

Universidade Federal da Paraíba

Centro de Ciências e Tecnologia

Coordenação de Pós-Graduação em Informática

**Aceleração de Métodos Iterativos para Solução de
Sistemas Lineares – avaliação crítica**

Carlos Vinícius da Costa Ramos

Campina Grande
1996

Carlos Vinícius da Costa Ramos

Aceleração de Métodos Iterativos para solução de Sistemas Lineares – avaliação crítica

Dissertação apresentada ao Curso de Mestrado em Informática da Universidade Federal da Paraíba, como requisito parcial para obtenção do título de Mestre em Informática.

Área de Concentração: Matemática Computacional

Orientador: Mário Toyotaro Hattori

Campina Grande
1996



R175a Ramos, Carlos Vinicius da Costa
Aceleracao de metodos iterativos para solucao de sistemas lineares : avaliacao critica / Carlos Vinicius da Costa Ramos. - Campina Grande, 1996.
68 f.

Dissertacao (Mestrado em Informatica) - Universidade Federal da Paraiba, Centro de Ciencias e Tecnologia.

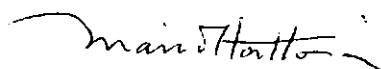
1. Matematica Computacional 2. Metodos Iterativos 3. Dissertacao I. Hattori, Mario Toyotaro, M.Sc. II. Universidade Federal da Paraiba - Campina Grande (PB)

CDU 519.6(043)


**ACELERAÇÃO DE MÉTODOS ITERATIVOS PARA SOLUÇÃO DE
SISTEMAS LINEARES - AVALIAÇÃO CRÍTICA**

CARLOS VINICIUS DA COSTA RAMOS

DISSERTAÇÃO APROVADA EM 09.08.96



PROF. MÁRIO TOYOTARO HATTORI, M.Sc
Presidente



PROF. BRUNO CORREIA DA NOBREGA QUEIROZ, M.Sc
Examinador



PROF. MAURO CAVALCANTE PEQUENO, Dr.
Examinador

CAMPINA GRANDE - PB

Dedicatória

À minha querida esposa Lourdinha e aos meus três filhos, Franksuel, Fabrícia e Frankálison, pela demonstração de carinho e compreensão nos momentos em que eu não podia dar atenção.

Agradecimentos

A Deus pela família maravilhosa que ele me deu.

Aos meus pais pelo esforço e dedicação. Ao meu pai, pelo exemplo do que é ser um homem íntegro e a minha mãe pelo amor infinito que se traduz nos pequenos gestos de renúncia e abnegação.

Aos meus queridos irmãos, cunhadas e cunhados, pelo apoio e solidariedade em todos os instantes de minha vida.

Ao meu orientador Mário Toyotaro Hattori, pela confiança em mim depositada pela demonstração de sincera amizade e pela paciência nos momentos difíceis do trabalho. Sua colaboração foi de vital importância para o meu enriquecimento pessoal e profissional.

Aos Professores e Funcionários do Departamento de Sistemas e Computação, pelo apoio que viabilizou esta pesquisa, Ao professor Bruno, Aninha, Alberto e Zeneide.

Às pessoas que durante todos estes anos me ajudaram, na colaboração técnica, troca de idéias, amizade, companheirismo e afeto,

aos Professores Antão, Peter, Hélio, Jacques e Hamurabi.

Aos colegas Kíssia, Kátia, Robson, Edjander, Eliane, Adeilton, Fábio, Reginaldo, Belo e Washington.

Resumo

O presente trabalho se propõe a comparar o desempenho dos métodos iterativos básicos de Jacobi e de Gauss-Seidel e dos métodos de aceleração de convergência, Sobre-Relaxação Sucessiva (SOR), de Chebyshev adaptativo, de Chebyshev não-adaptativo, o método dos Gradientes Conjugados e dos Gradientes Conjugados acelerado, para solução de sistemas de equações lineares.

Procurou-se estabelecer critérios práticos para decidir em que condições o emprego dos métodos de acelerações é recomendável em lugar dos métodos iterativos básicos quando as condições de aplicabilidade desses métodos são satisfeitas.

Abstract

The pupose of this work is to compare the performance of the basic Jacobi and Gauss-Seidel iterative methods, Successive over-relaxation, Chebyshev non-adaptative and adaptative convergence acceleration methods, the conjugate gradient and accelerated conjugate gradient methods for solution of systems of linear equations.

Provided applicability conditions for iterative methods are satisfied, practical criteria to decide under which conditions acceleration methods should be used instead basic iteration methods are sought.

Sumário

Capítulo I

Introdução	1
1.1 Métodos Iterativos	1
1.2 Objetivos do trabalho	2
1-3 Estrutura da dissertação	4

Capítulo II

Métodos iterativos básicos	6
2-1 Introdução	6
2-2 Convergência.....	7
2-3 Propriedades de matrizes	9
2-4 Métodos iterativos básicos	11
2-4.1 Método iterativo de Jacobi	11
2-4.2 Método iterativo de Gauss-Seidel.....	13

Capítulo III

Aceleração de métodos iterativos.....	16
3-1 Introdução	16
3-2 Método iterativo de Sobre-Relaxação Sucessiva (SOR)	17
3-3 Aceleração polinomial de métodos iterativos básicos.....	19
3-4 Aceleração de Chebyshev.....	22
3-5 Aceleração de Chebyshev adaptativa.....	26
3-6 Procedimento computacional	31

Capítulo IV

Método dos Gradientes Conjugados.....	36
4-1 Introdução	36
4-2 Método dos Gradientes Conjugados.....	36
4-3 Método dos Gradientes Conjugados Acelerado	40
4-4 Procedimento computacional	42

Capítulo V

Resultados e discussões.....	45
5-1 Introdução	45
5-2 Descrição das matrizes.....	45
5-2.1 Matriz simétrica	46
5-2.2 Matriz não-simétrica	46
5-3 Plano de testes	47
5-4 Observações sobre implementação	48
5-4.1 Processamento de matrizes por colunas.....	48
5-4.2 Algoritmo de Kahan	49
5-4.3 Critérios de interrupção das iterações	49
5-4.4 Escolha da norma a ser usada.....	51
5-4.5 Raio espectral	52
5-5 Resultados de testes.....	52
5-6 Observações sobre os resultados	61

Capítulo VI

Conclusões	63
Referências bibliográficas.....	67

Capítulo I

Introdução

1.1 Métodos Iterativos

Resolver sistemas de equações lineares cuja matriz dos coeficientes é esparsa, de ordem elevada e estruturada é muitas vezes um dos passos da solução de um problema maior. Exemplos típicos são os problemas que envolvem a resolução de sistemas de equações diferenciais parciais dos tipos parabólico, elíptico e hiperbólico.

Devido às dificuldades e, na maioria dos casos, à impossibilidade de se obter uma solução analítica de equações diferenciais, elas são resolvidas por métodos numéricos, o que envolve uma resolução de um sistema de equações lineares cuja matriz dos coeficientes possui estruturas especiais que podem ser exploradas. Solução de sistemas lineares com essas características ocorrem na solução de problemas de difusão de nêutrons, de escoamento de fluido,

de elasticidade, de transmissão de calor em sólidos, de previsão de tempo, além de problemas de engenharia [Young 71].

A resolução de sistemas lineares de grande porte e esparsos é praticamente impossível sem o uso do computador. Usualmente são utilizados métodos iterativos que não destroem a estrutura da matriz dos coeficientes, ao contrário dos métodos diretos que introduzem elementos não nulos nas posições que inicialmente continham elementos nulos. Métodos iterativos são mais adequados que os métodos diretos na solução de sistemas quando se procura minimizar o tempo de processamento e o espaço de armazenamento requerido no computador [Varga 62], devido à simplicidade e uniformidade das operações realizadas nas iterações. Contudo, deve-se chamar a atenção para o fato de que não há regras rígidas para escolher um método direto ou iterativo [Ralston 65]. Na verdade, existem controvérsias quanto a essas regras.

