, $\|\underline{A}\|_E$ é vantajosa sobre $\|\underline{A}\|_2$, apesar de não subordinada a $\|\underline{x}\|_2$, por ser mais fácil de ser calculada, o que também é verdade de $\|\underline{A}\|_\infty$ e $\|\underline{A}\|_1$. Apesar da desigualdade (2.49), $\|\underline{A}\|_E$ é freqüentemente uma boa aproximação de $\|\underline{A}\|_2$.

Uma importante propriedade das normas matriciais é que qualquer delas fornece uma delimitação superior dos módulos dos valores próprios da matriz.

Se λ é qualquer valor próprio de $\underline{A}$ e $\underline{x}$ é o vetor próprio correspondente, tem-se:

$$\underline{A}\,\underline{x} = \lambda\,\underline{x} \tag{2.55}$$

Aplicando normas cosistentes, tem-se:

$$\|\underline{A}\| \, \|\underline{x}\| \geq \|\underline{A}\,\underline{x}\| = \|\lambda\,\underline{x}\| = |\lambda| \cdot \|\underline{x}\|, \text{ de onde } \|\underline{A}\| \geq |\lambda| \tag{2.56}$$

Uma aplicação comum do conceito de norma matricial é que a condição necessária e suficiente para que uma seqüência de matrizes $\underline{A}_1, \underline{A}_2, \ldots, \underline{A}_k$, convirja para $\underline{A}$ é que:

$$\lim_{k \to \infty} \| \underline{A}_k - \underline{A} \| = 0 \tag{2.56}$$

de forma que, se $\underline{A}_k \to \underline{A}$, segue-se que $\| \underline{A}_k \| \to \| \underline{A} \|$.

3. NOÇÕES DO MÉTODO DOS ELEMENTOS FINITOS, PROCESSO DOS DESLOCAMENTOS, PARA ESTRUTURAS DE BARRAS DE COMPORTAMENTO LINEAR

3.1. Introdução

3.1.1. Histórico

O Método dos Elementos Finitos (MEF) pode ser considerado historicamente como uma contribuição original dos engenheiros estruturais à Física Matemática em geral, motivada pelo esforço de aplicação nessa área do computador eletrônico digital de programa armazenado para solução dos contínuos elásticos.

Entretanto, passados já 30 anos de sua primeira introdução nesta área, pode-se já traçar um quadro mais amplo de suas filiações matemáticas ([2], [21] e [22]). O processo se encaixa na família geral dos métodos de discretização dos problemas dos contínuos físicos e de aproximação de suas equações diferenciais governantes.

Descobrem-se suas raízes nos procedimentos com "*trial functions*" utilizados nos métodos de Rayleigh (1870) e Ritz (1909), bem como no tratamento de resíduos ponderados de soluções por séries de problemas elásticos propostos por Galerkin (1915).

Já em 1943, R. Courant apresenta solução para problema de torção com divisões triangulares e "*piecewise continuous functions*", que praticamente reproduz o MEF.

As primeiras contribuições originais na Engenharia Estrutural são de Argyris e Kelsey, em 1954, e no trabalho de 1956 de Turner, Clough, Martin e Topp para elasticidade plana. É ainda Ray Clough quem, em trabalho sobre aplicação à elasticidade plana, em 1960, pela primeira vez utiliza a expressão ELEMENTOS FINITOS. Daí para a frente, os progressos, tanto nas aplicações como na formalização matemática, se desenvolvem em ritmo espetacular, levando, ao final da década de 60, aos grandes programas comerciais aplicativos como o STARDYNE, NEAT, EASE, NASTRAN, SAP e outros. Tais

programas têm capacidade de atacar problemas de grande porte em elasticidade plana, espacial, placas, plasticidade, viscoelasticidade, solicitações dinâmicas, etc.

3.1.2. Equações Gerais

Na solução dos problemas de Engenharia Estrutural, pretende-se determinar os campos dos deslocamentos, deformações e tensões de uma estrutura que é essencialmente um corpo contínuo tridimensional. Esses campos podem ser representados simbolicamente pelos vetores $\underline{u}$, $\underline{\varepsilon}$ e $\underline{\sigma}$, respectivamente.

As deformações se relacionam aos deslocamentos por relações do tipo:

$$\underline{\varepsilon} = \underline{L}\ \underline{u} \tag{3.1}$$

(sendo $\underline{L}$ um operador linear apropriado).

Tensões e deformações se interrelacionam por *equações constitutivas* que definem o comportamento do material utilizado na forma:

$$\underline{\sigma} = \underline{D}\ \underline{\varepsilon} \tag{3.2}$$

sendo $\underline{D}$ uma matriz conveniente (obtida experimentalmente, por exemplo).

O corpo (estrutura) em análise estará submetido, em geral, a forças de massa $\underline{f}_m$ e a condições de contorno de deslocamentos $\underline{u} = \bar{\underline{u}}$ impostas em parte de sua superfície (vínculos), e de tensões superficiais $\underline{f}_s$ impostas em outra região, bem como, em geral, por forças $\underline{r}$ concentradas em pontos discretos.

Se for aplicado, agora, ao conjunto um campo virtual $\overset{*}{\underline{u}}$ de deslocamentos compatíveis com os vínculos, ter-se-á:

Deformações $\quad \overset{*}{\underline{\varepsilon}} = \underline{L}\ \overset{*}{\underline{u}}$ (3.3)

Trabalho virtual ao longo do deslocamento:

$$\int_V \overset{*}{\underline{\varepsilon}}^t \cdot \underline{\sigma}\, dV = \int_V \overset{*}{\underline{u}}^t \underline{f}_m\, dV + \int_S \overset{*}{\underline{u}}_s^t \underline{f}_s\, dS + \overset{*}{\underline{u}}^t \underline{r} \tag{3.4}$$

que pode ser escrito como a seguir, para um dado material:

$$\int_V (\underline{L}\,\overset{*}{\underline{u}})^t\,\underline{D}\,\underline{L}\,\underline{u}\,dV = \ldots \qquad (3.5)$$

No MEF, divide-se a estrutura contínua por linhas ou superfícies imaginárias em certo número de "elementos finitos". Embora haja infinitos pontos de contato no contorno desses elementos, serão aqui considerados apenas conectados por um conjunto discreto de pontos denominados *nós*, cujos *deslocamentos nodais* constituem o vetor $\underline{a}$, incógnita deste problema.

Os deslocamentos internos nos elementos serão aproximados, em função dos deslocamentos nodais, por

$$\underline{u} \simeq \hat{\underline{u}} = \underline{N}\,\underline{a} \qquad (3.6)$$

onde $\underline{N}$ são funções de coordenadas chamadas "funções de interpolação" ou "funções de forma", convenientemente escolhidas de modo a reproduzir uma dada componente de deslocamento nodal (isto é, assumir o valor 1), mantendo nulas as demais componentes neste nó e nos demais nós contíguos ao elemento.

As deformações internas nos elementos serão aproximadas por:

$$\hat{\underline{\varepsilon}} = \underline{L}\,\hat{\underline{u}} = \underline{L}\,\underline{N}\,\underline{a} = \underline{B}\,\underline{a} \qquad (3.7)$$

e as tensões aproximadas por

$$\hat{\underline{\sigma}} = \underline{D}\,\hat{\underline{\varepsilon}} = \underline{D}\,\underline{B}\,\underline{a} \qquad (3.8)$$

Pode-se agora reescrever as equações integrais de equilíbrio para deslocamentos nodais virtuais $\overset{*}{\underline{a}}$:

$$\int_V (\underline{B}\,\overset{*}{\underline{a}})^t\,\underline{D}\,\underline{B}\,\underline{a}\,dV = \int_V (\underline{N}\,\overset{*}{\underline{a}})^t\,\underline{f}_m\,dV + \int_S (\underline{N}\,\overset{*}{\underline{a}}_s)\;\; f_s dS + \overset{*t}{\underline{a}}_r\,\underline{r} \qquad (3.9)$$

em que agora $\underline{r}$ são forças concentradas nos nós escolhidos.

Como os deslocamentos $\overset{*}{\underline{a}}$ são arbitrários, pode-se fazer

$$\left(\int_V \underline{B}^t\,\underline{D}\,\underline{B}\,dV\right)\underline{a} = \int_V \underline{N}^t\,.\,\underline{f}_m\,dV + \int_S \underline{N}^t\,.\,\underline{f}_s\,.\,dS + r \qquad (3.10)$$

Serão agora identificados:

- Matriz de rigidez $\quad \underline{K} = \int_V \underline{B}^t\,\underline{D}\,\underline{B}\,dV$

- deslocamentos nodais $\underline{a}$ (incógnitas)
- vetor de cargas nodais $\underline{r}$
- esforços nodais equivalentes às forças de massa e contorno - $\underline{f}$,

com o que se chega à formulação das equações de equilíbrio do processo dos deslocamentos:

$$\underline{K}\,\underline{a} = \underline{r} - \underline{f} \qquad (3.11)$$

a resolver pelos processos numéricos ordinários.

Deve ficar claro que, devido à particular escolha das funções de forma, as integrais acima serão na realidade somatórias de integrais realizadas a nível dos elementos.

3.2. Aplicação: Estrutura Treliçada em Três Dimensões

3.2.1. Dedução da Matriz de Rigidez pelo MEF

Como aplicação da teoria desenvolvida na seção 3.12, deduzir-se-á a matriz de rigidez de barra de treliça espacial. De acordo com as hipóteses da Resistência dos Materiais, as barras de uma treliça se articulam nos nós e só trabalham a esforços axiais, sofrendo apenas deformações longitudinais. Para uma barra de comprimento L, seção transversal de área A, de material com módulo de Young E, escrever-se-á:

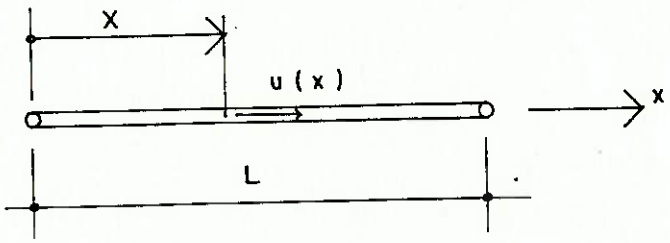


Fig. 3.1

Deslocamentos $\underline{u}$, como $u = u(x)$

Deformações: $\underline{\varepsilon} = \underline{L}\ \underline{u}$, como $\varepsilon_x(x) = \frac{d}{dx} u$

Tensões: $\underline{\sigma} = \underline{D}\ \underline{\varepsilon}$, como $\sigma_x(x) = E\ \varepsilon_x$

Serão agora aproximados os deslocamentos ao longo da barra em função dos deslocamentos nodais a_i, i = 1, ..., 6, por meio de funções de forma N_i, fazendo-se:

$$\underline{\hat{u}} = \underline{N}\,\underline{a}$$

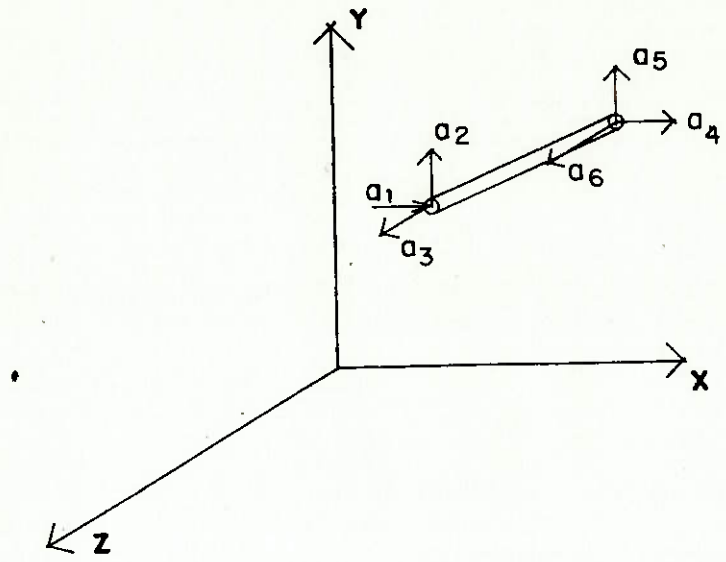


Figura 3.2

Obs.: Os cossenos diretores da barra serão designados por cx, cy e cz.

Ter-se-ão, para os deslocamentos do nó inicial da barra, funções do tipo da apresentada na Fig. 3.3 para a_1.

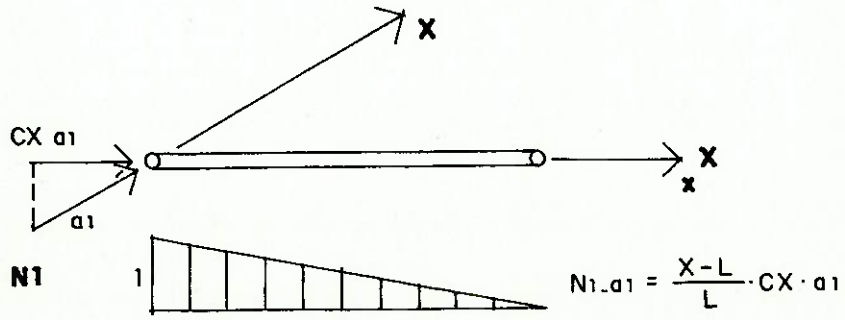


Figura 3.3

e, para os deslocamentos do nó final, funções do tipo apresentado na Fig. 3.4 para a_4:

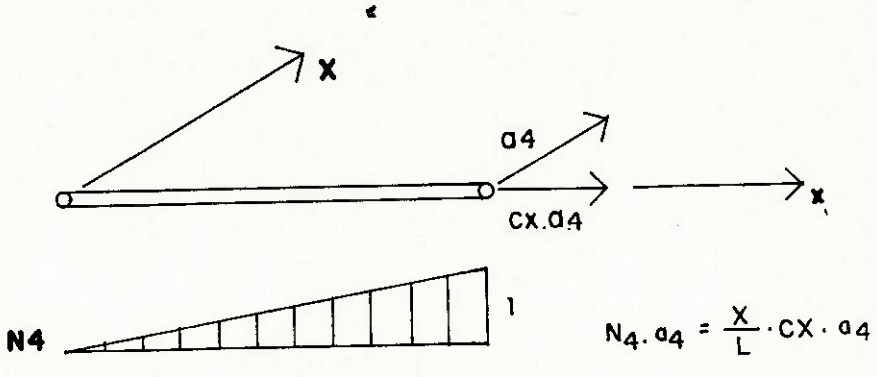


Figura 3.4

Em conjunto, escrever-se-iam:

- Deslocamentos aproximados, $\hat{\underline{u}} = \underline{N}\,\underline{a}$, como

$$\hat{u}(x) = \frac{1}{L}\left[(L-x)cx \quad (L-x)cy \quad (L-x)cz \quad x\cdot cx \quad x\cdot cy \quad x\cdot cz \right] \begin{Bmatrix} a_1 \\ \vdots \\ a_6 \end{Bmatrix}$$

- Deformações aproximadas, $\hat{\varepsilon} = \underline{L}\,\hat{u} = \underline{L}\,\underline{N}\,\underline{a} = \underline{B}\,a$, como

$$\hat{\varepsilon}_x(x) = \frac{d}{dx}\hat{u} = \frac{1}{L}\left[-cx \quad -cy \quad -cz \quad cx \quad cy \quad cz \right] \begin{Bmatrix} a_1 \\ \vdots \\ a_6 \end{Bmatrix}$$

- Tensões aproximadas, $\hat{\underline{\sigma}} = \underline{D}\,\hat{\underline{\varepsilon}} = \underline{D}\,\underline{B}$, como:

$$\hat{\sigma}_x(x) = E\,\hat{\varepsilon}_x = \frac{E}{L}\left[-cx \quad -cy \quad -cz \quad cx \quad cy \quad cz \right] \begin{Bmatrix} a_1 \\ \vdots \\ a_6 \end{Bmatrix}$$

Donde são extraídas as matrizes:

$$\underline{B}^t = \frac{1}{L} \begin{Bmatrix} -cx \\ -cy \\ -cz \\ cx \\ cy \\ cz \end{Bmatrix} \quad \text{e} \quad \underline{D}\,\underline{B} = \frac{E}{L}\left[-cx \quad -cy \quad -cz \quad cx \quad cy \quad cz \right]$$

Calcula-se agora a matriz de rigidez do elemento,

$$\underline{K}_e = \int_{V_e} \underline{B}^t \, \underline{D} \, \underline{B} \, dV$$

como:

$$\underline{K}_e = \frac{E}{L^2} \begin{Bmatrix} -cx \\ -cy \\ -cz \\ cx \\ cy \\ cz \end{Bmatrix} \begin{Bmatrix} -cx & -cy & -cz & cx & cy & cz \end{Bmatrix} \int_{V_e} dV$$

Para uma barra prismática, seu volume vale A L, obtendo a matriz de rigidez da barra:

$$\underline{K}_e = \frac{AE}{L} \begin{bmatrix} cx \cdot cx & cx \cdot cy & cx \cdot cz & -cx \cdot cx & -cx \cdot cy & -cx \cdot cz \\ & cy \cdot cy & cy \cdot cz & -cy \cdot cx & -cy \cdot cy & -cy \cdot cz \\ & & cz \cdot cz & -cz \cdot cx & -cz \cdot cy & -cz \cdot cz \\ & & & cx \cdot cx & cx \cdot cy & cx \cdot cz \\ & \text{simétrica} & & & cy \cdot cy & cy \cdot cz \\ & & & & & cz \cdot cz \end{bmatrix}$$

conforme se pode conferir em, por exemplo, *Gere e Weaver* (6).

3.2.2. Programa em BASIC para Microcomputador para Montagem da Matriz de Rigidez com Disposição Econômica da Memória

A seguir, lista-se um programa para microcomputador pessoal da linha TRS-80, na linguagem BASIC, para montagem da matriz de rigidez de treliças espaciais. No MEF, tais matrizes são simétricas e, em geral, bandeadas, isto é, os coeficientes não-nulos se agrupam em uma faixa irregular junto à diagonal principal. O contorno dessa faixa é conhecido por "linha do horizonte" ("*sky-line*", em inglês), pela semelhança com a silhueta de prédios contra o horizonte de uma cidade. Os elementos fora dessa linha não serão necessários ao cálculo, e é conveniente economizar essas posições de armazenagem, tendo em vista, em especial, a exiguidade típica das memórias das máquinas de pequeno porte. Na Fig. 3.5, apresenta-se um desenho típico.

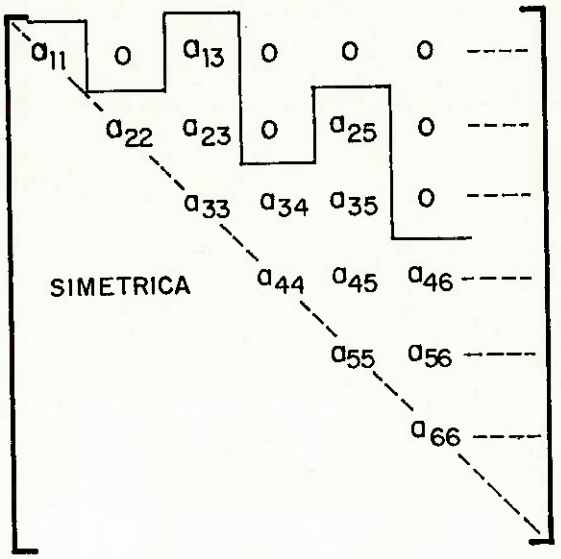


Figura 3.5

A origem dessa situação é o fato de que apenas os graus de liberdade pertencentes a nós interligados por elementos (barras) terão coeficientes de rigidez que podem ser não-nulos, devido à forma particular como as funções de interpolação do MEF são instituídas.

O algoritmo aqui utilizado armazena as colunas de coeficientes não-nulos, da diagonal principal para cima, de forma contínua em uma coluna única A, como se fosse um vetor. Para identificação dos elementos significativos, calcula-se a altura das colunas na sub-rotina COLUNA (instrução 400), verificando, para cada barra, a maior diferença entre os números dos graus de liberdade de suas extremidades. Na sub-rotina ENDEREÇO (instrução 500), estabelece-se um vetor MA que, a partir da altura das colunas, armazena o endereço dos elementos da coluna A que pertencem à diagonal principal.

O veículo de relacionamento dos graus de liberdade das extremidades de cada barra à numeração dos graus de liberdade dos nós da estrutura global é a matriz LM de 6 linhas (o número de graus de liberdade das extremidades de uma barra de treliça espacial) por NE colunas, sendo NE o número de

elementos (barras). Tal matriz é gerada na sub-rotina INPUT BARRAS (instrução 300).

