SILVIO IKUYO NABETA

Engenheiro Eletricista, Escola Politécnica da USP, 1983

SOLUÇÃO DE PROBLEMAS MAGNETOSTÁTICOS POR ELEMENTOS FINITOS UTILIZANDO O ICCG

Dissertação apresentada à Escola Politécnica da USP para obtenção do título de Mestre em Engenharia Elétrica.

Orientador: Prof. Dr. José Roberto Cardoso

São Paulo, 1990

Aos meus pais.

AGRADECIMENTOS

- Ao meu orientador Prof. Dr. José Roberto Cardoso, incansável incentivador durante todo o desenvolvimento deste trabalho;

- Ao Prof. Dr. Jean-Louis Coulomb do Institut National Polytechnique de Grenoble pelas discussões e sugestões na fase de elaboração do programa computacional;

- À Equacional Elétrica e Mecânica Ltda. pela gentileza em ceder os desenhos do motor analisado;

- Ao politécnico Valter Akira Honda pela tenacidade e dinamismo perante os inúmeros problemas computacionais surgidos;

- Ao engenheiro Douglas Ricardo de Freitas Clabunde pela excelente qualidade dos resultados gráficos;

- Ao analista de sistemas Carlos Hiroshi Sato pela orientação na utilização do editor "Chi-Writer";

- Ao engenheiro José Virgílio Meireles da Costa pelo auxílio na revisão do texto.

RESUMO

O objetivo deste trabalho é apresentar formalmente o Método dos Gradientes Conjugados com Pré-Condicionamento por Fatorização Incompleta de Cholesky (ICCG) para resolução de grandes sistemas, e aplicá-lo a um estudo magnetostático, desenvolvido pelo Método dos Elementos Finitos, de um motor síncrono de relutância

Esta apresentação consta ainda de uma comparação de desempenho do referido método com outros dois; o Método dos Gradientes Conjugados com Pré-Condicionamento por Sobre-relaxação Simétrica Sucessiva (SSOR-CG) e o Método de Decomposição de Cholesky, tanto a nível de resultados como a nível de tempo de processamento.

Foram calculados os auto-valores da matriz jacobiana para a análise da influência destes nos tempos de processamento dos três métodos utilizados.

As etapas de pré e pós processamentos foram realizadas através de programas já desenvolvidos pelo grupo de Elementos Finitos do Laboratório de Sistemas de Potência da Escola Politécnica da USP.

ABSTRACT

The aim of this work is to present formally the Incomplete Cholesky Conjugate Gradient Method (ICCG) used for solve large systems and its application in a magnetostatic study, which was developed by the Finite Element Method (FEM), of a synchronous reluctance motor.

This presentation also includes the comparison between the ICCG method and others two: the Symmetric Sucessive Overrelaxation Method (SSOR) and the Cholesky Method, pointing out their results and the processing time.

The eigenvalues of the jacobian matrix have been calculated in order to analyse their influence on the processing times of the three methods employed.

The pre and the post-processor stages were made using programs developed by the Finite Elements group in the Laboratory of Power Systems from Escola Politécnica da USP.

SUMÁRIO

CAPITULO 1: METODOS NUMERICOS DE RESOLUÇÃO DE SISTEMAS DE EQUAÇÕES RESULTANTES DO METODO DOS ELEMENTOS FINITOS

1.1 INTRODUÇÃO

Nestas últimas décadas, temos presenciado uma crescente evolução da informática e da disseminação do uso de computadores e microcomputadores.

Graças às potencialidades destas novas tecnologias em processar grandes quantidades de informações em velocidades consideráveis, introduziram-se profundas modificações nos meios de trabalho de engenheiros e pesquisadores.

Os métodos numéricos conquistaram uma posição de destaque como ferramentas poderosas e versáteis de análise de complexos problemas físicos.

Dentre estes métodos, o Método dos Elementos Finitos (MEF) tem-se apresentado com sucesso nas diversas áreas onde é utilizado, graças às virtudes de sua formulação que propiciam resultados acurados com relativa simplicidade de implantação.

O método consiste em transformar as equações diferenciais a derivadas parciais envolvidas no problema, em um sistema de equações algébricas lineares ou não-lineares a serem resolvidas por técnicas apropriadas.

Atualmente dá-se grande ênfase ao desenvolvimento de técnicas de resolução eficientes e rápidas, pois é nesta fase do MEF que se concentra grande parte do trabalho computacional, influenciando consideravelmente o custo de análise.

Apresentaremos neste trabalho um método para resolução de grandes sistemas lineares e que vem ganhando a credibilidade de vários pesquisadores. Recentemente desenvolvido, este método é

conhecido como Método dos Gradientes Conjugados com Pré-Condicionamento por Decomposição Incompleta de Cholesky (ICCG).

Além deste, constam outros dois métodos de resolução de sistemas lineares: o Método dos Gradientes Conjugados com Pré-Condicionamento por Sobre-Relaxação Simétrica Sucessiva (SSOR-CG) e o clássico Método da Decomposição de Cholesky, que terão os seus desempenhos comparados ao primeiro.

Para os sistemas não-lineares utilizaremos o tradicional Método de Newton-Raphson.

1.2 ESTADO DA ARTE

O Método dos Elementos Finitos (MEF) surgiu inicialmente como ferramenta de análise na Mecânica de Estruturas, principalmente na resolução de problemas provenientes da indústria aeronáutica.

Embora a formulação do método fosse conhecida desde 1909, através dos estudos de Ritz, e posteriormente ter sido modificada, em 1943, através da extensão proposta por Courant, foi apenas em 1960 que primeiramente se utilizou a terminologia Elementos Finitos no artigo publicado por Clough, "The Finite Element Method in Plane Stress Analysis".

Desde então, aliada ao advento dos computadores digitais, a utilização do MEF se difundiu enormemente. Hoje a sua aplicação engloba, entre outras, problemas tridimensionais, não-lineares e dinâmicos, abrangendo, também, áreas além da estrutural, tais como Mecânica dos Fluidos, Transferência de Calor e Eletromagnetismo.

Toda esta generalização de utilização do MEF trouxe consigo um aumento das dimensões dos sistemas algébricos lineares e não-lineares, exigindo o desenvolvimento de algoritimos mais eficazes de resolução, servindo-se, até mesmo, de técnicas de

armazenamento compacto que aproveitam as características intrínsecas das matrizes, tais como esparsidade e simetria.

Os métodos de resolução de sistemas lineares podem ser classificados do seguinte modo:

1. Método Direto,onde a solução é atingida após um número finito de operações conhecido com antecedência.

   Ex: Eliminação de Gauss, Banachievicz e Decomposição de Cholesky.

2. Método Iterativo, onde a solução é atingida através de um processo recursivo com uma precisão pré-determinada.

   Ex: Gauss-Seidel, Sobre-relaxação, Jacobi, Gradientes Conjugados.

Segundo Dhatt e Touzot [9], no princípio da utilização do MEF, os métodos iterativos eram os preferidos devido à simplicidade de programação e à menor alocação de memória. Com o decorrer do tempo, entretanto, os métodos diretos conquistaram espaço apoiando-se, principalmente, na vantagem de exigir menor número de operações para a obtenção da solução.

O inconveniente dos métodos diretos é a alteração da esparsidade da matriz resultante durante a decomposição da matriz original. Esta alteração resulta da introdução de elementos não-nulos em posições onde originalmente existiam zeros.

O aumento de elementos não-nulos acarreta um aumento do tempo de computação e torna complexa a utilização do armazenamento compacto.

Um fato de grande importância para o MEF é que esta introdução de elementos não-nulos está intimamente ligada à numeração global dos nós da malha, uma vez que esta influi na largura da banda da matriz.

O estudo de técnicas de ordenação de nós tem sido objeto de intensa pesquisa nestes últimos anos.

Algumas das técnicas mais recentes, tais como o Método Frontal, a Ordenação por Dissecações Sobrepostas (Nested Dissection Ordering) e o algoritimo de Cuthill-Mckee Reverso (RCM) podem ser encontradas com detalhes em George e Liu [12]; Duff, Erisman e Reid [10] e Axelsson [2].

No campo dos métodos iterativos o ICCG-Método dos Gradientes Conjugados com Pré-Condicionamento por Decomposição Incompleta de Cholesky tem se destacado pelos bons resultados apresentados a nível de rapidez de processamento, precisão e estabilidade numérica.

Este método é um aperfeiçoamento do Método dos Gradientes Conjugados descrito inicialmente por Hestenes e Stiefel em 1952 e que, devido às limitações computacionais da época, não suscitou maiores interesses.

O Método do Gradientes Conjugados baseia-se na resolução do sistema linear através da minimização iterativa de uma funcional quadrática utilizando as direções conjugadas.

Em 1970, Reid retoma o trabalho de Hestenes e Stiefel e demonstra a potencialidade do método em resolver grandes sistemas esparsos.

No sentido de conduzir a minimização de uma forma mais acelerada, Meijerink e Van der Vorst [5] propõe um pré-condicionamento da matriz do sistema, visando deixá-la próxima da matriz identidade, utilizando uma decomposição incompleta de Cholesky (IC).

Vários outros processos de pré-condicionamento têm sido propostos desde então, como o pré-condicionamento por sobre-relaxação simétrica sucessiva (SSOR) proposta por Axelsson [2].

Atualmente, observa-se, face ao contínuo aumento das dimensões

do sistema, que as pesquisas de aceleração de cálculo se orientam na direção de métodos que exploram uma melhor utilização das possibilidades computacionais, em particular os processamentos paralelo e vetorial.

CAPITULO 2: FORMULAÇÃO MATEMÁTICA DO PROBLEMA MAGNETOSTÁTICO PELO MÉTODO DOS ELEMENTOS FINITOS

2.1 INTRODUÇÃO

As Equações de Maxwell e as relações constitutivas são as bases para a formulação, pelo Método dos Elementos Finitos, do comportamento do campo eletromagnético em qualquer dispositivo eletromecânico.

Para o caso particular da magnetostática as Equações de Maxwell e as relações constitutivas de interesse são:

$$\nabla \times \vec{H} = \vec{J} \qquad [2.1]$$

$$\nabla . \vec{B} = 0 \qquad [2.2]$$

$$\vec{H} = \nu(B) . \vec{B} \qquad [2.3]$$

com $\vec{H}$: vetor intensidade magnética (A/m);

$\vec{B}$: vetor campo magnético (Wb/m^2);

$\vec{J}$: vetor densidade de corrente (A/m^2);

$\nu = 1/\mu$: relutividade magnética (m/H).

A partir de [2.2] é definido o vetor potencial magnético $\vec{A}$, tal que:

$$\vec{B} = \nabla \times \vec{A} \qquad [2.4]$$

Além desta relação é necessário impor outra condição para que $\vec{A}$ fique definido univocamente. Esta condição denominada "gauge" de Coulomb é a seguinte:

$$\nabla . \vec{A} = 0 \qquad [2.5]$$

Substituindo [2.4] em [2.3] e seu resultado em [2.1], obtemos:

$$\nabla \times \nu(B) . \nabla \times \vec{A} = \vec{J} \qquad [2.6]$$

A equação [2.6] é a Equação de Poisson não-linear para a magnetostática.

A análise a ser efetuada neste trabalho, como veremos, é tal que

pode ser representada por um modelo bidimensional. Neste caso vamos admitir que a densidade de corrente $\vec{J}$ é perpendicular ao plano de estudo XY, ou seja $\vec{J}=J.\vec{u}_z$, de modo que resultará um vetor potencial magnético:

$$\vec{A}=A.\vec{u}_z \qquad [2.7]$$

implicando, portanto, que [2.5] é automaticamente satisfeita.

Por outro lado, a Equação de Poisson não-linear [2.6], em vista das condições do modelo não-linear, pode ser reescrita na forma simplificada como segue:

$$\nabla.(\nu.\nabla A_0)+J=0 \qquad [2.8]$$

onde $\nabla=\frac{\partial}{\partial x}\vec{u}_x+\frac{\partial}{\partial y}\vec{u}_y$ e A_0 é a solução exata da componente z da Equacão de Poisson não-linear.

2.2 CONDIÇÕES DE CONTORNO

As condições de contorno encontradas nos limites do domínio são do tipo:

2.2.1 Condições de Dirichlet.

$$A_0=A_1 \text{ em } S_1 \qquad [2.9]$$

2.2.2 Condições de Neumann.

$$\frac{\partial A_0}{\partial n}=0 \text{ em } S_2 \qquad [2.10]$$

2.2.3 Condições de Periodicidade.

Algumas aplicações apresentam condições de periodicidade do tipo:

2.2.3.1 Condição Cíclica.

Neste caso, o potencial magnético num ponto da fronteira S_3 é igual a um ponto correspondente na fronteira S_4, isto é:

$$A(S_3)=A(S_4).$$

2.2.3.2 Condição Anti-cíclica.

Neste caso, o valor do potencial magnético num ponto da fronteira S_3 é oposto a um ponto correspondente na fronteira S_4, isto é:

$$A(S_3)=-A(S_4).$$

As condições de periodicidade não introduzem modificações na formulação matemática por Elementos Finitos. Suas utilizações permitirão reduzir ao máximo o domínio em estudo.

Note que $S=S_1 \cup S_2 \cup S_3 \cup S_4$.

2.3 FORMULAÇÃO BIDIMENSIONAL POR ELEMENTOS FINITOS

A resolução de [2.8], sujeita às condições de contorno expressas em 2.2, é realizada através da aplicação do Método dos Resíduos Ponderados, como segue [7]:

Se A_0 é a solução exata de [2.8], sujeita às condições de contorno já estabelecidas, a soma integral que se segue é identicamente nula, isto é:

$$\int_{\Delta} W_0 \cdot [\nabla \cdot (\nu \cdot \nabla A_0) + J] d\Delta + \int_{S_1} W_1 \cdot (A_0 - A_1) dS + \int_{S_2} W_2 \cdot \left(\frac{\partial A_0}{\partial n}\right) dS = 0 \quad [2.11]$$

onde as funções W's são denominadas funções de ponderação e sobre as quais não são impostas exigências de continuidade, excluídas apenas suas nulidades.

Como em todo método numérico, vamos procurar uma solução

aproximada de [2.8]. Seja A esta solução aproximada de modo que, substituída na Equação de Poisson e nas condições de contorno resulte nos seguintes Resíduos de Aproximação:

$$\left.\begin{aligned} R_o &= \nabla.(\nu.\nabla A)+J \text{ em } \Delta \\ R_1 &= A-A_1 \text{ em } S_1 \\ R_2 &= \frac{\partial A}{\partial n} \text{ em } S_2 \end{aligned}\right\} \qquad [2.12]$$

Desta forma, a soma integral indicada em [2.11], escrita em termos da solução aproximada A, resulta diferente de zero.

No Método dos Elementos Finitos vamos procurar a solução aproximada A que anula a referida integral.

Assim vamos impor:

$$\int_\Delta W_o.\nabla.(\nu.\nabla A+J)d\Delta + \int_{S_1} W_1.(A-A_1)dS + \int_{S_2} W_2.\left(\frac{\partial A}{\partial n}\right)dS = 0 \qquad [2.13]$$

Note que em [2.11] cada um dos termos da integral é identicamente nulo, ao passo que em [2.13] a soma é nula, de modo que os erros de aproximação serão distribuídos parte no domínio e parte na fronteira, segundo as funções de ponderação W's.

Aplicando o Teorema de Green em uma parte da primeira integral de [2.13], obtemos:

$$\int_\Delta W_o.\nabla.(\nu.\nabla A)d\Delta = -\int_\Delta \nu.\nabla W_o.\nabla A\ d\Delta + \oint_{S_1+S_2} \nu.W_o.\frac{\partial A}{\partial n}\ dS \qquad [2.14]$$

Substituindo em [2.13] obtemos a equação abaixo:

$$-\int_\Delta \nu\nabla W_o\nabla A\ d\Delta+\oint_{S_1+S_2} \nu W_o\frac{\partial A}{\partial n}\ dS+\int_\Delta W_o J\ d\Delta+\int_{S_1} W_1(A-A_1)dS+\int_{S_2} W_2\frac{\partial A}{\partial n}dS=0 \qquad [2.15]$$

A expressão [2.15] é denominada forma "fraca" do Método dos Resíduos Ponderados pois reduz o grau de continuidade da variável A, permitindo representá-la por funções mais simples impondo, entretanto, uma condição de continuidade para W_o, a qual deverá ter, no mínimo, a primeira derivada diferente de zero.

No Método dos Elementos Finitos impõe-se também que a solução do problema sobre S_1 seja exata, de forma que teremos $R_1=0$.

Em vista disto, a quarta integral de [2.15] desaparecerá.

Escolhendo-se $W_2=-\nu . W_0$ e expandindo-se a segunda integral em duas, uma sobre S_1 e outra sobre S_2, obteremos:

$$-\int_{\Delta} \nu . \nabla W_0 . \nabla A \, d\Delta + \int_{S_1} \nu . W \cdot \frac{\partial A}{\partial n} \, dS + \int_{\Delta} W_0 . J \, d\Delta = 0 \qquad [2.16]$$

A resolução pelo Método dos Elementos Finitos consiste em subdividir o domínio em pequenos subdomínios, denominados elementos finitos, sobre os quais as equações de comportamento dos campos são satisfeitas.

Assim sendo, supondo o domínio discretizado, a equação [2.16] pode ser escrita como segue:

$$\sum_{e=1}^{NE} \left\{ -\int_{\Delta^e} \nu . \nabla W_0 . \nabla A \, d\Delta + \int_{S_1^e} \nu . W_0 \cdot \frac{\partial A}{\partial n} \, dS + \int_{\Delta^e} W_0 . J \, d\Delta \right\} = 0 \qquad [2.17]$$

onde NE é o número total de elementos do domínio discretizado.

Para um melhor resultado impõe-se que a soma integral dentro das chaves seja nula, assim a expressão [2.16] é satisfeita em cada elemento obtendo:

$$-\int_{\Delta^e} \nu . \nabla W_0 . \nabla A \, d\Delta + \int_{S_1^e} \nu . W_0 \cdot \frac{\partial A}{\partial n} \, dS + \int_{\Delta^e} W_0 . J \, d\Delta = 0 \qquad [2.18]$$

A escolha adequada da função de ponderação W_0 leva aos vários métodos numéricos de resolução dos problemas de campos em geral.

Para o Método dos Elementos Finitos escolhe-se W_0 como sendo uma variação arbitrária da variável A (Galerkin), isto é:

$$W_0 = \delta A \qquad [2.19]$$

Escolhendo-se W_0 desta forma, a segunda integral sobre S_1^e desaparecerá, visto que nesta parte da fronteira a solução é imposta exata, resultando nula a variação nesta região.

Lembrando que em um ponto qualquer no interior do elemento o valor da variável A é obtida através de uma interpolação dos valores nos vértices do elemento, ou seja:

$$A=\sum_{i=1}^{n} N_i . A_i \qquad [2.20]$$

onde A_i: Potencial magnético no vértice i do elemento;

n: número de vértices do elemento.

Podemos, assim, escrever:

$$W_o=\sum_{j=1}^{n} N_j . \delta A_j \qquad [2.21]$$

Substituindo [2.20] e [2.21] em [2.18] resulta:

$$\sum_{j=1}^{n}\left\{-\sum_{i=1}^{n}\int_{\Delta^e}\nu . \nabla N_i . \nabla N_j . A_i \; d\Delta + \sum_{i=1}^{n}\int_{\Delta^e} N_j . J \; d\Delta\right\} \delta A_j = 0 \qquad [2.22]$$

sendo δA_j uma variação arbitrária da variável nodal, a solução de [2.22] só será possível se todos os seus coeficientes forem nulos. Esta afirmação pode ser expressa matricialmente sobre cada elemento como segue:

$$S^e . A^e = I^e \qquad [2.23]$$

onde:
$$\left.\begin{aligned} S^e_{ij} &= \int_{\Delta^e}\nu . \nabla N_i . \nabla N_j \; d\Delta \\ I^e_j &= \int_{\Delta^e} N_j . J \; d\Delta \end{aligned}\right\} \qquad [2.24]$$

S^e: matriz do elemento;

$A^e=[A_1 \; A_2 \ldots . A_n]^T$: vetor dos potenciais magnéticos.