Serão tratadas duas classes de métodos iterativos. A primeira, baseado na iteração linear em uma variável, para encontrar uma raiz de uma equação não linear

$$\mathbf{u} - g(\mathbf{u}) = \mathbf{0}, \mathbf{u}^{(m+1)} = g(\mathbf{u}^{(m)}).$$

A segunda, a dos gradientes conjugados, baseado na minimização da forma quadrática

$$Q(\mathbf{u}) = \frac{1}{2} \mathbf{u}^T A \mathbf{u} - \mathbf{u}^T \mathbf{b}.$$

1.2 Objetivos do trabalho

O objetivo do trabalho é estabelecer critérios práticos para a escolha de um método iterativo que possa ser aplicado com sucesso na solução de um sistema de equações lineares.

Para alcançar esse objetivo, inicialmente serão revistos os fundamentos da teoria dos métodos iterativos. Será feito um estudo de um método iterativo linear estacionário de primeiro grau que se caracteriza por fornecer a cada iteração uma aproximação do vetor solução satisfazendo a equação

$$\mathbf{u}^{(m)} = G\mathbf{u}^{(m-1)} + \mathbf{k}, \quad m = 1, 2, \dots \quad (1-2.1)$$

Na análise do método genérico, definido pela equação (1-2.1), não há preocupação quanto à forma da matriz G e do vetor \mathbf{k} . Escolhendo-se formas particulares de G e de \mathbf{k} , obtêm-se os métodos iterativos de Jacobi, de Gauss-Seidel e de Sobre-Relaxação Sucessiva. Todos os métodos serão aplicados a problemas comuns para observar qual deles produz o melhor resultado. As condições de convergência desses métodos, que variam de acordo com as propriedades da matriz dos coeficientes, são consideradas em seguida. A atenção será dada para o caso da matriz dos coeficientes ser real. Além disso, para limitar o escopo deste trabalho, só serão vistos os casos em que a matriz dos coeficientes tem, pelo menos, uma dentre as propriedades abaixo, o que garante a convergência do método:

1. Matriz estritamente diagonal dominante;
2. Matriz irredutível com diagonal dominante;
3. Matriz simétrica positiva definida e
4. Matriz consistentemente ordenada.

Serão discutidos, ainda, procedimentos gerais para aceleração de convergência dos métodos iterativos de Jacobi e de Gauss-Seidel (métodos iterativos básicos). Esses procedimentos de aceleração, chamados aceleração polinomial, envolvem a formação de uma nova sequência de vetores que são combinações lineares de aproximações obtidas pelo método básico. O procedimento polinomial é uma das muitas abordagens usadas para acelerar a convergência de métodos iterativos básicos.

A Aceleração de Chebyshev, que é uma otimização da aceleração polinomial, é obtida com um custo baixo e não requer armazenamento de grandes quantidades de dados adicionais na memória como na aceleração polinomial. A aceleração usa aproximações obtidas nas iterações $m-1$ e m para obter uma nova aproximação

$$\mathbf{u}^{(m+1)} = \rho_{m+1} \left\{ \gamma \left(G\mathbf{u}^{(m)} + \mathbf{k} \right) = (1 - \gamma)\mathbf{u}^{(m)} \right\} + (1 - \rho_{m+1})\mathbf{u}^{(m-1)}, \quad (1-2.2)$$

em que ρ_{m+1} e γ são parâmetros de aceleração e $m \geq 1$.

No procedimento (1-2.2) um teste será feito durante cada iteração para determinar se os parâmetros de aceleração satisfazem ou não as condições de convergência. Se não satisfizerem, o procedimento adaptativo faz uma nova estimativa para eles. As decisões tomadas pelo procedimento adaptativo serão baseadas na comparação da redução do erro observado nas iterações (1-2.2) com a redução esperada do erro quando o parâmetro estimado for aceitável.

O método dos Gradientes conjugados, uma outra classe de métodos, é baseado na minimização da forma quadrática

$$Q(\mathbf{u}) = \frac{1}{2} \mathbf{u}^T A \mathbf{u} - \mathbf{b}^T \mathbf{u}.$$

Desde que

$$Q(\mathbf{u}) = Q(\bar{\mathbf{u}}) + \frac{1}{2} (\mathbf{u} - \bar{\mathbf{u}})^T A (\mathbf{u} - \bar{\mathbf{u}}),$$

onde $\bar{\mathbf{u}} = A^{-1}\mathbf{b}$ é a solução do sistema linear $A\mathbf{u} = \mathbf{b}$, em que A é uma matriz simétrica positiva definida, resolver o problema $A\mathbf{u} = \mathbf{b}$ é equivalente ao problema de minimizar $Q(\mathbf{u})$.

1-3 Estrutura da dissertação

O trabalho contém 6 capítulos, dos quais o primeiro é uma introdução e os cinco restantes seu desenvolvimento.

O capítulo 2 trata dos métodos iterativos básicos. São definidos os métodos iterativos de Jacobi e de Gauss-Seidel, e as matrizes de iteração desses métodos são apresentadas. O tópico sobre convergência apresenta teoremas relacionando raio espectral e taxa de convergência com os autovalores das matrizes de iteração.

O capítulo 3 apresenta o método de Sobre-Relaxação Sucessiva (SOR) e o método de aceleração de Chebyshev. Uma breve apresentação do método SOR é feita enfatizando a consistência do método com o sistema $A\mathbf{u} = \mathbf{b}$. É apresentado de forma detalhada o método da aceleração de Chebyshev, finalizando o capítulo com um algoritmo global do método.

O capítulo 4 trata do método dos gradientes conjugados e do método dos gradientes conjugados acelerado com base na matriz de iteração de Jacobi, finalizando o capítulo com um algoritmo global do método da aceleração dos gradientes conjugados.

O capítulo 5 apresenta uma descrição das matrizes dos coeficientes dos sistemas usados em testes, um plano de testes, os resultados dos testes realizados e uma avaliação comparativa dos métodos.

O capítulo 6 apresenta conclusões e sugestões para trabalhos adicionais.

Capítulo II

Métodos iterativos básicos

2-1 Introdução

Este capítulo estuda uma classe de métodos iterativos para resolver sistemas de equações lineares

$$A\mathbf{u} = \mathbf{b}, \quad (2-1.1)$$

em que A é uma matriz real não singular $n \times n$ e \mathbf{b} é um vetor coluna real $n \times 1$ conhecido.

Todos os métodos considerados neste capítulo são métodos estacionários lineares de primeira ordem, que podem ser expressos na forma

$$\mathbf{u}^{(n+1)} = G\mathbf{u}^{(n)} + \mathbf{k}, \quad n = 1, 2, \dots, \quad (2-1.2)$$

em que G é a matriz de iteração real associado ao método dependente de A , e \mathbf{k} é um vetor conhecido que depende de A e de \mathbf{b} . O método é de primeiro grau se $\mathbf{u}^{(n+1)}$ depende unicamente de $\mathbf{u}^{(n)}$ e não de $\mathbf{u}^{(n-1)}, \dots, \mathbf{u}^{(0)}$. O método é linear se nem G nem \mathbf{k} dependem de $\mathbf{u}^{(n)}$, e é estacionário, se nem G nem \mathbf{k} dependem de n . Neste contexto, qualquer método da forma (2-1.2) será chamado método iterativo básico.

2-2 Convergência

O método iterativo definido por (2-1.2) é convergente se e somente se para qualquer aproximação inicial $\mathbf{u}^{(0)}$, a sequência gerada converge para algum $\bar{\mathbf{u}}$.