3.2.3. Listagem

```
10 CLS:CLEAR
20 DEFINT I-N
30 INPUT"N. NOS, N. SECOES, N. BARRAS";NP,NM,NE
40 GOSUB 200:REM INPUT NOS
50 PRINT"GRAUS DE LIBERDADE";NQ
60 GOSUB 300:REM INPUT BARRAS
70 GOSUB 900:REM INPUT CARGA
80 GOSUB 500:REM ENDERECO
90 GOSUB 600:REM MONTAGEM MATRIZ RIGIDEZ

200 REM SUB INPUT NOS
205 DIM X(3,NP), ID(3,NP)
210 FOR N=1 TO NP
215 PRINT"NO(";N;")",:INPUT"X,Y,Z";X(1,N),X(2,N),X(3,N)
220 NEXT N
225 INPUT"NUMERO NO C/ RESTRICAO(0 P/ SAIR)";NR
230 IF NR=0 THEN GOTO 245
235 INPUT"ENTRAR RESTR:DX,DY,DZ(0=LIVRE,1=FIXO)";ID(1,NR)
240 GOTO 225
245 NQ=0
250 FOR N=1 TO NP
255 FOR I=1 TO 3
260 IF ID(I,N)<>0 GOTO 270
265 NQ=NQ+1:ID(I,N)=NQ:GOTO275
270 ID(I,N)=0
275 NEXT I
280 NEXT N
285 RETURN


300 REM SUB INPUT BARRAS
305 DIM II(NE), JJ(NE), LM(6,NE), MT(NE), MH(NQ), MA(NQ+1)
310 INPUT"MODULO DE ELASTICIDADE";EM
315 FOR I=1 TO NM
320 PRINT"AREA SECAO TIPO";I;:INPUT AR(I)
325 NEXT I
330 PRINT"BARRA","NO INICIAL, NO FINAL, TIPO DE SECAO"
335 FOR N=1 TO NE
340 PRINT N,:INPUT I,J,MT
345 II(N)=I
350 JJ(N)=J
355 MT(N)=MT
360 FOR L=1 TO 3
365 LM(L,N)=ID(L,I):LM(L+3,N)=ID(L,J)
370 NEXT L
375 GOSUB 400:REM COLUNA
380 NEXT N
385 RETURN
```

```
400 REM SUB COLUNA
405 LS=NQ+1
410 FOR K=1 TO 6
415 IF LM(K,N)=0 GOTO 430
420 IF (LM(K,N)-LS)>=0 GOTO 430
425 LS=LM(K,N)
430 NEXT K
435 FOR K=1 TO 6
440 KK=LM(K,N)
445 IF KK=0 GOTO 460
450 ME=KK-LS
455 IF ME>MH(KK) THEN MH(KK)=ME
460 NEXT K
465 RETURN


500 REM SUB ENDERECO
505 MA(1)=1:MA(2)=2:MK=0
510 IF NQ=1 GOTO535
515 FOR I=2 TO NQ
520 IF MH(I)>MK THEN MK=MH(I)
525 MA(I+1)=MA(I)+MH(I)+1
530 NEXT I
535 MK=MK+1
540 NW=MA(NQ+1)-MA(1)
545 RETURN


600 REM SUB MONTAGEM MATRIZ
602 DIM S(22), A(NW)
605 FOR N=1 TO NE
610 MT=MT(N):X2=0
615 FOR L=1 TO 3
620 D(L)=X(L,II(N))-X(L,JJ(N))
625 X2=X2+D(L)*D(L)
630 NEXT L
635 XL=SQR(X2):XX=EM*AR(MT)*XL
640 FOR L=1 TO 3
645 ST(L)=D(L)/X2:ST(L+3)=-ST(L)
650 NEXT L
655 KL=0
660 FOR L=1 TO 6
665 YY=ST(L)*XX
670 FOR K=L TO 6
675 KL=KL+1:S(KL)=ST(K)*YY
680 NEXT K
685 NEXT L
690 GOSUB 800:REM SOMA BANDA
695 NEXT N
700 RETURN
```

```
800 REM SUB SOMA BANDA
805 NI=0
810 FOR I=1 TO 6
815 I1=LM(I,N)
820 IF I1<=0 GOTO 875
825 MI=MA(I1):KI=I
830 FOR J=1 TO 6
835 J1=LM(J,N)
840 IF J1<=0 GOTO 865
845 IJ=I1-J1
850 IF IJ<0 GOTO 865
855 KK=MI+IJ:KS=KI
860 IF J>=I THEN KS=J+NI
862 A(KK) = A(KK) + S(KS)
865 KI=KI+6-J
870 NEXT J
875 NI=NI+6-I
880 NEXT I
885 RETURN
```

4. SOLUÇÃO DE SISTEMAS DE EQUAÇÕES LINEARES SIMULTÂNEAS: O PROBLEMA TEÓRICO

4.1. Introdução

Embora o assunto básico deste trabalho seja a delimitação de erros na solução de equações de equilíbrio do MEF, neste capítulo será exposto, de forma breve, o problema geral da solução numérica de sistemas de equações lineares simultâneas de uma forma *teórica*, isto é, como se as operações envolvidas pudessem ser realizadas *sem erro de arredondamento*, a partir de parâmetros iniciais também *exatos*.

A álgebra matricial e a solução de equações simultâneas, embora tradicionalmente tratadas juntas, têm pouco em comum, cada assunto podendo ser estudado de forma estanque. Entretanto a concisão e elegância da notação matricial são importantes no desenvolvimento teórico, embora o trabalho efetivo da solução numérica tenha que ser feito sobre os elementos do sistema propriamente ditos.

Uma equação linear é uma expressão da forma:

$$a_1 x_1 + a_2 x_2 + \ldots + a_n x_n = b \qquad (4.1)$$

onde a_i e b são escalares, e x_i são as variáveis a determinar. Duas ou mais equações desse tipo formam um *sistema* de equações lineares simultâneas:

$$\begin{aligned} &a_{11} x_1 + a_{12} x_2 + \ldots + a_{1n} x_n = b \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ &a_{m1} x_1 + a_{m2} x_2 + \ldots + a_{mn} x_n = b_m \end{aligned} \qquad (4.2)$$

ou, simplesmente,

$$\sum_{k=1}^{n} a_{ik} x_k = b_i \qquad (4.3)$$

para m equações e n incógnitas.

Representa-se matricialmente como:

$$\underline{A}\,\underline{x} = \underline{b} \qquad (4.4)$$

onde $\underline{A}$ é a matriz dos coeficientes m · n, $\underline{x}$ é o vetor das incógnitas, e $\underline{b}$, das constantes.

Se $\underline{b}$ é nulo, o sistema se diz *homogêneo*, ocorrendo na análise estrutural no estudo de vibrações e instabilidade do equilíbrio.

Resolver o sistema é determinar um vetor $\underline{x}$, se existir, que satisfaça a todas as equações simultaneamente. Se essa solução existe, o sistema se diz *consistente*, caso contrário, *inconsistente*.

A verificação da consistência se faz construindo as matrizes dos coeficientes e a matriz aumentada pelas constantes:

$$\underline{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \qquad \underline{A}_b = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{pmatrix} \quad (4.5)$$

e calculando a ordem da maior matriz quadrada de determinante não-nulo que se pode extrair de cada uma delas $r(\underline{A})$ e $r(\underline{A}_b)$.

Se forem iguais essas ordens, a solução existirá, mas só será única se forem iguais ao número de equações. Em resumo, segundo (9), ter-se-á a Fig. 4.1.

Neste trabalho, interessará apenas o caso de m = n, que resulta em $\underline{A}$ quadrada, em que se terá solução única se e somente se o determinante de $\underline{A}$ for diferente de zero. A solução teórica é:

$$\underline{x} = \underline{A}^{-1}\,\underline{b} \qquad (4.6)$$

Têm-se assim três problemas de álgebra linear intimamente ligados: o cálculo do determinante de $\underline{A}$, o cálculo de sua inversa e a solução do sistema $\underline{A}\,\underline{x} = \underline{b}$. Os métodos numéricos para solução desses três problemas são em grande número, e trabalha-se intensamente em sua melhoria, pela necessidade de atacar com *precisão* e *velocidade* problemas de ordem cada vez maior. Em resumo, buscam-se programas *eficientes*. Foge do escopo deste trabalho o exame exaustivos desses métodos e avaliação de sua eficiência, atendo-se a uma apresentação rápida dos mais consagrados pelo uso.

Os métodos gerais de solução dividem-se em duas categorias: *exatos* e *iterativos*.

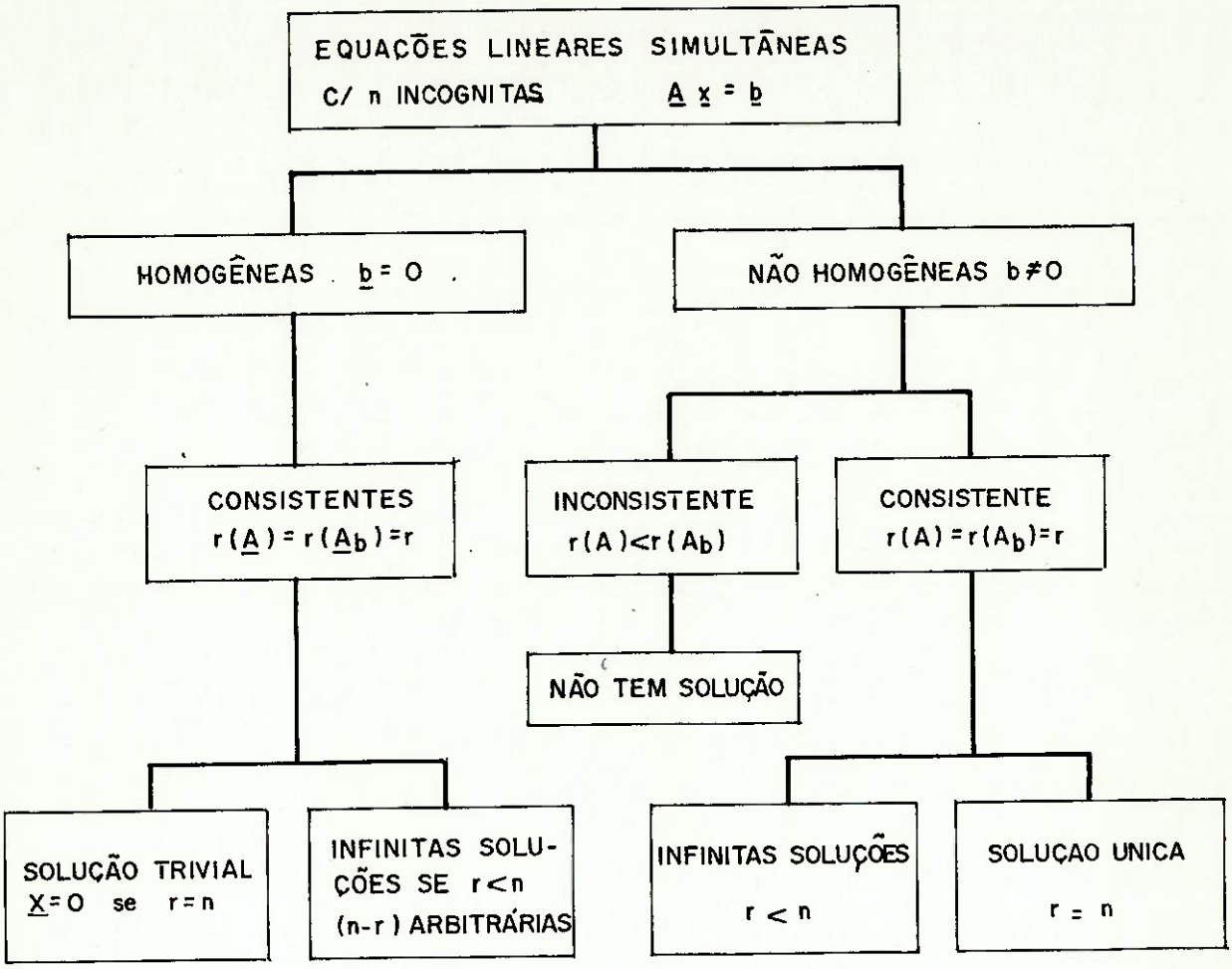


Fig. 4.1

Por *métodos exatos* entendem-se os que resolvem o problema em número finito e calculável de operações aritméticas simples. Se os dados iniciais forem exatos (por exemplo, são todos números racionais representados por frações ordinários) e se os cálculos forem feitos exatamente (por exemplo, pelas regras das operações com frações ordinárias), então a solução será também exata. Nos métodos exatos, o número de operações envolvidas depende apenas do algoritmo e da ordem da matriz e é, como se disse, avaliável.

Historicamente, o primeiro desses métodos, associado à eliminação de incógnitas, é o de Gauss, consistindo numa série de eliminações sucessivas pelas quais o sistema é reduzido a outro equivalente, de matriz triangular, de solução simples por retrosubstituição, sendo o determinante facilmente calculável pelo produto dos elementos da diagonal. Aparecem várias alternativas ao algoritmo de Gauss, principalmente no caso, aqui importante, de matriz simétrica positiva definida, embora o processo básico não exija

essa simetria e definição.

Os *métodos iterativos* consistem em encontrar um limite para soluções aproximadas sucessivas por algum processo uniforme a partir de valores iniciais arbitrários. Nesses métodos, são importantes a convergência e a razão de convergência. O mais antigo deles é o método de Gauss-Seidel, que converge se a matriz é positivo-definida. Essa convergência, em geral, é lenta, levando a recorrer a processos de "relaxação" para aceleração. Tais recursos são, em geral, pouco sistemáticos e de difícil programação, não havendo um procedimento universalmente eficiente, restringindo a aplicação dos métodos iterativos a alguns problemas especialmente favoráveis, como os de matriz de coeficientes quase diagonais e esparsas, ou com coeficientes muito pequenos, com pelo menos um grande em cada linha. Para mais detalhes, ver referência (8).

4.2. Métodos Exatos (por Eliminação)

4.2.1. A Decomposição Clássica A = LDU. Unicidade da Decomposição

Provar-se-á, de início, o "Teorema LDU", de *Turing* (18) e (19).

Teorema

Se todos os principais menores da matriz $\underline{A}$, de ordem n, são não-singulares, então há uma única matriz triangular, de diagonal unitária inferior $\underline{L}$ (em inglês, "*lower*"), uma única matriz diagonal $\underline{D}$, com elementos não-nulos na diagonal, e uma única matriz triangular, de diagonal unitária superior $\underline{U}$ (em inglês, "*upper*") tal que:

$$\underline{A} = \underline{L}\ \underline{D}\ \underline{U} \tag{4.7}$$

Chamar-se-á de d_k o elemento de $\underline{D}$ da linha k. O primeiro elemento dessa linha na equação matricial acima será:

$$a_{1k} = \ell_{11} \cdot d_1 \cdot u_{1k}$$

mas como foi feito

$$\ell_{11} = u_{11} = 1$$

ter-se-á:

$$d_1 = a_{11} \qquad \text{e} \qquad u_{1k} = \frac{a_{1k}}{d_1}$$

Suponha-se agora terem-se encontrado valores de ℓ_{ij}, u_{ij} com $j < i_0$ (ou seja, têm-se as primeiras $i_0 - 1$ linhas de $\underline{L}$ e colunas de $\underline{U}$), bem como os primeiros $i_0 - 1$ elementos diagonais d_k; suponha-se ainda que essas escolhas sejam únicas e que $d_k \neq 0$. Mostrar-se-á como a nova linha de $\underline{L}$ e a nova coluna de $\underline{U}$ e o próximo elemento diagonal $d_{i_0} \neq 0$ devem ser escolhidos para satisfazer as equações da próxima linha de $\underline{A} = \underline{L}\,\underline{D}\,\underline{U}$, e que esta escolha é única. As equações a satisfazer são:

$$\ell_{i_0 i_0} \cdot d_{i_0} \cdot u_{i_0 k} = a_{i_0 k} - \sum_{j<i_0} \ell_{i_0 j} \cdot d_j \cdot u_{jk} \quad (k \geq i_0) \tag{4.8}$$

$$\ell_{i_0 k} \cdot d_k \cdot u_{kk} = a_{i_0 k} - \sum_{j<k} \ell_{i_0 k} \cdot d_j \cdot u_{jk} \quad (k < i_0) \tag{4.9}$$

Os lados direitos das equações acima estão em função de quantidades já determinadas. Quando $k = i_0$, a primeira delas é satisfeita fazendo d_{i_0} = lado direito, com o que se determina d_{i_0}. As equações para $k > i_0$ podem então ser satisfeitas por um e um só conjunto de valores de u_{i_0}, se $d_{i_0} \neq 0$. As equações para $k < i_0$ também só se satisfazem para um e um só conjunto de valores $\ell_{i_0 k}$, desde que cada $d_k \neq 0$. O novo elemento diagonal d_{i_0} não é nulo, porque o menor principal de ordem i_0 de $\underline{A}$ é igual ao produto dos primeiros i_0 elementos diagonais d_k, completando a demonstração.

Caso se disponha da decomposição $\underline{A} = \underline{L}\,\underline{D}\,\underline{U}$, pode-se escrever o sistema $\underline{A}\,\underline{x} = \underline{b}$ como $\underline{L}\,\underline{D}\,\underline{U}\,\underline{x} = \underline{b}$, e sua solução se obtém nos seguintes dois passos:

$$\underline{L}\,\underline{y} = \underline{b} \qquad \text{e} \qquad \underline{DUx} = \underline{y} \tag{4.10}$$

em cada caso calculando a solução de um sistema de equações de matriz de coeficientes triangular, o que se consegue por simples retrosubstituição.

A montagem dessas matrizes triangulares se faz na seqüência conhecido do *algoritmo de Gauss*: adicionam-se múltiplos da primeira equação às outras todas, de forma a eliminar x_1 em todas menos na primeira. Em seguida, adicionam-se múltiplos da segunda às outras, para eliminar x_2 das equações exceto da primeira e da segunda, e assim por diante, até se obter a matriz triangular superior $\underline{DU}$ após $n - 1$ passos. As operações necessárias, em forma matricial, são:

$$(\underline{L}_{n-1})^{-1} \ldots (\underline{L}_2)^{-1} \cdot (\underline{L}_1)^{-1} \cdot \underline{A} = \underline{D}\,\underline{U} \tag{4.11}$$

onde:

$$(\underline{L}_i)^{-1} = \begin{bmatrix} 1 & & & \\ & \ddots & & \text{elementos não-mostrados nulos} \\ & & 1 & \\ & & -\ell_{i+1,i} & \ddots \\ & & -\ell_{i+2,i} & \ddots \\ & & -\ell_{ni} & 1 \end{bmatrix}; \quad \ell_{i+j,i} = \frac{a^{(i)}_{i+j,i}}{a^{(i)}_{ii}} \tag{4.12}$$

significa-se por $a^{(i)}$ que se estão utilizando no cálculo os elementos não da matriz original, mas da matriz modificada:

$$(\underline{L}_{i-1})^{-1} \ldots (\underline{L}_2)^{-1} \cdot (\underline{L}_1)^{-1} \cdot \underline{A} \tag{4.13}$$

Os elementos de L_i são obtidos trocando os sinais dos elementos fora da diagonal de $(\underline{L}_i)^{-1}$. Com isso, obtêm-se:

$$\underline{A} = \underline{L}_1 \cdot \underline{L}_2 \ldots \underline{L}_{n-1}\,\underline{D} \cdot \underline{U} \quad \text{ou} \quad \underline{A} = \underline{L}\,\underline{D}\,\underline{U} \tag{4.14}$$

em que

$$\underline{L} = \underline{L}_1\,\underline{L}_2 \ldots \underline{L}_{n-1} \tag{4.15}$$

É claro que, na prática, não se montam e guardam as matrizes intermediárias $\underline{L}_i$, operando-se, ao invés, sempre sobre os valores anteriores $a^{(i)}_{ij}$, substituindo no lugar por eles ocupados os novos valores $a^{(i+1)}_{ij}$.

O trabalho total de cálculo na solução de um sistema n · n, com a matriz completa e não-simétrica, segundo (5) e (18), é de cerca de $n^3/3$ multiplicações ou divisões no processo de eliminação, mais $n^2/2$ operações adicionais correspondentes às retrosubstituições finais (não se considerou nessa estimativa o tempo gasto em somas, subtrações e guarda e recuperações de valores, insignificante em face das operações de multiplicação e divisão).

É claro que, se já se resolveu um sistema e queira-se resolver outro com a mesma matriz de coeficientes e um novo vetor constante, a fase de eliminação não se repetirá, e bastarão as $n^2/2$ operações finais de retrosubstituição.

Se a intenção for executar a inversão de uma matriz, a seqüência indicada será, após a decomposição completada, inverter as matrizes $\underline{L}$ e $\underline{D}\,\underline{U}$ por retrosubstituição, para depois multiplicá-las para obtenção da inversa da matriz original, implicando em um trabalho total de cerca de n^3 multiplicações e divisões.

A solução de sistemas de equações utilizando a inversa da matriz de coeficientes é muito dispendiosa e raramente feita, em que pese o atrativo de que, uma vez obtida essa inversa, podem-se obter soluções para um número qualquer de vetores de constantes por simples multiplicação da inversa por cada um deles.

4.2.2. O Caso da Matriz Simétrica Positivo-Definida

Os problemas estruturais sempre levam a sistemas de equações de equilíbrio cuja matriz de coeficientes é simétrica e positivo-definida, em que $\underline{U} = \underline{L}^t$, fazendo com que se possa decompor a matriz na forma:

$$\underline{A} = \underline{L}\ \underline{D}\ \underline{L}^t \qquad (4.16)$$

sendo apenas necessário trabalhar e guardar os elementos de $\underline{D}\,\underline{L}^t$, resultando num trabalho final de cálculo em torno de $n^3/6$ multiplicações e divisões na decomposição, mais da ordem de $n^2/2$ operações na retrosubstituição final, segundo *(18)*.

Nos problemas normais do MEF, um número elevado dos coeficientes fora da diagonal principal é nulo, e os restantes em geral se agrupam numa faixa em torno daquela diagonal, reduzindo ainda mais drasticamente o número de operações a executar e o espaço de memória a utilizar, levando a algoritmos de grande eficiência, uma exemplo dos quais é detalhado no final deste capítulo, conforme referência *(3)*.

Como no caso da seção anterior, as retrosubstituições finais se fazem nas mesmas duas etapas:

$$\underline{L}\,\underline{y} = \underline{b} \quad \text{e} \quad \underline{D}\,\underline{L}^t\,\underline{x} = y \qquad (4.17)$$

ambos envolvendo apenas matrizes triangulares.

Cabe citar, por seu valor histórico, a clássica solução para sistemas simétricos devida a *Cholesky* (11) e (19), em que se decompõe a matriz de coeficientes em:

$$\underline{A} = \underline{\hat{L}}\ \underline{\hat{L}}^t \qquad (4.18)$$

de forma única, como se pode provar pela mesma seqüência de raciocínio utilizada no teorema de *Turing* (18). Tem-se:

$$\underline{\hat{L}} = \underline{L}\ \underline{D}^{1/2} \qquad (4.19)$$

É claro que na prática não se obtém $\underline{\hat{L}}$ a partir de $\underline{L}$ e $\underline{D}$ e sim de forma direta.

Embora a utilização do processo seja mais recomendável para sistemas positivo-definidos em que os elementos da diagonal principal são positivos, ele pode ser empregado em casos de matrizes simétricas não-positivo-definidas, envolvendo a raiz quadrada de números negativos, resultando em números imaginários puros, mas não complexos (9).

O grande senão do processo de Cholesky é envolver a operação de extração de raiz quadrada que, especialmente para computadores de poucos recursos, pode ser muito mais demorada e imprecisa que as demais operações elementares.

Como fecho da discussão dos processos exatos, cita-se o resultado de Klyuyev e Kokovkim-Shcherbac, apresentado em (18), de que o algoritmo de Gauss é o processo mais rápido de todos, quando se trabalha com linhas e colunas completas.

4.3. Sub-Rotinas para Microcomputador, em BASIC, para Solução de Sistemas com Matriz Simétrica, Positivo-Definida e Bandeada

4.3.1. Comentários

Lista-se, a seguir, exemplo de sub-rotinas para microcomputador pessoal da linha TRS-80, na linguagem BASIC, para solução de sistemas de equações de matriz de coeficientes simétrica, positivo-definida e bandeada.

O algoritmo básico é a eliminação de Gauss realizada na sub-rotina DECOMPOSIÇÃO LDLT (inicia-se na instrução 1000 da listagem) que triangula-

riza a matriz. A obtenção das incógnitas propriamente dita é feita por retrosubstituição na sub-rotina RETRO (instrução 1500). Esta segunda fase pode ser repetida para qualquer novo vetor carregamento sem se refazer a decomposição.

Em coerência com a forma econômica como se dispuseram os coeficientes na seção 3.2, também estas rotinas operam apenas com os coeficientes não-nulos, da diagonal principal para cima até a "linha do horizonte" ("*sky-line*"). Para utilização deste programa, é necessário que as colunas de elementos significativos da matriz já estejam estocadas de forma contínua em um vetor coluna A, acompanhado de um vetor MA que forneça o endereço dos coeficientes pertencentes à diagonal principal. Os detalhes da construção desses vetores já foram vistos em 3.2.

Além da economia de localização de memória nesse procedimento, é lógico que as operações são realizadas sobre um número bem menor de valores.