A introdução de [2.22] em [2.17] gera o sistema de equações não-linear:

$$\bar{S} . A = \bar{I} \qquad [2.25]$$

que é obtido a partir das matrizes dos elementos S^e [3].

A matriz $\bar{S}$, denominada matriz global do domínio, tem as propriedades da singularidade e da esparsidade. A introdução das condições de contorno sobre os vértices pertencentes a S_1 [7] levantam esta singularidade de modo que, após este procedimento, o sistema resultante é do tipo:

$$S . A = I \qquad [2.26]$$

com S: matriz global final NN X NN;

A: $[A_1 \; A_2 \ldots A_{NN}]^T$ NN X 1;

I: $[I_1\ I_2 \ldots I_{NN}]^T$ NN X 1;

NN: número de nós do domínio.

2.4 MODELAGEM DA NÃO-LINEARIDADE

A relutividade do material ferromagnético foi expressa através de uma curva polinomial cúbica conforme desenvolvida por Silvester, Cabayan, Browne [19].

Assim, definido o intervalo de densidade de fluxo e B_1 e B_2 os valores nas extremidades, assumimos:

$$x=\frac{B^2-B_1^2}{B_2^2-B_1^2},$$

onde: $B_1 \leq B \leq B_2,\ B_1 \leq B_2.$

Deste modo, o problema pode ser analisado como uma aproximação da relutividade em função de x, no intervalo $0 \leq x \leq 1$.

A expressão cúbica representante da relutividade será da forma:

$$\nu(x)=(2x^3-3x^2+1).\nu(0)+(-2x^3+3x^2).\nu(1)+(x^3-2x^2+x).\nu'(0)+(x^3-x^2).\nu'(1)$$

onde: $\nu(0)$ e $\nu(1)$ são os valores da relutividade nos extremos do intervalo e $\nu'(0)$ e $\nu'(1)$ são os valores da derivada da relutividade em relação a B^2 nestes pontos.

Os valores característicos, indicados nos extremos do intervalo, são obtidos a partir da construção gráfica da característica ν X B^2, a qual é subdividida em intervalos em cujos extremos são obtidos os valores da relutividade e de suas derivadas.

Este procedimento, exige a observação de algumas propriedades, a saber:

1. as derivadas devem ser contínuas para todos extremos do segmento, ou seja, deve-se verificar se a derivada no extremo direito de um segmento é igual à derivada obtida no extremo esquerdo do segmento contíguo. Isto se deve à expressão:

$$\frac{d\nu}{d(B^2)}=(6x^2-6x).\nu(0)+(-6x^2+6x).\nu(1)+(3x^2-4x+1).\nu'(0)+(3x^2-2x).\nu'(1)$$

2. a curva deve ser contínua em todos os extremos do segmento, isto é, através da expressão:

$$\frac{d^2\nu}{d(B^2)^2}=(12x-6).\nu(0)+(-12x+6).\nu(1)+(6x-4).\nu'(0)+(6x-2).\nu'(1)$$

o valor do extremo direito de um segmento, resulta no mesmo valor do extremo esquerdo do segmento contíguo.

3. A derivada em B=0 deve ser nula.

4. No segmento mais a direita (saturação) a derivada é constante.

Neste trabalho a característica de magnetização do material ferromagnético utilizado é mostrada na figura [2.1].

Após a construção da curva ν X B^2 a mesma foi segmentada em 5 intervalos, resultando nas seguintes expressões:

$\nu_1=280$ para $0\leq B^2\leq 0,6$;

$\nu_2=168x^3-173x^2+75x+280$ para $0,6\leq B^2\leq 1,2$;

$\nu_3=486x^3-429x^2+233x+350$ para $1,2\leq B^2\leq 1,8$;

$\nu_4=716x^3-899x^2+833x+640$ para $1,8\leq B^2\leq 2,4$;

$\nu_5=1183,3.B^2-1550$ para $B^2\geq 2,4$.

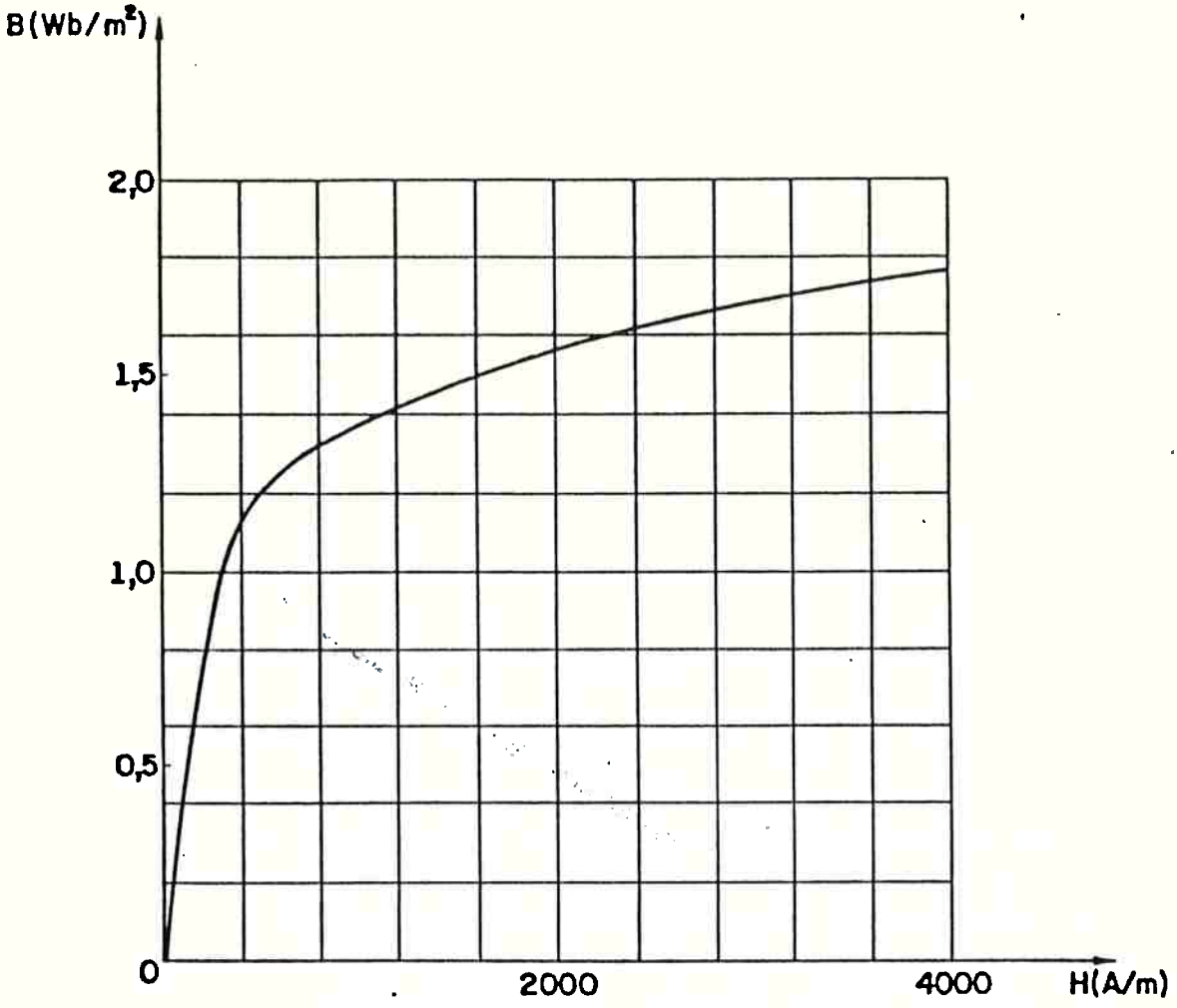

Figura 2.1: Característica de magnetização do Ferro-Silício.

2.5 RESOLUÇÃO DE SISTEMAS NÃO-LINEARES

Como visto na seção precedente, a formulação do problema magnetostático pelo MEF conduz-nos ao sistema não linear:

$$S.A=I,$$

onde A = vetor dos potenciais magnéticos nodais;

I = vetor associado às fontes de corrente;

S = matriz global.

A resolução deste sistema será através do Método de Newton-Raphson, analisado a seguir.

2.5.1 METODO DE NEWTON-RAPHSON

O sistema não linear [2.26] pode ser posto na forma:

$$R=S.A-I=0 \qquad [2.27]$$

Desenvolvendo [2.27] pela série de Taylor e truncando os termos de ordem superior a um, obtemos:

$$R(A^{k+1})=R(A^{k})+\frac{\partial R(A^{k})}{\partial A}\cdot \Delta A^{k}=0,$$

com $$\Delta A^{k}=A^{k+1}-A^{k}.$$

A matriz $P(A^{k})=\frac{\partial R(A^{k})}{\partial A}$ é denominada jacobiano do sistema.

Reordenando temos:

$$\left.\begin{aligned} P(A^{k}).\Delta A^{k}&=-R(A^{k})=S(A^{k}).A^{k}+I \\ A^{k+1}&=A^{k}+\Delta A^{k} \end{aligned}\right\} \qquad [2.28]$$

A solução do sistema será obtida através de um processo iterativo em [2.28], onde os termos gerais de P e R serão, conforme [13], da forma:

$$P^{\bullet}_{ij}=S^{\bullet}_{ij}+2.\sum_{k=1}^{n}\sum_{l=1}^{n}A^{\bullet}_{k}.A^{\bullet}_{l}.\int_{\Omega_e}(\nabla N_i.\nabla N_k)\cdot\frac{\partial \nu}{\partial B^{2}}\cdot(\nabla N_j.\nabla N_l)\,d\Omega_e,$$

$$R^{\bullet}_{i}=\sum_{k=1}^{n}S^{\bullet}_{ik}.A_k-I^{\bullet}_{k}=\int_{\Omega_e}\left(\sum_{k=1}^{n}\nu.(\nabla N_i.\nabla N_k).A_k-I.N_k\right)d\Omega_e.$$

A convergência do sistema será testada através da inequação:

$$\frac{\|\Delta A^{k}\|}{\|A^{k+1}\|}\leq \varepsilon,$$

onde ε é o valor do erro tolerável pré-estabelecido.

2.5.2 ASPECTOS COMPUTACIONAIS

O Diagrama Estruturado utilizado no Método de Newton-Raphson é apresentado na figura 2.2.

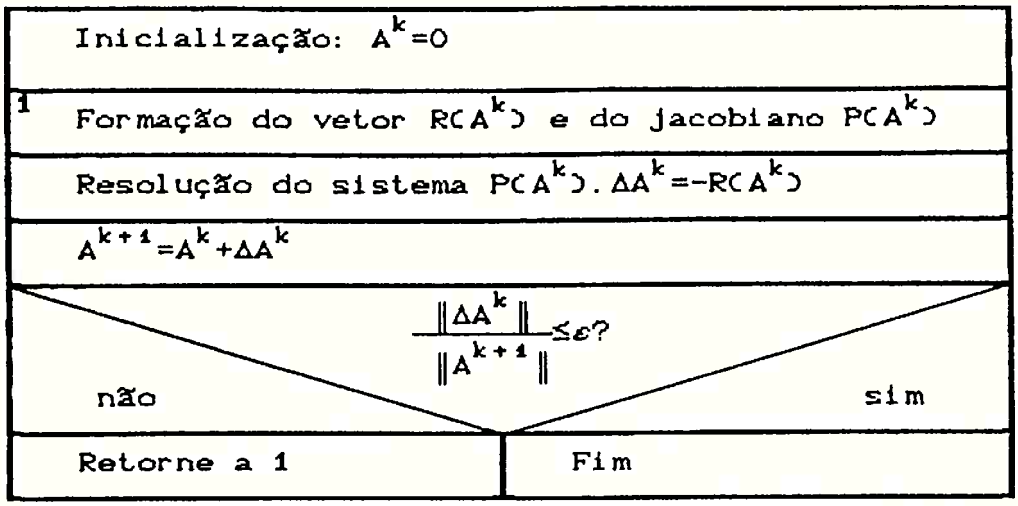

Figura 2.2: Diagrama Estruturado do Método de Newton-Raphson.

O armazenamento compacto da matriz P seguirá o Método da Lista Ordenada por Linha mostrado no exemplo da figura 2.3.

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

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
P	4	-1	-1	4	-1	-1	4	-1	4	-1	-1	4	-1	-1	4	-1	4	-1	4	-1	4
CN	1	2	4	2	3	5	3	6	4	5	7	5	6	8	6	9	7	8	8	9	9

RS	1	4	7	9	12	15	17	19	21
	1	2	3	4	5	6	7	8	9

Figura 2.3: Armazenamento compacto pelo Método da Lista Ordenada por Linha.

P é o vetor com os elementos não-nulos da matriz triangular superior de P.

CN é o vetor com os números da coluna dos elementos de P.

RS é o vetor com os posicionamentos dos elementos da diagonal principal.

CAPITULO 3: RESOLUÇÃO DE GRANDES SISTEMAS LINEARES ESPARSOS.

3.1 INTRODUÇÃO

A resolução do sistema não-linear [2.3] através do Método de Newton-Raphson introduz, por meio da expansão por série de Taylor, um sistema linear $P.\Delta A=-R$ onde P é o jacobiano do sistema.

As técnicas de resolução deste sistema linear compõem o objeto deste capítulo.

3.2 MÉTODOS ITERATIVOS

Os métodos iterativos apresentados são uma evolução do Método dos Gradientes Conjugados onde a resolução do sistema linear do tipo $H.x=b$ consiste em associar a este uma funcional quadrática da forma:

$$f(x)=\frac{1}{2}.x^T.H.x-b^T.x,$$

com $x \in \mathbb{R}^N$ e através da minimização desta, chegar à solução do sistema de equações dado.

Para a descrição dos métodos em questão faz-se necessária a apresentação formal do Método dos Gradientes Conjugados e de algumas definições, que permitirão uma compreensão mais suscinta deste.

3.2.1 FUNCIONAIS QUADRÁTICAS

Sejam: S um conjunto;

X um sub-conjunto de S;

$\mathbb{R}$ o conjunto dos números reais.

Definimos a funcional f em X como sendo a correspondência $f: X \Rightarrow \mathbb{R}$

Para o nosso caso: $S=\mathbb{R}^N$, o espaço vetorial de dimensão N;

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \quad x_i \in \mathbb{R}^N;$$ um elemento do sub-conjunto X.

Tomando-se então $x \in \mathbb{R}^N$ e $\xi > 0$; definimos $S(x,\xi)$ como sendo a vizinhança de x de raio ξ, onde:

$$S(x,\xi) = \{ y \in \mathbb{R}^N;\ 0 \leq \|x-y\| < \xi \};$$

sendo $\| \cdot \|$ a norma vetorial Euclidiana definida para todo $x \in \mathbb{R}^N$ com $\| x \| = \left[\sum_{i=1}^{N} x_i^2 \right]^{1/2}$.

DEFINIÇÃO 3.1:

Seja a funcional f definida em $X \subseteq \mathbb{R}^N$.

Então $\hat{x} \in X$ é dito como o minimizador local de f, se existir um $\xi > 0$ tal que $f(\hat{x}) \leq f(x)$, $\forall x \in S(\hat{x},\xi) \cap X$.

Se $f(\hat{x}) < f(x)$, $\forall x \in S(\hat{x},\xi) \cap X$, $x \neq \hat{x}$, então $\hat{x}$ será o minimizador local forte de f.

DEFINIÇÃO 3.2:

Seja a funcional f definida em $X \subseteq \mathbb{R}^N$.

Então $\hat{x} \in X$ será o minimizador global de f se $f(\hat{x}) \leq f(x)$, $\forall x \in X$.

Se $f(\hat{x}) < f(x)$ $\forall x \in X$, $x \neq \hat{x}$, então $\hat{x}$ será o minimizador global forte de f.

Antes de analisarmos a vizinhança do minimizador local $\hat{x}$, introduziremos o vetor gradiente e a matriz Hessiana.

O sub-conjunto X é dito aberto em $\mathbb{R}^N$, ou seja, todo $x \in X$ possui a propriedade de $S(x,\xi) \subset X$ para valores de ξ suficientemente pequenos.

Definimos então $C^m(X)$, onde m é um inteiro não negativo, como

sendo o conjunto de funcionais para as quais todas as derivadas parciais de ordem $\leq m$ existem e são contínuas em X.

DEFINIÇÃO 3.3

Seja $f \in C^1(X)$. O gradiente de f em $x \in X$ é o vetor

$$g(x) = \begin{bmatrix} \partial f/\partial x_1 \\ \vdots \\ \partial f/\partial x_N \end{bmatrix}$$

Seja $f \in C^2(X)$. A matriz Hessiana de f em $x \in X$ é a matriz simétrica:

$$H(x) = [h_{ij}]_{i,j=1}^{N}, \quad h_{i,j} = \partial^2 f/\partial x_i . \partial x_j$$

Analisaremos agora a vizinhança do minimizador local $\hat{x}$, utilizando para isso a expansão de Taylor.

Seja $f \in C^m(X)$ e para qualquer $x \in X$; $y \in \mathbb{R}^N$, $\|y\|=1$; considere a função $f(x+\tau . y)$; $0 \leq \tau \leq \xi$.

Pelo Teorema de Taylor, existe $\phi \in (0,1)$, tal que:

$$f(x+\tau . y) = \sum_{k=0}^{m-1} \frac{\tau^k}{k!} \cdot f^{(k)}(x,y) + \frac{\tau^m}{m!} \cdot f^{(m)}(x+\phi . \tau . y, y)$$

Desenvolvendo f para m=2 e substituindo $\tau . y$ por h obtemos:

$$f(x+h) = f(x) + g^T(x) . h + \frac{1}{2} \cdot h^T . H(x+\phi . h) . h$$

$$= f(x) + g^T(x) . h + \frac{1}{2} \cdot h^T . H(x) . h + R(x;h),$$

onde $R(x;h) = \frac{1}{2} \cdot h^T . [H(x+\phi . h) - H(x)] . h = \Phi(\|h^2\|)$.

Assim:

$$f(x+h) = f(x) + g^T(x) . h + \frac{1}{2} \cdot h^T . H(x) . h + \Phi(\|h\|^2) \qquad [3.1]$$

e analogamente para m = 1:

$$f(x+h) = f(x) + g^T(x) . h + \Phi(\|h\|) \qquad [3.2]$$

DEFINIÇÃO 3.4:

Seja $f \in C^1(X)$.

f é dita estacionária em $\hat{x} \in X$ se $g(\hat{x})=0$.

TEOREMA 3.1:

Seja $f \in C^1(X)$.

Se $\hat{x} \in X$ é um minimizador local de f, então f é estacionária em $\hat{x}$.

PROVA:

Substituindo em [3.2]: $x=\hat{x}$ e $h=-\alpha.g(\hat{x})$ onde α é uma variável real no intervalo $[0,\alpha_o]$, temos:

$$f(\hat{x}+h)=f(\hat{x})-\alpha.\|g(\hat{x})\|^2+\Phi(\alpha).$$

Considere que f é não estacionária em $\hat{x}$.

Então: $g(\hat{x})\neq 0$ e para um α suficientemente pequeno encontramos:

$$-\alpha.\|g(\hat{x})\|^2+\Phi(\alpha)<0$$

ou $f(\hat{x}+h)<f(\hat{x})$ e assim $\hat{x}$ não pode ser um minimizador local.