Definição 2-2.1 O raio espectral $S(A)$ de uma matriz A de ordem $n \times n$ é o maior dentre os módulos dos autovalores de A . Assim

$$S(A) = \max_{1 \leq i \leq n} |\lambda_i|.$$

Se os autovalores de A forem reais, denotamos por $m(A)$ e $M(A)$ algebricamente o menor e o maior autovalor de A , respectivamente, isto é,

$$m(A) = \min_{1 \leq i \leq n} \lambda_i \quad M(A) = \max_{1 \leq i \leq n} \lambda_i \quad (2-2.1)$$

O teorema a seguir é importante no estudo da convergência. Ele estabelece a condição que os autovalores de G devem satisfazer para haver convergência de um método iterativo.

Teorema 2-2.1 O método iterativo definido por (2-1.2) é convergente se e somente se o raio espectral da matriz G , $S(G)$, satisfaz a condição $S(G) < 1$ [Young 71].

O teorema 2-2.1 estabelece uma condição necessária e suficiente para a convergência do método. Entretanto, não garante a convergência para a solução única do sistema. Para se determinar quando isto ocorre, é necessária a introdução do conceito de consistência.

Definição 2-2.2 O método iterativo definido por (2-1.2) é consistente com a equação (2-1.1) se e somente se toda solução da equação (2-1.1) for solução da equação

$$(I - G)\mathbf{u} = \mathbf{k}, \quad (2-2.2)$$

em que I é a matriz identidade de ordem n .

O teorema seguinte estabelece que o método converge para a única solução do sistema (2-1.1) qualquer que seja a aproximação inicial.

Teorema 2-2.2 Se a equação (2-1.1) tiver solução e o método iterativo (2-1.2) for consistente com a equação (2-1.1) e convergente, então A é não singular e a convergência se dá para a única solução do sistema qualquer que seja a aproximação inicial [Young 71].

Se um método iterativo for consistente com o sistema (2-1.1), e se em alguma iteração, a solução $\bar{\mathbf{u}}$ desse sistema for obtida, em todas as iterações subsequentes o mesmo vetor $\bar{\mathbf{u}}$ será repetidamente encontrado. Esse fato evidencia também a importância do conceito de consistência.

Até o momento foram consideradas as condições sob as quais o método iterativo definido por (2-1.2) converge para a solução única do sistema $A\mathbf{u} = \mathbf{b}$, sem se preocupar com a velocidade de convergência. A taxa de convergência é um conceito que surge da necessidade de se analisar essa velocidade. Para medir a rapidez de convergência do método iterativo estacionário linear (2-1.2), o vetor erro $\boldsymbol{\varepsilon}^{(m)}$ é definido por

$$\boldsymbol{\varepsilon}^{(m)} \equiv \mathbf{u}^{(m)} - \bar{\mathbf{u}}. \quad (2-2.3)$$

Usando (2-1.2) com o fato de que $\bar{\mathbf{u}}$ também satisfaz a equação (2-2.2), tem-se

$$\begin{aligned} \boldsymbol{\varepsilon}^{(m)} &= (G\mathbf{u}^{(m+1)} + \mathbf{k}) - \mathbf{u} = G\mathbf{u}^{(m-1)} + (I - G)\mathbf{u} - \mathbf{u} \\ &= G\mathbf{u}^{(m-1)} - G\mathbf{u} + \mathbf{u} - \mathbf{u} = G(\mathbf{u}^{(m-1)} - \mathbf{u}) \\ &= G\boldsymbol{\varepsilon}^{(m-1)} = \dots = G^m\boldsymbol{\varepsilon}^{(0)}. \end{aligned} \quad (2-2.4)$$

Portanto, para qualquer norma, tem-se

$$\|\boldsymbol{\varepsilon}^{(m)}\| \leq \|G^m\| \cdot \|\boldsymbol{\varepsilon}^{(0)}\|. \quad (2-2.5)$$

Assim, $\|G^m\|$ dá uma medida pela qual a norma do erro foi reduzida depois de m iterações. Define-se taxa média de convergência de (2-1.2) por

$$R_m(G) \equiv -m^{-1} \log \|G^m\|. \quad (2-2.6)$$

Pode-se mostrar que, se $S(G) < 1$, então $\lim_{m \rightarrow \infty} (\|G^m\|)^{1/m} = S(G)$ [Young 71]. Isso permite definir taxa assintótica de convergência que é dada por

$$R_\infty(G) = \lim_{m \rightarrow \infty} R_m(G) = -\log(S(G)). \quad (2-2.7)$$

Observe que $R_m(G)$ depende da norma e $R_\infty(G)$ independe de qualquer norma.

Definição 2-2.3 Sejam G_1 e G_2 duas matrizes de iteração. A iteração (2-1.2), com $G = G_1$, é chamada assintoticamente mais rápida que a iteração com $G = G_2$, se $S(G_1) < S(G_2) < 1$.

Da definição de taxa de convergência conclui-se que a velocidade de convergência será tanto maior quanto menor for $S(G)$.

2-3 Propriedades de matrizes

Nesta seção são revistas algumas propriedades básicas de matrizes e notações da teoria de matrizes que serão usadas neste trabalho.

Definição 2-3.1 Uma matriz A de ordem $n \times n$ é simétrica positiva definida (SPD), se e somente se, A for simétrica e $\mathbf{v}^T A \mathbf{v} > 0$, para qualquer vetor de ordem $n \times 1$, $\mathbf{v} \neq 0$.

Definição 2-3.2 O método iterativo (2-1.2) é simetrizável se para alguma matriz W a matriz $W(I - G)W^{-1}$ é SPD. A matriz W é chamada matriz de simetrização.

Definição 2-3.3 Uma matriz A , de ordem $n \times n$ é irredutível, se $n = 1$ ou se $n > 1$ e dados dois subconjuntos não vazios S e T de W , o conjunto dos n primeiros inteiros positivos, satisfazendo $S + T = W$, existirem $i \in S$ e $j \in T$ tal que $a_{ij} \neq 0$.

Definição 2-3.4 Uma matriz A de ordem $n \times n$ é diagonal dominante se

$$1) |a_{ii}| \geq \sum_{j=1, j \neq i}^n |a_{ij}| = r_i, \quad i = 1, 2, \dots, n \text{ e}$$

$$2) \text{ se existe, pelo menos um } k \text{ tal que } |a_{kk}| > r_k.$$

A é estritamente diagonal dominante se

$$|a_{ii}| > \sum_{j=1, j \neq i}^n |a_{ij}|, \quad i = 1, 2, \dots, n.$$

Definição 2-3.5 Uma matriz A , $n \times n$, é consistentemente ordenada, se e somente se, existirem t subconjuntos disjuntos de W , o conjunto dos n primeiros inteiros positivos, tais que sua união seja igual a W e $j \in S_{k+1}$ quando $i \in S_k$ e $i > j$ enquanto que $j \in S_{k-1}$ se $i < j$.

As normas comumente utilizadas são

$$\|\mathbf{v}\|_2 = (\mathbf{v}^T \mathbf{v})^{1/2} \equiv \left[\sum_{i=1}^n (v_i)^2 \right]^{1/2}$$

chamada **norma euclidiana** que é equivalente ao comprimento ou distância da geometria analítica.

$$\|\mathbf{v}\|_1 = \sum_{i=1}^n |v_i|$$

chamada **norma 1** ou **uniforme**,

$$\|\mathbf{v}\|_{\infty} = \max_{1 \leq i \leq n} \{|v_i|\},$$

chamada **norma** ∞ ou de **Chebyshev**, e

$$\|\mathbf{v}\|_L = \|L\mathbf{v}\|_2,$$

em que L é uma matriz quadrada não singular chamada **norma** L .