Os ganhos em velocidade e precisão são óbvios, apesar da considerável sofisticação das verificações que o algoritmo tem que fazer para definição dos elementos sobre os quais operar.

4.3.2. Listagem

```
1000 REM SUB DECOMPOSICAO LDLT
1005 FOR N=1 TO NQ
1010 KN=MA(N):KL=KN+1:KU=MA(N+1)-1:KH=KU-KL
1015 IF KH<0 GOTO 1125
1020 IF KH=0 GOTO 1085
1025 K=N-KH:IC=0:KT=KU
1030 FOR J=1 TO KH
1035 IC=IC+1:KT=KT-1:KI=MA(K):ND=MA(K+1)-KI-1
1040 IF ND<=0 GOTO 1075
1045 IF IC>=ND THEN KK=ND ELSE KK=IC
1050 C=0
1055 FOR L=1 TO KK
1060 C=C+A(KI+L)*A(KT+L)
1065 NEXT L
1070 A(KT)=A(KT)-C
1075 K=K+1
1080 NEXT J
1085 K=N:B=0
1090 FOR KK=KL TO KU
1100 K=K-1:KI=MA(K):C=A(KK)/A(KI)
1105 B=B+C*A(KK)
1110 A(KK)=C
1115 NEXT KK
1120 A(KN)=A(KN)-B
1125 IF A(KN)<=0 PRINT"NAOPOSDEF EQ";N;" PIVO";A(KN):STOP
1130 NEXT N
1135 RETURN


1500 REM SUB RETRO
1505 FOR N=1 TO NQ
1510 KL=MA(N)+1:KU=MA(N+1)-1
1515 IF (KU-KL)<0 GOTO 1545
1520 K=N:C=0
1525 FOR KK=KL TO KU
1530 K=K-1:C=C+A(KK)*F(K)
1535 NEXT KK
1540 F(N)=F(N)-C
1545 NEXT N
1550 FOR N=1 TO NQ
1555 K=MA(N):F(N)=F(N)/A(K)
1560 NEXT N
1565 IF NQ=1 RETURN
1570 N=NQ
1575 FOR L=2 TO NQ
1580 KL=MA(N)+1:KU=MA(N+1)-1
1585 IF (KU-KL)<0 GOTO 1610
1590 K=N
1595 FOR KK=KL TO KU
1600 K=K-1:F(K)=F(K)-A(KK)*F(N)
1605 NEXT KK
1610 N=N-1
1615 NEXT L
1620 RETURN
```

5. ERROS DE ARREDONDAMENTO EM COMPUTAÇÃO DIGITAL

5.1. Introdução

Antes de enfocar a análise de erros e suas causas na solução numérica de sistemas de equações lineares simultâneas, vai-se, neste capítulo, abordar brevemente os conceitos básicos da análise de erros das soluções numéricas de problemas gerais de Matemática, devidos ao arredondamento na representação de números reais em um sistema de posição. No capítulo 6, o argumento será especializado ao caso dos sistemas de equações lineares.

A intenção final é determinar quantos dígitos significativos devem ser utilizados na representação de números em uma dada base ao longo de uma computação para obtenção de resultados dentro de certo nível de precisão, ou, o que é o mesmo, dado um certo número de dígitos colocados à disposição por um determinado equipamento, qual a precisão que se terá ao fim de um cálculo, e mesmo se o resultado terá qualquer significado. Esse procedimento equivale a estabelecer uma *delimitação de erro*.

5.2. Representação Digital de Números em Computadores

Os números com que se trabalha nos computadores digitais são chamados de *números digitais*, para distinguí-los dos números reais ordinários. Dois dos modos de representação disponíveis são: números de *ponto fixo* e de *ponto flutuante*.

Na representação de ponto fixo, observam-se as convenções (13):

a) Um número digital de ponto fixo é um agregado de dígitos de t casas, base B, com sinal.

b) O ponto (vírgula, na tradição brasileira) do número será colocado sempre na extrema esquerda (isto é, i = 0).

Assim,

$$x = s\,(a_1, a_2, \ldots, a_t) = s \cdot \sum_{i=1}^{t} B^{-i} \cdot a_i \qquad (5.1)$$

onde: $s = +1$ ou -1 e cada um dos elementos a_i pode assumir um valor entre 0 e B - 1, sendo B chamado de *base* (em geral B = 2 ou B = 10).

Pela posição adotada para a vírgula, é claro que $-1 \leq x \leq 1$.

Uma representação muito mais conveniente, disponível universalmente em todos os computadores modernos, a qual será utilizada neste trabalho, é o chamado formato de *ponto flutuante*. Consiste em uma fração com um número fixo de dígitos (t) após a vírgula, denominada mantissa (a), com sinal, multiplicada por uma base (B) elevada a um expoente (e):

$$x = s\ (0, a_1\ a_2\ \ldots\ a_t) \cdot B^e \qquad s \begin{cases} + \\ \text{ou} \\ - \end{cases} \qquad (5.2)$$

ou

$$x = s\ (a) \cdot B^e$$

sendo que:

- a mantissa (a) variará entre

$$B^{-1} \leq a \leq 1$$

- o primeiro dígito a_1 será sempre maior que zero;
- os demais dígitos de a_2 a a_t variarão entre 0 e B - 1;
- a base B em geral será 2 ou 10;
- o expoente variará entre limites que dependerão de cada máquina.

Para que se represente neste formato um número real qualquer, com número de dígitos significativos maior que t, será preciso arrendondá-lo, isto é, decidir qual será o último dígito significativo a_t segundo um *critério de arredondamento*.

Quando se trabalha na base 10, o critério de arredondamento mais indicado (15) seria:

a) se o próximo dígito à direita no número real for menor do que 5, o número é truncado, isto é, o algarismo de ordem mais baixa é conservado.

Exemplo para t = 3: 0,33333... 0,333

1,41421... 1,41

b) se o próximo dígito à direita de a_t no número real for maior ou igual a 5, o número será truncado, e adicionada ao resultado uma unidade ao algarismo de ordem mais baixa.

Exemplo para t = 3:	0,925925	0,926
	1,99746	2,00

Critério similar se aplicaria aos números de base binária (20), com a vantagem de só termos dois algarismos disponíveis: 0 ou 1. É, no entanto, corrente nos computadores simplesmente truncar os dígitos excedentes, sem qualquer alteração de a_t.

Como é de conhecimento geral, os computadores eletrônicos digitais aceitam e produzem respostas na base 10, mas, na verdade, internamente só podem trabalhar na base binária, representando números, sinais ou letras por seqüências de dígitos 0 ou 1, que são os dois únicos estados elétricos que pode reconhecer: ligado ou desligado. Dessa forma, vai-se depender, fundamentalmente, da máquina ("*hardware*", em inglês) para estabelecer o número de dígitos binários ("*bits*", em inglês) disponíveis para representação da mantissa e do expoente, sempre na base binária, e, em conseqüência, saber os limites de magnitude dos números representáveis.

A título de exemplo, será examinado o caso de algumas máquinas populares, segundo (17):

- No IBM-7094, cada número é representado, em "precisão simples", por uma palavra ("*byte*", em inglês), de 36 dígitos binários ("*bits*"):

  Primeiro dígito: sinal (0 para positivo, 1 para negativo)
  Oito dígitos seguintes: o expoente
  Demais dígitos: a mantissa

  É interessante notar que nessa máquina se utiliza o número 128 (10000000) como expoente de referência, equivalendo a e = 0, contando a partir dele, para frente ou para trás, a posição da vírgula, evitando-se, assim, o uso de expoentes negativos.

  Exemplo:

número decimal	binário equivalente	representação interna
+ 1,465	+ 1,01110111000 ...	0 10000001 101100001010 ...
+ 0,025	+ 0,00000110011 ...	0 01111011 110011000000 ...

- No IBM-360, a palavra ("*byte*") de 32 dígitos binários ("*bits*") é utilizada na representação de números em precisão simples, na forma:

  Primeiro dígito: sinal (0 para positivo, 1 para negativo)
  Sete dígitos seguintes: expoente
  24 dígitos seguintes: mantissa

  Se se operar nesta máquina com a chamada "*precisão dupla*", ter-se-á mais uma palavra inteira (32 "*bits*") adicional para a parte fracionária, que passa a dispor de 56 "*bits*". Em termos de números decimais, a precisão "simples" desse computador é de 6 dígitos decimais significativos, e a "dupla" é de 16 dígitios decimais significativos.

  Ainda nessa máquina, o expoente de referência é 64 (1000000, em representação binária).

- Os computadores comerciais de grande porte disponíveis no Brasil, de maior comprimento de palavra, são os da Control Data Corporation, com 14 e 28 dígitos decimais significativos em precisão simples e dupla, respectivamente.

- Nos microcomputadores de baixo custo mais populares, utiliza-se atualmente (*12*) palavra ("*byte*") de 8 dígitos binários ("*bits*"). Como exemplo, tome-se o TRS-80, no qual cada número é representado em precisão simples por 4 "*bytes*" (32 "*bits*") e, em precisão dupla, por 8 "*bytes*" (64 "*bits*"). Isso equivale a ter-se, em precisão simples, a possibilidade de até 7 dígitos decimais significativos, e, em precisão "dupla", até 16 dígitos decimais significativos, com grande sacrifício de espaço de memória.

5.3. Arredondamento na Computação Digital Elementar

A realização de operações aritméticas fundamentais por um computador digital difere, como se poderia antever, da computação com números reais, já que se estará operando com números digitais de precisão finita, e vão-se obter resultados com a mesma precisão finita.

Alguns autores, como *Von Neuman (13)*, chegam a chamar tais operações de "*pseudo-operações*", uma vez que do outro lado do sinal de igualdade da operação está um resultado que envolve um arredondamento, no caso geral, para sua representação digital. A igualdade só se satisfaria, a rigor, adicionando ao resultado o erro de arredondamento cometido na sua representação.

Os detalhes das operações aritméticas fundamentais em ponto flutuante diferem um pouco de um computador para outro, mas em geral são como se segue *(17)*, *(20)*. Sejam os operandos x_1 e x_2, onde:

$$x_1 = 2^{e_1} \cdot a_1 \text{ (ou } 10^{e_1} \cdot a_1) \quad , \quad x_2 = 2^{e_2} \cdot a_2 \text{(ou } 10^{e_2} \cdot a_2) \qquad (5.2)$$

• Na adição:

Calcula-se $e_1 - e_2$, supondo $x_1 > x_2$.

1) se $e_1 - e_2 > t$, então x_2 é muito pequeno e não afeta o resultado para t dígitos significativos:

$$x_1 + x_2 = x_1$$

2) se $e_1 - e_2 \leq t$, então a_2 tem sua vírgula acertada em $e_1 - e_2$ casas para a direita e a soma das mantissas é calculada exatamente com até 2t dígitos, sendo depois a vírgula acertada junto com o expoente para o padrão de representação, truncando-se os dígitos excedente a t na mantissa.

Se $x_2 > x_1$, basta inverter os papéis na exposição acima.

Em resumo, o valor computado da soma de dois números é sempre o que se obteria calculando a soma exata, e truncando-a para o número t de dígitos. Se a soma exata normalizada é $2^{e_3} \cdot a_3$ (ou $10^{e_3} \cdot a_3$), então é evidente que o módulo de erro está delimitado por $2^{e_3} \cdot 0{,}5 \cdot 2^{-t}$ (ou $10^{e_3} \cdot 0{,}5 \cdot 10^{-t}$). Uma forma mais usual de delimitação será o cálculo do erro relativo. Agora o módulo da soma exata está entre $0{,}5 \cdot 2^{e_3}$ e 2^{e_3}(ou $0{,}1 \cdot 10^{e_3}$ e 10^{e_3}), e, portanto, ter-se-á:

$$\text{valor calculado de } x_1 + x_2 = (x_1 + x_2)(1 + \varepsilon) \qquad (5.3)$$

$$|\varepsilon| \leq 2^{-t} \text{ (binário)} \qquad \text{ou} \qquad |\varepsilon| \leq 0{,}5 \cdot 10^{1-t} \text{ (decimal)} \qquad (5.4)$$

Para x_1 ou x_2 nulo, a soma não envolve arredondamento algum, e o erro é nulo.

Pode-se expressar o resultado na forma: o valor computado da soma de x_1 e x_2 é a soma exata de dois números $x_1(1 + \varepsilon)$ e $x_2(1 + \varepsilon)$ para algum valor de ε que satisfaça $|\varepsilon| \leq 2^{-t}$ (ou $|\varepsilon| \leq 0{,}5 \cdot 10^{1-t}$).

A subtração é exatamente paralela.

- Na multiplicação:

Calcula-se $e_3 = e_1 + e_2$, e o produto exato das mantissas $a_1 \cdot a_2$ com 2t casas, que está sempre no intervalo

$$\frac{1}{4} \leq a_1 \cdot a_2 \leq 1 \qquad \text{ou} \qquad (0{,}01 \leq a_1 \cdot a_2 \leq 1) \tag{5.5}$$

e será então normalizado, levando a vírgula para a esquerda e ajustando o expoente. O produto resultante com 2t dígitos é agora truncado de acordo. Se x_1 ou x_2 é nulo, o resultado também o é.

Como na soma ou na subtração, o produto computado é obtido arredondando o produto exato para t casas, e, portanto, ter-se-á:

$$\text{valor calculado de } x_1 \cdot x_2 = x_1 \cdot x_2 \ (1 + \varepsilon) \tag{5.6}$$

$$|\varepsilon| \leq 2^{-t} \text{ (binário)} \quad \text{ou} \quad |\varepsilon| \leq 0{,}5 \cdot 10^{1-t} \text{(decimal)} \tag{5.7}$$

Isso pode ser expresso na forma: o produto computado é o exato produto de dois números $x_1(1 + \varepsilon)$ e x_2, ou x_1 e $x_2(1 + \varepsilon)$, ou ainda, $x_1(1 + \varepsilon)^{1/2}$ e x $(1 + \varepsilon)^{1/2}$, onde $|\varepsilon| \leq 2^{-t}$ (ou $0{,}5 \cdot 10^{1-t}$).

- Na divisão:

Calcula-se $e_3 = e_2 - e_1$. a_1 é completado com t zeros para ser representado com 2t dígitos. Se $|a_1| > |a_2|$, a_1 é deslocado uma casa para a direita, e e_3 é aumentado de 1. Divide-se o número assim obtido por a_2 e tem-se um quociente já corretamente arredondado para t dígitos com módulo entre 1/2 e 1 (ou 0,1 e 1) em forma normalizada.

Se $x_1 = 0$ e $x_2 \neq 0$, então $x_1/x_2 = 0$. Se $x_2 = 0$, a operação não é realizada. Em geral, tem-se:

$$\text{valor computado de } x_1/x_2 = (x_1 \ / \ x_2) \ (1 + \varepsilon) \tag{5.8}$$

$$|\varepsilon| \leq 2^{1-t} \text{ (binário)} \quad \text{ou} \quad |\varepsilon| \leq 0{,}5 \cdot 10^{1-t} \text{ (decimal)} \tag{5.9}$$

que se pode expressar como: o quociente computado de x_1 e x_2 é o exato quociente de $x_1(1 + \varepsilon)$ e x_2 ou x_1 e $x_2(1 + \varepsilon)$ para algum ε satisfazendo $|\varepsilon| \leq 2^{-t}$ (ou $0{,}5 \cdot 10^{1-t}$).

De um modo geral, essas operações simples têm pequeno erro relativo, devendo ainda entender-se com clareza que o que se estabeleceu acima foram os *limites máximos* que esses erros podem atingir, não sendo provável que ocorram com freqüência.

5.4. Conceitos Básicos e Convenções da Análise de Erros na Computação Digital

Duas formulações para análise de erros são utilizadas e se dizem prospectiva ("*forward*", em inglês) e retrospectiva ("*backward*", em inglês).

- Análise Prospectiva:

  Numa seqüência de cálculo, passa-se por uma série de equações, em cada fase obtendo-se uma nova quantidade x em função de quantidades anteriormente calculadas ou de parâmetros iniciais $a_1, a_2, \ldots, a_n$

  $$x = f(a_1, a_2, \ldots, a_n) \tag{5.10}$$

  a função devendo sempre ser uma sucessão de operações fundamentais. Devido aos arredondamentos nas operações, o *valor computado* de x será diferente do que seria obtido exatamente. Na análise prospectiva, chama-se de $\bar{x}$ o valor computado de x, e procura-se delimitar o erro $\bar{x} - f(a_i)$. Comparam-se, assim, x e $\bar{x}$.

  Percebe-se, nessa comparação, a óbvia fraqueza do método, uma vez que no caso geral não se dispõe do valor exato de x, e pode-se ainda notar que, se se tiver tido o trabalho de calculá-lo, não haverá muito sentido na análise de erro.

- Análise Retrospectiva:

  Aqui não se está preocupado em comparar o valor exato com o computado, mas sim em mostrar que o valor computado é a solução exata de $f(a_1 + \varepsilon_1, a_2 + \varepsilon_2, \ldots, a_n + \varepsilon_n)$ para algum valor de ε_i, e dar delimitações para esses ε_i.

Como neste enfoque não se tem nunca a oportunidade de calcular o valor exato x, não há sequer a necessidade de diferenciá-lo do valor computado $\bar{x}$, que é o único que se terá, tornando desnecessário utilizar a barra superior na representação. Assim, um dado processo será representado por:

$$x = f(a_1 + \varepsilon_1,\ a_2 + \varepsilon_2,\ \ldots,\ a_n + \varepsilon_n) \qquad (5.11)$$

seguido das inequações satisfeitas por ε_i.

Pode-se objetar que este enfoque é incompleto, por não dar uma estimativa da diferença entre a solução exata e a computada, mas sempre se terá um estágio final em que essa diferença será avaliada.

5.5. Limitação Básica da Computação Digital

Haverá, em geral, um limite definido na precisão de cálculo possível quando se está trabalhando com números representados por t dígitos significativos.

Qualquer cálculo tem por finalidade encontrar um conjunto de números-resposta a partir de um grupo de números-dados.

Se o problema é, como se terá no próximo capítulo, resolver o sistema de equações $\underline{A}\ \underline{x} = \underline{b}$, os elementos a_{ij} da matriz $\underline{A}$ e b_i do vetor $\underline{b}$ são os dados, e os elementos x_i de $\underline{x}$ são as respostas.

Pode acontecer de os parâmetros iniciais serem exatamente conhecidos mas não exatamente representáveis com os dígitos disponíveis. Essas restrições a t dígitos iniciais já impede qualquer precisão final melhor do que esta mesmo antes de qualquer cálculo.

Pode ser, por exemplo, que a matriz dos coeficientes $\underline{A}$ de um sistema já seja resultado de um produto de duas outras matrizes, cujos elementos são números de t dígitos. Nesse caso, os elementos de $\underline{A}$ teriam que ser escritos com no mínimo 2t dígitos para exata representação.

Como já foi comentado, se os dados são oriundos de observação física, carregam em si toda a imprecisão inerente a esses processos. O computador estará resolvendo, assim, um problema aproximado ao real, com o agravante

de tais erros serem de ordem muito mais grosseira que os devidos ao arredondamento a t dígitos.

5.6. Problemas Mal-Condicionados

Da seção prévia, conclui-se que, a menos que se dispusesse de dados iniciais exatamente definidos e representáveis, começam-se os cálculos com o que se chama uma aproximação de t dígitos do problema real.

Vai-se considerar agora a magnitude do efeito de tais erros iniciais sobre a resposta. Se perturbações iniciais relativamente pequenas levam a erros comparativamente grandes na resposta, diz-se que há um *problema mal-condicionado* numericamente.

Tais problemas são, por todos os aspectos, indesejáveis, pois, para garantir uma certa precisão na resposta, é preciso partir de parâmetros iniciais muito mais precisos. Isso pode ser mesmo impossível, na prática, pela própria limitação da máquina com que se trabalha ou pelo instrumental de medida de dados.

Deve ser colocado aqui, com bastante clareza, que a condição do problema será verificada com respeito ao tipo de resposta que se quiser. Em outras palavras não se pode dizer simplesmente que uma certa matriz $\underline{A}$ é *mal-condicionada* sem se saber a que fim ela se destina. Vai depender de se ela faz parte da solução de um sistema $\underline{A}\,\underline{x} = \underline{b}$, por exemplo, ou se se está procurando seus valores e vetores próprios. Ela pode ser mal-condicionada num processo e não no outro.

5.7. Conceito de Número de Condição

As perguntas são: pode-se saber de antemão se o problema é ou não mal-condicionado? E em que grau?

Se se quiser *quantificar* a sensibilidade de cada uma das respostas $x_1, x_2, \ldots, x_n$ de um problema a variações nos dados $a_1, a_2, \ldots, a_m$ do mesmo, ter-se-á que calcular as seguintes razões de variação:

$$k_{ij} = \frac{\partial x_i}{\partial a_j} \qquad (i = 1, \ldots, n; \quad j = 1, \ldots, m) \qquad (5.12)$$

É claro que os problemas da prática gerariam, no caso geral, uma quantidade grande demais desses valores para serem úteis ao analista, levando a estimativas mais modestas de conjunto que serão denominados *números de condição*.

O conceito de número de condição não deve ser muito rígido, de forma a poder-se defini-lo de várias formas alternativas, tais como as seguintes, propostas por *Wilkinson (20)*:

a) Comparação da magnitude absoluta do total dos erros da solução com a magnitude do total dos desvios nos dados:

$$\sqrt{\sum \Delta x_i^2} \leq K \sqrt{\sum \Delta a_i^2} \qquad (5.13)$$

b) Comparação do erro absoluto em cada uma das respostas com o total dos desvios nos dados:

$$|\Delta x_j| \leq K_j \sqrt{\sum \Delta a_i^2} \qquad (5.14)$$

c) Comparação do erro relativo de cada resposta com o erro relativo nos parâmetros:

$$\left| \frac{\Delta x_j}{x_j} \right| \leq K_j \sqrt{\sum \left(\frac{\Delta a_i}{a_i}\right)^2} \qquad \text{para } x_j \neq 0 \qquad (5.15)$$

d) Comparação do erro relativo de cada resposta e a resposta total com o erro relativo nos parâmetros e o total deles:

$$\frac{|\Delta x_j|}{\sqrt{\sum x_j^2}} \leq K_j \sqrt{\frac{\sum \Delta a_i^2}{\sum a_i^2}} \qquad (5.16)$$

Como já foi várias vezes repetido, se não for possível representar os m parâmetros iniciais a_i com os t dígitos disponíveis, já se estará introduzindo m valores a_i' tais que:

$$\left| \frac{a_i' - a_i}{a_i} \right| \leq 2^{-t} \qquad \text{(binário)} \tag{5.17}$$

Dispondo de um número de condição do tipo (b), ter-se-ia, então:

$$| \Delta x_j | \leq \frac{1}{2} \cdot K_j \cdot m^{1/2} \cdot 2^{-t} \tag{5.18}$$

Erros dessa ordem são inerentes à computação digital, mesmo para processos muito bem condicionados. Se K_j é aproximadamente 2^{K_j}, então haverá erros de ordem de $2^{K_j - t}$. Como se observa, os números de condição não dependem, em geral, do número de dígitos disponíveis à representação, mas de características intrínsecas aos dados do problema a resolver.