A condição de f estacionária em $\hat{x}$ é necessária mas não suficiente para $\hat{x}$ ser um minimizador local.

A condição suficiente é dada no próximo teorema.

TEOREMA 3.2:

Seja $f \in C^2(X)$ e seja f estacionária em X.

Então $\hat{x}$ é um minimizador local forte de f e a matriz Hessiana $H(\hat{x})$ é definida positiva.

PROVA:

Introduzindo $\hat{x}$ e $g(\hat{x})=0$ na expressão [3.1] obtemos:

$$f(\hat{x}+h)=f(\hat{x})+\frac{1}{2}.h^T.H(\hat{x}).h+\Phi\left(\|h\|^2\right)$$

Se $H(\hat{x})$ é definida positiva, então existe um número positivo λ_1 tal que: $h^T.H(\hat{x}).h\geq\lambda_1.\|h\|^2$, $\forall h \in \mathbb{R}^N$.

Então: $f(\hat{x}+h)-f(\hat{x})\geq\frac{1}{2}.\lambda_1.\|h\|^2+\Phi\left(\|h\|^2\right)$

Como $f(\hat{x}+h)-f(\hat{x})>0$ para valores suficientemente pequenos de h, temos que $f(\hat{x}+h)>f(\hat{x})$ em alguma vizinhança de $\hat{x}$, ou seja, $\hat{x}$ é um minimizador local forte de f.

DEFINIÇÃO 3.5:

Seja $f \in C^m(X)$, $x \in X \subseteq \mathbb{R}^N$ e $y \in \mathbb{R}^N$, onde $\|y\|=1$.

A derivada direcional de ordem m de f em x na direção y será:

$$f^{(m)}(x,y) \equiv \left. \frac{d^m f(x+\tau . y)}{d\tau^m} \right|_{\tau=0}$$

Assim para: $f \in C^1(X)$: $f^{(1)}(x,y)=g^T(x).y$ [3.3]

$f \in C^2(X)$: $f^{(2)}(x,y)=y^T.H(x).y$ [3.4]

Da desigualdade de Cauchy-Schwarz, ou seja,

$$|x^T.y| \leq \|x\|.\|y\| \quad , \quad \forall x,y \in \mathbb{R}^N$$

e da expressão [3.3] obtemos:

$$\max_{y,\|y\|=1} \left| f^{(1)}(x,y) \right| = \left| f^{(1)}(x,\hat{y}) \right| = \left\| g(x) \right\|;$$

$$\hat{y} \equiv g(x) / \|g(x)\|$$

Esta expressão e [3.3] provam que $f^{(1)}(x,y)=0$ para todas as direções y, se e somente se $g(x) = 0$, ou seja, se e somente se x é o ponto estacionário de f.

Além disso, é certo de [3.4] que $f^{(2)}(x,y)>0$ para todas as direções y, se e somente se $H(x)$ é definida positiva.

Com base nestas observações podemos reformular os teoremas precedentes da seguinte forma:

TEOREMA 3.3:

Seja $f \in C^1(X)$.

Se $\hat{x} \in X$ é um minimizador local de f, então $f^{(1)}(x,y)=0$ para todas as direções y.

TEOREMA 3.4:

Seja $f \in C^2(X)$ e suponha que para algum $\hat{x} \in X$ temos $f^{(1)}(x,y)=0$ para todas as direções y.

Então $\hat{x}$ é um minimizador local forte de f se $f^{(2)}(x,y) \geq 0$ para todas as direções de y.

DEFINIÇÃO 3.6:

Seja f definida em $X \subseteq \mathbb{R}^N$ e seja $k \in R_f$, onde R_f é a extensão de f, ou seja, $R_f=\{k \in \mathbb{R};\ f(x)=k \text{ para algum } x \in X\}$.

Então o conjunto $L_k= \{x \in X;\ f(x)=k\}$ define o nível de superfície de f.

Supondo $\hat{x} \in X$ como o minimizador local de f e $f(\hat{x})=k$, então a intersecção de L_k e a vizinhança $S(\hat{x},\xi)$ consiste de um único ponto $\hat{x}$ se ξ é suficientemente pequeno.

Para valores de k ligeiramente superiores a $\hat{k}$, os níveis de superfícies geralmente circundam $\hat{x}$ e assumem a forma de elipsóides quando $k \to \hat{k}$, conforme mostrado na figura 3.1.

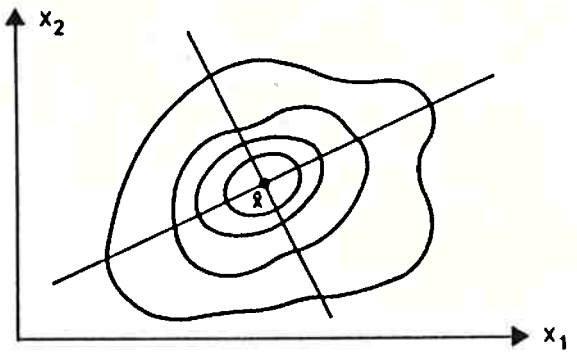

Figura 3.1 - Níveis de superfície na vizinhança do minimizador local

Um fato bem conhecido é que se x pertence ao nível de superfície L_k, então o vetor gradiente $g(x)$ é perpendicular a L_k em x e aponta na direção a qual a funcional aumenta mais rapidamente.

Geralmente, métodos numéricos utilizados na determinação do minimizador local forte de uma funcional se apresentam melhor quando os níveis de superfície na vizinhança do minimizador são esferas e pior quando existem distorções pronunciadas destas.

A grandeza $D_k=\inf_{y \in S_k}\{\sup_{x \in L_k}\|x-y\| / \inf_{x \in L_k}\|x-y\|\} \geq 1$, onde S_k é o conjunto de todos os pontos interiores a L_k; é uma medida desta distorção de L_k da forma esférica.

No caso de esferas; $D_k=1$.

DEFINIÇÃO 3.7:

Uma funcional quadrática possui a forma:

$$f(x)=\frac{1}{2}\cdot x^T.H.x-b^T.x+c \ , \quad x \in \mathbb{R}^N \qquad [3.5]$$

onde H é uma matriz simétrica N x N; $b \in \mathbb{R}^N$ e $c \in \mathbb{R}$.

O gradiente e a matriz Hessiana de [3.5] são facilmente determinados, resultando:

$$g(x)=H.x-b,$$

$$H(x)=H.$$

Portanto, a funcional quadrática tem a propriedade de que a matriz Hessiana é constante.

Analisando a expressão [3.5] mais detalhadamente, observamos que $\hat{x}$ será um ponto estacionário de f se o gradiente se igualar a zero.

Ou seja: $H.x-b=0$.

Se a matriz H é não singular, fato que assumiremos daqui em diante, então $\hat{x}$ é determinada unicamente por $\hat{x}=H^{-1}.b$.

Através de uma simples transformação, podemos reescrever [3.5] da seguinte forma:

$$f(x)=\frac{1}{2}\cdot(x-\hat{x})^T.H.(x-\hat{x})+\hat{c} \ , \quad x \in \mathbb{R}^N,$$

$$\text{onde} \quad \hat{c}=-\frac{1}{2}\cdot b^T.\hat{x}+c.$$

Seja o conjunto $\{\lambda_i, v_i\}_{i=1}^N$ as auto-soluções de H, isto é,

$$H.v_i=\lambda_i.v_i \ ; \ i=1,2,\ldots N.$$

Desde que H é simétrica, os auto-valores são reais e ordenados: $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_N$ e os auto-vetores satisfazem a orto-normalidade: $v_i^T.v_j=\delta_{i,j}$; $i,j=1,2,\ldots N$.

Definindo $\Lambda=\text{diag}(\lambda_1,\lambda_2,\ldots\lambda_N)$ e $V=[v_1,v_2,\ldots v_N]$ observamos que V é uma matriz ortogonal, ou seja, $V^{-1}=V^T$ e $H.V=V.\Lambda$.

Introduzindo uma nova variável $z = V^T.(x-\hat{x})$, temos:

$$\tilde{f}(z)\equiv f(V.z+\hat{x})=\frac{1}{2}\cdot z^T.V^T.H.V+\hat{c} \ = \ \frac{1}{2}\cdot z^T.\Lambda.z+\hat{c}$$

$$\text{ou, } \tilde{f}(z)=\frac{1}{2}\cdot\sum_{k=1}^{N}\lambda_i.z_i^2+\hat{c} \quad , \quad z\in\mathbb{R}^N.$$

Uma vez que $\tilde{f}(z)=f(x)$ sob a transformação $z=V^T.(x-\hat{x})$, podemos direcionar nossa atenção para $\tilde{f}(z)$.

Se H é definida positiva, então todos os seus auto-valores são positivos e a amplitude de $\tilde{f}$ é $[\hat{c},\omega]$,ou seja, z=0 é o minimizador global forte de $\tilde{f}$.

O nível de superfície L_k, para $k>\hat{c}$ é a elipsóide :

$$\sum_{i=1}^{N}\lambda_i.z_i^2=\hat{k}$$

onde $\hat{k}=2.(k-\hat{c})>0$, como mostrados na figura 3.2.

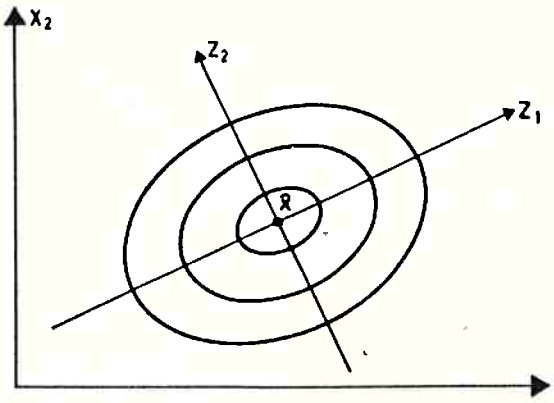

Figura 3.2 - Níveis de superfície (N=2) da funcional quadrática com a matriz Hessiana definida positiva.

A medida de distorção D_k assumirá o valor:

$$D_k\equiv K(H)=\lambda_N/\lambda_1$$

onde K(H) é conhecido como número de condição espectral de H.

3.2.2 METODO DOS GRADIENTES CONJUGADOS

Muitos métodos numéricos iterativos de determinação do minimizador da funcional f são da forma:

$$x^{k+1}=x^k+\tau_k.d^k \qquad [3.6]$$

onde: d^k é a direção de busca;

τ^k é escolhida de forma a minimizar $f(x)$ em um intervalo da linha que passa por x e tem a direção de d_k.

Temos, então, dois problemas distintos associados à expressão

[3.6]: a escolha de d_k e a inspeção de $f(x)$ na linha $x^{k+1}=x^k+\tau_k.d^k$ onde $-\infty<\tau<+\infty$.

DEFINIÇÃO 3.8:

Suponhamos que para uma dada funcional f e para os vetores x e d existe $\tau_0>0$ tal que:

$$f(x+\tau.d)<f(x) \quad , \quad 0<\tau\leq\tau_0.$$

Neste caso então d é dita direção descendente de f em x.

TEOREMA 3.5:

Sejam $f \in C^1(\mathbb{R}^N)$ e $g(x)$ o gradiente de f em x.

Se o vetor d satisfaz $g^T(x).d<0$, então d é uma direção descendente de f em x.

PROVA:

Da expressão [3.2] temos:

$$f(x+\tau.d)=f(x)+\tau.g^T(x).d+\Phi(\tau).$$

Como assumimos $g^T(x).d<0$, teremos: $\tau.g^T(x).d+\Phi(\tau)<0$ para pequenos valores de τ.

Assim $f(x+\tau.d)<f(x)$.

Se $g^T(x).d=0$, então não podemos determinar se d é uma direção descendente de f em x sem informações adicionais.

TEOREMA 3.6:

Seja $f \in C^2(\mathbb{R}^N)$ e suponhamos que $g^T(x).d=0$ e $d^T.H(x).d=0$ para algum x e d.

Então d é uma direção descendente de f em x.

PROVA:

Da expressão [3.1]

$$f(x+\tau d)=f(x)+\tau.g^T(x).d+\frac{1}{2}.\tau.d^T.H(x).d+\Phi(\tau^2).$$

Como $g^T(x).d=0$ e $d^T.H(x).d=0$, temos $f(x+\tau.d)<f(x)$.

TEOREMA. 3.7:

Seja $f \in C^1(\mathbb{R}^N)$.

Entre todas as direções de busca d em um ponto x, a direção que minimiza f mais rapidamente na vizinhança de x é $d=-g(x)$.

PROVA:

A prova deste teorema é análogo à apresentada na definição 3.5, onde temos a minimização de $g^T(x).y$ para todo g tal que $\|y\|=1$.

Desde que $|g^T(x).y| \leq \|g(x)\|$, o minimo é obtido para $y=-g(x)/\|g(x)\|$.

Consideraremos agora o problema da determinação de τ_k, dados x^k e d^k, na minimização de $f(x)$ sobre a linha $x=x^k+\tau.d^k$, $-\infty<\tau<+\infty$ para $\tau=\tau_k$.

Os procedimentos de determinação de τ_k são conhecidos como linha de busca.

Seja a funcional quadrática;

$$f(x)=\frac{1}{2}.x^T.H.x-b^T.x+c$$

aplicando uma transformação conveniente temos:

$$f(x+\tau.d)=\frac{1}{2}.\tau^2.d^T.H.d+\tau.d^T.g(x)+\tilde{c}$$

onde $\tilde{c}$ é independente de τ.

Se $d\neq 0$, então $d^T.H.d>0$ e $f(x+\tau.d)$ é uma parábola ascendente na variável τ.

Minimizando então $f(x+\tau.d)$ em relação a τ obtemos:

$$\tau=-d^T.g(x)/d^T.H.d$$

Desta forma, estabelecido d^k, τ_k assumirá, para $f(x)$ quadrática, a expressão:

$$\tau_k=-(d^k)^T.g(x^k)/(d^k)^T.H.d^k \qquad [3.7]$$

Uma vez determinados d^k e τ_k, temos então que o minimizador $\hat{x}=H^{-1}.b$ da funcional quadrática $f(x)=\frac{1}{2}.x^T.H.x-b^T.x+c$, $x \in \mathbb{R}^N$, será determinado através de operações iterativas do tipo:

$$x^{k+1}=x^k+\tau^k.d^k \quad , \quad k=1,2,\ldots,$$

onde: $\tau_k = -(d^k)^T . g^k / (d^k)^T . H . d^k$

e $g^k = g(x^k) = H . x^k - b.$

Observamos ainda que o gradiente em x^{k+1} é ortogonal à direção d^k.

Este fato é deduzido da seguinte formulação:

$$x^{k+1} = x^k + \tau_k . d^k$$

Multiplicando ambos os lados por H e subtraindo b:

$$g^{k+1} = g^k + \tau_k . H . d^k . \qquad [3.8]$$

Assim: $(d^k)^T . g^{k+1} = (d^k)^T . g^k + \tau_k . (d^k)^T . H . d^k$

Substituindo τ_k:

$$(d^k)^T . g^{k+1} = 0 \qquad [3.9]$$

Suponhamos agora que as direções de busca d sejam determinadas iterativamente conforme:

$$d^{k+1} = -g^{k+1} + \beta_k . d^k \qquad [3.10]$$

ou seja, de uma combinação entre a direção anterior d^k e o gradiente no ponto x^{k+1} com β_k ainda a ser determinado.

Substituindo (k+1) por k na expressão [3.10] e multiplicando ambos os lados por g^k observamos que:

$$(g^k)^T . d^k = -\|g^k\|^2 + \beta_{k-1} . (g^k)^T . d^{k-1}$$

De [3.9] temos que $(g^k)^T . d^{k-1} = 0$.

Assim $(g^k)^T . d^k = -\|g^k\|^2$.

Se $d^k = 0$ então $\|g^k\| = 0$, implicando em $g^k = H . x^k - b = 0$ e $x^k = \hat{x}$.

Portanto, teremos a direção de busca igual a zero somente quando o minimizador $\hat{x}$ for encontrado. Desta forma também eliminamos o problema de indefinição de τ_k quando $d^k = 0$

A estratégia a ser adotada para a determinação de β_k será a partir da minimização, em cada iteração, do erro $\|x - \hat{x}\|_H$ sobre um certo sub-conjunto de $\mathbb{R}^N$.

DEFINIÇÃO 3.9:

O produto interno de energia e a norma de energia correspondentes a uma matriz H definida positiva são; respectivamente:

$$(x,y)_H = x^T.H.y \quad ; \quad x,y \in \mathbb{R}^N,$$

$$e \qquad \|x\|_H = (x,x)_H^{1/2} = (x^T.H.x)^{1/2} \quad ; \quad x \in \mathbb{R}^N.$$

utilizando a definição 3.9 e a relação $(x-\hat{x}) = H^{-1}.g$ obtemos:

$$\|x-\hat{x}\|_H = \|g\|_{H^{-1}} \qquad [3.11]$$

Podemos então conduzir a minimização do erro $\|x-\hat{x}\|_H$ através da minimização de $\|g\|_{H^{-1}}$.

Aplicando recursivamente as expressões:

$$g^{k+1} = g^k + \tau_k.H.d^k$$

$$e \qquad d^{k+1} = -g^{k+1} + \beta_k.d^k$$

observamos que para qualquer escolha de $\beta_k = 0,1,2\ldots$, o gradiente g^k tem a forma:

$$g^k = g^0 + \sum_{L=1}^{k} \alpha_L^{(k)}.H^L.g^0$$

onde:

$$\alpha_k^{(k)} = (-1)^k.\prod_{i=1}^{k-1} \tau_i \neq 0.$$

DEFINIÇÃO 3.10:

Sejam: $S_k = \{H.g^0, H^2.g^0, H^k.g^0\}$ um sub-espaço de $\mathbb{R}^N$ com dimensão igual ao número de vetores linearmente independentes no conjunto $H.g^0$, $H^2.g^0$,...e $T_k = \{g \in \mathbb{R}^N;\ g=g^0+h\ ;\ h \in S_k\}$ um sub-conjunto de $\mathbb{R}^N$, onde $g^k \in T_k$.

TEOREMA 3.9:

O parâmetro β_k será dado por:

$$\beta_k = (g^{k+1})^T.H.d^k / (d^k)^T.H.d^T,$$

se impusermos a condição $\|g^k\|_{H^{-1}} = \min_{g \in T_k} \|g\|_{H^{-1}}$.

Esta relação implica ainda em:

$$(g^k)^T . g^l = 0 \quad ; \quad l \neq k; \qquad [3.12]$$

$$(d^k)^T . H . d^l = 0 \quad ; \quad l \neq k. \qquad [3.13]$$

PROVA:

A condição $\|g^k\|_{H^{-1}} = \min_{g \in T_k} \|g\|_{H^{-1}}$ é equivalente a:

$$\|g^0+h^k\|_{H^{-1}} = \min_{h \in S_k} \|g^0+h\|_{H^{-1}}, \text{ onde } h^k = g^k - g^0.$$

Ou seja, temos que encontrar em S_k um vetor h^k o mais próximo possível de $(-g^0)$, de forma que o erro entre eles, medido na norma de H^{-1}, seja o mínimo.

Demonstra-se que h^k existe; é único e tem a propriedade de tornar o erro ortogonal, em relação ao produto interno, a todo h em S_k, isto é :

$$(g^0+h^k)^T . H^{-1} . h = 0 \quad , \quad \forall h \in S_k.$$

Desta forma, a solução para $\min_{g \in T_k} \|g\|_{H^{-1}}$ é $g^k = g^0 + h^k$, que satisfaz $(g^k)^T . H^{-1} . h = 0$, $\forall h \in S_k$.

Para $g \in T_{k-1}$, o vetor $h = H.g$ pertence a S_k.

Logo g^k segue a propriedade:

$$(g^k)^T . g = 0 \quad , \quad \forall g \in T_{k-1} \; ;$$

demonstrando-se assim a primeira relação.