2-4 Métodos iterativos básicos

Na seção anterior foram tratados métodos iterativos genéricos definidos pela equação (2-1.2) sem especificar a forma da matriz G e do vetor \mathbf{k} . A escolha desses parâmetros define métodos particulares tais como o de Jacobi e o de Gauss-Seidel.

Dada a equação (2-1.1), o vetor solução $\bar{\mathbf{u}}$ existe e é único se, e somente se, A for não singular e esse vetor solução é dado por

$$\bar{\mathbf{u}} = A^{-1}\mathbf{b}.$$

Assume-se daqui para frente que a matriz A seja não singular e os elementos da diagonal principal $a_{i,i}$ sejam todos números não nulos.

A matriz A pode ser decomposta em

$$A = D - L - U,$$

em que $D = \text{diag} \{a_{1,1}, a_{2,2}, \dots, a_{n,n}\}$, L e U são, respectivamente, matrizes triangulares estritamente inferior e superior de ordem $n \times n$, cujos elementos são aqueles, respectivamente, abaixo e acima da diagonal principal da matriz A .

2-4.1 Método iterativo de Jacobi

De acordo com a equação (2-1.1), o sistema de equações pode ser reescrito na forma

$$D\mathbf{u} = (L + U)\mathbf{u} + \mathbf{b}. \quad (2-4.1.1)$$

Desde que os elementos $a_{i,i}$ de A não sejam nulos, obtém-se o seguinte método iterativo, derivado de (2-4.1.1):

$$u_i^{(m+1)} = - \sum_{j=1, j \neq i}^n \frac{a_{i,j}}{a_{i,i}} u_j^m + \frac{b_i}{a_{i,i}}, \quad 1 \leq i \leq n, \quad m \geq 0. \quad (2-4.1.2)$$

em que $u_i^{(0)}$ são estimativas iniciais dos componentes da solução única $\bar{\mathbf{u}}$ de (2-1.1). Na notação matricial, (2-4.1.2) pode ser expresso na forma:

$$\mathbf{u}^{(m+1)} = D^{-1}(L + U)\mathbf{u}^{(m)} + D^{-1}\mathbf{b}, \quad m \geq 0. \quad (2-4.1.3)$$

Este é o método iterativo de Jacobi.

Observando a correspondência entre os elementos das equações (2-1.2) e (2-4.1.3)

$$G = D^{-1}(L + U) \quad \text{e} \quad \mathbf{k} = D^{-1}\mathbf{b}. \quad (2-4.1.4)$$

A matriz $B = D^{-1}(L + U)$ é chamada matriz de iteração de Jacobi associada a matriz A . Além disso, tem-se:

$$(I - G) = (I - B) = I - D^{-1}(L + U) = I - D^{-1}(D - A) = D^{-1}A.$$

Considere o seguinte sistema linear:

$$\begin{aligned} 3u_1 + u_2 + u_3 &= 5 \\ u_1 + 2u_2 &= 3 \\ u_1 + \frac{1}{2}u_2 + 2u_3 &= 6. \end{aligned} \quad (2-4.1.5)$$

Rescrevendo o sistema $A\mathbf{u} = \mathbf{b}$

$$\begin{aligned}
 u_1^{(m+1)} &= -\frac{1}{3}u_2^{(m)} - \frac{1}{3}u_3^{(m)} + \frac{5}{3}, \\
 u_2^{(m+1)} &= -\frac{1}{2}u_1^{(m)} + \frac{3}{2}, \\
 u_3^{(m+1)} &= -\frac{1}{2}u_1^{(m)} - \frac{1}{4}u_2^{(m)} + 3,
 \end{aligned}$$

que é da forma $\mathbf{u}^{(m+1)} = B\mathbf{u}^{(m)} + \mathbf{k}$ em que

$$B = \begin{pmatrix} 0 & -\frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{2} & 0 & 0 \\ -\frac{1}{2} & -\frac{1}{4} & 0 \end{pmatrix}, \quad \mathbf{k} = \left(\frac{5}{3}, \frac{3}{2}, 3\right)^T \quad \text{e} \quad S(B) \cong 0,63.$$

2-4.2 Método iterativo de Gauss-Seidel

Examinando o método iterativo de Jacobi (2-4.1.3), observa-se que em geral deve-se salvar todos os componentes do vetor $\mathbf{u}^{(m)}$ enquanto calcula os componentes do vetor $\mathbf{u}^{(m+1)}$. Intuitivamente parece mais atraente usar as últimas estimativas $u_i^{(m+1)}$ do componente u_i do vetor solução único $\bar{\mathbf{u}}$, em todos componentes subsequentes, para obter o método iterativo

$$u_i^{(m+1)} = -\sum_{j=1}^{i-1} \frac{a_{i,j}}{a_{i,i}} u_j^{(m+1)} - \sum_{j=i+1}^n \frac{a_{i,j}}{a_{i,i}} u_j^m + \frac{b_i}{a_{i,i}}, \quad 1 \leq i \leq n, \quad m \geq 0. \quad (2-4.2.1)$$

Este é o método iterativo de Gauss-Seidel e tem a vantagem de não exigir o armazenamento simultâneo de dois vetores $\mathbf{u}^{(m+1)}$ e $\mathbf{u}^{(m)}$ em cada passo como no método iterativo de Jacobi. Na notação matricial, (2-4.2.1) fica

$$\mathbf{u}^{(m+1)} = D^{-1}L\mathbf{u}^{(m+1)} + D^{-1}U\mathbf{u}^{(m)} + D^{-1}\mathbf{b}$$

Como $(D - L)$ é uma matriz triangular inferior, $\det(D - L) \neq 0$ e, portanto, inversível. Logo,

$$\mathbf{u}^{(m+1)} = (D - L)^{-1}U\mathbf{u}^{(m)} + (D - L)^{-1}\mathbf{b}.$$

Assim o método de Gauss-Seidel pode ser expresso na forma

$$\mathbf{u}^{(m+1)} = \mathcal{L}\mathbf{u}^{(m)} + \mathbf{k}, m \geq 0, \quad (2-4.2.2)$$

em que

$$\mathcal{L} \equiv (D - L)^{-1}U, \quad \mathbf{k} \equiv (D - L)^{-1}\mathbf{b}$$

A matriz \mathcal{L} é chamada matriz de iteração de Gauss-Seidel associada à matriz A . Além disso, tem-se:

$$\begin{aligned} (I - G) &= (I - \mathcal{L}) = I - (D - L)^{-1}U = I - (D - L)^{-1}[(D - L) - A] \\ &= I - (D - L)^{-1}(D - L) + (D - L)^{-1}A = (D - L)^{-1}A. \end{aligned}$$

Para o sistema (2-4.1.5), obtém-se a iteração de Gauss-Seidel

$$\begin{aligned} u_1^{(m+1)} &= -\frac{1}{3}u_2^{(m)} - \frac{1}{3}u_3^{(m)} + \frac{5}{3}, \\ u_2^{(m+1)} &= -\frac{1}{2}u_1^{(m+1)} + \frac{3}{2}, \\ u_3^{(m+1)} &= -\frac{1}{2}u_1^{(m+1)} - \frac{1}{4}u_2^{(m+1)} + 3. \end{aligned}$$

Substituindo $u_1^{(m+1)}$ e $u_2^{(m+1)}$ no lado direito pelos valores $u_1^{(m)}$, $u_2^{(m)}$, $u_3^{(m)}$ observa-se que a matriz de iteração de Gauss-Seidel é:

$$\mathcal{L} = \begin{pmatrix} 0 & -\frac{1}{3} & -\frac{1}{3} \\ 0 & \frac{1}{6} & \frac{1}{6} \\ 0 & \frac{1}{8} & \frac{1}{8} \end{pmatrix}, \quad \mathbf{k} = \left(\frac{5}{3}, \frac{4}{6}, 2\right)^T \quad \text{e} \quad S(\mathcal{L}) \equiv 0,29.$$

Observa-se que \mathcal{L} nunca é formada explicitamente no cálculo dos vetores $\mathbf{u}^{(m)}$.