6. SOLUÇÃO DE SISTEMAS DE EQUAÇÕES LINEARES SIMULTÂNEAS: O PROBLEMA PRÁTICO DA ANÁLISE DE ERROS

6.1. Conceito de Sistema Mal-Condicionado

Na solução numérica de sistemas de equações lineares simultâneas, com representação de números num sistema de número fixo de dígitos significativos, mesmo utilizando métodos "exatos" como os estudados no capítulo 4, está-se sujeito a erros de solução devidos a:

a) Arredondamento nas operações elementares, fazendo desaparecer dígitos significativos. Tal problema, como se viu, pode ser contornado, pelo menos teoricamente, pela utilização de aritmética de maior precisão (precisão "dupla", por exemplo).

b) Perturbações nos dados iniciais dentro dos limites da precisão adotada. Em detalhe: a solução teórica do sistema $\underline{A}\,\underline{x} = \underline{b}$ é $\underline{x} = \underline{A}^{-1}\,\underline{b}$, e existe se e somente se o determinante de $\underline{A}$ não é nulo. Há casos em que pequenas perturbações na matriz, dentro dos limites de precisão da representação numérica, podem levar o determinante a ser nulo ou não, comprometendo assim a própria certeza da existência ou não de solução única.

Outras situações a que se pode ser levado: um sistema pode ter a propriedade de um vetor solução $\underline{x}$ ser diferente da solução real, mesmo que, substituído no sistema original, resulte em resíduos pequenos com relação ao vetor conhecido $\underline{b}$; pequenas perturbações em $\underline{b}$ podem levar a grandes variações em $\underline{x}$.

Esses fenômenos têm sido conhecidos desde os tempos de Gauss para matrizes de ordem baixa. Com o advento dos computadores, tornou-se viável e necessário abordar problemas de grande porte, motivando os primeiros estudos, nessa área, do que Turing chamou pela primeira vez (em (19)) de "*sistemas mal-condicionados*".

Várias questões se impõem:

1) Pode-se saber de antemão se se está lidando com um sistema particularmente sensível?

2) Essa sensibilidade é uma característica intrínseca da matriz?

Para melhor dramatizar a idéia, considere-se o seguinte exemplo, sugerido por Rutishauser, citado por *Todd (18)*:

$$\begin{pmatrix} 10 & 1 & 4 & 0 \\ & 10 & 5 & -1 \\ & & 10 & 7 \\ & & & 9 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{pmatrix} \qquad (6.1)$$

Vai-se testar a sensibilidade da solução do sistema acima, de matriz simétrica positivo-definida, a variações nos elementos do vetor constante e da matriz propriamente dita.

De início, será adotado um vetor $\underline{b}$, calculando-se a solução exata, e, em seguida, serão impostas perturbações de ordem 10^{-3}, 10^{-2} e 10^{-1} em seus elementos, calculando-se em cada caso a solução $\underline{x}$:

$$\begin{aligned} \underline{b}_1 &= (15 \quad , \quad 15 \quad , \quad 26 \quad , \quad 15 \quad) \\ \underline{b}_2 &= (15{,}001 \quad , \quad 14{,}999 \quad , \quad 25{,}999 \quad , \quad 15{,}001) \\ \underline{b}_3 &= (15{,}01 \quad , \quad 14{,}99 \quad , \quad 25{,}99 \quad , \quad 15{,}01 \quad) \\ \underline{b}_4 &= (15{,}1 \quad , \quad 14{,}9 \quad , \quad 25{,}9 \quad , \quad 15{,}1 \quad) \end{aligned} \qquad (6.2)$$

As soluções para cada um serão:

$$\begin{aligned} \underline{x}_1 &= (1 \quad , \quad 1 \quad , \quad 1 \quad , \quad 1 \quad) \\ \underline{x}_2 &= (1{,}497 \quad , \quad 1{,}91 \quad , \quad -0{,}44 \quad , \quad 2{,}208 \quad) \\ \underline{x}_3 &= (5{,}97 \quad , \quad 8{,}91 \quad , \quad -13{,}4 \quad , \quad 13{,}08 \quad) \\ \underline{x}_4 &= (50{,}7 \quad , \quad 80{,}1 \quad , \quad -143{,}0 \quad , \quad 121{,}8 \quad) \end{aligned} \qquad (6.3)$$

As diferenças são, realmente, notáveis, mesmo em se levando em conta que os elementos da matriz são dados exatamente, sem nenhum problema de arredondamento envolvido.

Vamos agora impor uma perturbação de ordem 10^{-3}, 10^{-2} e 10^{-1} no primeiro dos elementos da matriz, resolvendo o sistema para um mesmo vetor $\underline{b} = (15,\ 15,\ 26,\ 15)$:

$$\begin{aligned}
&\text{para } a_{11} = 10 && \text{obtém-se } \underline{x}_1 = (1 \quad , \ 1 \quad , \ 1 \quad , \ 1 \quad)\\
&\text{para } a_{11} = 10{,}001 && \text{obtém-se } \underline{x}_2 = (0{,}905 \ , \ 0{,}85 \ , \ 1{,}275 \ , \ 0{,}77 \ )\\
&\text{para } a_{11} = 10{,}01 && \text{obtém-se } \underline{x}_3 = (0{,}488 \ , \ 0{,}185 \ , \ 2{,}483 \ , -0{,}244)\\
&\text{para } a_{11} = 10{,}1 && \text{obtém-se } \underline{x}_4 = (0{,}087 \ , -0{,}452 \ , \ 3{,}644 \ , -1{,}217)
\end{aligned} \tag{6.4}$$

Novamente constatam-se diferenças notáveis no resultado.

Uma interessante forma de visualizar o problema, lembrada por *Livesley* em *(10)*, é do ponto de vista geométrico: sistemas mal-condicionados correspondem a hiperplanos quase paralelos, ou, na geometria de duas dimensões, a retas quase paralelas. Por exemplo, os sistemas:

$$\begin{aligned}
x_1 + 2x_2 &= 3 & 1{,}001x_1' + 2x_2' &= 3\\
3{,}001x_1 + 6x_2 &= 9{,}001 & 3x_1' + 6x_2' &= 9
\end{aligned} \tag{6.5}$$

cujos coeficientes diferem por no máximo 10^{-3}, possuem, respectivamente, as soluções:

$$\underline{x} = (1\ ,\ 1) \qquad \text{e} \qquad \underline{x}' = (-1{,}5\ ,\ 1{,}5)$$

Estendendo essa argumentação geométrica para três dimensões, ter-se-ia:

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} \qquad \text{ou} \qquad \begin{aligned} \underline{a}_1^t \underline{x} &= b_1 \\ \underline{a}_2^t \underline{x} &= b_2 \\ \underline{a}_3^t \underline{x} &= b_3 \end{aligned} \tag{6.6}$$

Se os vetores $\underline{a}_1$, $\underline{a}_2$ e $\underline{a}_3$ não são coplanares, isto é $\underline{A}^{-1} \neq 0$, eles constituem uma base de referência em que b_i são as componentes do vetor $\underline{x}$ procurado. Resta saber se esse sistema de referência são eixos "bons" ou "maus" para representação do vetor. Se um dos vetores, $\underline{a}_3$ por exemplo, cair quase no plano de $\underline{a}_1$ e $\underline{a}_2$, os eixos serão "maus", levando a grande imprecisão na determinação das componentes buscadas.

Observa-se que o determinante de um sistema mal-condicionado é, em geral, pequeno, isto é, próximo de zero. Na matriz de Rutishauser vale 1, nos dois sistemas acima, - 0,001 e 0,006, respectivamente. Entretanto esta condição, como observa *Turing* em (19), por si só não caracteriza o mal-condicionamento, já que se podem construir sistemas com determinantes arbitrariamente pequenos, dividindo uma das equações por uma constante. A magnitude relativa dos elementos da matriz tem também importância, como se pode ver do exemplo:

$$\begin{pmatrix} 20 & 0 & 0 \\ & 1 & 0 \\ & & 0{,}05 \end{pmatrix} \quad \text{e} \quad \begin{pmatrix} 20 & 0 & 0 \\ & 0{,}2 & 0 \\ & & 0{,}25 \end{pmatrix} \qquad (6.7)$$

são matrizes de mesmo determinante, mas a primeira é "pior" condicionada que a segunda.

6.2. Números de Condição de uma Matriz

Para caracterização do condicionamento numérico de matrizes de coeficientes de sistemas de equações lineares simultâneas, vários autores têm definido diferentes quantidades, genericamente denominadas *números de condição*, no sentido dado na conceituação do Capítulo 5. Como regra geral, tais números crescem em proporção ao grau de mal-condicionamento do sistema. São quase todos escritos em função das *normas matriciais* vistas no Capítulo 2.

Os números de *Turing* (19) são:

$$\text{- Número N} = 1/n \; \|\underline{A}\|_E \cdot \|\underline{A}^{-1}\|_E = n(\underline{A}) \qquad (6.8)$$

$$\text{- Número M} = n \cdot M(\underline{A}) \cdot M(\underline{A}^{-1}) = m(\underline{A}) \qquad (6.9)$$

Há razoável concordância entre as duas medidas, embora o número M tenda a ser maior, especialmente com matrizes quase diagonais.

Todd define ainda, em (18):

$$\text{Número P} = \frac{\text{máximo valor próprio de } \underline{A}}{\text{mínimo valor próprio de } \underline{A}} = p(\underline{A}) \qquad (6.10)$$

$$\text{Número H} = \|\underline{A}\|_2 \cdot \|A^{-1}\|_2 = h(\underline{A}) \tag{6.11}$$

É fácil mostrar que:

$$\text{Número H} = \sqrt{\frac{\text{maior valor próprio de } \underline{A}\,\underline{A}^t}{\text{menor valor próprio de } \underline{A}\,\underline{A}^t}} \tag{6.12}$$

De onde se tira que

$$h(\underline{A}) = \sqrt{p\,(\underline{A}\,\underline{A}^t)} \tag{6.13}$$

De um modo geral, tem-se que:

$$p(\underline{A}) \leq h(\underline{A}) \tag{6.14}$$

e para o caso particular de matrizes simétricas, tem-se o importante resultado:

$$p(\underline{A}) = h(\underline{A}) \tag{6.15}$$

A dificuldade prática de cálculo de alguns desses números reside na determinação de valores próprios máximos e mínimos de matrizes, operação sempre trabalhosa. Por isso vai-se fazer uso dos fatos de que: estimativas bastante boas da norma espectral são obtidas da norma euclidiana; deve-se lembrar que qualquer das normas é sempre um limitante superior dos valores próprios da matriz; é possível estimar o maior valor próprio da forma utilizável, usando a norma linha, coluna ou a euclidiana.

Já o cálculo do valor próprio mínimo não é tão imediato, e no programa apresentado ao final deste capítulo ele é obtido por iteração inversa após a decomposição $\underline{L}\,\underline{D}\,\underline{L}^t$ de $\underline{A}$.

Alguns exemplos de números de condição:

a) As matrizes ortogonais são as melhor condicionadas, com números de condição todos unitários.

b) Matriz de *Rutishauser (18)*:

$$p(\underline{R}) = h(\underline{R}) = \text{máximo valor próprio/mínimo valor próprio} =$$

$$= \frac{19,1225}{0,0005343} = 35790$$

c) Matriz de *Wilson* (5):

$$\underline{W} = \begin{pmatrix} 10 & 7 & 8 & 7 \\ & 5 & 6 & 5 \\ & & 10 & 9 \\ & & & 10 \end{pmatrix}$$

$$p(\underline{W}) = h(\underline{W}) = \frac{30,28868}{0,01015005} = 2984$$

d) Exemplo de *Bathe*(3):

$$\underline{B} = \begin{pmatrix} 4,855 & -4 & 1 & 0 \\ & 5,855 & -4 & 1 \\ & & 5,855 & -4 \\ & & & 4,855 \end{pmatrix}$$

$$p(\underline{B}) = h(\underline{B}) = \frac{12,9452}{0,000898} = 14416$$

Turing, em seu trabalho (19), dá uma interpretação dentro da teoria das probabilidades para seus números de condição N e M.

Se se considerarem os elementos a_{ij} da matriz $\underline{A}$ como os valores médios que esses elementos, como variáveis independentes, assumiriam, com mesma variância, pequena em comparação com os elementos em si, então os números de Turing mostram quantas vezes a razão do quadrado médio dos erros nas incógnitas pelo quadrado médio das próprias incógnitas excede a razão do quadrado médio das perturbações nos coeficientes do sistema pelo quadrado médio dos próprios coeficientes.

Uma outra interpretação probabilística, dada por *Faddeev* em (5), é que o número H de Todd dá a razão entre o maior e o menor semi-eixo de um elipsóide de dispersão de um vetor cujas componentes são os erros nas incógnitas.

6.3. Delimitação de Erros na Solução de Sistemas de Equações

6.3.1. Sensibilidade a Perturbações no Vetor Independente

Num sistema

$$\underline{A}\ \underline{x} = \underline{b} \tag{6.16}$$

perturbações $\Delta\underline{b}$ motivarão perturbações $\Delta\underline{x}$ na solução, de forma:

$$\underline{A}\ (\underline{x} + \Delta\underline{x}) = \underline{b} + \Delta\underline{b} \tag{6.17}$$

$$\underline{A} \cdot \Delta\underline{x} = \Delta\underline{b} \tag{6.18}$$

$$\Delta\underline{x} = \underline{A}^{-1} \cdot \Delta\underline{b} \qquad \text{se det } \underline{A} \neq 0 \tag{6.19}$$

Utilizando agora normas consistentes, tem-se:

$$\|\Delta\underline{x}\| = \|\underline{A}^{-1} \cdot \Delta\underline{b}\| \leq \|\underline{A}^{-1}\| \cdot \|\Delta\underline{b}\| \tag{6.20}$$

$$\frac{\|\Delta\underline{x}\|}{\|\Delta\underline{b}\|} \leq \|\underline{A}^{-1}\| \tag{6.21}$$

Para estimar a sensibilidade relativa, faz-se:

$$\|\underline{b}\| = \|\underline{A}\ \underline{x}\| \leq \|\underline{A}\| \cdot \|\underline{x}\| \tag{6.22}$$

$$\|\underline{x}\| \geq \|\underline{b}\| \cdot \|\underline{A}\|^{-1} \tag{6.23}$$

Donde:

$$\frac{\|\Delta\ \underline{x}\|}{\|\underline{x}\|} \leq \frac{\|\underline{A}^{-1}\| \cdot \|\ \Delta\ \underline{b}\|}{\|\underline{A}\|^{-1}\ \|\underline{b}\|} = \|\underline{A}\| \cdot \|\underline{A}^{-1}\| \cdot \frac{\|\ \Delta\ \underline{b}\|}{\|\ \underline{b}\|} \tag{6.24}$$

Fica assim demonstrado que $\|\underline{A}\| \cdot \|\underline{A}^{-1}\|$ definida na seção 6.2 como um dos *números de condição*, é de fato a quantidade decisiva na amplificação dos erros. O número de condição mais utilizado é calculado com a norma espectral:

$$h(\underline{A}) = \|\underline{A}\|_2 \cdot \|\underline{A}^{-1}\|_2 \tag{6.25}$$

sendo, por essa razão, denominado *número de condição espectral*.

Quando esse número é muito grande, o resultado obtido para delimitação de erros é, em geral, pessimista para a maioria das perturbações $\Delta\underline{b}$ em $\underline{b}$, mas sempre haverá algum $\underline{b}$ e $\Delta\underline{b}$ para os quais a delimitação obtida é realista.

6.3.2. Sensibilidade a Perturbações na Matriz de Coeficientes

Admitir-se-á que no sistema

$$\underline{A}\ \underline{x} = \underline{b} \tag{6.26}$$

perturbações $\underline{E}$ na matriz levam a erros $\Delta\underline{x}$ na solução:

$$(\underline{A} + \underline{E})\ (\underline{x} + \Delta\underline{x}) = \underline{b} \tag{6.27}$$

$$(\underline{A} + \underline{E})\underline{x} + (\underline{A} + \underline{E})\Delta\underline{x} = \underline{b} \tag{6.28}$$

$$\underline{A}\ \underline{x} + \underline{E}\ \underline{x} + (\underline{A} + \underline{E})\Delta\underline{x} = \underline{b} \tag{6.29}$$

$$(\underline{A} + \underline{E})\Delta\underline{x} = -\ \underline{E}\ \underline{x} \tag{6.30}$$

Mesmo que $\underline{A}$ não seja singular, o que naturalmente se assume, $\underline{A} + \underline{E}$ poderia sê-lo, se não se restringir $\underline{E}$.

Escrevendo

$$\underline{A} + \underline{E} = \underline{A}\ (\underline{I} + \underline{A}^{-1}\ \underline{E}) \tag{6.31}$$

fica evidente que $\underline{A} + \underline{E}$ será não-singular se $\underline{I} + \underline{A}^{-1} \cdot \underline{E}$ também o for. Se λ_i são os autovalores de $\underline{A}^{-1}\ \underline{E}$, tem-se:

$$\|\underline{A}^{-1} \cdot \underline{E}\| \geq \lambda_i \tag{6.32}$$

e desde que os valores próprios de $\underline{I} + \underline{A}^{-1} \cdot \underline{E}$ são $1 + \lambda_i$, a matriz é não-singular se:

$$\|\underline{A}^{-1} \cdot \underline{E}\| < 1 \tag{6.33}$$

Assumindo essa condição, e escrevendo $\underline{F} = \underline{A}^{-1}\ \underline{E}$, tem-se:

$$\Delta\underline{x} = -\ (\underline{A} + \underline{E})^{-1} \cdot \underline{E} \cdot \underline{x} = -\ (\underline{I} + \underline{F})^{-1} \cdot \underline{A}^{-1} \cdot \underline{E} \cdot \underline{x} \tag{6.34}$$

Escrevendo $\underline{G} = (\underline{I} + \underline{F})^{-1}$, tem-se:

$$\underline{I} = \underline{G} + \underline{F}\,\underline{G}$$

e aplicando normas:

$$1 \geq \|\underline{G}\| - \|\underline{F}\| \cdot \|\underline{G}\| \tag{6.35}$$

para qualquer norma, exceto a euclidiana, em que a normal da matriz identidade não é unitária, como se sabe. Desde que $\|\underline{F}\| < 1$,

$$\underline{G} \leq \frac{1}{1 - \|\underline{F}\|} \tag{6.36}$$

com o que se pode ter agora o erro na solução delimitado por:

$$\Delta\underline{x} = - \underline{G} \cdot \underline{A}^{-1} \cdot \underline{E} \cdot \underline{x} \tag{6.37}$$

$$\|\Delta\underline{x}\| \leq \frac{\|\underline{A}^{-1} \cdot \underline{E}\| \, \|\underline{x}\|}{1 - \|\underline{A}^{-1} \cdot \underline{E}\|} \quad \text{para} \quad \|\underline{A}^{-1}\| \, \|\underline{E}\| < 1 \tag{6.38}$$

Embora tenha sido dito que esta prova não vale para a norma euclidiana, em que $\|\underline{I}\|_E = n^{1/2}$ e não 1, se o resultado obtido é verdadeiro para a norma espectral, é verdadeiro também para a euclidiana, já que sempre

$$\|\underline{A}\|_2 \leq \|\underline{A}\|_E \tag{6.39}$$

A delimitação do erro relativo pode agora ser escrita na forma

$$\frac{\|\Delta\,\underline{x}\|}{\|\underline{x}\|} \leq \frac{\|\underline{A}\| \cdot \|\underline{A}^{-1}\| \cdot (\|\underline{E}\| \cdot \|\underline{A}\|^{-1})}{1 - \|\underline{A}\| \cdot \|\underline{A}^{-1}\| \cdot (\|\underline{E}\| \cdot \|\underline{A}\|^{-1})} \tag{6.40}$$

onde se exprime o erro relativo procurado em $\underline{x}$ em termos do erro relativo

$$(\|\underline{E}\| \cdot \|\underline{A}\|^{-1})$$

em $\underline{A}$.

De novo, o número de condição $\|\underline{A}\| \cdot \|\underline{A}^{-1}\|$ é a quantidade decisiva.

6.3.3. Efeito do Arredondamento dos Elementos da Matriz

A delimitação de erros de solução devidos a perturbações na matriz, obtida na seção anterior, pode agora ser aplicada para delimitar a ordem de grandeza dos erros de solução que se terá, devido à obrigatoriedade inicial de se escreverem os coeficientes da matriz arredondados para t dígitos significativos, mesmo que esses coeficientes sejam, por outros aspectos, *exatos*, sem perturbações.

Sejam e_{ij}, elementos de $\underline{E}$, os erros de arredondamento nos coeficientes a_{ij} de $\underline{A}$, delimitados por:

$$\begin{aligned} |e_{ij}| &\leq 2^{-t} \cdot |a_{ij}| \qquad \text{(binário)} \\ \|\underline{E}\|_E &\leq 2^{-t} \|\underline{A}\|_E \qquad \text{(binário)} \end{aligned} \tag{6.41}$$

É possível substituir essa delimitação no resultado da seção anterior:

$$\frac{\|\Delta \underline{x}\|}{\|\underline{x}\|} \quad \frac{2^{-t} \cdot \|\underline{A}\|_E \cdot \|\underline{A}^{-1}\|_E}{1 - 2^{-t} \|\underline{A}\|_E \cdot \|\underline{A}^{-1}\|_E} \tag{6.42}$$

A menos que $2^{-t} \cdot \|\underline{A}\|_E \cdot \|\underline{A}^{-1}\|_E$ seja menor que 1, o erro será grande.

Do estudo das normas matriciais, pode-se comparar, para a norma espectral:

$$\|\underline{E}\|_2 \leq 2^{-t} \cdot n^{1/2} \cdot \|\underline{A}\|_2 \tag{6.43}$$

podendo-se então escrever, em termos do número de condição espectral, a delimitação:

$$\frac{\|\Delta \underline{x}\|_2}{\|\underline{x}\|_2} \leq \frac{2^{-t} \cdot n^{1/2} \cdot h(\underline{A})}{1 - 2^{-t} \cdot n^{1/2} \cdot h(\underline{A})} \tag{6.44}$$

Tem-se agora que:

$$2^{-t} \cdot n^{1/2} \cdot h(\underline{A}) \ll 1 \tag{6.45}$$

se se quiser ter pequenos erros de solução relativos.

Este resultado final sugere que, quando se usar aritmética do ponto flutuante com t dígitos binários significativos, não haverá, em geral, possibili-

dade de se obter uma solução sequer aproximada para um sistema de equações para o qual o número de condição espectral seja

$$h(\underline{A}) \geq 2^t \cdot n^{-1/2} \tag{6.46}$$

o que é, de fato, verdadeiro.

6.4. Sub-rotinas para Microcomputador, em BASIC, para Estimar o Número de Condição de um Sistema

6.4.1. Comentários

Listam-se, a seguir, sub-rotinas para microcomputador pessoal da linha TRS-80, na linguagem BASIC, para estimar o número de condição de sistema de equações cuja matriz de rigidez é simétrica, positivo-definida e bandeada, como as que ocorrem no MEF.