A segunda relação obtemos de [3.8]; [3.10]; [3.12]:

$$(d^k)^T . H . d^l = (H . d^k)^T . d^l = \tau_k^{-1} . (g^{k+1} - g^k)^T . d^l =$$

$$= \tau_k^{-1} . (g^{k+1} - g^k)^T . (-g^l + \beta_{l-1} . d^{l-1}) =$$

$$= (\beta_{l-1} / \tau_k) . (g^{k+1} - g^k)^T . d^{l-1}.$$

Por indução obtemos:

$$(d^k)^T . H . d^l = \tau_k^{-1} . \left(\prod_{i=0}^{l-1} \beta_i \right) . (g^{k+1} - g^k)^T . d^0 = 0$$

uma vez que $d^0 = -g^0$.

Finalmente a expressão de β_k pode ser demonstrada desenvolvendo:

$$0 = (d^k)^T . H . d^l = (-g^{k+1} + \beta_k \, d^k)^T . H . d^k.$$

isolando β_k: $\quad \beta_k = (g^{k+1})^T . H . d^k / (d^k)^T . H . d^k.$

Podemos simplificar as expressões de β_k e τ_k expandindo d^k em:

$$d^k=-g^k+\beta_{k-1}.d^{k-1}=-g^k+\beta_{k-1}.\left(-g^{k-1}+\beta_{k-2}.d^{k-2}\right) =$$

$$= -g^k-\beta_{k-1}.g^{k-1}+\beta_{k-1}.\beta_{k-2}.\left(-g^{k-2}.\beta_{k-3}.d^{k-3}\right) =....$$

e reescrevendo [3.8] como $H.d^k=\tau_k^{-1}.(g^{k+1}-g^k)$ obtemos:

$$\beta_k=\left(g^{k+1}\right)^T.g^{k+1}/\left(g^k\right)^T.g^k \qquad [3.14]$$

Utilizando a mesma expansão de d^k encontramos que:

$$\left(d^k\right)^T.g^k=\left(-g^k\right)^T.g^k$$

seguindo-se de [3.7]:

$$\tau_k=\left(g^k\right)^T.g^k/\left(d^k\right)^T.H.d^k \qquad [3.15]$$

3.2.2.1 ANÁLISE DE CONVERGÊNCIA

TEOREMA 3.10:

O Método dos Gradientes Conjugados possui a seguinte propriedade: $x^m=\hat{x}$ para algum $m\leq N$, onde N é a dimensão de H.

PROVA:

Suponhamos que o contrário é verdadeiro.

Então teríamos $g^k\neq 0$ para k=0,1.. N e aplicando [3.12] obteríamos N+1 vetores de dimensão N mutuamente ortogonais.

Uma vez que esta situação é impossível, o teorema está provado.

Este teorema garante um término finito ao Método dos Gradientes Conjugados em *m* iterações, sendo $m\leq N$.

No contexto da prática computacional, duas observações devem ser feitas a esta propriedade:

A primeira se refere aos erros de truncamento que ocorrem no processo e podem permitir iterações superiores a *m*.

A segunda é que *m* pode ser de valor tão elevado que o tempo requerido para as *m* iterações é inaceitável.

Uma vez que as matrizes resultantes do Método dos Elementos Finitos são geralmente de grandes dimensões, esta segunda

observação afeta prioritariamente o problema.

Analisaremos então a relação entre a precisão e o número de iterações associados ao sistema.

Como visto anteriormente, o vetor gradiente g^k produzido pelo Método dos Gradientes Conjugados tem a propriedade:

$$\|g^k\|_{H^{-1}} = \min_{g \in T_k} \|g\|_{H^{-1}}$$

onde o elemento típico de T_k é da forma:

$$g = g^0 + \sum_{l=1}^{k} \alpha_l . H^l . g^0 .$$

Seja Π_k^1 o conjunto de polinômios P_k de grau k tal que $P_k(0)=1$.

A variável independente de P_k, será por conveniência, um escalar ou uma matriz NxN.

Consideremos o conjunto:

$$\tilde{T}_k = \{ g \in \mathbb{R}^N ;\ g = P_k(H) . g^0 ;\ P_k \in \Pi_k^1 \}$$

$\tilde{T}_k$ um sub-conjunto de T_k contendo g^k.

Assim:

$$\|g^k\|_{H^{-1}} = \min_{g \in \tilde{T}_k} \|g\|_{H^{-1}} = \min_{P_k \in \Pi_k^1} \|P_k(H) . g^0\|_{H^{-1}}$$

$$= \min_{P_k \in \Pi_k^1} \left[(g^0)^T . H^{-1} . P_k(H)^2 . g^0 \right]^{1/2} \qquad [3.16]$$

Utilizaremos este resultado para provar o teorema fundamental da convergência no Método dos Gradientes Conjugados.

TEOREMA 3.11:

Suponhamos que para um conjunto S contendo todos os auto-valores de H, para um M>0 e para $\tilde{P}_k(\lambda) \in \Pi_k^1$ o seguinte é verdadeiro.

$$\max_{\lambda \in s} |\tilde{P}_k(\lambda)| \leq M$$

Então: $$\|x^k - \hat{x}\|_H \leq M . \|x^0 - \hat{x}\|_H \qquad [3.17]$$

PROVA:

Sejam as auto-soluções de H $\{ \lambda_i, v_i \}_{i=1}^{N}$ ordenados em $0<\lambda_1 \leq \lambda_2 \leq \ldots \leq$, e com os autos-vetores ortonormais, isto é: $v_i^T . v_j = \delta_{ij}$.

O gradiente inicial tem a expansão:

$$g^0 = \sum_{i=1}^{N} a_i . v_i \quad ; \quad a_i = v_i^T . g^0 .$$

Um cálculo direto mostra que:

$$(g^0)^T . H^{-1} . P_k(H)^2 . g^0 = \sum_{i=1}^{N} a_i^2 . \lambda_i^{-1} . P_k(\lambda_i)^2 .$$

Daí, por [3.16] temos:

$$\| g^k \|^2_{H^{-1}} = \min_{P_k \in \Pi_k^1} \left[\sum_{i=1}^{N} a_i^2 . \lambda_i^{-1} . P_k(\lambda_i)^2 \right] .$$

Uma vez que assumimos no início que:

$$|\tilde{P}_k(\lambda_i)| \leq M,$$

temos:

$$\| g^k \|^2_{H^{-1}} \leq M^2 . \sum_{i=1}^{N} a_i^2 . \lambda_i^{-1} = M^2 . \| g^0 \|^2_{H^{-1}}$$

Da identidade $\| x^k - \hat{x} \|_H = \| g^k \|_{H^{-1}}$ obtemos [3.17].

Podemos agora relacionar um conjunto S contendo todos os auto-valores de H e procurar um polinômio $\tilde{P}_k \in \Pi_k^1$ tal que $M \equiv \max_{\lambda \in S} |\tilde{P}_k(\lambda)|$ seja pequeno.

Este valor de M poderá, então, ser usado em [3.17].

Supomos, até aqui, que todos os auto-valores são reais e positivos.

Assim temos $Q=[\lambda_1, \lambda_N]$ e devemos procurar $\tilde{P}_k \in \Pi_k^1$ com a seguinte propriedade:

$$\max_{\lambda_1 \leq \lambda \leq \lambda_N} |\tilde{P}_k(\lambda)| = \min_{P_k \in \Pi_k^1} \left[\max_{\lambda_1 \leq \lambda \leq \lambda_N} |P_k(\lambda)| \right]$$

A solução deste problema é conhecida como sendo:

$$\tilde{P}_k(\lambda) = \frac{T_k . [(\lambda_N + \lambda_1 - 2\lambda) / (\lambda_N - \lambda_1)]}{T_k . [(\lambda_N + \lambda_1) / (\lambda_N - \lambda_1)]}$$

Onde T_k é o polinômio de Chebyshev de grau k [ver teorema A1 do Apêndice]

A propriedade: $$\max_{\lambda_1 \le \lambda \le \lambda_2} |\tilde{P}_k(\lambda)| = T_k \cdot [(\lambda_N+\lambda_1)/(\lambda_N-\lambda_1)]^{-1} \qquad [3.18]$$

juntamente com o teorema 3.11 e o item A5 do Apêndice levam-nos ao seguinte teorema:

TEOREMA 3.12:

Para o método dos gradientes conjugados temos a seguinte estimativa de erro:

$$\|x^k-\hat{x}\|_H \le T_k \cdot [(\lambda_N+\lambda_1)/(\lambda_N-\lambda_1)]^{-1} \cdot \|x^0-\hat{x}\|_H,$$

mais ainda, se $p(\xi)$ é definido para algum $\xi>0$ como o menor inteiro k tal que:

$$\|x^k-\hat{x}\|_H \le \xi \cdot \|x^0-\hat{x}\|_H \quad ; \quad \forall x^0 \in \mathbb{R}^N,$$

então $$p(\xi) \le \frac{1}{2} \cdot \sqrt{K(H)} \cdot \ln(2/\xi)+1 \qquad [3.19]$$

Devido à generalidade e à simplicidade, a expressão [3.19] é de grande utilidade possibilitando o uso do número de condição espectral $K(H)$ para estimar a taxa de convergência.

Segundo Axelsson [2], faz-se necessário observar que dependendo da distribuição dos auto-valores, a expressão [3.19] pode ser pessimista.

3.2.2.2 PRÉ - CONDICIONAMENTO

Seja C uma matriz definida positiva fatorada na forma $C=E.E^T$ e seja a funcional

$$f(x) = \frac{1}{2} \cdot x^T.H.x - b^T.x + c \quad ; \quad x \in \mathbb{R}^N$$

onde H é definida positiva.

Definiremos uma segunda funcional quadrática $\tilde{f}(y)$ pela transformação $y=E^T.x$, ou seja, $\tilde{f}(y) = f(E^{-T}.y) = \frac{1}{2} \cdot y^T.\tilde{H}.y - \tilde{b}^T.y + \tilde{c}$,

onde $\tilde{H}=E^{-1}.H.E^{-T}$; $\tilde{b}=E^{-1}.b$; $\tilde{c}=c$ e $\tilde{H}$ é simétrica.

Mais ainda:

Se $y^T.\tilde{H}.y=x^T.H.x$ e $x^T.H.x>0$, $\forall x\neq 0$, pois H é definida positiva, então $y^T.\tilde{H}.y>0$, $\forall y\neq 0$, implicando em $\tilde{H}$ definida positiva.

Aplicando a transformação de similaridade em $\tilde{H}$ [Ver Albrecht] obtemos:

$$E^{-T}.\tilde{H}.E^{T}=E^{-T}.E^{-1}.H=C^{-1}.H$$

Concluimos daí que $\tilde{H}$ e $C^{-1}.H$ possuem os mesmos auto-valores e o número de condição espectral fica definido por $K(\tilde{H})=\tilde{\lambda}_N/\tilde{\lambda}_1$, onde $0<\tilde{\lambda}_1\leq\tilde{\lambda}_2\leq\ldots\leq\tilde{\lambda}_N$.

A matriz C é chamada matriz pré-condicionadora e $\tilde{H}$ matriz pré-condicionada.

Uma vez que $y^k=E^T.x^k$; $k=0,1,\ldots$ e $\hat{y}=E^T.x$ temos $\|y^k-\hat{y}\|_{\tilde{H}}=\|x^k-\hat{x}\|_H$ e para $p(\xi)$, o menor número de iterações para $\|x^k-\hat{x}\|_H\leq\xi.\|x^0-\hat{x}\|_H$:

$$p(\xi)\leq\frac{1}{2}.\sqrt{K(\tilde{H})}.\ln(2/\xi)+1$$

Portanto, se C tem a propriedade de tornar $K(\tilde{H})<K(H)$, então o Método dos Gradientes Conjugados Pré-Condicionado possui uma taxa de convergência mais rápida do que o método não pré-condicionado.

Em geral, uma boa matriz pré-condicionadora possui as seguintes propriedades:

1. $K(\tilde{H})$ é significativamente menor que $K(H)$;
2. os elementos de C são rapidamente determinados e não requerem armazenamento excessivo comparado com H;
3. o sistema $C.h^k=g^k$ pode ser resolvido mais eficientemente que $H.x=b$.

É importante notar que a matriz C possui várias fatorizações da forma $C=E.E^T$ e que utilizaremos uma fatorização mais geral da forma $C=F.G^{-1}.F^T$, onde F é uma matriz triangular inferior e G uma matriz diagonal.

Assim na k-ésima iteração do sistema $C.h^k=g^k$ no método pré-condicionado teremos:

$$F.\tilde{h}^k=g^k \text{ , obtendo-se } \tilde{h}^k$$

$$\text{e } F^T.h^k=G.\tilde{h}^k \text{ , obtendo-se } h^k.$$

Uma vez que as expressões são formadas por matrizes triangulares, as soluções das mesmas são de fácil realização; satisfazendo assim a propriedade 3 descrita anteriormente.

Com relação à propriedade 2 podemos afirmar que:

a. As entradas de F e G também são entradas de H e portanto imediatamente obtíveis;

b. Os armazenamentos de F e G são similares ao armazenamento requerido por H.

3.2.2.2.1 MÉTODOS ITERATIVOS ESTACIONÁRIOS

A fonte das matrizes pré-condicionadoras é a classe de Métodos Iterativos Estacionários de solução para o sistema $H.\hat{x}=b$.

Seja $H=M+R$ [3.20]

uma decomposição de H tal que M é não singular.

Seja x^0 um vetor arbitrário e consideremos a seqüência $x^0, x^1, x^2 \ldots$ gerada pela resolução do sistema

$$M.x^{k+1}=-R.x^k+b \text{ , } k=0,1,2\ldots \quad [3.21]$$

Para determinarmos quando esta seqüência converge para $\hat{x}$ combinamos a relação $M.\hat{x}=-R.x^k+b$ com [3.21] obtendo:

$$M.(x^{k+1}-\hat{x})=-R.(x^k-\hat{x}),$$

ou $$(x^{k+1}-\hat{x})=B^k.(x^0-\hat{x}), \text{ com } B\equiv M^{-1}.R \quad [3.22]$$

Sejam $\{\zeta_i, w_i\}_{i=1}^N$ as auto-soluções de B onde os auto-vetores são linearmente independentes.

Então para x^0 temos a expressão $x^0-\hat{x}$ na forma:

$$x^0-\hat{x}=c_1.w_1+c_2.w_2+\ldots+c_N.w_N,$$

que substituindo em [3.22] fornece:

$$x^k-\hat{x}=c_1.\zeta_1^k.w_1+c_2.\zeta_2^k.w_2+\ldots+c_N.\zeta_N^k.w_N.$$

Esta expressão mostra que:

$$\lim_{k\to\infty} x^k=\hat{x},\ \forall x^0\in\mathbb{R}^N \iff \rho(B)<1,$$

onde $\rho(B)=\max_{1\le i\le N}|\zeta_i|$ é o raio espectral de B.

A desigualdade $\rho(B)<1$ é, portanto, a condição necessária e suficiente para a convergência da seqüência x^0, x^1, x^2... para $\hat{x}$.

A expressão [3.21] é conhecida como Método Iterativo Estacionário, correspondente à decomposição da matriz H conforme [3.20], para resolução do sistema $H.\hat{x}=b$.

Suponhamos agora que H e M são definidas positiva e tomemos M como a matriz pré-condicionadora C no Método dos Gradientes Conjugados.

Analisaremos, então, o número da condição espectral resultante desta substituição, $C^{-1}.H=M^{-1}.H=I-B$.

Observamos que os auto-valores de $C^{-1}.H$ são: $\tilde{\lambda}_i=1-\zeta_i$, $i=1,2..N$.

Se os auto-valores de B são ordenados por :

$$-1<\zeta_N\le\ldots\le\zeta_2\le\zeta_1<1\ ,$$

$$\text{então,}\quad 0<\tilde{\lambda}_1\le\tilde{\lambda}_2\le\ldots\le\tilde{\lambda}_N.$$

e o número de condição espectral K(H) resulta:

$$K(\tilde{H})=\tilde{\lambda}_N/\tilde{\lambda}_1=(1-\zeta_N)/(1-\zeta_1)\le[1+\rho(B)]/[1-\rho(B)] \qquad [3.23]$$

Utilizando esta inequação podemos comparar as taxas de convergência do Método Iterativo Estacionário e do Método dos Gradientes Conjugados Pré-Condicionados.

Sejam $\hat{k}_1$ e $\hat{k}_2$ os números de iterações requeridos pelos dois métodos, respectivamente, para obtermos $\|x^k-\hat{x}\|_H<\xi.\gamma$, onde ξ é positivo e $x^0-\hat{x}=\sum_{i=1}^{N}c_i.w_i$; $\gamma=\sum_{i=1}^{N}|c_i|.\|w_i\|_H$.

Assim para $\hat{k}_1$ temos:

$$\hat{k}_1=[\ln(1/\xi)/\ln(1/\rho(B))]$$

e para $\hat{k}_2$:

$$\hat{k}_2=\frac{1}{2}\cdot\sqrt{K(\tilde{H})}\,.\ln(2/\xi).$$

Dividindo $\hat{k}_2$ por $\hat{k}_1$ e utilizando [3.23]:

$$\frac{\hat{k}_2}{\hat{k}_1}=\frac{1}{2}\cdot f[\rho(B)].[\ln(2/\xi)/\ln(1/\xi)] \qquad [3.24]$$

onde $f[\rho(B)]$ pode ser posta na forma:

$$f(\zeta)=\ln(1/\zeta).\sqrt{(1+\zeta)/(1-\zeta)}\ ,\ 0<\zeta<1,$$

ou
$$f(\zeta)=\sqrt{2.(1-\zeta)}+\Phi[(1-\zeta)^{3/2}],\ \zeta\to 1 \qquad [3.25]$$

Uma vez que $\lim_{\zeta\to 1} f(\zeta)=0$, a expressão [3.24] indica que o Método dos Gradientes Conjugados Pré-Condicionado é mais rápido que o Método Iterativo Estacionário quando $\rho(B)$ é próximo da unidade.

3.2.2.2.2 PRÉ-CONDICIONAMENTO SSOR

O método da sobre-relaxação simétrica sucessiva (SSOR) para resolução do sistema $H.\hat{x}=b$ é o procedimento iterativo de dois passos definido pelo seguinte algorítmo [2]:

$$x_i^{(k+1/2)}=(1-\omega).x_i^{(k)}-\frac{\omega}{h_{ii}}\cdot\left(\sum_{j=1}^{i-1}h_{ij}\,x_j^{(k+1/2)}+\sum_{j=i+1}^{N}h_{ij}\,x_j^{(k)}-b_i\right) \qquad [3.26a]$$

$i=1,2,\ldots.N.$

$$x_i^{(k+1)}=(1-\omega).x_i^{(k+1/2)}-\frac{\omega}{h_{ii}}\left(\sum_{j=1}^{i-1}h_{ij}\,x_j^{(k+1/2)}+\sum_{j=i+1}^{N}h_{ij}\,x_j^{(k+1)}-b_i\right) \qquad [3.26b]$$

$i=N,\ N-1,\ldots.1.$

Seja H definida positiva, decomposta na forma:

$$H=D+L+L^T$$

onde: $d_{ij}=\begin{cases}h_{ij}\text{ para } i=j\\ 0\text{ para } i\neq j\end{cases}$

$$l_{ij}=\begin{cases}h_{ij}\text{ para } i>j\\ 0\text{ para } i\leq j\end{cases}$$

D = Matriz Diagonal;

L = Matriz Triangular Inferior.