Para o sistema (2-4.1.5), foi obtido $S(B) \cong 0,63$ e $S(\mathcal{L}) \cong 0,29$. Então, a iteração de Gauss-Seidel converge assintoticamente mais rápida do que o método de Jacobi.

Definidos esses métodos iterativos básicos pode-se concluir pela construção das matrizes de iteração, que uma condição necessária para que qualquer um dos métodos seja aplicável, é que todos os elementos da diagonal principal dessa matriz dos coeficientes sejam diferentes de zero.

Por outro lado, conforme foi visto, para os métodos em questão, é possível expressar \mathbf{k} por $M\mathbf{b}$ e $(I - G)$ por MA , onde M é uma matriz que pode ser identificada em cada situação. Assim considerando-se o sistema (2-1.1), tem-se que $(I - G)A^{-1}\mathbf{b} = MAA^{-1}\mathbf{b} = M\mathbf{b} = \mathbf{k}$ e $A^{-1}\mathbf{b}$ é também solução de (2-1.2). Isto mostra que o método definido pela equação (2-1.2) é consistente com o sistema (2-1.1) de acordo com a definição de consistência.

Com base no exposto acima, pode-se afirmar que os métodos iterativos de Jacobi e de Gauss-Seidel são consistentes com o sistema (2-1.1) que tem solução única.

Capítulo III

Aceleração de métodos iterativos

3-1 Introdução

Neste capítulo serão discutidos uma variante do método iterativo de Gauss-seidel, o método iterativo de Sobre-Relaxação Sucessiva (SOR) e os procedimentos gerais para aceleração da convergência de métodos iterativos básicos. Esses procedimentos de aceleração, chamados de aceleração polinomial, envolvem a formação de uma nova sequência de vetores que são combinações lineares de aproximações obtidas pelo método iterativo básico. O procedimento polinomial é uma das muitas abordagens para a aceleração da convergência de métodos iterativos básicos.

O procedimento polinomial geral é definido supondo apenas que o método básico seja consistente. De qualquer maneira, mais tarde, logo depois de considerar a Aceleração de Chebyshev, assumir-se-á que os métodos básicos também são simetrizáveis .

3-2 Método iterativo de Sobre-Relaxação Sucessiva (SOR)

Um terceiro método iterativo para resolver o sistema de equações lineares (2-1.1) está relacionado com o método iterativo de Gauss-Seidel. É o método de Sobre-Relaxação Sucessiva (SOR) que é definido por

$$u_i^{(m+1)} = \omega \left\{ - \sum_{j=1}^{i-1} \frac{a_{i,j}}{a_{i,i}} u_j^{(m+1)} - \sum_{j=i+1}^n \frac{a_{i,j}}{a_{i,i}} u_j^{(m)} + \frac{b_i}{a_{i,i}} \right\} + (1-\omega)u_i^{(m)},$$

$$1 \leq i \leq n, \quad m \geq 0. \quad (3-2.1)$$

Aqui ω é um número real conhecido como fator de relaxação. Quando $\omega = 1$, o método SOR se reduz ao método de Gauss-Seidel. Se $\omega > 1$ ou $\omega < 1$, tem-se sobre-relaxação ou sub-relaxação, respectivamente.

De acordo com a equação (3-2.1), o método SOR pode ser expresso na forma

$$\mathbf{u}^{(m+1)} = \omega(D^{-1}L\mathbf{u}^{(m+1)} + D^{-1}U\mathbf{u}^{(m)} + D^{-1}\mathbf{b}) + (1-\omega)\mathbf{u}^{(m)}, \quad (3-2.2)$$

reescrevendo (3-2.2)

$$D\mathbf{u}^{(m+1)} - \omega L\mathbf{u}^{(m+1)} = \omega U\mathbf{u}^{(m)} + \omega \mathbf{b} + (1-\omega)D^{-1}\mathbf{u}^{(m)}$$

$$(D - \omega L)\mathbf{u}^{(m+1)} = (\omega U + (1-\omega)D^{-1})\mathbf{u}^{(m)} + \omega \mathbf{b}.$$

Como $(D - \omega L)$ é uma matriz triangular inferior, $\det(D - \omega L) \neq 0$ e, portanto, inversível. Logo,

$$\mathbf{u}^{(m+1)} = (D - \omega L)^{-1}(\omega U + (1-\omega)D^{-1})\mathbf{u}^{(m)} + (D - \omega L)^{-1}\omega \mathbf{b}.$$

Esta equação pode ser reescrita na forma

$$\mathbf{u}^{(m+1)} = \mathcal{L}_\omega \mathbf{u}^{(m)} + \mathbf{k}_\omega, \quad m \geq 0, \quad (3-2.3)$$

3-3 Aceleração polinomial de métodos iterativos básicos

Suponha o método iterativo básico $\mathbf{u}^{(n+1)} = G\mathbf{u}^{(n)} + \mathbf{k}$ (2-1.2) consistente com o problema matricial $A\mathbf{u} = \mathbf{b}$ com A não singular. Considere a seqüência $\{\mathbf{w}^{(m)}\}$ de aproximações geradas pelas iterações do método básico (2-1.2), isto é, dado $\mathbf{w}^{(0)}$, a seqüência $\{\mathbf{w}^{(m)}\}$ é gerada por

$$\mathbf{w}^{(m)} = G\mathbf{w}^{(m-1)} + \mathbf{k}, m = 1, 2, \dots \quad (3-3.1)$$

De acordo com a equação (2-2.4), o vetor erro $\bar{\boldsymbol{\epsilon}}^{(m)} = \mathbf{w}^{(m)} - \bar{\mathbf{u}}$, associado com a m -ésima iteração de (3-3.1) satisfaz

$$\bar{\boldsymbol{\epsilon}}^{(m)} = G^m \bar{\boldsymbol{\epsilon}}^{(0)}. \quad (3-3.2)$$

Como forma de acelerar a convergência das aproximações $\mathbf{w}^{(m)}$, considere uma nova seqüência de vetores $\{\mathbf{u}^{(m)}\}$ determinado por uma combinação linear

$$\mathbf{u}^{(m)} = \sum_{i=0}^m v_{m,i} \mathbf{w}^{(i)}, \quad m = 0, 1, 2, \dots \quad (3-3.3)$$

A única restrição imposta aos números reais $v_{m,i}$, é

$$\sum_{i=0}^m v_{m,i} = 1, \quad m = 0, 1, 2, \dots \quad (3-3.4)$$

Essa condição é imposta para garantir que $\mathbf{u}^{(m)} = \bar{\mathbf{u}}$ para todo $m \geq 0$, sempre que o vetor inicial $\mathbf{w}^{(0)}$ for igual à solução $\bar{\mathbf{u}}$. Como um exemplo, escrevendo $v_{m,i} = 1/(m+1)$, $m \geq 0$, os vetores $\mathbf{u}^{(m)}$ representam a média aritmética dos vetores $\mathbf{w}^{(i)}$, $0 \leq i \leq m$.