Estas sub-rotinas calculam uma estimativa do número de condição espectral, que, para matrizes simétricas, é dado pelas expressões:

$$h(\underline{A}) = \|\underline{A}\|_2 \cdot \|\underline{A}^{-1}\|_2 \tag{6.47}$$

ou

$$p(\underline{A}) = h(\underline{A}) = \frac{\lambda_{máx}}{\lambda_{mín}} \tag{6.48}$$

onde:

$\|\underline{A}\|_2$ é a norma espectral da matriz $\underline{A}$

$\|\underline{A}^{-1}\|_2$ é a norma espectral da inversa de $\underline{A}$

$\lambda_{máx}$ é o maior valor próprio de $\underline{A}$

$\lambda_{mín}$ é o menor valor próprio de $\underline{A}$

O algoritmo que foi adotado estima a norma espectral de $\underline{A}$ pela norma euclidiana, $\|\underline{A}\|_E$, com boa aproximação em geral. A sub-rotina correspondente foi denominada NORMA (instrução 900) e opera sobre a matriz $\underline{A}$ estocada em forma de um vetor coluna que reúne as colunas de coeficientes não-nulos, da diagonal principal para cima até a "linha do horizonte". Deve-se dispor ainda de um vetor MA que armazene os endereços dos elementos da diagonal principal. Essa disposição de memória foi discutida em 3.2.

O cálculo do menor autovalor de A é feito pela sub-rotina CONDIÇÃO (instrução 1300). O algoritmo obtém seu autovetor correspondente por iteração inversa a partir de um vetor de teste inicial de elementos todos unitários. Para tanto, supõe-se que já se dispuseram em um vetor coluna A os elementos da matriz triangularizada pela sub-rotina DECOMPOSIÇÃO LDLT (ver 4.3). Em cada passo, o vetor de teste é retrosubstituído pela sub-rotina RETRO (ver 4.3), obtendo-se um novo vetor de teste, que é, em seguida, normalizado de forma a ter comprimento unitário. É calculado também o coeficiente de Rayleigh em cada etapa, que, como se sabe (ver, e.g., *Bathe* (3)), fornece aproximação excelente do autovalor procurado.

6.4.2. Listagem

```
900 REM SUB NORMA
905 D=0
910 FOR I=1 TO NW
915 D=D+2*A(I)*A(I)
920 NEXT I
925 FOR I=1 TO NQ
930 K=MA(I)
935 D=D-A(K)*A(K)
940 NEXT I
945 D=SQR(D)
950 RETURN
```

```
1300 REM SUB CONDICAO
1302 DIM F(NQ), E(NQ)
1305 FOR I=1 TO NQ
1310 F(I)=1
1315 NEXT I
1320 J=1
1325 FOR I=1 TO NQ
1330 E(I)=F(I)
1335 NEXT I
1340 G1=0:G2=0
1345 GOSUB 1500
1350 FOR I=1 TO NQ
1355 G1=G1+F(I)*F(I)
1360 G2=G2+F(I)*E(I)
1365 NEXT I
1370 H2=G2/G1
1375 H3=ABS(H2-H1)/H2
1380 IF H3<=.0001 GOTO 1420
1385 H1=H2
1390 G1=SQR(G1)
1395 FOR I=1 TO NQ
1400 F(I)=F(I)/G1
1405 NEXT I
1410 J=J+1
1415 IF J>=6 GOTO 1420 ELSE GOTO 1325
1420 PRINT"MENOR VALOR PROPRIO=";H2
1425 PRINT"NUMERO DE CONDICAO";D/H2
1430 RETURN
1500 REM SUB RETRO
1505 FOR N=1 TO NQ
1510 KL=MA(N)+1:KU=MA(N+1)-1
1515 IF (KU-KL)<0 GOTO 1545
1520 K=N:C=0
1525 FOR KK=KL TO KU
1530 K=K-1:C=C+A(KK)*F(K)
1535 NEXT KK
1540 F(N)=F(N)-C
1545 NEXT N
1550 FOR N=1 TO NQ
1555 K=MA(N):F(N)=F(N)/A(K)
1560 NEXT N
1565 IF NQ=1 RETURN
1570 N=NQ
1575 FOR L=2 TO NQ
1580 KL=MA(N)+1:KU=MA(N+1)-1
1585 IF (KU-KL)<0 GOTO 1610
1590 K=N
1595 FOR KK=KL TO KU
1600 K=K-1:F(K)=F(K)-A(KK)*F(N)
1605 NEXT KK
1610 N=N-1
1615 NEXT L
1620 RETURN
```

7. CONSIDERAÇÕES SOBRE A ORIGEM FÍSICO DO MAL-CONDICIONAMENTO NUMÉRICO DE FORMULAÇÕES DE DESLOCAMENTOS DO MEF PARA ESTRUTURAS RETICULADAS DE COMPORTAMENTO LINEAR

7.1. Interpretação Física dos Valores Próprios da Matriz de Rigidez de uma Estrutura

7.1.1. Equação do Movimento de uma Estrutura

Quando o carregamento de uma estrutura, discretizada pelo MEF em N graus de liberdade, varia com o tempo, sua equação de movimento pode ser formulada a partir do equilíbrio dinâmico das forças associadas a esses graus de liberdade. Em geral, haverá quatro tipos de forças relacionadas a cada i-ésimo grau de liberdade:

f_{I_i} = forças de inércia

f_{A_i} = forças de amortecimento

f_{E_i} = forças elástica

$p_i(t)$ = cargas externas aplicadas, variáveis no tempo

No conjunto, ter-se-á, em forma vetorial, a equação de equilíbrio:

$$\underline{f}_I + \underline{f}_A + \underline{f}_E = \underline{p}(t) \qquad (7.1)$$

Essas forças se relacionam com as componentes de deslocamento $\underline{x}$ por:

- Forças elásticas: consideradas linearmente proporcionais ao deslocamento.

$$\begin{Bmatrix} f_{E_1} \\ f_{E_2} \\ \cdots \\ f_{E_n} \end{Bmatrix} = \begin{bmatrix} K_{11} & K_{12} & \cdots & K_{1n} \\ K_{21} & K_{22} & \cdots & K_{2n} \\ & & \cdots & \\ K_{n1} & & \cdots & K_{nn} \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{Bmatrix} \quad \text{ou} \quad \underline{f}_E = \underline{K}\,\underline{x} \qquad (7.2)$$

A matriz K, *matriz de rigidez*, é formalmente obtida por:

$$\underline{K} = \int_V \underline{B}^t\, \underline{D}\, \underline{B}\, dV \qquad (7.3)$$

e seus elementos têm a interpretação de:

K_{ij} = componente de força correspondente à coordenada i devido a um deslocamento unitário ao longo do grau de liberdade j, mantidos nulos todos os demais n - 1 deslocamentos possíveis.

- Forças de amortecimento: consideradas linearmente proporcionais à velocidade de deslocamento $\dot{\underline{x}}$ (o ponto sobre $\underline{x}$ indica, segundo a notação de Newton, derivação no tempo).

$$\begin{Bmatrix} f_{A_1} \\ f_{A_2} \\ \vdots \\ f_{A_n} \end{Bmatrix} \quad \begin{Bmatrix} C_{11} & C_{12} & \dots & C_{1n} \\ C_{21} & C_{22} & \dots & C_{2n} \\ \dots & \dots & \dots & \dots \\ C_{n1} & \dots & \dots & C_{nn} \end{Bmatrix} \begin{Bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \vdots \\ \dot{x}_n \end{Bmatrix} \quad \text{ou} \quad \underline{f_A} = \underline{C}\,\dot{\underline{x}} \qquad (7.4)$$

A matriz C, *matriz de amortecimento*, é formalmente dada por:

$$\underline{C} = \int_V c \quad \underline{N}^t \, \underline{N} \, dV \qquad (c = c(x, y, z)) \qquad (7.5)$$

e seus elementos têm a interpretação de:

C_{ij} = componente de força correspondente à coordenada i devido a velocidade unitária de movimento ao longo do grau de liberdade j.

- Forças de inércia: linearmente proporcionais à aceleração $\ddot{\underline{x}}$ segundo as leis do movimento de Newton.

$$\begin{Bmatrix} f_{I_1} \\ f_{I_2} \\ \vdots \\ f_{I_n} \end{Bmatrix} \quad \begin{Bmatrix} M_{11} & M_{12} & \dots & M_{1n} \\ M_{21} & M_{22} & \dots & M_{2n} \\ \dots & \dots & \dots & \dots \\ M_{n1} & \dots & \dots & M_{nn} \end{Bmatrix} \begin{Bmatrix} \ddot{x}_1 \\ \ddot{x}_2 \\ \vdots \\ \ddot{x}_n \end{Bmatrix} \quad \text{ou} \quad \underline{f_I} = \underline{M}\,\ddot{\underline{x}} \qquad (7.6)$$

A matriz M, *matriz de massa*, é formalmente fornecida por:

$$\underline{M} = \int_V m \quad \underline{N}^t \, \underline{N} \, dV \quad (m = m(x, y, z)) \qquad (7.7)$$

e seus elementos têm a interpretação de:

M_{ij} = componente de força correspondente à coordenada i devida a aceleração unitária do movimento ao longo do grau de liberdade j.

Em resumo, a equação matricial do movimento da estrutura é retirada do equilíbrio dinâmico do sistema e é expressa por:

$$\underline{M}\ \ddot{\underline{x}} + \underline{C}\ \dot{\underline{x}} + \underline{K}\ \underline{x} = \underline{p}(t) \qquad (7.8)$$

7.1.2. Decomposição Modal da Equação do Movimento

A técnica mais geralmente utilizada para solução da equação do movimento (7.8) é a chamada *decomposição modal*, a seguir resumida.

Desconsiderando o amortecimento, ao se colocar em movimento de vibrações livres uma estrutura, ter-se-ia a equação de movimento

$$\underline{M}\ \ddot{\underline{x}} + \underline{K}\ \underline{x} = \underline{0} \qquad (7.9)$$

e admitindo que, no movimento resultante, os pontos da estrutura se deslocam em fase numa forma elástica $\underline{\bar{x}}$, com amplitude variando harmonicamente, ter-se-á:

$$\underline{x}(t) = \underline{\bar{x}}\ e^{i\omega t} \qquad \text{e} \qquad \ddot{\underline{x}}(t) = -\omega^2\ \underline{\bar{x}}\ e^{i\omega t} \qquad (7.10)$$

onde ω = freqüência de vibração, resultando em

$$\underline{K}\ \underline{\bar{x}} = \omega^2\ \underline{M}\ \underline{\bar{x}} \qquad (7.11)$$

em que é reconhecido um problema de valores e vetores próprios do tipo

$$\underline{K}\ \underline{\bar{x}} = \lambda\ \underline{M}\ \underline{\bar{x}} \qquad (7.12)$$

em que

$$\begin{Bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{Bmatrix} = \begin{Bmatrix} \omega_1^2 \\ \omega_2^2 \\ \vdots \\ \omega_n^2 \end{Bmatrix} \qquad (7.13)$$

sendo ω_i^2 o quadrado da i-ésima *freqüência natural da vibração do sistema*, ou seja, uma das freqüências em que os pontos da estrutura vibram em fase,

mantendo uma elástica de forma $\underline{\hat{x}}_i$ constante, em que apenas a amplitude varia.

Normalizando o vetor deslocamento $\underline{\hat{x}}_i$ de um dado modo natural i, ter-se-á a forma nodal $\underline{\phi}_i$ desse modo, e, para o conjunto dos n modos, a matriz modal

$$\underline{\Phi} = (\underline{\phi}_1 \quad \underline{\phi}_2 \quad ..... \quad \underline{\phi}_n) \tag{7.14}$$

Pode ser demonstrado (ver, e.g., *Clough (4)*) que quaisquer dois modos $\underline{\hat{x}}_i$ e $\underline{\hat{x}}_j$ (ou $\underline{\phi}_i$ e $\underline{\phi}_j$) são ortogonais a $\underline{M}$ e $\underline{K}$, ou seja:

$$\underline{\hat{x}}_i^t \, \underline{M} \, \underline{\hat{x}}_j = 0 \quad \text{ou} \quad \underline{\phi}_i^t \, \underline{M} \, \underline{\phi}_j = 0 \quad \text{para} \quad i \neq j \tag{7.15}$$

$$\underline{\hat{x}}_i^t \, \underline{K} \, \underline{\hat{x}}_j = 0 \quad \text{ou} \quad \underline{\phi}_i^t \, \underline{K} \, \underline{\phi}_j = 0 \quad \text{para} \quad i \neq j \tag{7.16}$$

Será admitido, ainda, que se tenha amortecimento do tipo de Rayleigh, cuja matriz de amortecimento seja combinação linear de $\underline{M}$ e $\underline{K}$, tal que se possa assim estender as condições de ortogonalidade a $\underline{C}$.

A *técnica da decomposição modal* das equações do movimento consiste em escrever o vetor deslocamento $\underline{x}$ como uma combinação linear dos vetores modais $\underline{\phi}_i$, cujos coeficientes y_i seriam as amplitudes modais, isto é:

$$\underline{x} = \underline{\phi}_1 \, y_1 + \underline{\phi}_2 \, y_2 + \ldots + \underline{\phi}_n \, y_n \quad \text{ou} \quad \underline{x} = \underline{\Phi} \, \underline{y} \tag{7.17}$$

que, substituída na equação do movimento (7.8), dá:

$$\underline{M} \, \underline{\Phi} \, \ddot{\underline{y}} + \underline{C} \, \underline{\Phi} \, \dot{\underline{y}} + \underline{K} \, \underline{\Phi} \, \underline{y} = \underline{p}(t) \tag{7.18}$$

Pré-multiplicando pelo i-ésimo vetor modal $\underline{\Phi}_i^t$, ter-se-á:

$$\underline{\Phi}_i^t \, \underline{M} \, \underline{\Phi} \, \ddot{\underline{y}} + \underline{\phi}_i^t \, \underline{C} \, \underline{\Phi} \, \dot{\underline{y}} + \underline{\phi}_i^t \, \underline{K} \, \underline{\Phi} \, \underline{y} = \underline{\phi}_i^t \, \underline{p}(t) \tag{7.19}$$

mas, devido à ortogonalidade dos modos, desaparecerão todas as equações exceto a i-ésima:

$$\underline{\phi}_i^t \, \underline{M} \, \underline{\phi}_i \, \ddot{y}_i + \underline{\phi}_i^t \, \underline{C} \, \underline{\phi}_i \, \dot{y}_i + \underline{\phi}_i^t \, \underline{K} \, \underline{\phi}_i \, y_i = \underline{\phi}_i^t \, \underline{P}(t) \tag{7.20}$$

ou

$$M_i \, \ddot{y}_i + C_i \, \dot{y}_i + K_i \, y_i = P_i(t) \tag{7.21}$$

constituindo uma equação de movimento independente para cada modo i, em que

M_i, C_i, K_i, e P_i são, respectivamente, as componentes modais da massa, amortecimento, rigidez e carga, ou simplesmente a massa, amortecimento, rigidez e carga modais, obtidas por:

$$M_i = \underline{\phi}_i^t \underline{M} \underline{\phi}_i \qquad C_i = \underline{\phi}_i^t \underline{C} \underline{\phi}_i \qquad K_i = \underline{\phi}_i^t \underline{K} \underline{\phi}_i \qquad P_i = \underline{\phi}_i^t \underline{p}(t)$$

Torna-se, assim, possível descrever o movimento do sistema como uma superposição de formas elásticas modais com amplitudes y_i dadas por equações de movimento de um grau de liberdade do tipo (7.21).

7.1.3. Significado Físico dos Valores Próprios da Matriz de Rigidez

Está-se, agora, capacitado a analisar o significado dos valores próprios da matriz de rigidez que desempenham um papel de grande relevância na avaliação de seu condicionamento numérico, como já se viu no capítulo precedente.

Tome-se a equação das vibrações livres não-amortecidas (7.11) e reescreva-se para um determinado modo de vibração:

$$\underline{K} \cdot \underline{\bar{x}}_i = \omega_i^2 \cdot \underline{M} \cdot \underline{\bar{x}}_i \tag{7.22}$$

ou, dividindo por uma amplitude de referência, isto é, normalizando:

$$\underline{K} \underline{\phi}_i = \omega_i^2 \underline{M} \underline{\phi}_i \tag{7.23}$$

Pré-multiplicando pela transposta da forma modal $\underline{\phi}_i^t$, e invocando a ortogonalidade dos modos, obtém-se:

$$\underline{\phi}_i^t \cdot \underline{K} \cdot \underline{\phi}_i = \omega_i^2 \cdot \underline{\phi}_i^t \underline{M} \cdot \underline{\phi}_i \tag{7.24}$$

$$K_i = \omega_i^2 \cdot M_i \tag{7.25}$$

ou

$$K_i = \lambda_i \cdot M_i$$

com o que se podem reescrever as equações de movimento modais (7.21) como sendo:

$$\ddot{y} + \frac{C_i}{M_i} \dot{y} + \lambda_i \cdot y = \frac{P_i}{M_i} \tag{7.26}$$

com o que fica bem claro que os valores próprios obtidos na equação (7.12),

λ_i, podem ser caracterizados como a *RIGIDEZ* de cada modo natural de deslocamento da estrutura.

Argumento: o problema de obterem-se os valores próprios da matriz de rigidez de uma estrutura para utilizá-los posteriormente na avaliação de seu condicionamento numérico é formalmente dado por:

$$\underline{K}\,\underline{\hat{X}} = \lambda \cdot \underline{\hat{X}} \qquad (7.27)$$

que em nada difere do problema de obtenção das freqüências naturais de vibração enunciado em (7.12) para uma matriz de massa unitária, isto é, $\underline{M} = \underline{I}$, com o que se teria:

$$\underline{K}\,\underline{\hat{X}} = \lambda \cdot \underline{I} \cdot \underline{\hat{X}} \qquad (7.28)$$

A simples inspeção conjunta de (7.27), (7.28) e (7.25) leva a concluir que os valores próprios da matriz de rigidez, λ_i, correspondem à rigidez de cada modo de deslocamento, relacionando sua amplitude à componente de carga respectiva.

Assim, o *menor valor próprio* corresponde ao modo de deslocamento *mais flexível* (variações pequenas no carregamento modal respectivo, levando a grandes variações na amplitude de deslocamento). O *maior valor próprio* corresponde ao modo de deslocamento *mais rígido* (variações no carregamento modal, implicando em pequena variação na amplitude do correspondente modo).

Percebe-se daí a influência sobre o condicionamento numérico de um determinado problema estrutural da existência de direções em que cargas relativamente pequenas produziriam grandes deslocamentos, ou de peças que, por sua rigidez desproporcional, levariam a deslocamentos relativamente pequenos face a outros da estrutura.

Torna-se, assim, importante para a concepção ou discretização de uma estrutura manter em mente e controlar, se possível, os efeitos da existência de tais extremos de flexibilidade e/ou rigidez.

Na seção que se segue, isto será melhor exemplificado.

7.2. Exemplos da Influência do Modelo Físico no Condicionamento

Será apresentada, nesta seção, uma série de exemplos de estruturas em que se procurará mostrar a origem física do problema de mal-condicionamento das equações de equilíbrio do MEF. Os comentários, conclusões e recomendações originados dos exemplos são apresentados na seção seguinte, 7.3, onde se procurará generalizar as idéias.

7.2.1. Exemplo: Viga Balcão como Pórtico Espacial e como Grelha

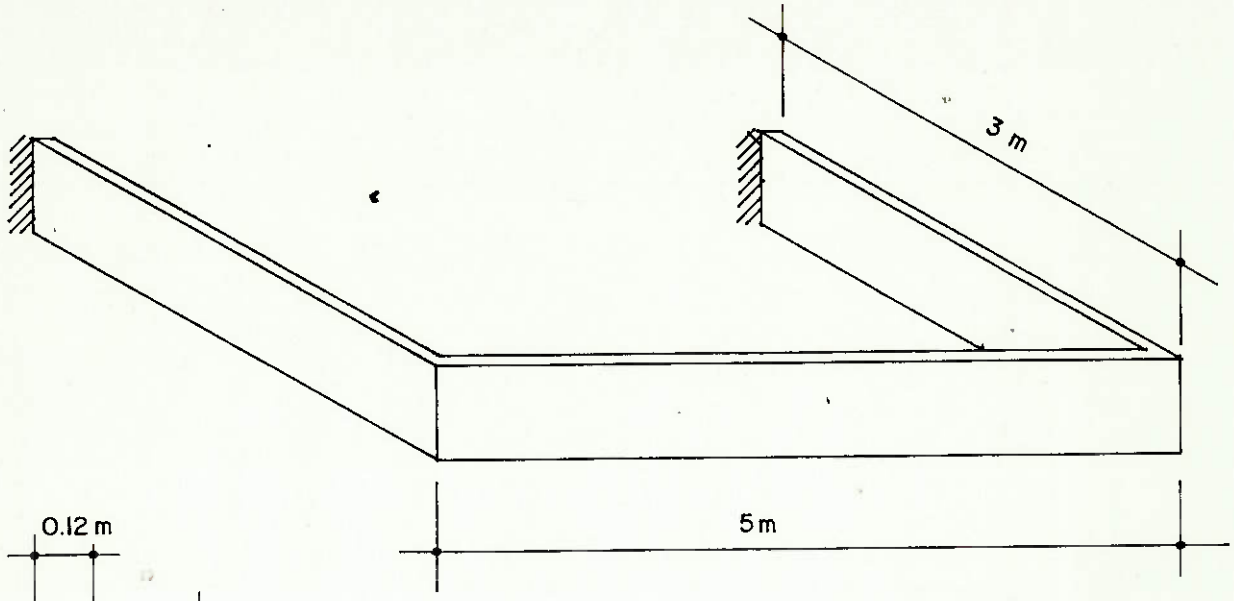
Como primeiro exemplo, obter-se-á para a viga balcão da Fig. 7.1 sua matriz de rigidez, primeiro como pórtico espacial e depois como grelha, com os correspondentes valores próprios máximo e mínimo e respectivos vetores próprios. Com esses elementos, ter-se-á o número de condição em cada caso, e a interpretação física dos valores próprios extremos como rigidez de seus respectivos modos de deslocamento, mais flexível e mais rígido, que serão visualizados pelo desenho de suas deformadas.

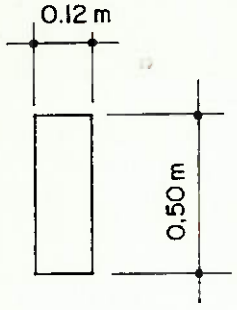
Na Fig. 7.1, são apresentados os dados geométricos da estrutura-exemplo. A Fig. 7.2 mostra os 12 graus de liberdade nodais admitidos em sua concepção como pórtico espacial. A matriz de rigidez 12 x 12 correspondente a esses graus de liberdade é apresentada na Fig. 7.3.