Desta forma, [3.26a] e [3.26b] podem ser reformuladas como seguem:

$$x^{(k+1/2)}=(1-\omega).x^{(k)}-\omega.D^{-1}.\left(L.x^{(k+1/2)}+L^T.x^k-b\right)$$

$$x^{(k+1)}=(1-\omega).x^{(k+1/2)}-\omega.D^{-1}.\left(L.x^{(k+1/2)}+L^T.x^{(k+1)}-b\right)$$

ou, após a eliminação de $x^{(k+1/2)}$ temos: $x^{(k+1)}=B.x^k+M^{-1}.b$, [3.27]

onde:

$$B=\left(\frac{1}{\omega}.D+L^T\right)^{-1}.\left[\left(\frac{1}{\omega}-1\right).D-L\right].\left(\frac{1}{\omega}D+L\right)^{-1}\left[\left(\frac{1}{\omega}-1\right).D-L^T\right]$$

$$M=\frac{1}{2-\omega}.\left(\frac{1}{\omega}.D+L\right).\left(\frac{1}{\omega}.D\right)^{-1}.\left(\frac{1}{\omega}.D+L\right)^T.$$

Multiplicando a equação [3.27] por M e fazendo R=-M.B obtemos a forma [3.21].

Consideraremos então a matriz pré-condicionadora como sendo:

$$C=\frac{1}{2-\omega}.\left(\frac{1}{\omega}.D+L\right).\left(\frac{1}{\omega}.D\right)^{-1}.\left(\frac{1}{\omega}.D+L\right)^T \qquad [3.28]$$

onde C é definida positiva se e somente se $0<\omega<2$ e portanto restingiremos seu valor a este intervalo.

TEOREMA 3.13:

O número de condição espectral $K(\tilde{H})$ associado com o Método dos Gradientes Conjugados Pré-Condicionado utilizando C da forma [3.28] satisfaz $K(\tilde{H})\leq F(\omega)$, onde:

$$F(\omega)=\frac{1+[(2-\omega)^2/4\omega].\mu+\omega.\delta}{2-\omega},\quad 0<\omega<2; \qquad [3.29]$$

$$\mu=\max_{x\neq o} x^T.D.x/x^T.D.x;$$

$$\delta=\max_{x\neq o}\frac{x^T.(L.D^{-1}.L^T-(1/4).D).x}{x^T.H.x}\geq-\frac{1}{4}.$$

Mais ainda, se

$$\left\|D^{-1/2}.L.D^{-1/2}\right\|_\infty\leq\frac{1}{2},\quad\left\|D^{-1/2}.L^T.D^{-1/2}\right\|_\infty\leq\frac{1}{2} \qquad [3.30]$$

então: $-\frac{1}{4}\leq\delta\leq 0$.

PROVA:

$K(H)=\tilde{\lambda}_N/\tilde{\lambda}_1$, onde $\tilde{\lambda}_N$ e $\tilde{\lambda}_1$ são, respectivamente, o maior e o menor auto-valor de $C^{-1}.H$.

Estes auto-valores possuem a seguinte propriedade:

$$\tilde{\lambda}_1=\min_{x\neq 0} R(x), \quad \tilde{\lambda}_N=\max_{x\neq 0} R(x) \qquad [3.31]$$

onde $R(x)$ é o quociente de Rayleigh definido por:

$$R(x)=x^T.H.x/x^T.C.x, \quad x\neq 0.$$

Mostraremos primeiro que $\tilde{\lambda}_N\leq 1$.

Sejam $A=\left(\frac{2}{\omega}-1\right).D$; $V=\left(1-\frac{1}{\omega}\right).D+L$, com A definida positiva para $0<\omega<2$.

A matriz C em [3.28] pode ser escrita na forma:

$$C=(A+V).A^{-1}.(A+V)^T=A+V+V^T+V.A^{-1}.V^T=H+V.A^{-1}.V^T$$

e para qualquer x e $y\equiv V^T.x$ temos:

$$x^T.C.x=x^T.H.x+y^T.A^{-1}.y.$$

Como C, H e A^{-1} são definidas positiva segue de [3.31] que $\tilde{\lambda}_N\leq 1$.

A matriz C pode também ser expressa como:

$$C=(2-\omega)^{-1}.\left[H+\frac{1}{4\omega}.(2-\omega)^2.D+\omega.\left(L.D^{-1}.L^T-\frac{1}{4}.D\right)\right].$$

Efetuando $x^T.C.x$ e utilizando as desigualdades:

$$x^T.D.x\leq\mu.x^T.H.x,$$

$$\text{e } x^T.\left(L.D^{-1}.L^T-\frac{1}{4}.D\right).x\leq\delta.x^T.H.x\ ,$$

obtemos que $R(x)\geq\frac{1}{F(\omega)}$, $\forall x\neq 0$ onde $F(\omega)$ é dado por [3.29].

Assim, devido a [3.31] $\tilde{\lambda}_1\geq\frac{1}{F(\omega)}$ e este resultado, juntamente com $\tilde{\lambda}_N\leq 1$ implica em $K(\tilde{H})\leq F(\omega)$.

Como D e H são definidas positiva, temos que $\mu>0$.

Na análise de δ fazemos $D=H-L-L^T$ e, uma vez que L^T é singular, achamos x de tal forma que $L^T.x=0$.

Assim: $x^T.\left(L.D^{-1}.L^T-\frac{1}{4}.D\right).x=-\frac{1}{4}.x^T.D.x$, implicando em $\delta\geq\frac{1}{4}$.

Assumindo as relações [3.30] como satisfeitas, fazemos:

$$x^T.L.D^{-1}.L^T.x=y^T.\tilde{L}.\tilde{L}^T.y;\quad x^T.D.x=y^T.y,$$

onde $y=D^{1/2}.x$ e $\tilde{L}=D^{-1/2}.L.D$ e utilizando as desigualdades

encontramos:

$$y^T.\tilde{L}.\tilde{L}^T.y/y^T.y \leq \rho.(\tilde{L}.\tilde{L}^T) \leq \|\tilde{L}.\tilde{L}^T\|_\infty \leq \|\tilde{L}\|_\infty.\|\tilde{L}^T\|_\infty \leq \frac{1}{4}.$$

Daí: $x^T.L.D^{-1}.L^T.x-\frac{1}{4}.x^T.D.x = y^T.\tilde{L}.\tilde{L}^T.y-\frac{1}{4}.y^T.y \leq 0$, portanto $\delta \leq 0$.

3.2.2.2.2.1 TAXA DE CONVERGÊNCIA COM UM PRÉ-CONDICIONAMENTO SSOR ÓTIMO

Analisando [3.29], observamos que:

$$\min_{0<\omega<2} F(\omega)=F(\omega^*)=\sqrt{\left(\frac{1}{2}+\delta\right).\mu}+\frac{1}{2}, \qquad [3.32a]$$

onde

$$\omega^*=2/\left[1+\left(2/\sqrt{\mu}\right).\sqrt{\frac{1}{2}+\delta}\right] \qquad [3.32b]$$

Mostraremos, então, que $\mu \leq K(H)$.

Para este propósito, introduziremos o quociente de Rayleigh para H: $\tilde{R}(x)=x^T.H.x/x^T.x$, $x\neq 0$, que possui as propriedades:

$$\lambda_1=\min_{x\neq 0}\tilde{R}(x), \quad \lambda_N=\max_{x\neq 0}\tilde{R}(x).$$

Além disso, pelo fato de $\tilde{R}(e_i)=h_{ii}$, $e_i=[0.....\underset{1.....i}{1}0....0]^T$ temos:

$$\max_{1\leq i\leq N} h_{ii} \leq \lambda_N.$$

Agora, $\mu=\max \dfrac{x^T.D.x}{x^T.H.x} \leq \dfrac{\max_{x\neq 0}(x^T.D.x/x^T.x)}{\min_{x\neq 0}(x^T.H.x/x^T.x)}$, onde $d=\max_{1\leq i\leq N} d_{ii}$.

Mas $d_{ii}=h_{ii}$, seguindo daí que $d<\lambda_N$ e $\mu \leq K(H)$.

Assim alcançamos a expressão:

$$\min_{0<\omega<2} K(\tilde{H})(\omega) \leq \sqrt{\left(\frac{1}{2}+\delta\right).K(H)}+\frac{1}{2}, \qquad [3.33]$$

que é o melhor resultado para o pré-condicionamento SSOR.

Observamos que, sempre que o lado direito da expressão for menor que $K(H)$, o método dos Gradientes Conjugados com Pré-Condicionamento SSOR com um ótimo valor de ω possui uma alta taxa de convergência em relação ao Método dos Gradientes Conjugados simples.

Mais ainda, se $K(H)>>1$ e se H satisfaz [3.30], então o

pré-condicionamento SSOR permite uma taxa de convergência muito maior.

3.2.2.2.3 PRÉ-CONDICIONAMENTO POR FATORIZAÇÃO INCOMPLETA

A relação H=M+R permite-nos eliminar R no Método Iterativo Estacionário [3.21].

Utilizando a identidade $-R.x^k+b=M.x^k+b-H.x^k$, observamos que [3.21] é equivalente a resolver:

$$\left.\begin{aligned} M.\delta^k &= b-H.x^k \\ x^{k+1} &= x^k+\delta^k \end{aligned}\right\} \qquad [3.34]$$

$$\text{para } k=0,1,2\ldots$$

Algumas vezes as equações [3.34] são mais vantajosas para a formulação computacional. É o caso, por exemplo, quando $M=U^T.U$, onde U é uma aproximação dos fatores de Cholesky de H.

O procedimento definido por [3.34] é conhecido como refinamento iterativo. O erro $(H-U^T.U)$ será devido exclusivamente a erros de arredondamento ou à fatorização incompleta de H. Este último caso será o objeto de discussão, visto que a decomposição de H fornece a matriz pré-condicionadora para o Método dos Gradientes Conjugados, assim como para o Método Iterativo Estacionário.

Nosso interesse estará voltado para o caso onde H é de grandes dimensões e esparsa e que devido a isto, o armazenamento dos elementos não-nulos da fatorização U de Cholesky de H ocupa maior área de memória que os elementos não-nulos de H.

Para qualquer matriz H com estrutura simétrica $(h_{ij}\neq 0 \rightarrow h_{ji}\neq 0)$ e sem zeros na diagonal principal, seja:

$$m(i)=\min\{j;\ (1\leq j\leq i) \wedge (h_{ij}\neq 0)\},\ i=1,2..N,$$

ou seja, $h_{i,m(i)}$ éo primeiro elemento não nulo na linha i.

Então o envelope de H é o conjunto de pares índices:

$$S=\{(i,j) \cup (j,i);\ m(i)\leq j\leq i,\ 1\leq i\leq N\},$$

conforme ilustrado na figura 3.3.

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

Figura 3.3. Envelope S de uma matriz simétrica.

Em matrizes provenientes de problemas de elementos finitos é comum este envelope ser largo e esparsamente completo por elementos não-nulos.

Durante a fatorização muitos elementos não-nulos são introduzidos na matriz U nos lugares que, originariamente em H, eram zeros e este fato torna complexa a implementação computacional desta fatorização no tocante ao armazenamento.

Podemos reduzir esta complexidade utilizando uma fatorização incompleta de H.

A idéia básica desta fatorização incompleta é escolher um sub-conjunto J de S, incluindo os pares (1,1),(2,2),...(N,N) da diagonal principal e modificar o procedimento da fatorização, de forma que, os elementos não-nulos de U estejam restritos a J.

Existem várias maneiras de executar esta modificação. O método adotado neste trabalho consiste em assumir J largo o suficiente de forma a incluir todos os pares índices dos elementos não-nulos de H, isto é, $S_H \subseteq J$, onde $S_H=\{(i,j);\ h_{ij}\neq 0\}$ implicando em $(h_{ij}=0) \notin J$.

Assim teremos para a decomposição incompleta de Cholesky (IC) de H:

$$u_{ii}=\sqrt{h_{ii}} \quad , \quad u_{ij}=\frac{h_{ij}}{u_{ii}} \qquad [3.35a]$$

$$u_{ii}=\sqrt{\left(h_{ii}-\sum_{j=1}^{i-1} u_{ji}^2\right)} \qquad [3.35b]$$

$$u_{ik} = \begin{cases} \frac{1}{u_{ii}} \cdot \left(h_{ik} - \sum_{j=1}^{i-1} u_{ji} \cdot u_{jk} \right), & [(i,k) \in J] \\ 0, & [(i,k) \notin J] \end{cases} \qquad [3.35c]$$

3.2.3 ASPECTOS COMPUTACIONAIS

O Diagrama Estruturado para o Método dos Gradientes Conjugados com Pré-Condicionamento é mostrado na figura 3.4.

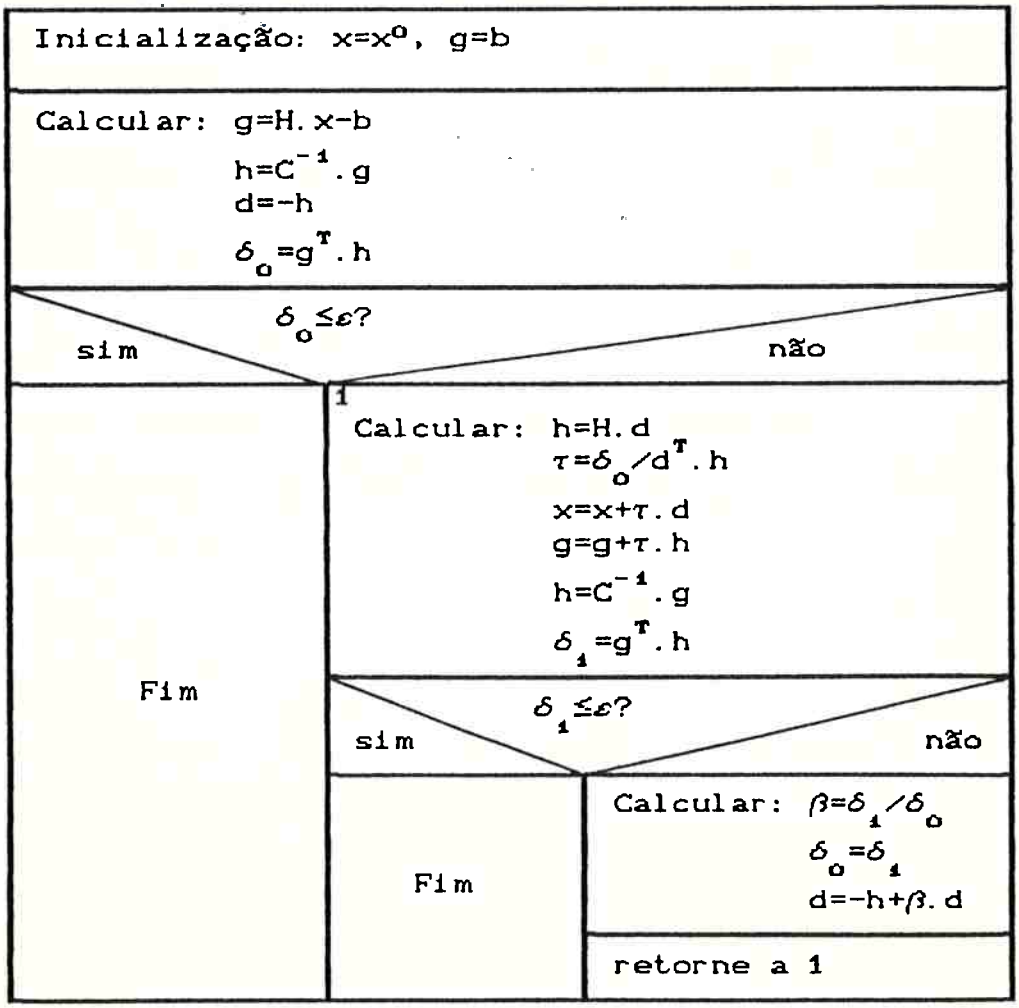

Figura 3.4: Diagrama Estruturado do Método dos Gradientes Conjugados com Pré-Condicionamento.

Os vetores x^0 e b são dados de entrada do programa

A matriz pré-condicionadora C segue a expressão [3.28] para o método SSOR e a forma $C=U^T.U$ para o método IC onde U é a decomposição de H segundo [3.35].

A expressão $h=C^{-1}.g$ deve ser interpretada como a solução de C.h=g. Observamos assim que, a inversão de C não é necessária, uma vez que C em ambos os métodos, é uma composição de matrizes triangulares permitindo o cálculo do vetor h por substituições simples.

O método SSOR não exige o armazenamento da matriz C visto que esta é obtida diretamente de H; diferentemente do método IC, onde C é obtida através de uma transformação de H e portanto exige um armazenamento individual.

O vetor h é utilizado para armazenar tanto H.d quanto $C^{-1}.g$.

3.3 MÉTODOS DIRETOS

Os métodos diretos são assim conhecidos por resolverem o sistema H.x=b com um número finito de operações.

O método mais conhecido é o Método de Eliminação de Gauss, que consiste em fatorar a matriz H em duas matrizes triangulares , ou seja: H=L.U,

onde: L = matriz triangular inferior;

U = matriz triangular superior.

O problema H.x=b fica, dessa forma, reduzido a dois sistemas simples: L.y=b e U.x=y, cujas soluções são facilmente obtíveis, uma vez que L e U são matrizes triangulares.

3.3.1 METODO DA DECOMPOSIÇÃO DE CHOLESKY

Supondo que H seja uma matriz real e simétrica obtemos uma simplificação do Método de Eliminação de Gauss, pois, neste caso, existe uma fatorização da forma:

$$H=U^T.U,$$

onde U é uma matriz triangular superior de elementos reais ou imaginários puros e U^T sua matriz transposta.

TEOREMA 3.14:

Seja H simétrica com elementos reais.

Então H é definida positiva se, e somente se, os elementos de U são reais.

PROVA:

1) Se U é real então: $x^T.H.x=x^T.U^T.U.x=y^T.y>0$, pois $y=U.x$ é real para todo x real, $x\neq 0$.

2) Se H é definida positiva então para todo x real, $x\neq 0$: $x^T.H.x=x^T.U^T.U.x>0$, donde U.x é real, conseqüentemente U é real.

Como H no nosso caso é definida positiva, garantimos a decomposição com elementos reais em U.

3.3.2 ASPECTOS COMPUTACIONAIS

O Diagrama Estruturado utilizado para o Método da Decomposição de Cholesky é mostrado na figura 3.5.

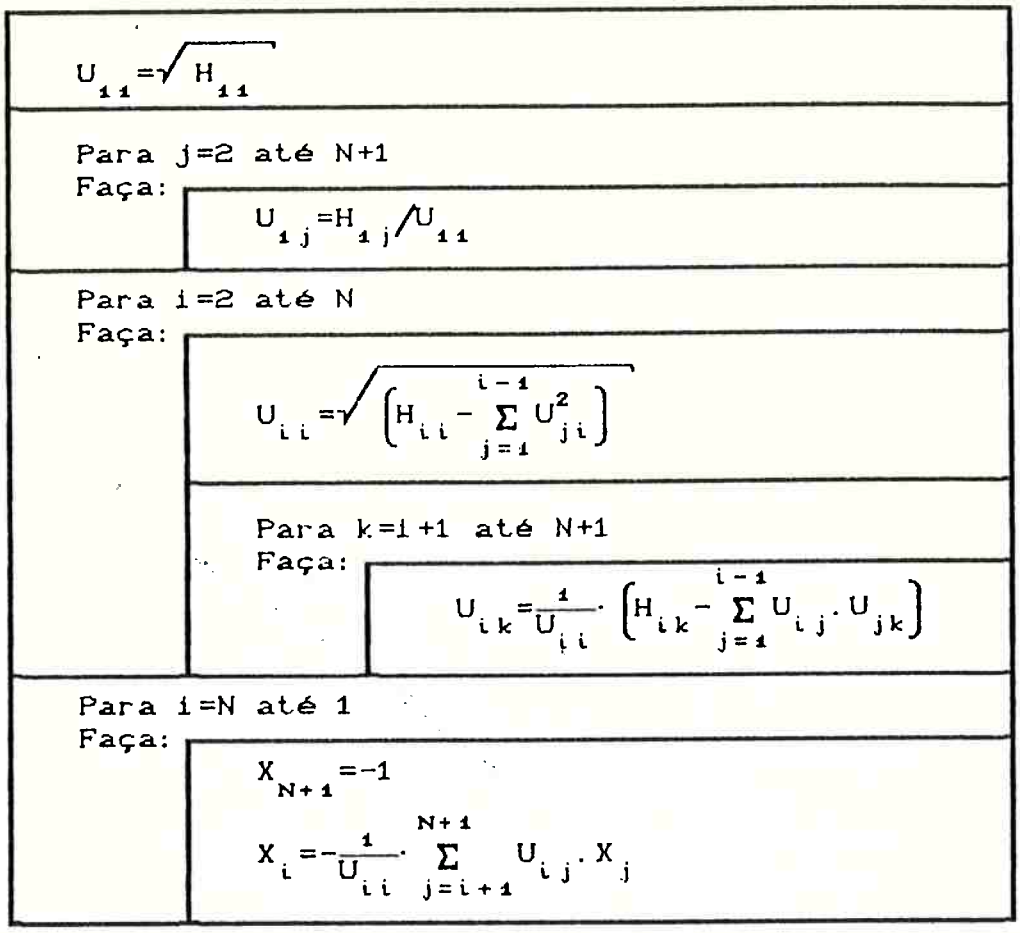

Figura 3.5: Diagrama Estruturado do Método da Decomposição de Cholesky.