Se $\boldsymbol{\epsilon}^{(m)} \equiv \mathbf{u}^{(m)} - \bar{\mathbf{u}}$ denota o vetor erro associado ao vetor $\mathbf{u}^{(m)}$ de (3-3.3) e (3-3.4) resulta

$$\begin{aligned}\boldsymbol{\varepsilon}^{(m)} &= \sum_{i=0}^m v_{m,i} \mathbf{w}^{(i)} - \bar{\mathbf{u}} = \sum_{i=0}^m v_{m,i} \mathbf{w}^{(i)} - \sum_{i=0}^m v_{m,i} \bar{\mathbf{u}} \\ &= \sum_{i=0}^m v_{m,i} (\mathbf{w}^{(i)} - \bar{\mathbf{u}}) = \sum_{i=0}^m v_{m,i} \bar{\boldsymbol{\varepsilon}}^{(i)}.\end{aligned}$$

Usando (3-3.2) pode-se expressar $\boldsymbol{\varepsilon}^{(m)}$ na forma

$$\boldsymbol{\varepsilon}^{(m)} = \sum_{i=0}^m v_{m,i} \bar{\boldsymbol{\varepsilon}}^{(i)} = \sum_{i=0}^m v_{m,i} G^i \bar{\boldsymbol{\varepsilon}}^{(0)} = \left(\sum_{i=0}^m v_{m,i} G^i \right) \bar{\boldsymbol{\varepsilon}}^{(0)}.$$

De (3-3.3) e (3-3.4) resulta $\bar{\boldsymbol{\varepsilon}}^{(0)} = \boldsymbol{\varepsilon}^{(0)}$. Assim, pode-se expressar $\boldsymbol{\varepsilon}^{(m)}$ na forma

$$\boldsymbol{\varepsilon}^{(m)} = Q_m(G) \boldsymbol{\varepsilon}^{(0)}, \quad (3-3.5)$$

em que $Q_m(G)$ é o polinômio matricial $Q_m(G) = v_{m,0}I + v_{m,1}G + \dots + v_{m,m}G^m$. Se $Q_m(x) = v_{m,0} + v_{m,1}x + \dots + v_{m,m}x^m$ é o polinômio algébrico associado, então a condição (3-3.4) requer que $Q_m(1) = 1$. Essa condição é a única restrição imposta para $Q_m(x)$.

Devido a forma (3-3.5) associada ao vetor erro, o procedimento de (3-3.1) e (3-3.3) é chamado de *Método de Aceleração Polinomial* aplicado ao método básico (3-3.1). Esse procedimento é chamado método semi-iterativo com respeito ao método iterativo (3-2.1) [Varga 62].

O alto custo de operações aritméticas envolvidas e a grande quantidade de dados armazenados na memória usando (3-3.3) para obter $\mathbf{u}^{(m)}$, implica na necessidade de buscar alternativas menos dispendiosas para a computação de $\mathbf{u}^{(m)}$. Mostrar-se-á que uma computação simples de $\mathbf{u}^{(m)}$ é possível sempre que o polinômio $Q_m(x)$ satisfizer a relação de recorrência

$$\begin{aligned}Q_0(x) &= 1, \\ Q_1(x) &= \gamma_1 x - \gamma_1 + 1, \\ Q_{m+1}(x) &= \rho_{m+1}(\gamma_{m+1}x + 1 - \gamma_{m+1})Q_m(x) + (1 - \rho_{m+1})Q_{m-1}(x), \quad m \geq 1\end{aligned} \quad (3-3.6)$$

com $\gamma_1, \rho_2, \gamma_2, \dots$ números reais. Note que $Q_m(x)$, definido por (3-3.6), satisfaz $Q_m(1) = 1$ para todo $m \geq 0$. Observe que o conjunto das seqüências polinomiais $\{Q_m(x)\}$ satisfazendo (3-3.6) é grande.

Teorema 3-3.1 Suponha que o método básico (3-3.1) seja consistente. Se a seqüência $\{Q_m(x)\}$ for dada por (3-3.6), então as aproximações $\mathbf{u}^{(m)}$ de (3-3.3) podem ser obtidas usando a relação de três termos, assim chamada por envolver aproximações fornecidas por três iterações consecutivas.

$$\mathbf{u}^{(1)} = \gamma_1 (G\mathbf{u}^{(0)} + \mathbf{k}) + (1 - \gamma_1)\mathbf{u}^{(0)},$$

$$\mathbf{u}^{(m+1)} = \rho_{m+1} \left\{ \gamma_{m+1} (G\mathbf{u}^{(m)} + \mathbf{k}) + (1 - \gamma_{m+1})\mathbf{u}^{(m)} \right\} + (1 - \rho_{m+1})\mathbf{u}^{(m-1)},$$

para $m \geq 1$ (3-3.7)

Reciprocamente, qualquer procedimento iterativo com aproximações $\mathbf{u}^{(m)}$ definidas por (3-3.7) é equivalente ao procedimento polinomial (3-3.1) e (3-3.3), com os polinômios dados por (3-3.6) [Young 81].

Definição 3-3.2 O raio espectral virtual da matriz $Q_m(G)$ é definido por

$$\bar{S}(Q_m(G)) \equiv \max_{m(G) \leq x \leq M(G)} |Q_m(x)| \quad (3-3.8)$$

em que, $m(G)$ e $M(G)$ denota, respectivamente, algebricamente o menor e o maior autovalor de G .

O método polinomial particular, obtido pela escolha da seqüência polinomial $\{Q_m(x)\}$ tal que $\bar{S}(Q_m(G))$, $m = 1, 2, \dots$ seja minimizado, é chamado *método de aceleração polinomial de Chebyshev*.

Na secção seguinte será explicado que, na verdade, o polinômio matricial $Q_m(x)$ que minimiza $\bar{S}(Q_m(G))$ pode ser definido em função do polinômio de Chebyshev.

3-4 Aceleração de Chebyshev

Primeiramente será mostrado que o polinômio matricial $Q_m(G)$ que minimiza $\bar{S}(Q_m(G))$ é única e pode ser definida em termos de polinômios de Chebyshev. Para qualquer inteiro não negativo m , o polinômio de Chebyshev de grau m em w pode ser definido por

$$\begin{aligned} T_m(w) &= \frac{1}{2} \left[(w + \sqrt{w^2 - 1})^m + (w - \sqrt{w^2 - 1})^m \right] \\ &= \frac{1}{2} \left[(w - \sqrt{w^2 - 1})^m + (w + \sqrt{w^2 - 1})^m \right] \\ &= \cosh(n \cosh^{-1} w), \quad w > 1 \\ &= \cos(n \cos^{-1} w), \quad -1 \leq w \leq 1. \end{aligned} \tag{3-4.1}$$

e satisfaz os três termos da relação de recorrência

$$T_0(w) = 1, \quad T_1(w) = w,$$

$$T_{m+1}(w) = 2wT_m(w) - T_{m-1}(w), \quad m \geq 1. \tag{3-4.2}$$

As propriedades fundamentais do polinômio de Chebyshev que serão usadas são enunciadas no seguinte teorema.

Teorema 3-4.1 Seja m um inteiro fixo e d qualquer real fixo tal que $d > 1$. Se

$$H_m(w) = T_m(w)/T_m(d)$$

em que $T_m(w)$ é o polinômio de Chebyshev (3-4.2), então

$$H_m(d) = 1 \quad \text{e} \quad \max_{-1 \leq w \leq 1} |H_m(w)| = \frac{1}{T_m(d)}.$$

Além disso, se $Q(w)$ for qualquer polinômio de grau menor ou igual a m tal que $Q(d)=1$ e

$$\max_{-1 \leq w \leq 1} |Q(w)| \leq \max_{-1 \leq w \leq 1} |H_m(w)|,$$

então

$$Q(w) = H_m(w)$$

[Young 81].