Finalmente, nas Fig. 7.4 e 7.5, têm-se os valores próprios máximo e mínimo e os respectivos vetores próprios, obtidos por iteração, e sua representação gráfica.

Em seguida, na Fig. 7.6, apresenta-se a mesma viga balcão como grelha, com seis graus de liberdade convencionais, resultando na matriz de rigidez da Fig. 7.7 e nos vetores próprios representados nas Fig. 7.8 e 7.9.

Comentários e conclusões sobre o exemplo estão englobados na seção seguinte, 7.3.





$E = 2.000.000$ tf/m

$A = 0,06$ m^2

$I_{máx} = 0,00125$ m^4

$I_{mín} = 0,000072$ m^4

$I_{tor} = 0,000244$ m^4

$G = 0,4 \quad E$

Fig. 7.1 - Viga Balcão: Geometria

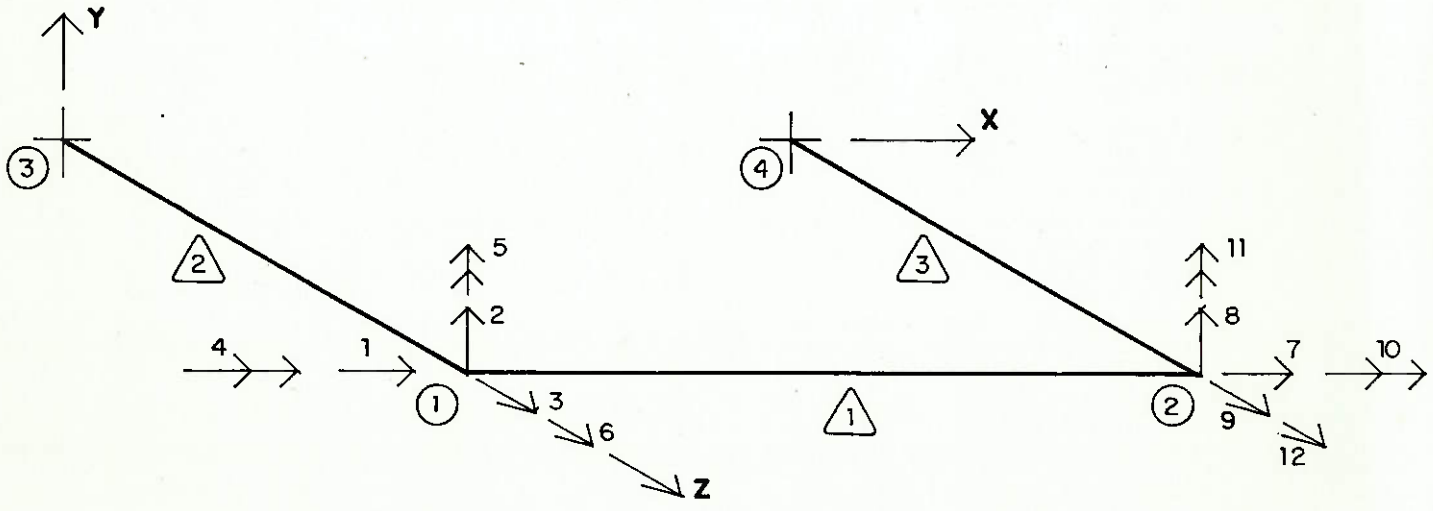


Fig. 7.2 - Graus de Liberdade da Viga Balcão considerada como Pórtico Espacial

40.013,824	0	0	0	34,56	0	-13,824	0	0	0	34,56	0
	1351,11	0	0	0	-1666,67	0	-240	0	-600	0	0
		24.064	-600	96	0	0	0	24.000	0	0	0
			2065	0	0	0	600	0	1000	0	0
				307,2	0	34,56	0	0	0	57,6	0
					3372,33	0	0	0	0	0	-39
						40.013,824	0	0	0	-34,56	0
							1351,11	0	600	0	-1666,67
								24.064	0	96	0
		Simétrica							2065	0	0
										307,2	0
											3372,33

Fig. 7.3 - Matriz de Rigidez da Viga Balcão como Pórtico Espacial

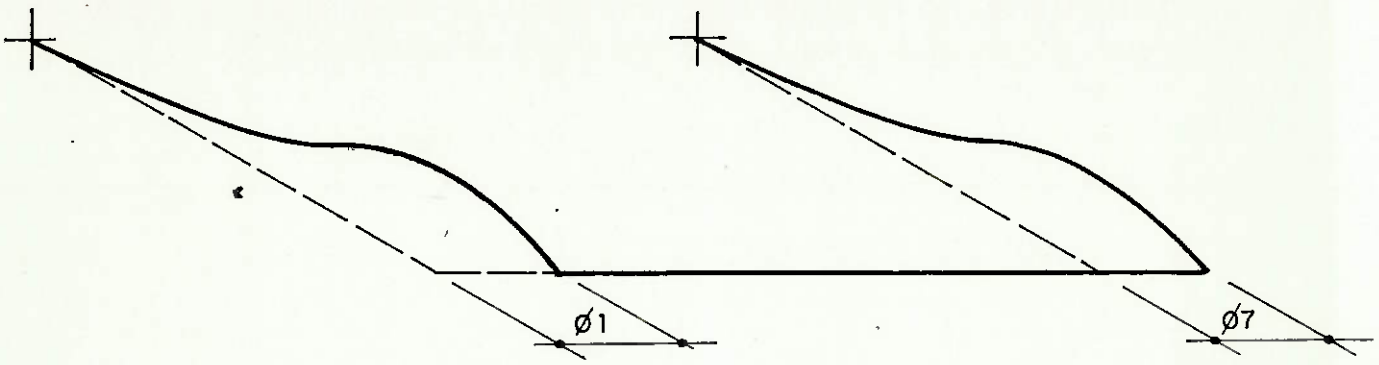


Fig. 7.4 - Menor Valor Próprio e Correspondente Vetor Próprio da Viga Balcão como Pórtico Espacial

ϕ_1 = 0,708349	ϕ_5 = 0,000002	ϕ_9 = 0,000000	λ_{min} = 36
ϕ_2 = 0,000000	ϕ_6 = 0,000000	ϕ_{10} = 0,000000	
ϕ_3 = 0,000000	ϕ_7 = 0,705861	ϕ_{11} = 0,000002	
ϕ_4 = 0,000000	ϕ_8 = 0,000000	ϕ_{12} = 0,000000	

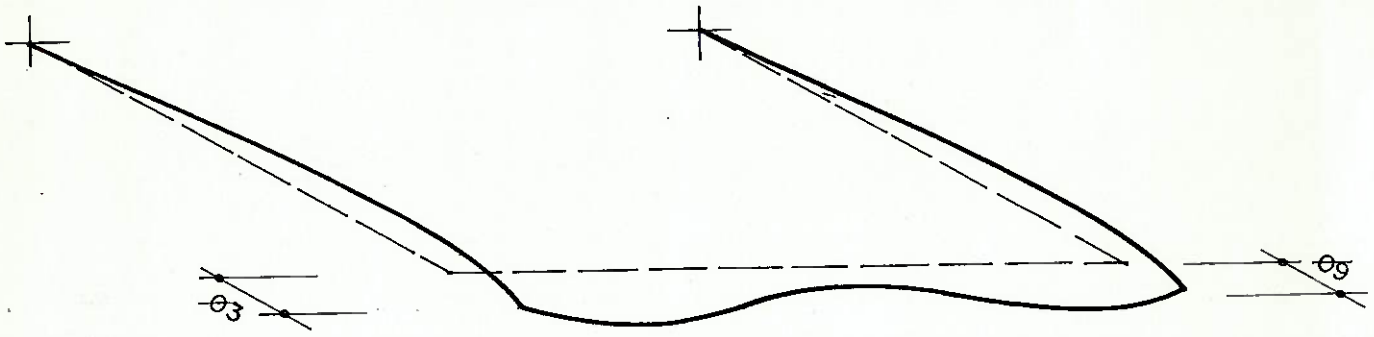


Fig. 7.5 - Maior Valor Próprio e Correspondente Vetor Próprio da Viga Balcão como Pórtico Espacial

ϕ_1 = 0,000342	ϕ_5 = - 0,198139	ϕ_9 = 0,678778	$\lambda_{máx}$ = 74590
ϕ_2 = 0,001132	ϕ_6 = 0,000603	ϕ_{10} = 0,000097	
ϕ_3 = 0,678778	ϕ_7 = - 0,000342	ϕ_{11} = - 0,198139	
ϕ_4 = 0,000097	ϕ_8 = 0,000669	ϕ_{12} = 0,000351	

$$\text{nº condição} = \frac{\lambda_{máx}}{\lambda_{mín}} = 2074$$

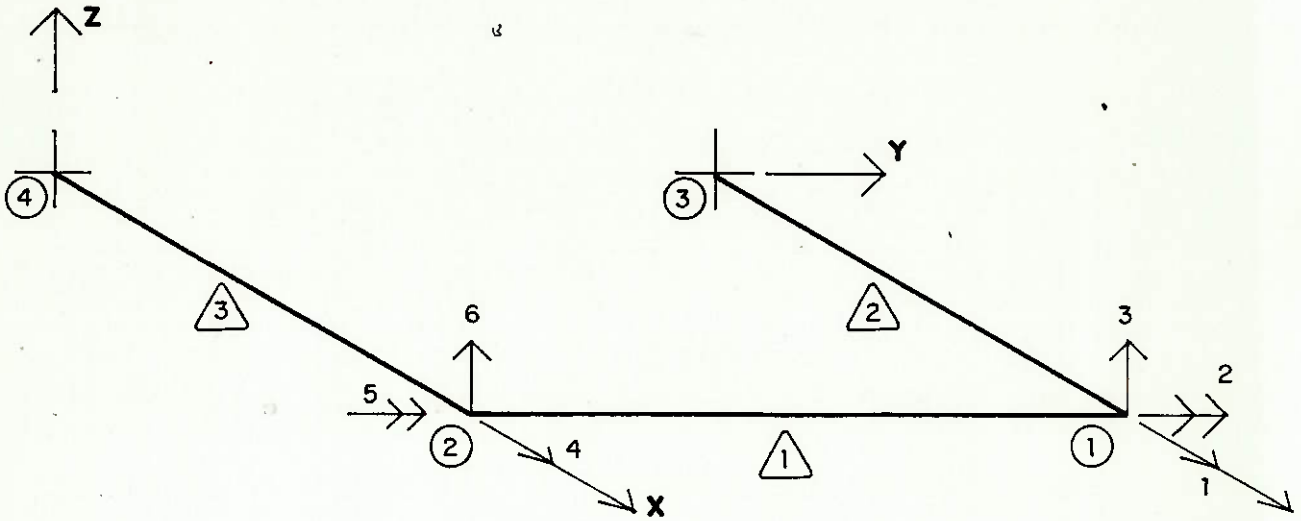


Fig. 7.6 - Viga Balcão como Grelha

$$\begin{bmatrix} 2065 & 0 & -600 & 1000 & 0 & 600 \\ & 3372{,}33 & 1666{,}67 & 0 & -39 & 0 \\ & & 1351{,}11 & -600 & 0 & -240 \\ & & & 2065 & 0 & 600 \\ & \text{Simétrica} & & & 3372{,}33 & 1666{,}67 \\ & & & & & 1351{,}11 \end{bmatrix}$$

Fig. 7.7 - Matriz de Rigidez da Viga Balcão como Grelha

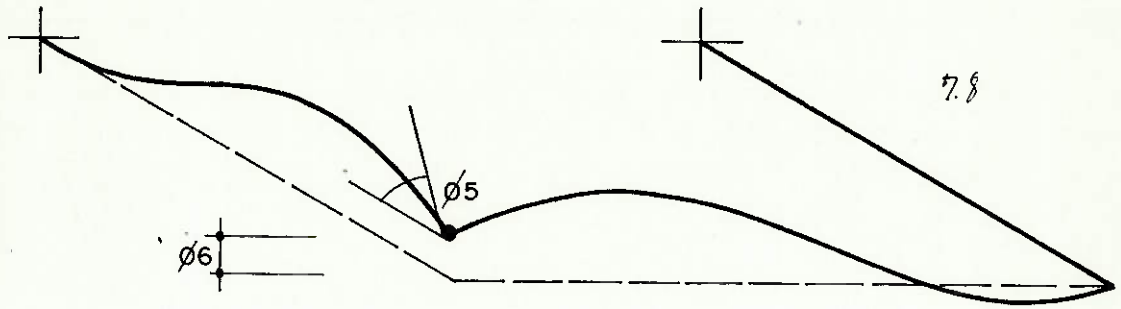


Fig. 7.8 - Maior Valor Próprio e Respectivo Vetor Próprio para a Viga Balcão como Grelha

$\phi_1 = 0{,}202650$ $\phi_4 = 0{,}202650$

$\phi_2 = 0{,}158047$ $\phi_5 = 0{,}797015$ $\lambda_{máx} = 4388{,}5$

$\phi_3 = 0{,}003550$ $\phi_6 = 0{,}507584$

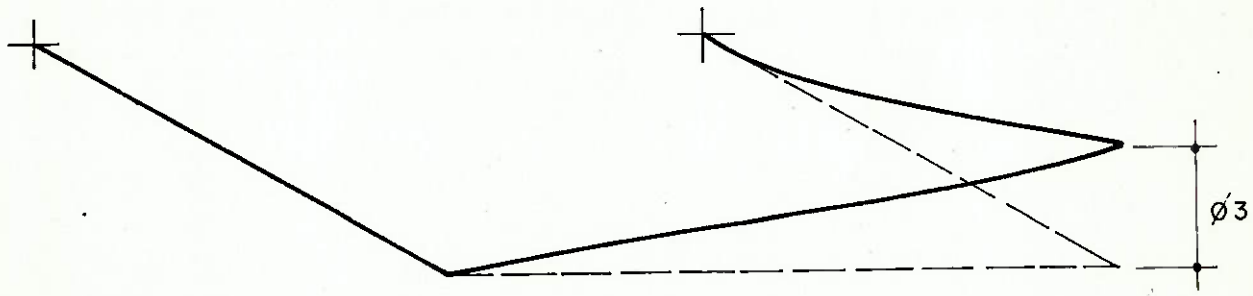


Fig. 7.9 - Menor Valor Próprio e Respectivo Vetor Próprio para a Viga Balcão como Grelha

$\phi_1 = 0{,}164646$ $\phi_4 = 0{,}164646$

$\phi_2 = -0{,}452545$ $\phi_5 = -0{,}045870$ $\lambda_{mín} = 217$

$\phi_3 = 0{,}856205$ $\phi_6 = 0{,}076123$

nº de condição $\lambda_{máx}/\lambda_{mín} = 20$

7.2.2. Exemplo: Viga em Balanço

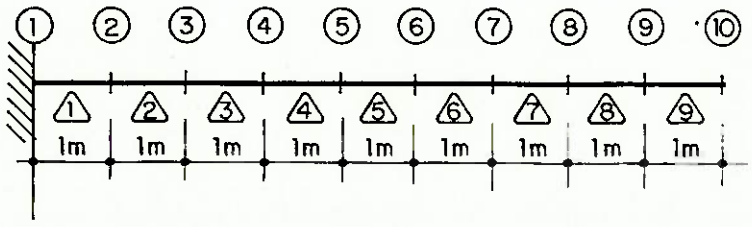
Um exemplo clássico de mal-condicionamento é uma viga em balanço quando discretizada por um grande número de nós ao longo de seu eixo.

Melosh (11) faz um estudo exaustivo do problema com elementos de barra, placa, paralelepípedos, etc, incluindo variação de rigidez.

Apresenta-se, como um exemplo adicional, uma viga em balanço, discretizada em nove elementos prismáticos planos, atribuindo-se, a princípio, três graus de liberdade por nó e, posteriormente, apenas deslocamento vertical e rotação em cada nó.

A rigor, bastariam dois graus de liberdade para caracterizar completamente o estado de deslocamento da viga, a saber: a flecha e a rotação da extremidade livre. A divisão em vários trechos acaba levando a uma severa situação de mal-condicionamento, que, no entanto, não se origina da consideração da rigidez axial da barra considerada no primeiro cálculo, mas sim da flexibilidade do modo fundamental de deslocamento face à rigidez relativa dos elementos em que foi dividida.

Os dados geométricos adotados no exemplo são os mesmos de problema analisado em classe, no Curso de Pós-Graduação de Análise Matricial de Estruturas, ministrado na Escola Politécnica da USP pelo Prof. Victor M. de Souza Lima.



E = 540 A = 0,5 I = 0,09

Fig. 7.10 - Dados Geométricos

1a. solução: três graus de liberdade por nó
nº de condição = 70.325

2a. solução: dois graus de liberdade por nó
nº de condição = 64.947

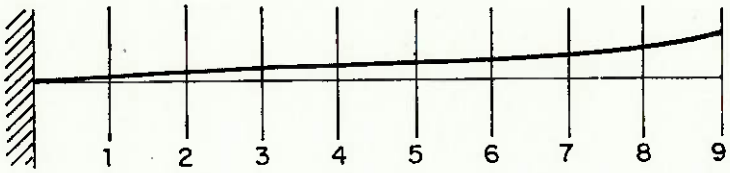


Fig. 7.11 - Valor Próprio Mínimo e Respectivo Vetor (representados apenas os deslocamentos verticais)

ϕ_1 = 0,011662 ϕ_4 = 0,159033 ϕ_7 = 0,405731
ϕ_2 = 0,044349 ϕ_5 = 0,234344 ϕ_8 = 0,496824 $\lambda_{mín}$ = 0,07
ϕ_3 = 0,094608 ϕ_6 = 0,317484 ϕ_9 = 0,589084

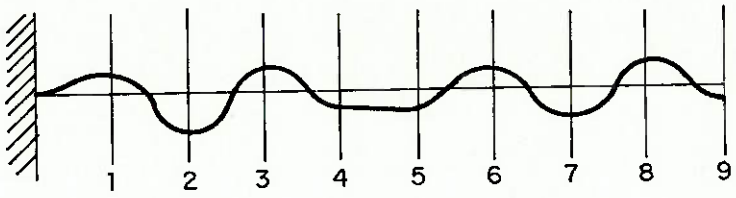


Fig. 7.12 - Valor Próprio Máximo e Respectivo Vetor (representados apenas os deslocamentos verticais)

ϕ_1 = 0,317528 ϕ_4 = -0,119985 ϕ_7 = -0,438198
ϕ_2 = - 0,470406 ϕ_5 = -0,121030 ϕ_8 = 0,405362 $\lambda_{máx}$ = 4578
ϕ_3 = 0,357218 ϕ_6 = 0,323840 ϕ_9 = -0,174726

7.2.3. Exemplos: Treliças Planas

Apresenta-se, a seguir, uma série de exemplos que basicamente são traves treliçadas de banzos paralelos, em balanço, constituídas de módulos quadrados subdivididos em triângulos retângulos isósceles de catetos de 1 m de comprimento.

Para análise de alguns dos vários efeitos sobre o condicionamento, os exemplos se sucedem como se segue:

a) para análise do efeito do comprimento em balanço são calculados, como treliças planas, os números de condição para traves de 1 m, 2 m e 3 m de comprimento;

b) para análise do efeito de rigidez relativa das barras, calcula-se, para cada um dos três comprimentos indicados, como treliças planas, o número de condição para banzos e diagonais com mesma seção e para diagonais de área de seção de 1/5 da área dos banzos e montantes.

O efeito da consideração das treliças como estruturas espaciais, levando em conta o contraventamento entre elas, será analisado para a mesma série de exemplos, no Capítulo 8, que é integralmente dedicado a esse assunto.

1º Exemplo

comprimento = 1 m
E = 21.000.000 tf/m
seção dos banzos, montante e diagonal: $A = 0{,}0005\ m^2$
valor próprio máximo = 29.539
valor próprio mínimo = 1241
número de condição = 24

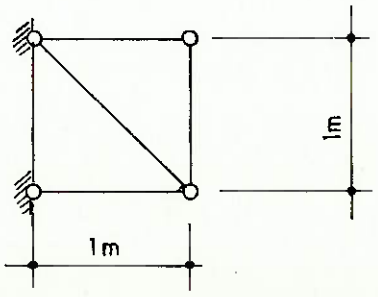


Fig. 7.13

2º Exemplo

comprimento = 1 m

E = 21.000.000 tf/m^2

seção dos banzos e montante: A = 0,0005 m^2

área da seção da diagonal: A = 0,0001 m^2

número de condição = 146

3º Exemplo

comprimento = 2 m

E = 21.000.000 tf/m^2

área da seção dos banzos, montante e diagonal = 0,0005 m^2

valor próprio máximo = 57.661

valor próprio mínimo = 328

número de condição = 176

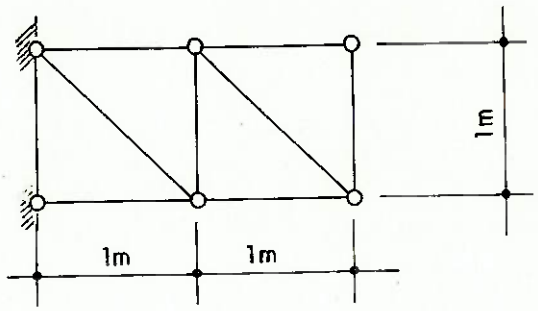


Fig. 7.14

4º Exemplo

comprimento = 2 m

E = 21.000.000 tf/m^2

área da seção de banzos e montante = 0,0005 m^2

área da seção das diagonais = 0,0001 m^2

número de condição = 787

5º Exemplo

comprimento = 3 m

E = 21.000.000 tf/m^2

área da seção de banzos, montantes e diagonais = 0,0005 m^2

valor próprio máximo = 76.007

valor próprio mínimo = 120

número de condição = 632

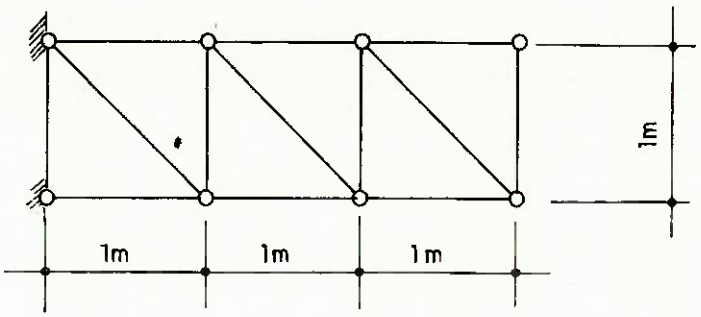


Fig. 7.15

6º Exemplo

comprimento = 3 m

E = 21.000.000 tf/m^2

área da seção de banzos e montantes = 0,0005 m^2

área da seção de diagonais = 0,0001 m^2

número de condição = 2145

7.3. Conclusões

Dos exemplos da seção anterior, pode-se chegar a algumas conclusões de ordem geral sobre a origem física do mal-condicionamento numérico das equações de equilíbrio do Método dos Deslocamentos.

Observa-se, em resumo que:

a) a existência de modos de deslocamento muito flexíveis face a outros muito rígidos na mesma estrutura leva a números de condição altos, ou seja, a mau-condicionamento.