Os elementos H com índice N+1 para a coluna correspondem aos elementos do vetor b, ou seja, $H_{1,N+1}=b_1$, $H_{2,N+1}=b_2$,...

O surgimento de elementos não-nulos em U em posições, onde originalmente em H tinhamos zeros, torna complexa a utilização do armazenamento compacto, deteriorando a velocidade de processamento do programa e aumentando a memória alocada para o armazenamento.

CAPITULO 4: APLICAÇÃO E RESULTADOS

4.1 INTRODUÇÃO

A análise do desempenho do método ICCG foi realizada sobre um motor síncrono trifásico de relutância ,cuja seção de interesse é mostrada na figura 4.1.a.

O sistema computacional utilizado foi o IBM 4381 do Centro de Computação Eletrônica da USP e a linguagem utilizada foi o FORTRAN.

As densidades de corrente impressas nos condutores do estator foram distribuidas nas ranhuras de maneira uniforme, levando-se em consideração o devido fator de enchimento, e com intensidades correspondentes ao instante em que uma das fases tem valor máximo e as outras duas metade deste valor e negativo.

A distribuição de correntes nas ranhuras foi realizada de forma a se ter o fluxo magnético gerado alinhado com o eixo direto do motor síncrono de relutância, para que o efeito da saturação do material ferromagnético fosse sentido em sua totalidade.

A razão da escolha ter recaído sobre um motor se deve ao fato de evidências já comprovadas de um mau comportamento do ICCG na solução de problemas magnetostáticos com a presença de pequenos entreferros [17], fato este não observado neste trabalho.

Além do ICCG, o mesmo caso foi processado utilizando-se os métodos SSOR-CG e Decomposição de Cholesky o que permitiu a comparação de desempenho entre os três a nível de resultados, tempo de processamento e distribuição gráfica das linhas de campo.

Na etapa de pré-processamento, a discretização do domínio foi elaborada automaticamente utilizando-se o algoritmo de Delaunay.

Na etapa de pós-processamento foi utilizado um programa desenvolvido pelo pesquisador Douglas Ricardo de Freitas Clabunde

do Laboratório de Sistemas de Potência da Escola Politécnica da USP. Ambos estavam implementados em microcomputador compatível com o IBM-PC.

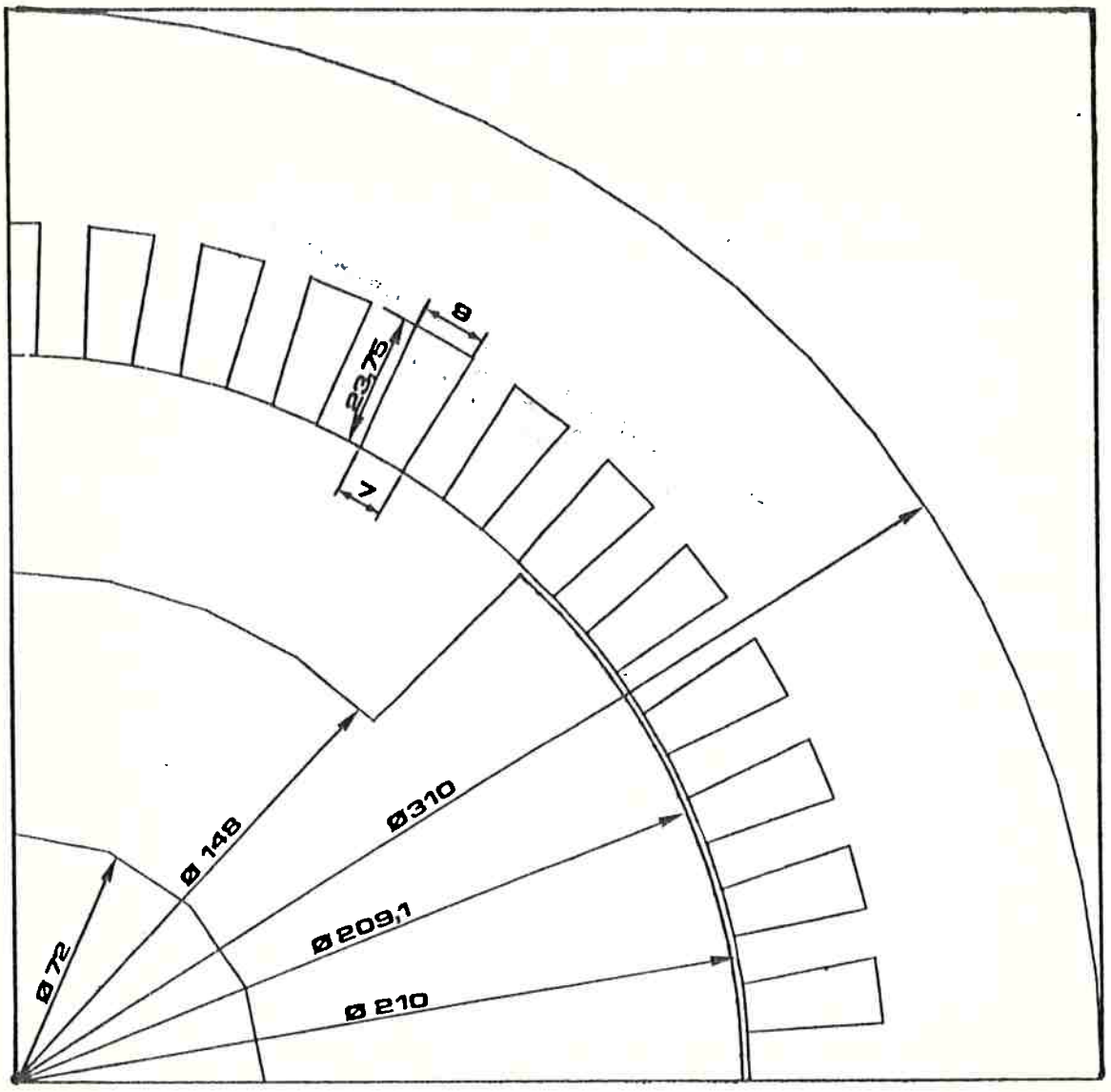

Figura 4.1.a: Motor síncrono de relutância (dimensões em mm)

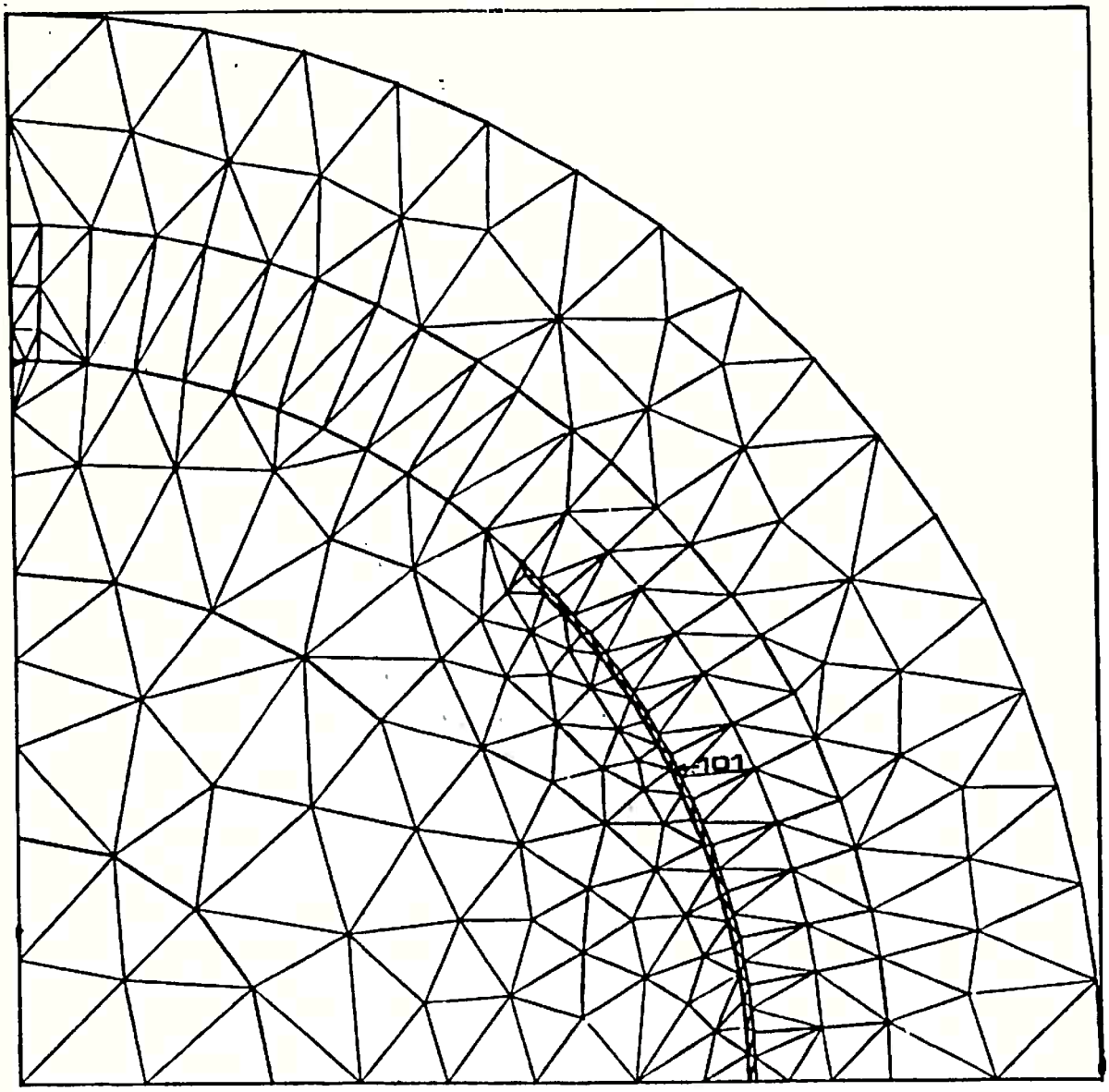

Figura 4.1.b: Domínio triangularizado segundo algoritmo de Delaunay.

4.2 RESULTADOS

O erro tolerável teve seu valor fixado em 10^{-4}, tanto para o Método de Newton-Raphson quanto para os Métodos Iterativos ICCG e SSOR-CG. Ainda no Método SSOR-CG, o parâmetro ω foi estabelecido em 0,8, valor este obtido por otimização empírica.

A nível de comparação dos resultados foi observado um desvio máximo de 10,1% no nó 101 para o ICCG e 29,6% no mesmo nó para o SSOR-CG, ambos em relação ao Método da Decomposição de Cholesky (indicado na figura 4.1.b).

A tabela 4.1 mostra os valores absolutos do potencial magnético neste ponto segundo os três métodos.

METODO	CHOLESKY	SSOR-CG	ICCG
potencial magnético(Wb/m)	-0,00128980180	-0,00167124788	-0,00142026646

Tabela 4.1: Valores do potencial magnético no ponto 101.

Os tempos de processamento observados estão indicados na tabela 4.2 abaixo.

METODO	CHOLESKY	SSOR-CG	ICCG
TEMPO CPU (S)	3751,944	653,081	593,115

Tabela 4.2: Tempo total gasto de CPU.

Nas figuras 4.2.a, 4.2.b e 4.2.c são apresentadas graficamente as distribuições das linhas de campo para os respectivos métodos de resolução. Como se pode verificar não são observadas diferenças sensíveis entre elas.

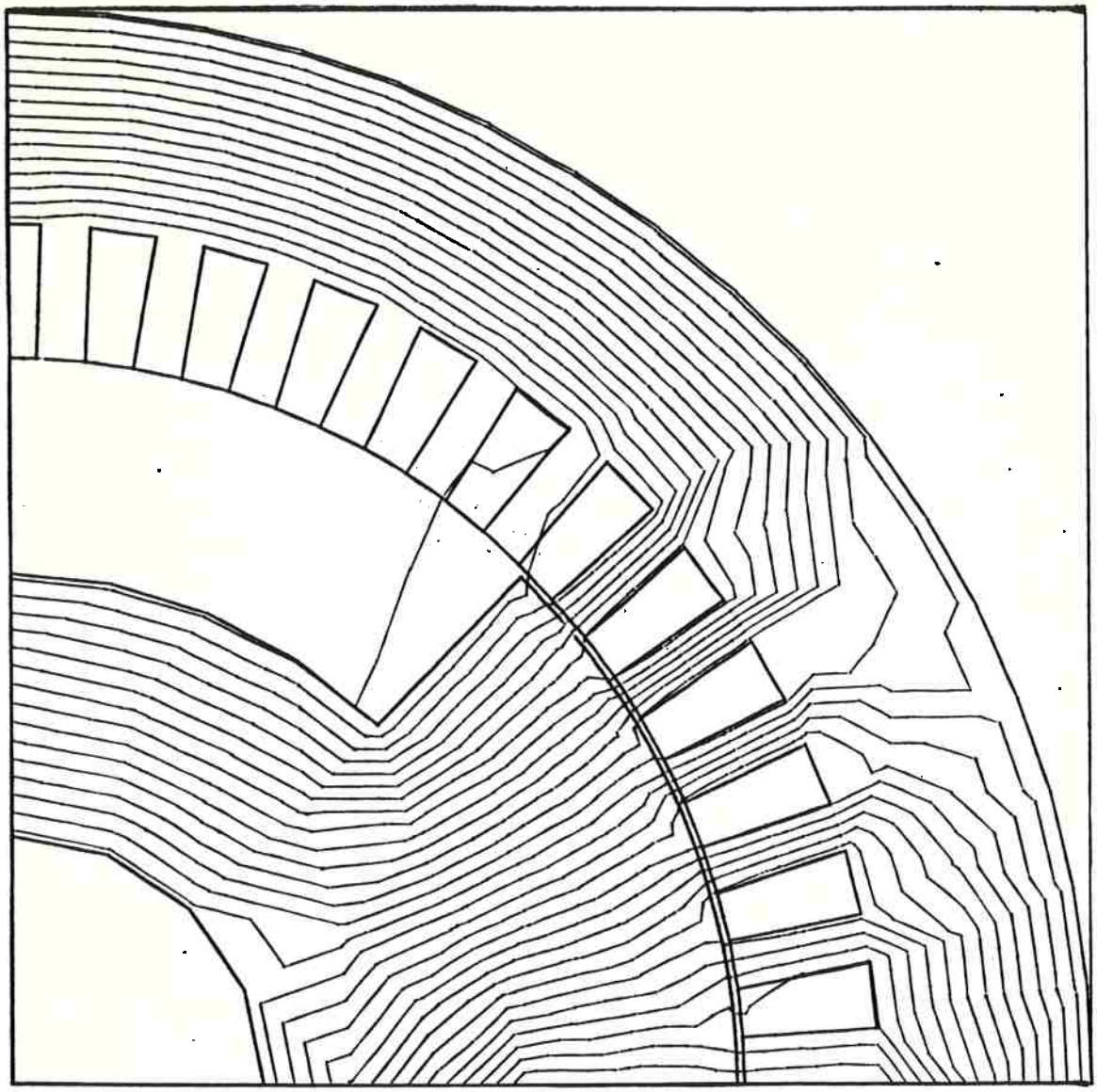

Figura 4.2.a: Linhas de campo resultantes do Método da Decomposição de Cholesky. 30 equipotenciais: max=0,05404720 Wb/m, min=-0,05645078 Wb/m.

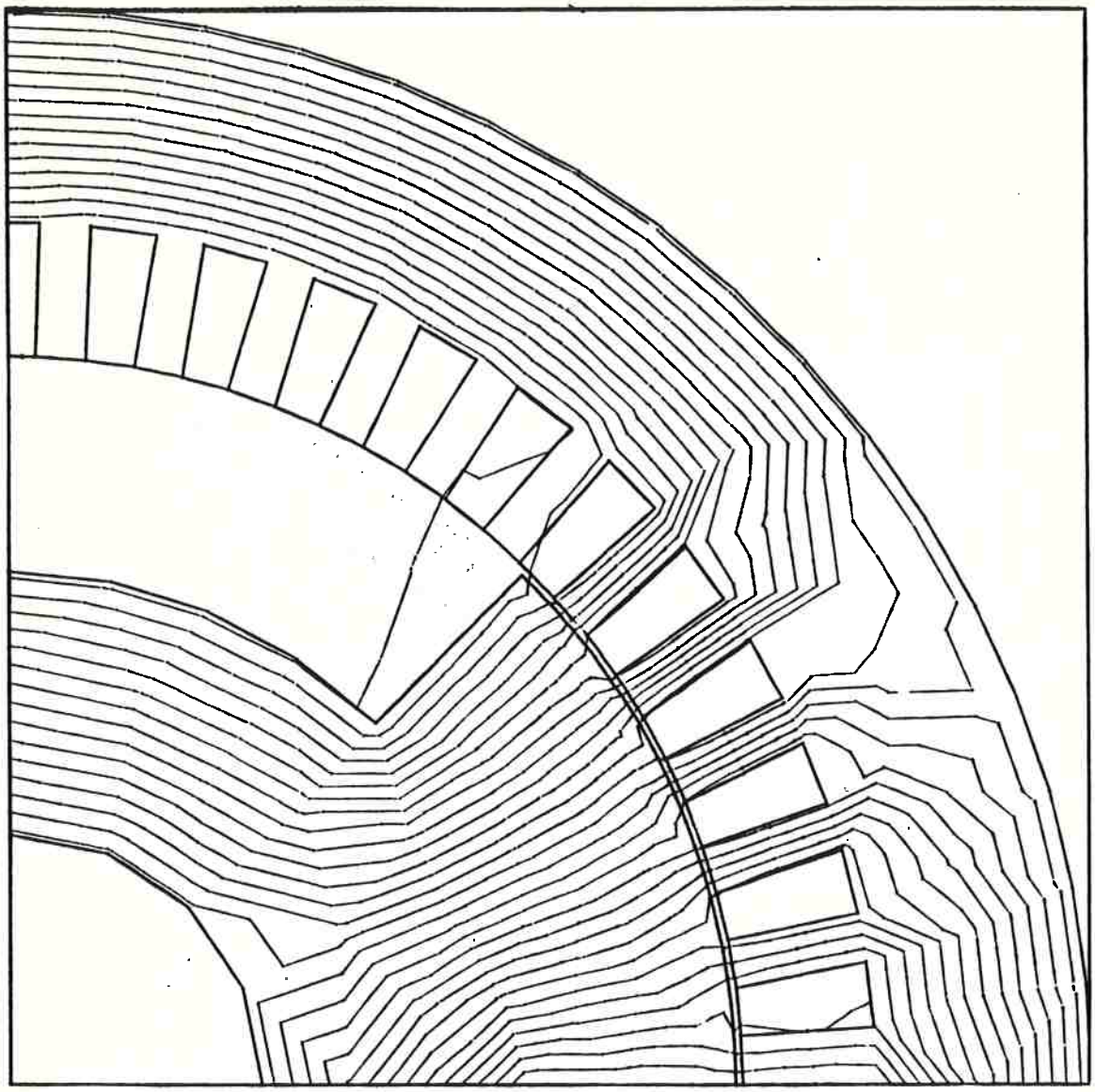

Figura 4.2.b: Linhas de campo resultantes do Método SSOR-CG. 30 equipotenciais: max=0,05823114 Wb/m, min=-0,05666643 Wb/m.