Retornando agora ao problema de minimizar $\bar{S}(Q_m(G))$, procura-se um polinômio $P_m(x)$ tal que $P_m(1) = 1$ e

$$\max_{m(G) \leq x \leq M(G)} |P_m(x)| \leq \max_{m(G) \leq x \leq M(G)} |Q_m(x)|, \quad (3-4.3)$$

em que $Q_m(G)$ é qualquer polinômio de grau menor ou igual a m satisfazendo $Q_m(1) = 1$. A existência de cada polinômio é assegurada pelo teorema 3-4.1. Seja

$$w(x) \equiv (2x - M(G) - m(G)) / (M(G) - m(G)) \quad (3-4.4)$$

uma transformação linear que mapeia o intervalo $m(G) \leq x \leq M(G)$ no intervalo $-1 \leq w \leq 1$ e $H_m(w) = T_m(w(x)) / T_m(w(1))$. Como o método básico (2-1.2) é simetrizável, $M(G) < 1$. Observe que $w(x)$ é uma função linear crescente em x . Logo, $w(1) > 1$. Se agora $P_m(x)$ for definido como

$$P_m(x) \equiv T_m \left(\frac{2x - M(G) - m(G)}{M(G) - m(G)} \right) / T_m \left(\frac{2 - M(G) - m(G)}{M(G) - m(G)} \right), \quad (3-4.5)$$

então

$$\max_{m(G) \leq x \leq M(G)} |P_m(x)| = \max_{-1 \leq w \leq 1} |H_m(w)|.$$

Pelo teorema 3-4.1 $P_m(x)$ definido por (3-4.5) é o único polinômio que satisfaz (3-4.3).

Teorema 3-4.2 Seja \mathcal{J}_m o conjunto de polinômios $\{Q_m(x)\}$ de grau menor ou igual a m que satisfaz $Q_m(1) = 1$. Então, o polinômio $P_m(x)$ de (3-4.5) é o único polinômio no conjunto \mathcal{J}_m que satisfaz

$$\max_{m(G) \leq x \leq M(G)} |P_m(x)| \leq \max_{m(G) \leq x \leq M(G)} |Q_m(x)|$$

para qualquer $Q_m(x) \in \mathcal{J}_m$ [Young 81].

Considere agora os aspectos computacionais e de convergência da aceleração polinomial, quando $P_m(x)$ for aplicado ao método iterativo básico (2-1.2). Da equação (3-4.2), pode-se mostrar que o polinômio $P_m(x)$ satisfaz a relação de recorrência

$$P_0(x) = 1, \quad P_1(x) = \gamma x - \gamma + 1,$$

$$P_{m+1}(x) = \rho_{m+1}(\gamma x + 1 - \gamma)P_m(x) + (1 - \rho_{m+1})P_{m-1}(x) \text{ para } m \geq 1, \quad (3-4.6)$$

em que

$$\gamma = 2/(2 - M(G) - m(G)), \quad (3-4.7)$$

$$\rho_{m+1} = 2w(1)T_m(w(1))/T_{m+1}(w(1)), \quad (3-4.8)$$

e $w(x)$ é definido por (3-4.4). Logo, pelo teorema 3-3.1 as aproximações do procedimento polinomial baseado em $P_m(x)$, podem ser obtidas usando a relação

$$\mathbf{u}^{(m+1)} = \rho_{m+1} \left\{ \gamma (G\mathbf{u}^{(m)} + \mathbf{k}) + (1 - \gamma)\mathbf{u}^{(m)} \right\} + (1 - \rho_{m+1})\mathbf{u}^{(m-1)}. \quad (3-4.9)$$

O método definido por (3-4.9) com γ e ρ_{m+1} dados por (3-4.7) e (3-4.8), respectivamente, será chamado procedimento de aceleração de Chebyshev.

Outra vez fazendo uso da relação de recorrência do polinômio de Chebyshev (3-4.2), os parâmetros ρ_{m+1} de (3-4.8) podem ser escritos na forma computacionalmente conveniente

$$\rho_1 = 1, \quad \rho_2 = \left(1 - \frac{1}{2}\sigma^2\right)^{-1},$$

$$\rho_{m+1} = \left(1 - \frac{1}{4}\sigma^2\rho_m\right)^{-1}, \quad m \geq 2, \quad (3-4.10)$$

em que

$$\sigma = 1/w(1) = (M(G) - m(G))/(2 - M(G) - m(G)). \quad (3-4.11)$$

Agora, a taxa de convergência do procedimento de Chebyshev será examinado. O raio espectral virtual de $P_m(G)$ é dado por

$$\bar{S}(P_m(G)) = [T_m(w(1))]^{-1} = [T_m(1/\sigma)]^{-1}.$$

Usando (3-4.1) e uma pequena arrumação algébrica, $T_m(1/\sigma)$ pode ser escrito na forma

$$T_m\left(\frac{1}{\sigma}\right) = \frac{1+r^m}{2r^{m/2}}, \quad (3-4.12)$$

em que

$$r = \left(1 - \sqrt{1 - \sigma^2}\right) / \left(1 + \sqrt{1 - \sigma^2}\right). \quad (3-4.13)$$

Assim, obtém-se

$$\bar{S}(P_m(G)) = 2r^{m/2} / (1+r^m).$$

Desta forma, a taxa de convergência para o método de Chebyshev pode ser escrita assim

$$\begin{aligned}
R_m(P_m(G)) &= -(1/2)\log(2r^{m/2}/(1+r^m)) \\
&= -\frac{1}{2}\log r - \frac{1}{m}\log\left(\frac{2}{1+r^m}\right).
\end{aligned}$$

A taxa assintótica de convergência pode ser expressa na forma

$$R(P_m(G)) = R_\infty(P_m(G)) = -\frac{1}{2}\log r.$$

3-5 Aceleração de Chebyshev adaptativa

Agora será estudado o processo de aceleração de Chebyshev, em que as estimativas m_E e M_E são usadas no lugar de $m(G)$ e $M(G)$, respectivamente. Usando essas estimativas, o polinômio de Chebyshev normalizado (3-4.5), será escrito como

$$P_{m,E}(x) \equiv T_m\left(\frac{2x - M_E - m_E}{M_E - m_E}\right) / T_m\left(\frac{2 - M_E - m_E}{M_E - m_E}\right) = \frac{T_m(w_E(x))}{T_m(w_E(1))},$$

em que

$$w_E(x) = (2x - M_E - m_E) / (M_E - m_E).$$

Se

$$\gamma = 2/(2 - M_E - m_E),$$

$$\sigma_E = (M_E - m_E) / (2 - M_E - m_E)$$

e

$$P_{m+1} = 2w_E(1)T_m(w_E(1))/T_{m+1}(w_E(1)),$$

resulta como antes, que $P_{m,E}(x)$ satisfaz a relação de recorrências (3-4.6), em que as estimativas m_E e M_E substituem $m(G)$ e $M(G)$, respectivamente [Young 81]. Assim, pelo teorema 3-3.1, as

$$\delta^{(m)} = (G - I)\epsilon^{(m)}. \quad (3-5.6)$$

Observe que, sendo $\mathbf{k} = (I - G)\bar{\mathbf{u}}$, (3-5.6) será obtida diretamente de (3-5.5). Além disso, sendo o método simetrizável, $(I - G)$ é não singular. Portanto,

$$\epsilon^{(m)} = (G - I)^{-1}\delta^{(m)}.$$

Para qualquer norma $\|\cdot\|$

$$\|\epsilon^{(m)}\| = \|(G - I)^{-1}\| \|\delta^{(m)}\|$$

e, em particular,

$$\|\epsilon^{(m)}\|_w \leq (1 - M(G))^{-1} \|\delta^{(m)}\|_w \quad (3-5.7)$$

Assim, um limite superior para $\|\epsilon^{(m)}\|_w$ pode ser obtido de $\|\delta^{(m)}\|_w$, contanto que $M(G)$ seja conhecido.