Foi o que se viu no exemplo 7.2.1, em que o deslocamento lateral da viga balcão, considerada como pórtico espacial, tal como mostrado na Fig. 7.4, é obtido com pequeno esforço nessa direção, enquanto que a deformação axial das barras, mostrada na Fig. 7.5, exigiria esforços consideráveis nessa direção.

Viu-se, também, no exemplo da viga em balanço, que a rigidez à deformação axial da mesma não influi significativamente, sendo mais importante a grande flexibilidade do modo fundamental da barra, mostrado na Fig. 7.11, face à rigidez relativa à flexão dos outros elementos em que a viga foi subdividida.

Nos exemplos de treliças planas, a redução da rigidez das diagonais ou o aumento do balanço resultavam sempre em números de condição maiores pela maior deslocabilidade de seus nós que essas ações provocam.

b) se o carregamento aplicado não for predominantemente no sentido dos modos de deformação críticos, será de se esperar que os erros não sejam da ordem prevista pelo número de condição, ou, mais claramente, o modelo físico não será adequado.

Assim, na viga balcão de 7.2.1, se o carregamento for vertical, a estrutura se comportará como grelha, cujo condicionamento numérico, no exemplo, será muitíssimo melhor.

Das duas observações acima pode-se concluir e recomendar que, se for possível modelar as estruturas de forma a eliminar modos de deformação muito flexíveis ou muito rígidos, quando o carregamento não tiver componentes significativos nesses modos, estar-se-á melhorando a precisão de trabalho e, provavelmente, diminuindo o volume do mesmo.

Tipicamente, a consideração da deformação axial nas estruturas de barras pode levar a modos de deslocamento muito rígidos face aos deslocamentos de flexão. Há, obviamente, limites a esses raciocínios. A não-consideração da deformabilidade axial de barras de pórticos, desejável do ponto de vista

do condicionamento, dificultaria a programação do processo e cortaria em parte a generalidade do processo. Alternativamente, poder-se-ia tentar aumentar a rigidez à flexão. Em estruturas aporticadas de edifícios altos, poder-se-ia pensar na introdução de travamentos que reduzissem a flexibilidade lateral. É notório, por exemplo, o efeito contraventante que os painéis de alvenaria têm nesses edifícios. Já foi sugerido que seu efeito poderia ser levado em conta como se fossem barras diagonais trabalhando a compressão, melhorando de forma extraordinária o condicionamento do problema.

Um outro aspecto muitíssimo importante emerge das observações feitas no exemplo da viga balcão: calculada como pórtico espacial, obtiveram-se os modos de deslocamento das Fig. 7.4 e 7.4, de rigidez muito disparatada, mas nos quais um carregamento vertical usual não teria componentes. Modelada com grelha, a estrutura passa a ter um condicionamento bom, com a vantagem adicional de um menor número de graus de liberdade e correspondente trabalho de cálculo.

Em (8), *Jennings e Malik* apresentam a estrutura de cobertura reproduzida na Fig. 7.19, que, calculada como pórtico espacial (seis graus de liberdade por nó), resultou em um número de condição de 1.600.000. O exemplo é, no entanto, enganoso quanto ao condicionamento da estrutura, pois ela é muito flexível a esforços normais a vigas principais em balanço, dada a total inexistência de contraventamento entre elas. Se remodelada para treliça espacial com a introdução de contraventamentos que evitem a hipostaticidade, estima-se que o número de condicionamento cairia, provavelmente dividido por 100, além do trabalho de cálculo bem menor, devido aos três graus de liberdade a menos por nó. Os próprios autores afirmam ter como carregamento apenas esforços de gravidade. Nesse caso, a estrutura poderia ainda ser considerada como uma série de treliças planas, com consideráveis vantagens de precisão e tempo.

Chega-se aqui a uma conclusão e a uma recomendação no sentido de que, às vezes, é contraproducente uma sofisticação maior do modelo. Em outras palavras, se, por exemplo, um pórtico espacial puder ser subdividido em planos para os carregamentos considerado, isso deve ser feito, com resulta-

dos provavelmente melhores; se, também, uma estrutura treliçada puder ser representada em planos, não deverá, via de regra, ser calculada como espacial, podendo mesmo surgir situações de hipostaticidade não pressentidas.

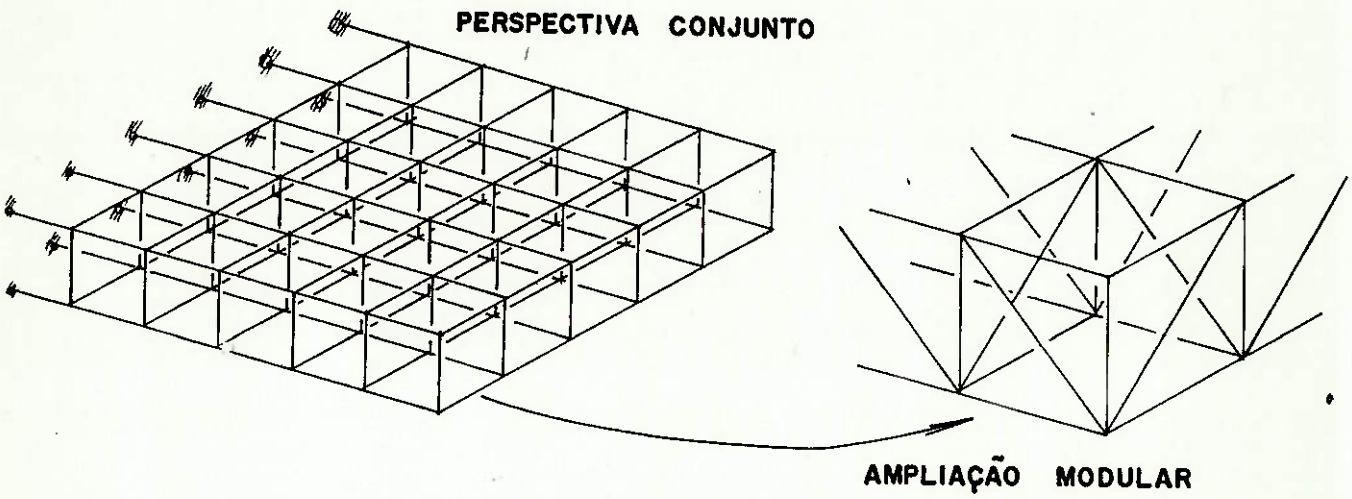


Fig. 7.19

8. UMA APLICAÇÃO: O NÚMERO DE CONDIÇÃO COMO ELEMENTO DE AVALIAÇÃO DO CONTRAVENTAMENTO DE ESTRUTURAS TRELIÇADAS

8.1. Introdução ao Conceito

Será utilizado o exposto nas conclusões e recomendações extraídas dos exemplos de origem física do condicionamento das equações de equilíbrio do MEF apresentadas no Capítulo 7.

Viu-se que, se uma estrutura treliçada — por exemplo, uma cobertura — puder ter suas treliças principais representadas em planos, quando seu carregamento assim o permitir, dever-se-á optar por calculá-la como estrutura plana e não espacial, para melhor condicionamento do problema.

No entanto, assim procedendo, restariam a avaliar as peças de contraventamento, cuja função mesma é responder às pequenas solicitações não contidas nesses planos, de estimação difícil e precária.

É interessante verificar o fato de que, ao se calcular uma estrutura dessas características como treliça espacial, estão-se introduzindo modos de deslocamento muito flexíveis, correspondentes aos deslocamentos normais ao plano das treliças principais, pelos quais respondem as peças de contraventamento, geralmente longas e de reduzida seção.

Embora o carregamento principal, neste caso, seja coplanar às tesouras principais, esses modos de deslocamento muito flexíveis, normais a elas, acabam respondendo por elevação no número de condição do problema.

Desse fato surge a idéia de comparar os números de condição de uma estrutura treliçada do tipo descrito, quando considerada como estrutura plana e quando calculada como espacial com rigidez variável de contraventamento. Ter-se-ia, assim, uma forma de avaliar a eficiência desse sistema de contraventamento em reproduzir a situação do modelo plano em que os nós das treliças principais têm o deslocamento transversal restrito.

É lógico que existe o aspecto econômico do problema a respeitar, e haveria, assim, um valor aceitável para a relação entre os números de condição dos

modelos plano e espacial em que se teria um contraventamento eficiente e econômico.

Cabe lembrar, ainda, nesta altura, que, embora na prática se deixe de contraventar alguns nós (por exemplo, os do banzo inferior de uma tesoura de cobertura), isso equivaleria, num cálculo como estrutura espacial, em matriz de rigidez singular, que poderia ser interpretado como possuindo um número de condição infinito.

8.2. Programa para Microcomputador, em BASIC, para Cálculo do Número de Condição de uma Estrutura Treliçada

É fornecida, a seguir, listagem de um programa simples, em lingugagem BASIC, para microcomputador pessoal, para estimativa do número de condição da matriz de rigidez de estruturas treliçadas espaciais (ou planas, como um caso particular).

O algoritmo estoca a matriz de rigidez como uma coluna, incluindo apenas os elementos da diagonal principal e os significativos da meia banda superior.

O maior valor próprio da matriz de rigidez é estimado pela norma euclidiana da mesma, e o menor valor próprio é calculado por iteração inversa a partir da matriz já triangularizada por Gauss, o que seria um passo intermediário necessário (e o mais custoso em tempo) na seqüência usual da solução de uma estrutura pelo Método dos Deslocamentos.

```
5 CLS:CLEAR
10 PRINT"CONDICAO MATRIZ RIGIDEZ TRELICA":PRINT
12 DEFINT I-N
15 INPUT"N. NOS, N. SECOES, N. BARRAS";NP,NM,NE
20 GOSUB 200:REM INPUT NOS
25 DIM MH(NQ)
30 GOSUB 300:REM INPUT BARRAS
35 DIM MA(NQ+1)
40 GOSUB 500: REM ENDERECO
45 MM=NW/NQ
50 PRINT"GRAUS DE LIBERDADE";NQ
55 PRINT"ELEMENTOS DA MATRIZ";NW
60 PRINT"MAXIMA LARGURA DE BANDA";MK
65 PRINT"LARGURA DE BANDA MEDIA";MM
70 DIM A(NW)
75 GOSUB 600:REM MONTAGEM MATRIZ
80 PRINT"MATRIZ DE RIGIDEZ PRONTA"
85 GOSUB 900:REM NORMA
90 PRINT"NORMA MATRIZ RIGIDEZ=";D
95 PRINT"INICIO DECOMPOSICAO  LDLT"
100 GOSUB 1000:REM DECOMPOSICAO LDLT
105 PRINT"DECOMPOSICAO COMPLETA"
110 INPUT"CARREGAMENTO, CONDICAO, FIM(1,2,3)";M
115 ON M GOTO 120,125,150
120 GOSUB 1200:REM CARREGAMENTO
123 GOTO 110
125 GOSUB 1300:REM CONDICAO
130 GOTO 110
150 END
200 REM SUB INPUT NOS
205 DIM X(3,NP), ID(3,NP)
210 FOR N=1 TO NP
215 PRINT"NO(";N;")",:INPUT"X,Y,Z";X(1,N),X(2,N),X(3,N)
220 NEXT N
225 INPUT"NUMERO NO C/ RESTRICAO(0 P/ SAIR)";NR
230 IF NR=0 THEN GOTO 245
235 INPUT"ENTRAR RESTR:DX,DY,DZ(0=LIVRE,1=FIXO)";ID(1,NR)
240 GOTO 225
245 NQ=0
250 FOR N=1 TO NP
255 FOR I=1 TO 3
260 IF ID(I,N)<>0 GOTO 270
265 NQ=NQ+1:ID(I,N)=NQ:GOTO275
270 ID(I,N)=0
275 NEXT I
280 NEXT N
285 RETURN
300 REM SUB INPUT BARRAS
305 DIM II(NE), JJ(NE), LM(6,NE), MT(NE)
310 INPUT"MODULO DE ELASTICIDADE";EM
315 FOR I=1 TO NM
320 PRINT"AREA SECAO TIPO";I;:INPUT AR(I)
325 NEXT I
330 PRINT"BARRA","NO INICIAL, NO FINAL, TIPO DE SECAO"
335 FOR N=1 TO NE
340 PRINT N,:INPUT I,J,MT
345 II(N)=I
350 JJ(N)=J
355 MT(N)=MT
360 FOR L=1 TO 3
365 LM(L,N)=ID(L,I):LM(L+3,N)=ID(L,J)
370 NEXT L
375 GOSUB 400:REM COLUNA
```

```
380 NEXT N
385 RETURN
400 REM SUB COLUNA
405 LS=NQ+1
410 FOR K=1 TO 6
415 IF LM(K,N)=0 GOTO 430
420 IF (LM(K,N)-LS)>=0 GOTO 430
425 LS=LM(K,N)
430 NEXT K
435 FOR K=1 TO 6
440 KK=LM(K,N)
445 IF KK=0 GOTO 460
450 ME=KK-LS
455 IF ME>MH(KK) THEN MH(KK)=ME
460 NEXT K
465 RETURN
500 REM SUB ENDERECO
505 MA(1)=1:MA(2)=2:MK=0
510 IF NQ=1 GOTO535
515 FOR I=2 TO NQ
520 IF MH(I)>MK THEN MK=MH(I)
525 MA(I+1)=MA(I)+MH(I)+1
530 NEXT I
535 MK=MK+1
540 NW=MA(NQ+1)-MA(1)
545 RETURN
600 REM SUB MONTAGEM MATRIZ
602 DIM S(22)
605 FOR N=1 TO NE
610 MT=MT(N):X2=0
615 FOR L=1 TO 3
620 D(L)=X(L,II(N))-X(L,JJ(N))
625 X2=X2+D(L)*D(L)
630 NEXT L
635 XL=SQR(X2):XX=EM*AR(MT)*XL
640 FOR L=1 TO 3
645 ST(L)=D(L)/X2:ST(L+3)=-ST(L)
650 NEXT L
655 KL=0
660 FOR L=1 TO 6
665 YY=ST(L)*XX
670 FOR K=L TO 6
675 KL=KL+1:S(KL)=ST(K)*YY
680 NEXT K
685 NEXT L
690 GOSUB 800:REM SOMA BANDA
695 NEXT N
700 RETURN
800 REM SUB SOMA BANDA
805 NI=0
810 FOR I=1 TO 6
815 I1=LM(I,N)
820 IF I1<=0 GOTO 875
825 MI=MA(I1):KI=I
830 FOR J=1 TO 6
835 J1=LM(J,N)
840 IF J1<=0 GOTO 865
845 IJ=I1-J1
850 IF IJ<0 GOTO 865
855 KK=MI+IJ:KS=KI
860 IF J>=I THEN KS=J+NI
862 A(KK) = A(KK) + S(KS)
865 KI=KI+6-J
```

```
870 NEXT J
875 NI=NI+6-I
880 NEXT I
885 RETURN
900 REM SUB NORMA
905 D=0
910 FOR I=1 TO NW
915 D=D+2*A(I)*A(I)
920 NEXT I
925 FOR I=1 TO NQ
930 K=MA(I)
935 D=D-A(K)*A(K)
940 NEXT I
945 D=SQR(D)
950 RETURN
1000 REM SUB DECOMPOSICAO LDLT
1005 FOR N=1 TO NQ
1010 KN=MA(N):KL=KN+1:KU=MA(N+1)-1:KH=KU-KL
1015 IF KH<0 GOTO 1125
1020 IF KH=0 GOTO 1085
1025 K=N-KH:IC=0:KT=KU
1030 FOR J=1 TO KH
1035 IC=IC+1:KT=KT-1:KI=MA(K):ND=MA(K+1)-KI-1
1040 IF ND<=0 GOTO 1075
1045 IF IC>=ND THEN KK=ND ELSE KK=IC
1050 C=0
1055 FOR L=1 TO KK
1060 C=C+A(KI+L)*A(KT+L)
1065 NEXT L
1070 A(KT)=A(KT)-C
1075 K=K+1
1080 NEXT J
1085 K=N:B=0
1090 FOR KK=KL TO KU
1100 K=K-1:KI=MA(K):C=A(KK)/A(KI)
1105 B=B+C*A(KK)
1110 A(KK)=C
1115 NEXT KK
1120 A(KN)=A(KN)-B
1125 IF A(KN)<=0 PRINT"NAOPOSDEF EQ";N;" PIVO";A(KN):STOP
1130 NEXT N
1135 RETURN
1200 REM SUB CARREGAMENTO
1205 FOR I=1 TO NQ
1210 PRINT"F(";I;")=";:INPUT F(I)
1215 NEXT I
1220 GOSUB 1500
1225 FOR I=1 TO NQ
1230 PRINT"F(";I;")=";F(I)
1235 NEXT I
1240 RETURN
1300 REM SUB CONDICAO
1302 DIM F(NQ), E(NQ)
1305 FOR I=1 TO NQ
1310 F(I)=1
1315 NEXT I
1320 J=1
1325 FOR I=1 TO NQ
1330 E(I)=F(I)
1335 NEXT I
1340 G1=0:G2=0
1345 GOSUB 1500
1350 FOR I=1 TO NQ
```

```
1355 G1=G1+F(I)*F(I)
1360 G2=G2+F(I)*E(I)
1365 NEXT I
1370 H2=G2/G1
1375 H3=ABS(H2-H1)/H2
1380 IF H3<=.0001 GOTO 1420
1385 H1=H2
1390 G1=SQR(G1)
1395 FOR I=1 TO NQ
1400 F(I)=F(I)/G1
1405 NEXT I
1410 J=J+1
1415 IF J>=6 GOTO 1420 ELSE GOTO 1325
1420 PRINT"MENOR VALOR PROPRIO=";H2
1425 PRINT"NUMERO DE CONDICAO";D/H2
1430 RETURN
1500 REM SUB RETRO
1505 FOR N=1 TO NQ
1510 KL=MA(N)+1:KU=MA(N+1)-1
1515 IF (KU-KL)<0 GOTO 1545
1520 K=N:C=0
1525 FOR KK=KL TO KU
1530 K=K-1:C=C+A(KK)*F(K)
1535 NEXT KK
1540 F(N)=F(N)-C
1545 NEXT N
1550 FOR N=1 TO NQ
1555 K=MA(N):F(N)=F(N)/A(K)
1560 NEXT N
1565 IF NQ=1 RETURN
1570 N=NQ
1575 FOR L=2 TO NQ
1580 KL=MA(N)+1:KU=MA(N+1)-1
1585 IF (KU-KL)<0 GOTO 1610
1590 K=N
1595 FOR KK=KL TO KU
1600 K=K-1:F(K)=F(K)-A(KK)*F(N)
1605 NEXT KK
1610 N=N-1
1615 NEXT L
1620 RETURN
```

8.3. Exemplos

8.3.1. Exemplos: Traves Treliçadas de Banzos Paralelos em Balanço

Volta-se, nesta seção, aos mesmos exemplos de 7.2.3. Considera-se agora que, em cada caso, ou seja, traves de 1 m, 2 m e 3 m de comprimento em balanço, tenha-se um par de treliças espaçadas de 3 m. Entre elas haverá terças com a mesma seção dos banzos, montantes e diagonais, no sentido perpendicular. Além disso, ter-se-á o contraventamento necessário à estabilidade do conjunto.

O efeito de rigidez desse contraventamento é testado em dois casos:

a) contraventamento com seção de área igual às demais peças;

b) contraventamento com seção de área de 1/5 das demais peças.

Nota-se que, na primeira hipótese, obviamente muito antieconômica, os números de condição calculados pouco diferirão dos obtidos anteriormente para treliça plana.

Na segunda hipótese, com um contraventamento de seção mais próxima da que se faria normalmente, o número de condição da estrutura espacial será da ordem de três a quatro vezes o da treliça plana correspondente.