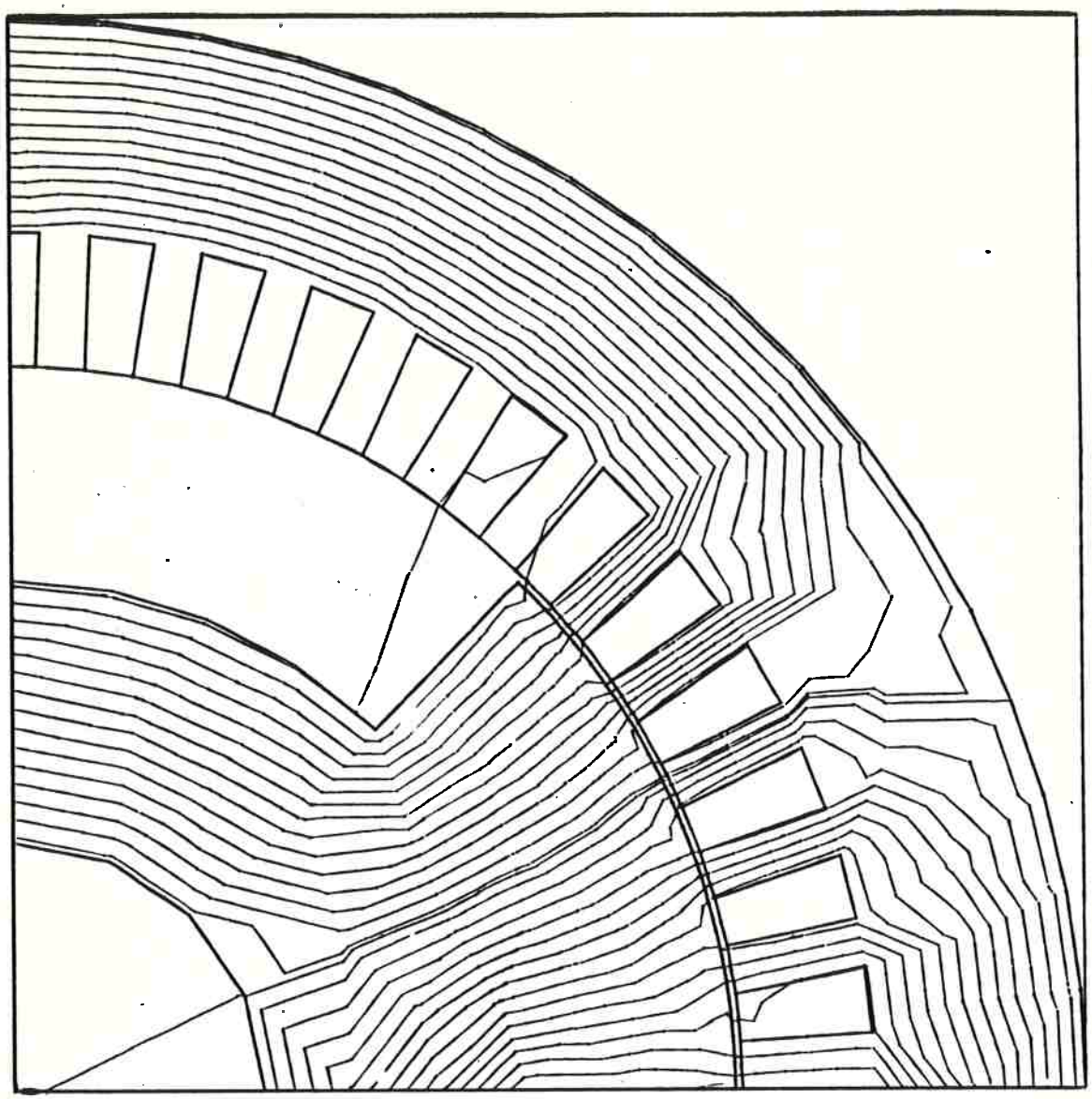

Figura 4.2.c: Linhas de campo resultantes do Método ICCG. 30 equipotenciais: max=0,05427600 Wb/m, min=-0,05670520 Wb/m.

As figuras 4.3.a, 4.3.b e 4.3.c mostram as distribuições dos auto-valores para a primeira matriz jacobiana montada no processamento, evidenciando que o pré-condicionamento aproxima os auto-valores da unicidade e diminui o número de condição espectral $\mathbb{K}$. A tabela 4.3 mostra os auto-valores mínimos e máximos para cada pré-condicionamento e os respectivos números de condição espectral $\mathbb{K}$.

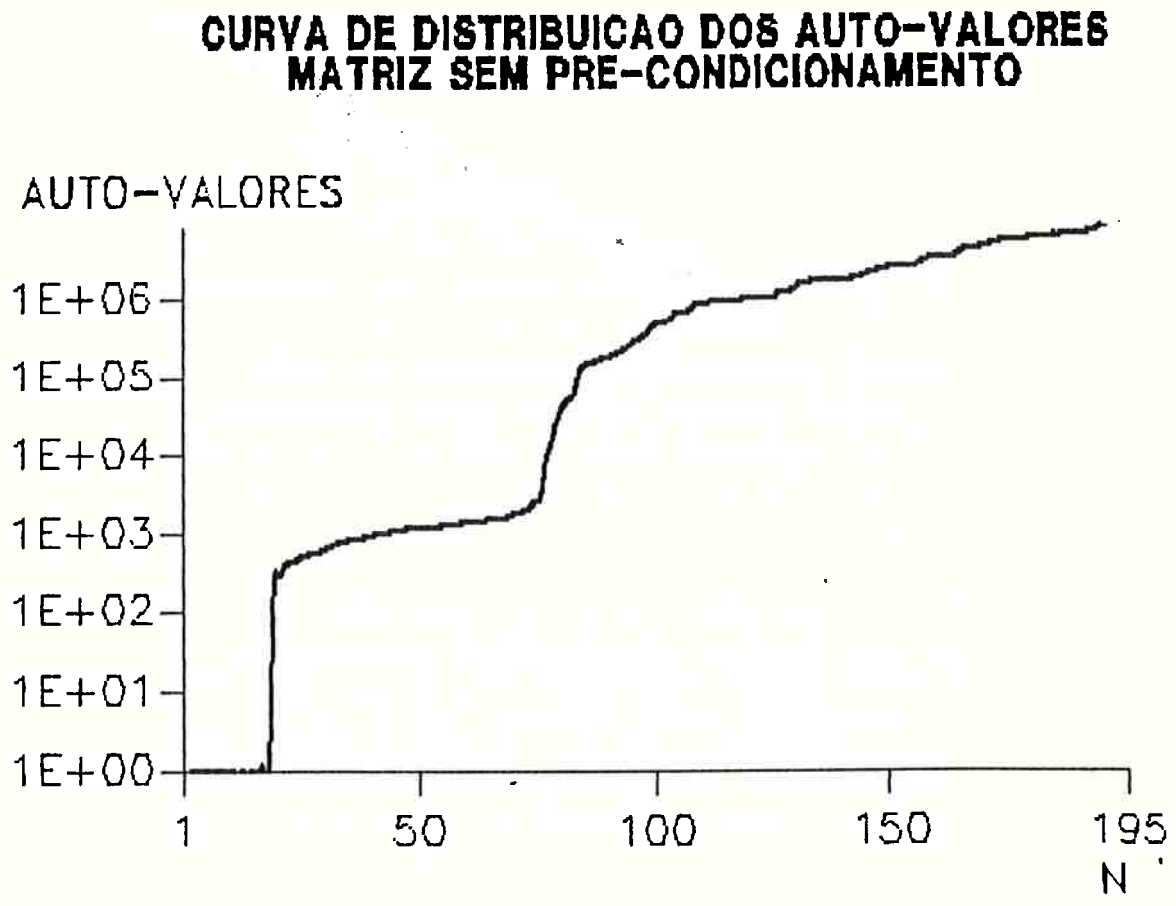

Figura 4.3.a: Distribuição dos auto-valores para a matriz sem pré-condicionamento.

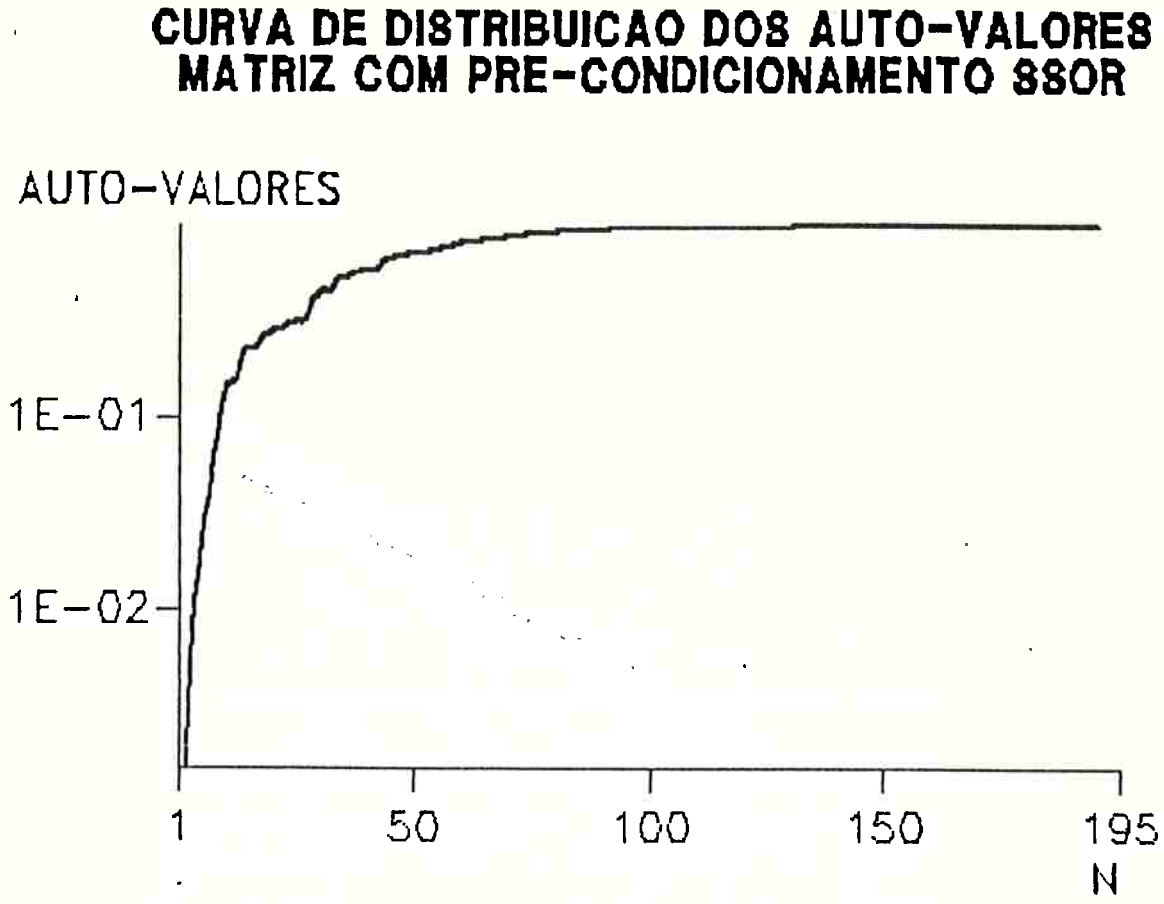

Figura 4.3.b: Distribuição dos auto-valores para o pré-condicionamento SSOR.

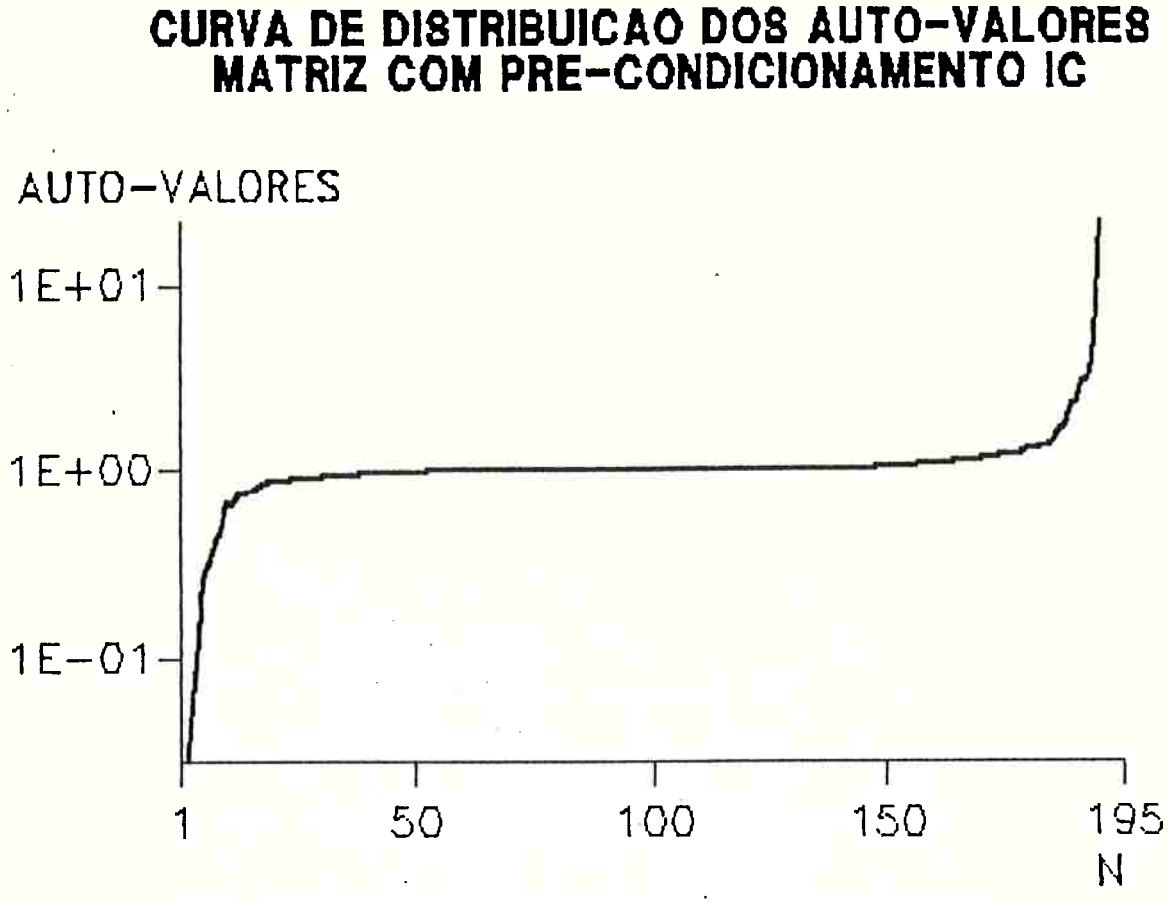

Figura 4.3.c: Distribuição dos auto-valores para o pré-condicionamento IC.

METODO	matriz sem pre-condicion.	pre-condicion. SSOR	pre-condicion. IC
auto-valor mínimo (λmin)	1,000	0,0016	0,0288
auto-valor máximo (λmax)	8.202.373,000	0,9999	22,0540
numero condição espectral ($\lambda max/\lambda min$)	8.202.373,000	624,9375	765,7640

Tabela 4.3: Auto-valores mínimos e máximos e números de condição espectral K da primeira matriz jacobiana.

CAPITULO 5: CONSIDERAÇÕES FINAIS

5.1 CONCLUSÕES

Os resultados obtidos mostram que o Método dos Gradientes Conjugados com Pré-condicionamento por Decomposição Incompleta de Cholesky (ICCG) foi aplicado com sucesso na análise magnetostática do motor em questão.

Os gráficos de distribuição de auto-valores comprovaram na prática a importância do pré-condicionamento e as boas taxas de convergência, tanto do ICCG como do SSOR-CG, evidenciadas nos tempos requeridos de CPU pelos dois métodos.

Estes gráficos ratificaram também os resultados alcançados por Watanabe [21] na análise de Problemas de Correntes Marítimas por Elementos Finitos utilizando o ICCG.

No seu trabalho, o pesquisador japonês decompõe a matriz A do sistema de maneira incompleta na forma $L.D.L^T$, onde L é uma matriz triangular inferior e D é a matriz diagonal.

Os elementos de L seguem a expressão :

$$L_{ji}=A_{ji}-\sum_{k=1}^{i-1} L_{jk}.L_{ik}.D_k,\quad i=1,2,\ldots N;\quad j=i,i+1,\ldots N$$

A decomposição incompleta depende da escolha do conjunto P, que é o conjunto dos elementos para os quais $L_{ij}=0$.

Quatro conjuntos foram escolhidos:

$P_A=\{(i,j) \mid |i-j|\neq 0,1,m,\ 1\leq i\leq N,\ 1\leq j\leq N\}$,

$P_B=\{(i,j) \mid |i-j|\neq 0,1,k,\ 1\leq i\leq N,\ 1\leq j\leq N,\ 3\leq k\leq m\}$,

$P_C=\{(i,j) \mid |i-j|\neq 0,1,\ldots k-1,k,\ 1\leq i\leq N,\ 1\leq j\leq N,\ 1\leq k\leq m\}$,

$P_D=\{(i,j) \mid a_{ij}=0,\ 1\leq i\leq N,\ 1\leq j\leq N\}$.

As distribuições dos auto-valores para os quatro tipos de P, assim como para a matriz não pré-condicionada, são mostradas a seguir.