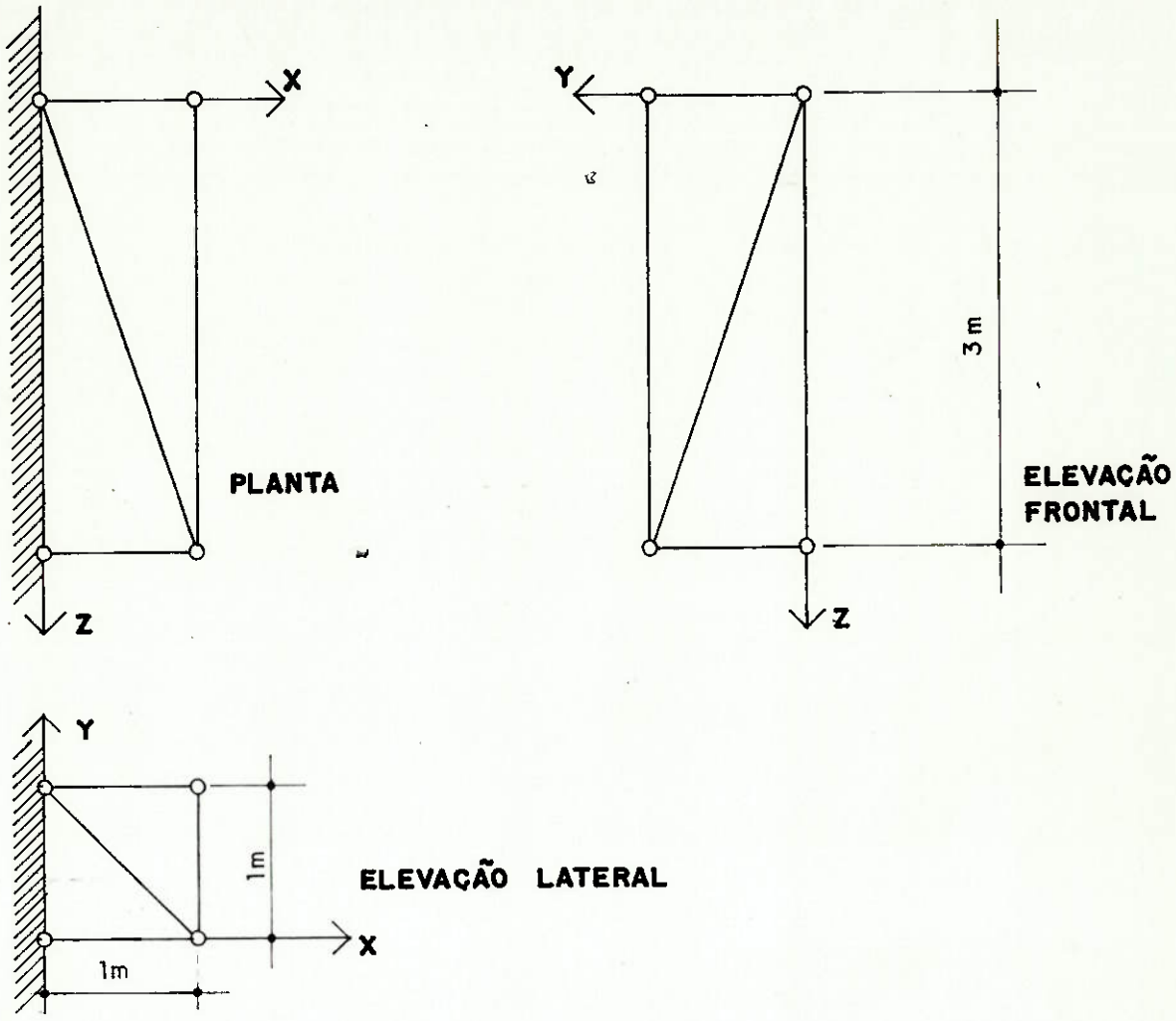


Fig. 8.1 - Par de Treliças de um Módulo

E = 21.000.000 tf/m^2

área de seção de banzos, montantes, diagonais e terças = 0,0005 m^2

1a. hipótese de contravento: seção igual às demais peças

$\lambda_{máx}$ = 42.406 $\lambda_{mín}$ = 1130

nº de condição = 40

2a. hipótese de contravento: seção = 0,0001 m^2

$\lambda_{máx}$ = 43.240 $\lambda_{mín}$ = 284

nº de condição = 152

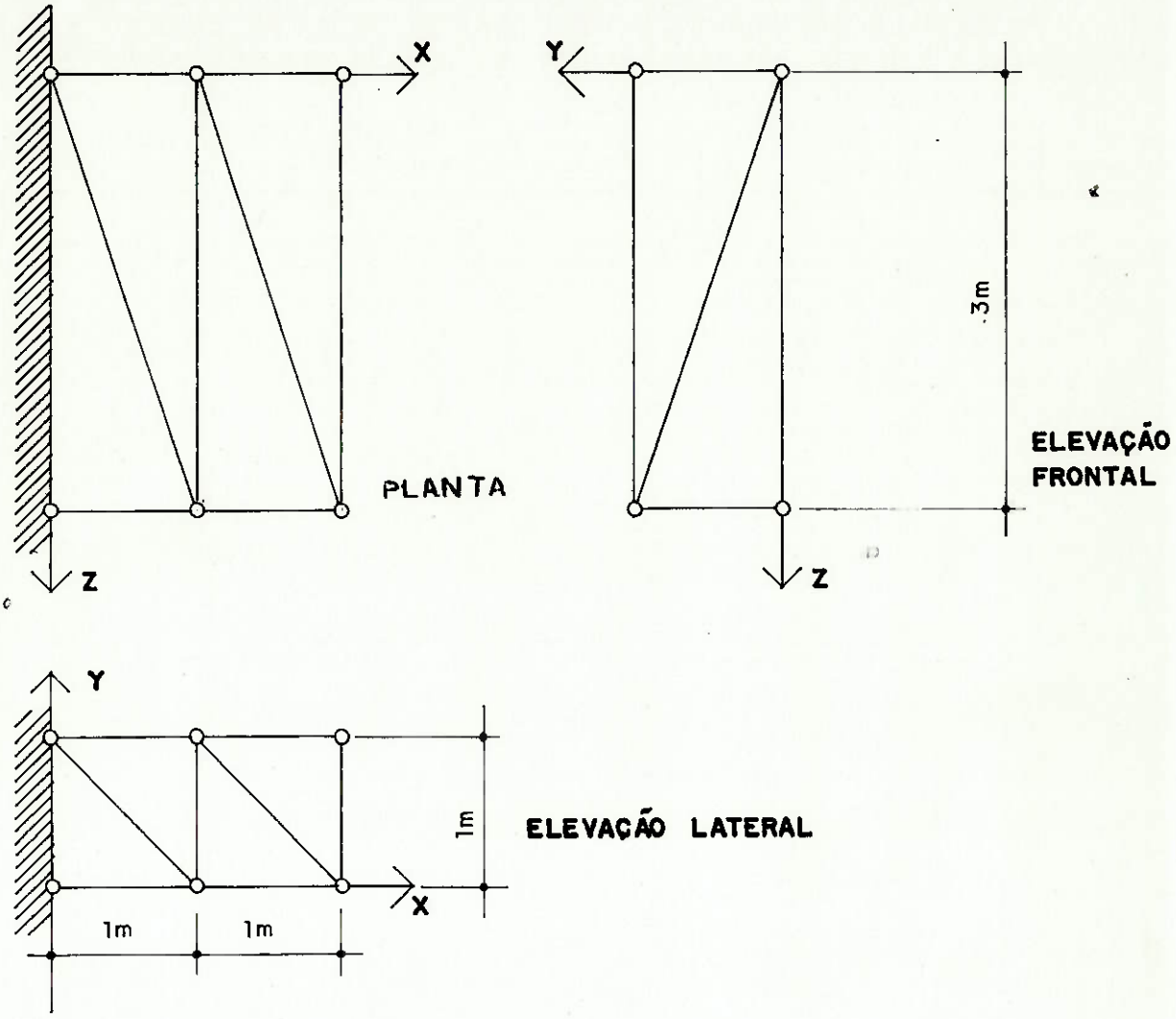


Fig. 8.2 - Par de Treliças de Dois Módulos.

E = 21.000.000 tf/m^2

área da seção de banzos, montantes, diagonais e terças = 0,0005 m^2

1a. hipótese de contravento: seção igual às demais peças

$\lambda_{máx}$ = 83.163 $\lambda_{mín}$ = 328

nº de condição = 262

2a. hipótese de contravento: seção = 0,0001 m^2

$\lambda_{máx}$ = 83.192 $\lambda_{mín}$ = 103

nº de condição = 808

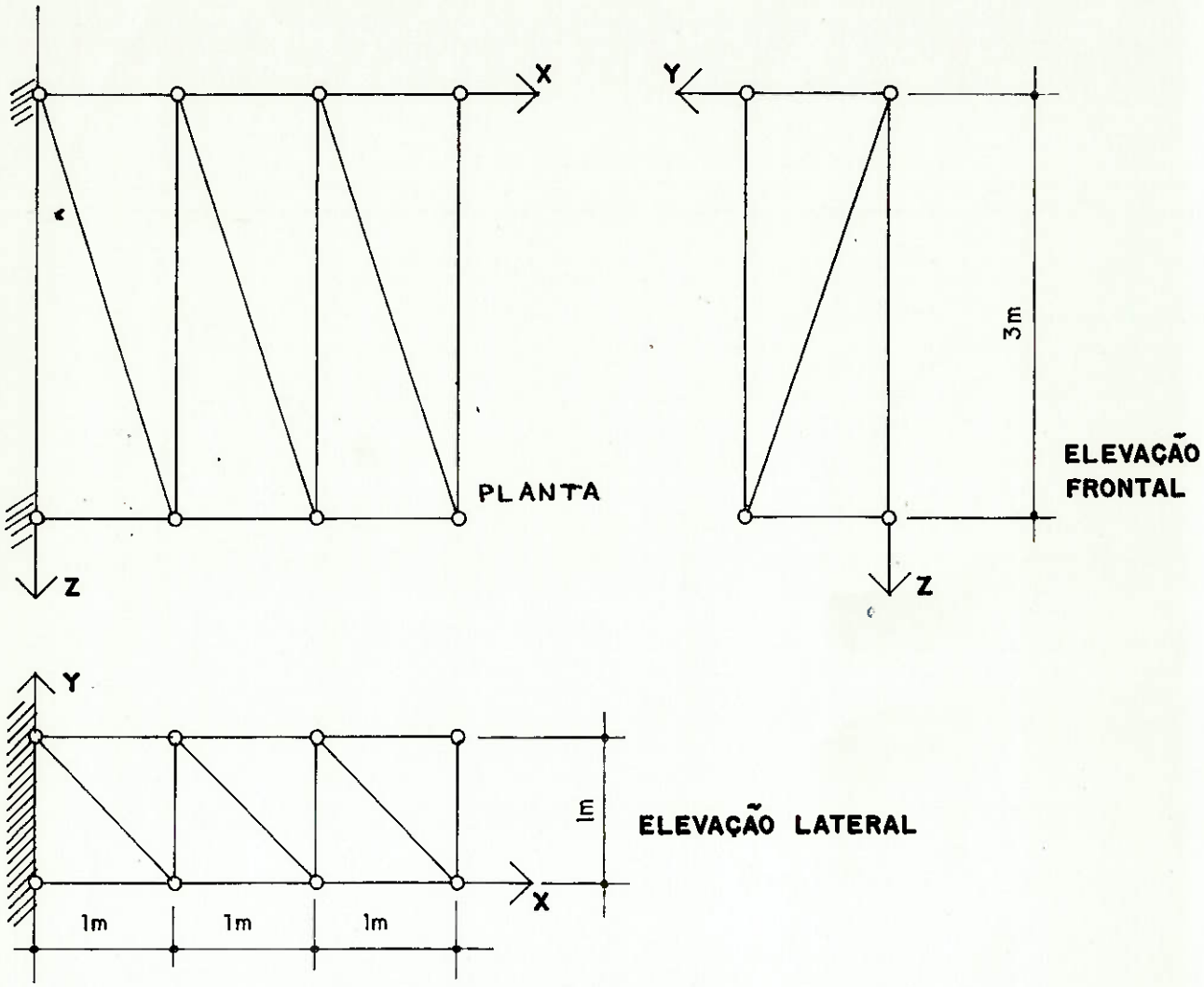


Fig. 8.3 - Par de Treliças de Três Módulos

E = 21.000.000 tf/m²

área da seção de banzos, montantes, diagonais e terças = 0,0005 m²

1a. hipótese de contravento: seção igual às demais peças

$\lambda_{máx}$ = 113.158 $\lambda_{mín}$ = 121

nº de condição = 931

2a. hipótese de contravento: seção = 0,0001 m²

$\lambda_{máx}$ = 109.417 $\lambda_{mín}$ = 52

nº de condição = 2120

8.3.2. Exemplo: Cobertura de Duas Águas

Como último exemplo, ver-se-á uma típica estrutura metálica treliçada de suporte a uma cobertura de duas águas.

Como nas anteriores, será calculada a tesoura principal como treliça plana, obtendo-se seu número de condição, e depois será feito o mesmo para um conjunto de duas treliças como estrutura espacial, com duas hipóteses de contraventamento, em que a segunda seria a que reproduziria melhor o que a prática adota, e que acusa uma relação de cerca de quatro vezes entre os números de condição obtidos.

• 1a. hipótese: tesoura principal como treliça plana

seção dos banzos = 0,0025 m^2
seção de diagonais e montantes = 0,0020 m^2
$\lambda_{máx}$ = 21.316
$\lambda_{mín}$ = 90
número de condição = 236

• 2a. hipótese: conjunto de duas tesouras como estrutura espacial

seção das terças = 0,0014 m^2
seção do contraventamento = 0,0007 m^2
$\lambda_{máx}$ = 278.969
$\lambda_{mín}$ = 787
número de condição = 354

• 3a. hipótese: conjunto de duas tesouras como estrutura espacial

seção das terças = 0,0014 m^2
seção do contraventamento = 0,00012 m^2
$\lambda_{máx}$ = 271.458
$\lambda_{mín}$ = 236
número de condição = 1151

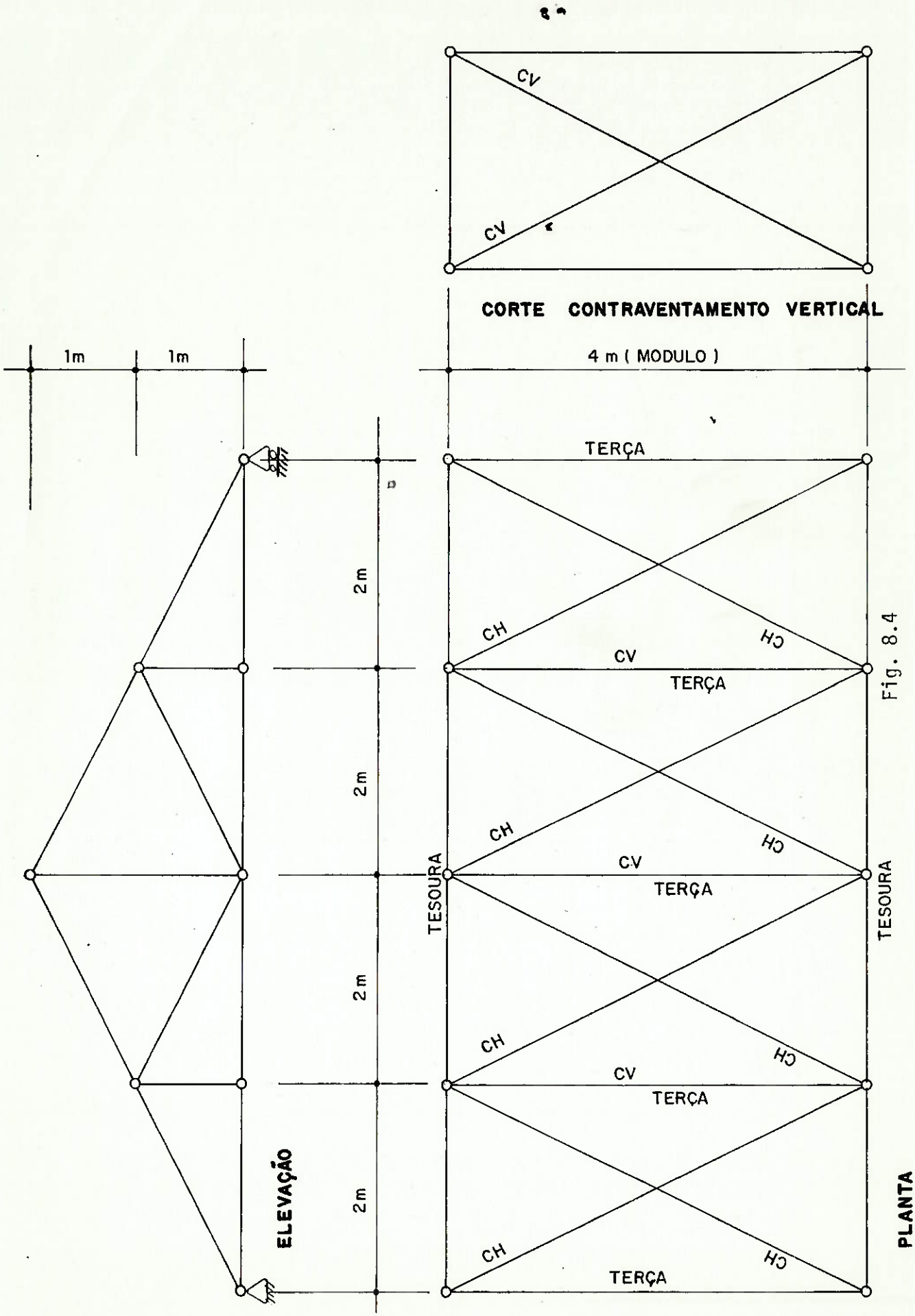

CV
CV
CORTE CONTRAVENTAMENTO VERTICAL
1m
1m
4 m (MODULO)
TERÇA
2 m
2 m
2 m
2 m
CH
CH
CV
TERÇA
CH
CH
CV
TERÇA
CH
CH
CV
TERÇA
CH
CH
TERÇA
TESOURA
TESOURA
ELEVAÇÃO
PLANTA

Fig. 8.4

9. CONCLUSÕES E SUGESTÕES

9.1. Conclusões

• Neste trabalho foram adotados como elementos de medida do condicionamento dos sistemas de equações os chamados *números de condição*. Em particular, utilizou-se o *número de condição espectral*, definido, para matriz de coeficientes $\underline{A}$, como:

$$h(\underline{A}) = \|\underline{A}\|_2 \quad \|\underline{A}^{-1}\|_2 \tag{9.1}$$

onde

$\|\underline{A}\|_2$ é a norma espectral de $\underline{A}$

$\|\underline{A}^{-1}\|_2$ é a norma espectral da inversa de $\underline{A}$

ou, de outra forma, para matrizes simétricas, dado por:

$$p(\underline{A}) = \frac{\lambda_{máx}}{\lambda_{mín}} \tag{9.2}$$

onde

$\lambda_{máx}$ é o maior autovalor de $\underline{A}$

$\lambda_{mín}$ é o menor autovalor de $\underline{A}$

• Os sistemas mal-condicionados, ou seja, propensos a amplificação ao longo do processo de solução de erros iniciais, possuem números de condição elevados.

O fato de um sistema ter número de condição de valor alto não autoriza, de momento, afirmar seu mau-condicionamento.

Os números de condição em geral não são função do volume de operações envolvido no processo de solução, o que levanta questões quanto a seu poder de avaliar os efeitos do arredondamento nesses cálculos. Entretanto, em caso de dúvida é conveniente, se possível, que se evitem números de condição grandes.

• Os vetores próprios da matriz de rigidez de uma estrutura, K, correspondem a formas de deslocamento em conjunto *similares* aos modos naturais de vibração da estrutura (para massa unitária associada a cada parâmetro de

deslocamento).

Os autovalores λ_i são caracterizados como a *rigidez* de cada um desses modos, relacionando a amplitude y_i de uma dada forma de deslocamento com a respectiva componente de carregamento modal P_i, como se em cada caso se tivesse um sistema de um grau de liberdade.

• O valor próprio máximo corresponde à forma de deslocamentos mais rígida (variação grande em sua componente de carregamento resultando em pequenas mudanças de amplitude). O valor próprio mínimo corresponde à forma de deslocamentos mais flexível (variações pequenas na componente de carregamento levando a variação grande de amplitude).

• Uma estrutura que possua formas de deslocamento muito rígidas coexistindo com outras muito flexíveis *poderá* estar sujeita a mau-condicionamento de suas equações de equilíbrio, já que possuirá número de condição elevado.

• O fato mesmo de os autovalores terem interpretação de rigidez das formas de deslocamento leva a dizer-se que as amplitudes das perturbações nos deslocamentos são função da magnitude das perturbações na correspondente componente de carregamento. Por outra forma: se o carregamento não for predominantemente na direção das formas com rigidez crítica, o número de condição obtido poderá não espelhar convenientemente possíveis tendências a amplificação de erros. O modelo utilizado pode ter sido mais refinado que o necessário: se ele não possuísse essas formas de deslocamento teria equações de equilíbrio em menor número e, provavelmente, mais bem-condicionadas.

9.2. Sugestões

Na intenção de se evitarem números de condição elevados, ou seja, coexistência em uma mesma estrutura de formas de deslocamento muito flexíveis e muito rígidas, vale sugerir:

- em pequenos pórticos, a consideração da deformação axial das barras pode, conforme cita *Livesley* em (*10*), ser causa de dificuldades com os sistemas de equações resultantes. Em um programa para microcomputador de uso restrito, poder-se-ia pensar em ignorá-la;

- uma grelha, que, por convenção, tem carregamento apenas normal a seu plano, nada ganha se calculada como pórtico espacial, em que aparecem modos de deslocamento muito flexíveis ou muito rígidos contidos em seu plano;

- um edifício de múltiplos andares, a pouca rigidez do deslocamento em conjunto, acompanhando o andar mais alto, face à grande dificuldade de se obter o deslocamento de apenas um andar intermediário, sugere que se obteria melhor condicionamentocom algum travamento, por exemplo, considerando a contribuição dos painéis de alvenaria;

- cuidado deve ser tomado em estruturas de carregamento auto-equilibrado em que, para evitar hipostaticidade do conjunto, se introduzem molas ou barras de fixação cuja rigidez deve ser criteriosamente adotada para que não se chegue a formas de deslocamento excessivamente rígidas ou flexíveis;

- às vezes, é contraproducente uma sofisticação no modelo. Se um pórtico espacial puder ser dividido em planos, para o carregamento considerado, esse modelo será mais favorável; o mesmo se pode dizer de uma estrutura treliçada que possa ser considerada uma seqüência de treliças planas.

Uma sugestão deste trabalho é que no caso de tesouras de cobertura paralelas o cálculo do número de condição de cada treliça como estrutura plana e do conjunto como espacial pode dar uma idéia do grau de adequabilidade do contraventamento, para comparação com parâmetros práticos e econômicos de projeto.

BIBLIOGRAFIA

(1) Albrecht, P.
"Análise Numérica - um Curso Moderno"
Rio de Janeiro, Livros Técnicos e Científicos, 1977.

(2) André, J. C.
"Introdução ao Método dos Elementos Finitos para Estruturas de Comportamento Linear"
São Paulo, EPUSP, 1975.

(3) Bathe, K. J., e Wilson, E. L.
"Numerical Methods in Finite Element Analysis"
Englewood Cliffs, Prentice-Hall, 1976.

(4) Clough, R. W., e Penzien, J.
"Dynamics of Structures"
Tokyo, McGraw-Hill/Kogakusha, 1975.

(5) Faddeev, D. K., e Faddeeva, V. N.
"Computational Methods of Linear Algebra"
San Francisco, Freeman, 1963.

(6) Gere, J. M., e Weaver, W.
"Análise de Estruturas Reticuladas"
Rio de Janeiro, Ed. Guanabara Dois, 1981.

(7) Irons, B., e Ahamad, S.
"Techniques of Finite Elements"
Chichester, John Wiley & Sons, 1968.

(8) Jennings, A., e Malik, G. M.
"The Solution of Sparse Linear Equations by the Conjugate Gradient Method"
Int. Journal for Num. Meth. in Eng., 12, pp. 141-158, 1978.

(9) Kardestuncer, H.
"Elementary Matrix Analysis of Structures"
New York, McGraw-Hill, 1974.

(10) Livesley, R. K.
"Matrix Methods of Structural Analysis"
London, Pergamon Press, 1975.

(11) Melosh, R. J.
"Manipulation Errors in Finite Element Analysis"
in "Recent Advances in Matrix Methods of Structural Analysis"
(Gallagher, Uamada & Oden, Ed.), Huntsville, 1971.

(12) Mirshavka, V.
"BASIC sem Segredos"
São Paulo, Nobel, 1983.

(13) von Neumann, J.; e Goldstine, H. H.
"Numerical Inverting of Matrices of High Order"
Bull. Am. Math. Soc., 53, pp. 1021-1099, 1947.

(14) Salvetti, D. D.
"Elementos de Programação"
São Paulo, Cia. Editora Nacional, 1972.

(15) Salvetti, D. D.
"Tópicos de Cálculo Numérico"
São Paulo, SBC, FCA, 1982.

(16) Shah, J. M.
"Ill-Conditioned Stiffness Matrices"
Proceedings, ASCE, Journal of the Structural Div., vol. 92,
No. ST6, pp. 443-457, 1966.

(17) Sheid, P.
"Introdução à Ciência dos Computadores"
São Paulo, McGraw-Hill, 1971.

(18) Todd, J.
"Basic Numerical Mathematics" - vol. 2, "Numerical Algebra"
Stuttgart, BirkHauser, 1972.

(19) Turing, A. M.
"Rounding-Off Errors in Matrix Processes"
Quart. Journal Mech. Applc. Math., 1, pp. 287-308, 1948.

(20) Wilkinson, J. M.
"Rounding Errors in Algebraic Processes"
Englewood Cliffs, Prentice-Hall, 1963.

(21) Zagottis, D. L.
"Conceituação do Método dos Elementos Finitos"
São Paulo, EPUSP, 1977.

(22) Zienkiewicz, O. C.
"The Finite Element Method in Engineering Science"
London, McGraw-Hill, 1971.