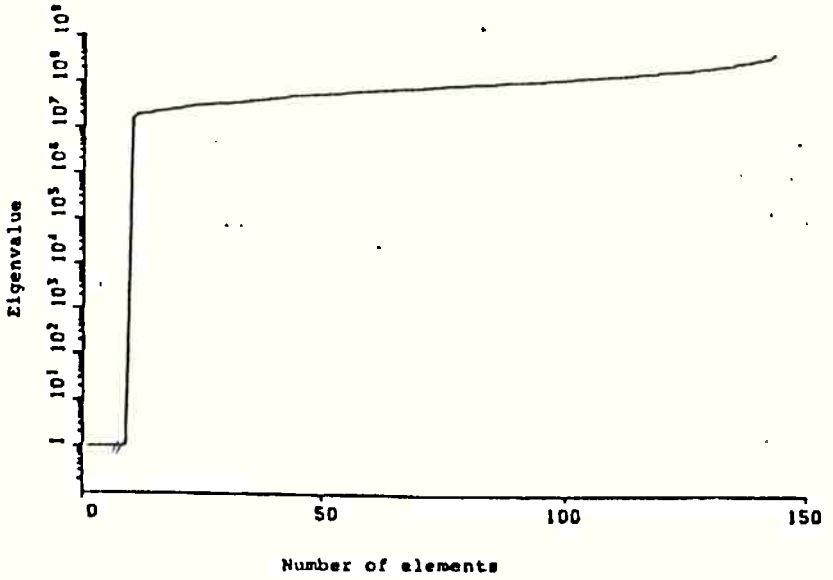

Figura 4.4.a: Distribuição dos auto-valores da matriz sem pré-condicionamento

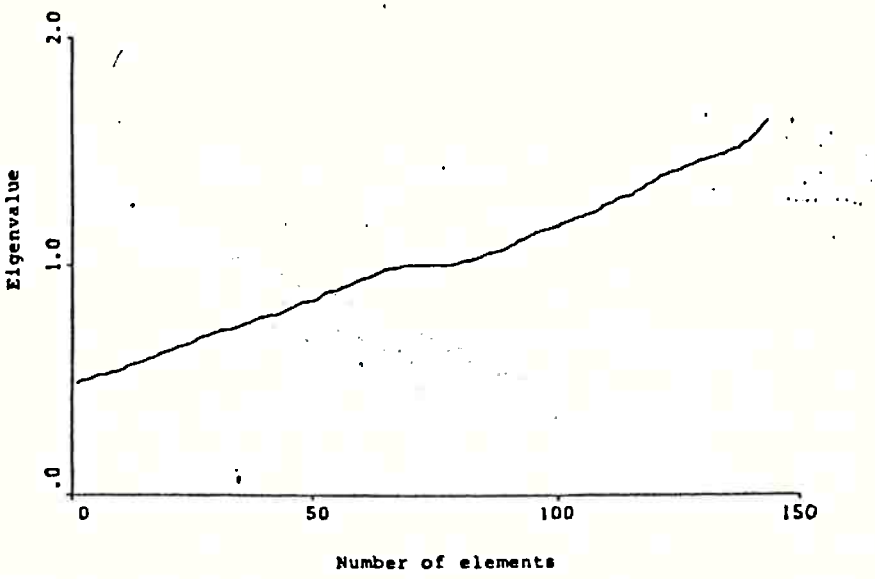

Figura 4.4.b: Distribuição dos auto-valores considerando P_A.

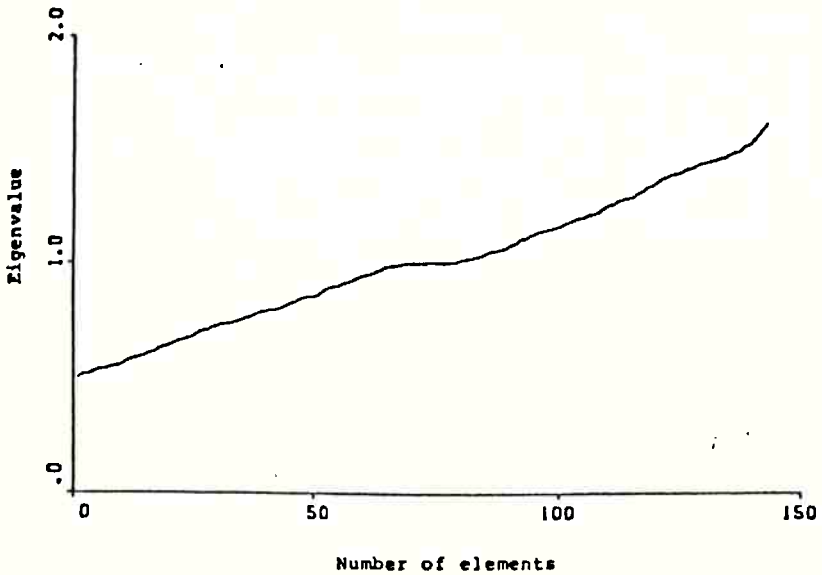

Figura 4.4.c: Distribuição dos auto-valores considerando P_B.

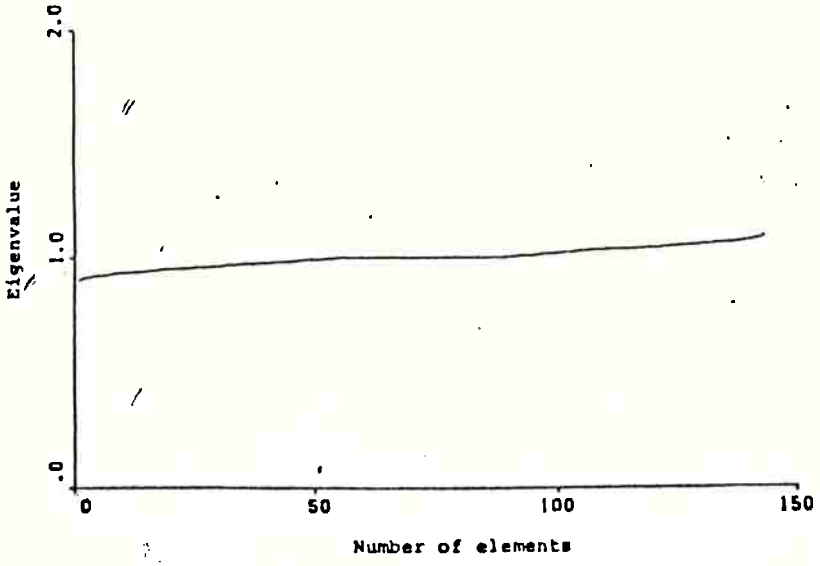

Figura 4.4.d: Distribuição dos auto-valores considerando P_C.

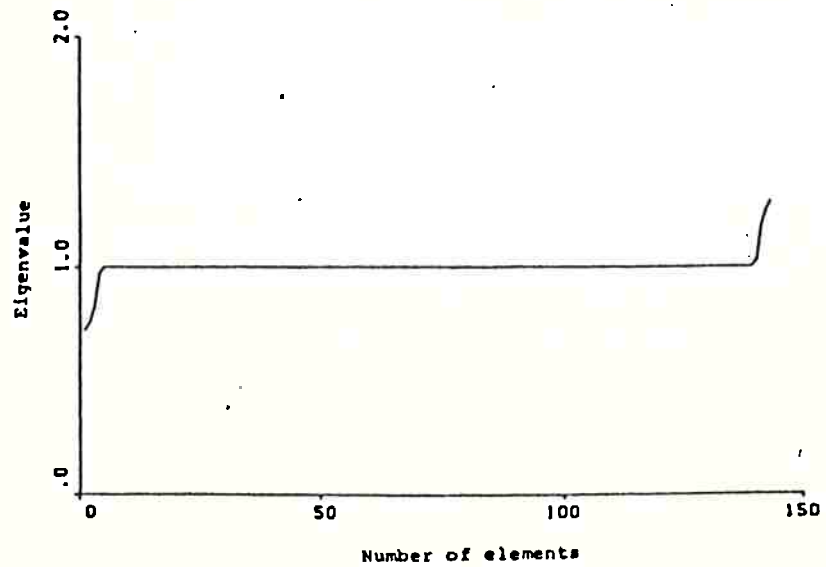

Figura 4.4.e: Distribuição dos auto-valores considerando P_D.

A nível de precisão, os resultados alcançados, tanto pelo ICCG como pelo SSOR-CG, foram comparados com os resultados do Método da Decomposição de Cholesky, uma vez que estes últimos incorporavam tão somente erros devidos a arredondamento. Desta comparação observamos que os resultados do ICCG se fixaram dentro de uma margem aceitável, apresentando uma variação relativa máxima de 10,1%, enquanto o SSOR-CG apresentou 29,6%.

Embora o método SSOR-CG tenha oferecido bons resultados, é conveniente ressaltar a dificuldade de se fixar o fator de aceleração ω para cada caso processado. Neste trabalho ω foi fixado em 0,8, valor este alcançado após várias tentativas e análise dos resultados obtidos.

As dificuldades encontradas durante a elaboração do programa computacional se centraram nas operações envolvendo matrizes com armazenamento compacto.

Devido às características peculiares de endereçamento deste tipo de armazenamento, todo tratamento matricial se tornou complexo, exigindo atenção extra.

5.2 DESENVOLVIMENTOS FUTUROS

O bom desempenho apresentado pelo método ICCG abre perspectivas para a utilização deste em outras aplicações, como, por exemplo, nas análises dinâmicas e naquelas em regime permanente senoidal.

Em particular, para as análises dinâmicas, Watanabe [21] afirma:

"O método ICCG é um Método iterativo que trabalha eficientemente com um bom valor inicial. Em problemas dependentes do tempo, a solução atingida num certo passo de tempo t pode ser utilizada como valor inicial para a solução do próximo passo $t+\Delta t$, quando Δt é pequeno, o que torna, portanto, o ICCG adequado para problemas

dependentes do tempo."

A utilização do ICCG em regimes permanente senoidal prevê algumas modificações na sua formulação, dado que nestes tipos de regime as matrizes envolvidas exibem componentes reais e complexos.

Segundo Sabonnadière [18], para estes casos, com matrizes não singulares, reais ou complexas, o método aplicável é o Bi-gradientes Conjugados com Pré-condicionamento.

Finalizando, embora o pré-condicionamento IC exposto neste trabalho tenha se apresentado satisfatório, é importante analisar outros tipos de pré-condicionamento, como os citados por Axelsson [2] e Watanabe [21], visando sempre um método mais eficiente.

APÊNDICE A: POLINOMIAIS DE CHEBYSHEV

Os polinomiais de Chebyshev possuem várias utilizações na análise numérica. Nossa breve apresentação neste apêndice se prenderá às necessidades da seção 3.2.2.1 em particular ao teorema 3.12.

A função $\cos(k.\theta)$, $-\pi\leq\theta\leq\pi$, onde k é um inteiro não-negativo, pode ser expressa como um polinomial de grau k em termos de $\cos\theta$, como pode ser verificado pela identidade trigonométrica:

$$\cos[(k+1).\theta]=2.\cos(\theta).\cos(k.\theta)-\cos[(k-1).\theta].$$

A relação $\cos(k.\theta)=T_k[\cos(\theta)]$ [A.1]

define o polinomial de Chebyshev de grau k.

Efetuando uma transformação de variáveis $x=\cos(\theta)$, $-\pi\leq\theta\leq\pi$, vemos que: $T_0(x)=1$, $T_1(x)=x$, $T_{k+1}(x)=2.x.T_k(x)-T_{k-1}(x)$, $k=2,3,\ldots$

Os seis primeiros polinomiais de Chebyshev são:

$$T_0(x)=1,\ T_1(x)=x,\ T_2(x)=2.x^2-1,\ T_3(x)=4.x^3-3.x,$$
$$T_4(x)=8.x^4-8.x^2+1,\ T_5(x)=16.x^5-20.x^3+5.x.$$

Evidentemente o domínio de definição destes polinomiais pode ser extendido do intervalo $-1\leq x\leq 1$ para valores arbitrários de x.

Consideremos a função $\cosh(k.\theta)=\frac{1}{2}\cdot\left(e^{k\theta}+e^{-k\theta}\right)$, $\theta\geq 0$.

Das relações: $\cosh(0)=1=T_0[\cosh(\theta)]$, $\cosh(\theta)=T_1[\cosh(\theta)]$ e da identidade: $\cosh[(k+1).\theta]=2.\cosh(\theta).\cosh(k.\theta)-\cosh[(k-1).\theta]$, obtemos: $\cosh(k.\theta)=T_k[\cosh(\theta)]$.

Combinando este resultado com A.1, alcançamos as seguintes expressões para $T_k(x)$:

$$T_k(x)=\begin{cases}\cos\left[k.\cos^{-1}(x)\right], & \text{para } -1\leq x\leq 1, \quad [A.2a]\\ \cosh\left[k.\cosh^{-1}(x)\right], & \text{para } x\geq 1, \quad [A.2b]\\ (-1)^k.\cosh\left[k.\cosh^{-1}(-x)\right], & \text{para } x\leq 1, \quad [A.2c]\end{cases}$$

ou equivalentemente:

$$T_k(x)=\frac{1}{2}\cdot\left[\left(x+\sqrt{x^2-1}\right)^k+\left(x-\sqrt{x^2-1}\right)^k\right],\ \text{onde}\ -\infty<x<\infty. \qquad [A.2d]$$

Analisando A.2a observamos que:

$$\max_{-1\le x\le 1}|T_k(x)|=1$$

e $T_k(x_i)=(-1)^i$, $i=0,1,\ldots k$, onde $x_i=\cos(i.\pi/k)$.

Para $|x|>1$ o valor de $|T_k(x)|$ aumenta rapidamente com $|x|$. O seguinte teorema reflete o fato de que num certo sentido a taxa de crescimento é máxima entre polinomiais de grau k.

TEOREMA A.1:

Seja Π_k^1 o conjunto de polinomiais de grau k com a propriedade $P_k(0)=1$ e seja $b>a>0$.

Então $$\max_{a\le x\le b}|\tilde{P}_k(x)|=\min_{P_k\in\Pi_k^1}\max_{a\le x\le b}|P_k(x)|,$$

onde $$\tilde{P}_k(x)=T_k.\left(\frac{b+a-2.x}{b-a}\right)\Big/T_k.\left(\frac{b+a}{b-a}\right).$$

PROVA:

Suponhamos que para algum $P_k\in\Pi_k^1$ temos:

$$\max_{a\le x\le b}|P_k(x)|<\max_{a\le x\le b}|\tilde{P}_k(x)| \qquad [A.4]$$

e seja $r(x)=\tilde{P}_k(x)-P_k(x)$.

O polinomial $T_k[(b+a-2.x)/(b-a)]$ é a generalização de $T_k(-x)$ no intervalo $a\le x\le b$. Seu valor absoluto não excede a unidade neste intervalo e de [A.3] observamos que oscila entre 1 e -1 no conjunto de k+1 pontos $a=\tilde{x}_0<\tilde{x}_1<\ldots.<\tilde{x}_{k-1}<\tilde{x}_k=b$, onde $\tilde{x}_i=\frac{1}{2}.[b+a-(b-a).x_i]$.

Segue de [A.4] que $r(x)$ atinge alternadamente valores positivos e negativos neste conjunto e, portanto, tem k zeros no intervalo $a<x<b$.

Mais ainda $r(0)=\tilde{P}_k(0)-P_k(0)=0$.

Portanto $r(x)$é um polinomial de grau $\le k$ possuindo no pior dos casos k+1 zeros. Como esta situação é impossível, o teorema está provado.

O teorema 3.12 da seção 3.2.2.1 requer a determinação do menor inteiro k tal que:

$$T_k[(b+a)/(b-a)]>1/\varepsilon$$

para um dado valor de ε. Utilizando [A.2d] encontramos para qualquer $\alpha>1$:

$$T_k\left(\frac{\alpha+1}{\alpha-1}\right)=\frac{1}{2}\cdot\left[\left(\frac{\sqrt{\alpha}+1}{\sqrt{\alpha}-1}\right)^k+\left(\frac{\sqrt{\alpha}-1}{\sqrt{\alpha}+1}\right)^k\right]>\frac{1}{2}\cdot\left(\frac{\sqrt{\alpha}+1}{\sqrt{\alpha}-1}\right)^k.$$

Portanto $k>\ln(2/\varepsilon)/\ln[(\sqrt{\alpha}+1)/(\sqrt{\alpha}-1)]\Rightarrow T_k\cdot\left(\frac{\alpha+1}{\alpha-1}\right)>\frac{1}{\varepsilon}$.

Uma vez que $\ln\left[\left(\sqrt{\alpha}+1\right)/\left(\sqrt{\alpha}-1\right)\right]>2/\sqrt{\alpha}$, $\alpha>1$, obtemos, fazendo $\alpha=b/a$, $k>\frac{1}{2}\cdot\sqrt{b/a}.\ln(2/\varepsilon)\Rightarrow T_k[(b+a)/(b-a)]>1/\varepsilon$. [A.5]

BIBLIOGRAFIA

[1] ALBRECHT, P. Análise numérica: um curso moderno. Rio de Janeiro, Livros Técnicos e Científicos, Pontifícia Universidade Católica do Rio de Janeiro, 1973.

[2] AXELSSON, O.; BARKER, V.A. Finite element solution of boundary value problems: theory and computation. Florida, Academic Press, 1984.

[3] BASTOS, J.P.A., Eletromagnetismo e cálculo de campos. Florianópolis, Editora da UFSC, 1989.

[4] BATHE, K.J. Finite element procedures in engineering analysis. Englewoods Cliffs, Prentice Hall, 1982.

[5] BATHE, K.J.; WILSON, E.L. Numerical methods in finite element analysis. Englewoods Cliffs, Prentice Hall, 1976.

[6] BOLDRINI, J.L.; COSTA, S.I.R.; RIBEIRO, V.L.F.F.; WETZLER, H.G. Álgebra linear. São Paulo, Harper e Row, 1978.

[7] CARDOSO, J.R. Problemas de campos eletromagnéticos estáticos e dinâmicos: uma abordagem pelo método dos elementos finitos. São Paulo, 1985. 159p. Tese (Doutorado) - Escola Politécnica, Universidade de São Paulo.

[8] COLEMAN, T.F. Large sparse numerical optimization. Lectures Notes in Computer Science. Berlin, Springer-Verlag, 1984.

[9] DHATT, G.; TOUZOT, G. Une présentation de la méthode des elements finis. 2.ed. Paris, Editeur Paris, 1984.

[10] DUFF, I.S.; ERISMAN, A.M., REID, J.K. Direct methods for sparse matrices. Oxford, Clarendon Press, 1986.

[11] DURAND, E. Solutions numériques des équations algébriques. Paris, Masson, 1972. V.2: Systèmes de plusieurs équations, valeurs propres des matrices.

[12] GEORGE, A.; LIU, J.W.H. Computer solution of large sparse positive definite systems. Englewoods Cliffs, Prentice, 1981.

[13] LEBENSZTAJN, L. Desenvolvimento de pré e pós processadores para o método dos elementos finitos aplicados à conversão eletromecânica de energia. São Paulo, 1989. 96p. Dissertação (Mestrado) - Escola Politécnica, Universidade de São Paulo.

[14] MANTEUFFEL, T.A. An incomplete factorization technique for positive definite linear systems. Mathematics of Computation, Menasha, v.34, n.150, p.473-97, Apr. 1980.

[15] MEIJERINK, J.A.; VAN DER VORST, H.A. An iterative solution method for linear systems of which the coefficient matrix is a symmetric M-matrix. Mathematics of Computation, Menasha, v.31, n.137, p.148-62, Jan. 1977.

[16] MORALES, J.L.; SANCHES, F.J. Solución eficiente de sistemas de ecuaciones lineales grandes y huecos. Boletin IIE, Col.del Valle, v.13, n.1, p.19-27, Ene./Feb. 1989.

[17] RAIZER, A. Contribuição à elaboração de um sistema tridimensional de cálculo de campos elétricos e magnéticos, utilizando a técnica dos elementos finitos. Florianópolis, 1987. 144p. Dissertação (Mestrado) - Universidade Federal de Santa Catarina.

[18] SABONNADIERE, J.C.; COULOMB, J.L. Éléments finis et CAO. Paris, Hermes, 1986.

[19] SILVESTER, P.P.; CABAYAN, H.S.; BROWNE, B.T. Efficient techniques for finite element analysis of electric machines. IEEE Transactions on Power Apparatus and Systems, New York, v.92, n.4, p.1274-81, July/Aug. 1973.

[20] SILVESTER, P.P.; FERRARI, R.L. Finite element for

electrical engineers. Cambridge, Cambridge University Press, 1983.

[21] WATANABE, M.; NAKAJIMA, H.; MORI, M. A finite element solution of a tidal current problem in the Seto inland sea by using the ICCG method. International Journal for Numerical Methods in Engineering, Chichester, v.21, p.1427-45, 1985.

[22] ZIENKIEWICZ, O.C. The finite element method. 3.ed. London, McGraw-Hill, 1977.

[23] ZIENKIEWICZ, O.C.; MORGAN, K. Finite elements and approximation. New York, John Wiley, 1983.