A relação (3-4.2), pode ser expressa em termos de $\delta^{(m)}$ como

$$\delta^{(m)} = P_{m,E}(G)\delta^{(0)}.$$

Assim, a norma- w de $\delta^{(m)}$ satisfaz

$$\|\delta^{(m)}\|_w \leq S(P_{m,E}(G))\|\delta^{(0)}\|_w \leq \bar{S}(P_{m,E}(G))\|\delta^{(0)}\|_w. \quad (3-5.8)$$

O procedimento de parada das iterações será baseado na desigualdade (3-5.7) e o procedimento de estimativa de parâmetros na desigualdade (3-5.8). Para o procedimento dado aqui, assume-se que a estimativa M_E para $M(G)$ poderá sofrer mudanças durante o processo iterativo. A seguinte notação será usada:

- $P_{p,E}(x)$ denota o polinômio de Chebyshev normalizado de grau p usando a mais recente estimativa M_E de $M(G)$;
- q denota a última iteração na qual a estimativa M_E foi usada;
- m denota a iteração.

Se a estimativa corrente M_E de $M(G)$ satisfizer $M_E < M(G)$, resulta de (3-4.8) que

$$\|\delta^{(m)}\|_w / \|\delta^{(q)}\|_w > 2r^{p/2} / (1 + r^p),$$

em que $p = m - q$ e r é dado por (3-4.13).

Se os parâmetros de aceleração não forem satisfatórios, uma nova estimativa deve ser obtida. Nas condições da hipótese (3-5.4), requer uma nova estimativa M'_E para $M(G)$. Para obter a nova estimativa M'_E , novamente a desigualdade (3-5.8) é usada. Como $M_E < M(G)$, (3-5.8) pode ser expresso na forma

$$\|\delta^{(m)}\|_w / \|\delta^{(q)}\|_w \leq \bar{S}(P_{p,E}(G)) = P_{p,E}(M(G)) = T_p(w_E(M(G))) / T_p(w_E(1)).$$

A nova estimativa M'_E para $M(G)$ será o maior real x que satisfaz a equação de Chebyshev

$$T_p(w_E(x)) / T_p(w_E(1)) = \|\delta^{(m)}\|_w / \|\delta^{(q)}\|_w. \quad (3-5.9)$$

O seguinte teorema resume as afirmações acima.

Teorema 3-5.1 Seja

$$B \equiv \|\delta^{(m)}\|_w / \|\delta^{(q)}\|_W$$

e

$$Q \equiv 1/T_p(w_E(1)) = 2r^{P/2}/(1+r^P),$$

em que r é dado por (3-4.13). Se as condições da hipótese (3-5.4) forem satisfeitas e se $B > Q$, então, a maior solução real M'_E para a equação de Chebyshev (3-5.9) é dada por

$$M'_E = \frac{1}{2} \left[M_E + m_E + \frac{1}{2} (M_E - m_E) (Y^2 + 1) / Y \right],$$

com Y um número real positivo dado por

$$Y = \left[(B/Q) + \sqrt{(B/Q)^2 - 1} \right]^{1/P}.$$

Além disso, M'_E satisfaz $M_E < M'_E \leq M(G)$ [Young 81].

Para o teste de parada, [Young 81] recomenda

$$\frac{1}{1 - M'_E} \frac{\|\delta^{(m)}\|_\beta}{\|\mathbf{u}^{(m+1)}\|_\eta} \leq \zeta,$$

em que ζ é o grau de aproximação desejado, M'_E é a melhor estimativa de $M(G)$, e $\|\cdot\|_\beta$ e $\|\cdot\|_\eta$ são normas. Na implementação foi utilizado outro teste de parada que será descrito no capítulo 5.

3-6 Procedimento computacional

Um procedimento adaptativo para a aceleração de Chebyshev baseado nas discussões acima pode ser descrito pelo **Algoritmo 3-5.1** [Young 81]. O procedimento estima parâmetros adaptativos utilizando a norma-2 do pseudo-resíduo δ . As variáveis e constantes usadas são:

- n número de iterações executadas,
- p é o grau do polinômio de Chebyshev usado,
- s é o número de diferentes estimativas M_E usadas para $M(G)$,
- τ_s Limite superior para s -ésima estimativa M_E para $M(G)$, e
- P^* Cada polinômio de Chebyshev deve ter grau pelo menos maior que p^* antes de mudar a estimativa M_E .

As entradas necessárias para o algoritmo são:

- ζ a tolerância desejada,
- m_E estimativa para o menor autovalor de G ; m_E deve satisfazer a condição $m_E < m(G)$,
- M_E estimativa para o maior autovalor de G ; M_E deve satisfazer a condição $m_E \leq M_E < 1$,
- F parâmetro estratégico, cujo valor assumido é 0.7 se F estiver fora do intervalo $[.1, .9]$,
- u_Γ aproximação inicial do vetor solução,
- d parâmetro estratégico, cujo valor assumido é 0.1 se d estiver fora do intervalo $[.001, .8]$,
- IE Variável de controle que transmite informações a respeito da estimativa dos autovalores:
 - $IE = -1$ implica que a estimativa inicial para o menor e o maior autovalor deve ser calculada pela subrotina,
 - $IE = 0$ implica que as estimativas para o menor e o maior autovalor são dados de entrada e a subrotina poderá atualizar seu valor,
- $ILIM$ o limite do número de iterações para obter a estimativa inicial do menor autovalor. ($ILIM$ deve ser pelo menos 8).

Algoritmo 3-5.1 - Procedimento adaptativo para a aceleração de Chebyshev usando a norma-2.

Entrada: $(\zeta, M_E, m_E, F, \mathbf{u}_\Gamma, d, IE, ILIM)$ $\{M_E \text{ deve estar no intervalo } m_E < M_E < 1\}$.

$n := 0; p := -1; M'_E := M_E; \mathbf{u}_\phi := \mathbf{0}; s := 0;$

$\tilde{m}_E := m_E; DELNP := 1.0; R := 1.0;$

if $IE \geq 0$, **then continue**

else

Begin

$M'_E := \tau_1; m_E := \tau_1;$

End

Próxima iteração

$n := n + 1; p := p + 1; DELNO := DELNP;$

if $p := 0$, **then** {inicializa o novo polinômio}

Begin

$s := s + 1; T := 0; \text{if } M'_E > \tau_s, \text{ then } M'_E := \tau_s; \text{ else continue;}$

$M_E := M'_E, \rho := 1.0; \gamma = 2 / (2 - M_E - m_E);$

$\sigma_E := (M_E - m_E) / (2 - M_E - m_E);$

$r := (1 - \sqrt{1 - \sigma_E^2}) / (1 + \sqrt{1 - \sigma_E^2});$

$p^* := [\log(d) / \log(r)];$

if $p^* < 6$, **then** $p^* = 6$; **else continue;**

if $IE \geq 0$, **then continue;** **else** $p^* = 8$;

End

else {Continua geração do polinômio}

Begin

if $p = 1$, **then** $\rho := 1 / (1 - \frac{1}{2} \sigma_E^2)$; **else** $\rho := 1 / (1 - \frac{1}{4} \sigma_E^2 \rho)$;

End

Nova Iteração

$$\delta := G\mathbf{u}_r + \mathbf{k} - \mathbf{u}_r; \text{DELNP} := \|\delta\|_2; \text{DELNE} := \|\delta\|_p;$$

$$\mathbf{u} := \rho(\lambda\delta + \mathbf{u}_r) + (1 - \rho)\mathbf{u}_\phi; \text{YUN} := \|\mathbf{u}\|_\eta;$$

$$\mathbf{u}_\phi := \mathbf{u}_r; \mathbf{u}_r := \mathbf{u};$$

Cálculo da Nova Estimativa M'_E :

$$\text{RO} := R; R := \text{DELNP}/\text{DELNO};$$

if $p \leq 2$, then

Begin

 if $p = 0$, then $\text{DELNPI} := \text{DELNP}$; else continue;

 Go to Próxima Iteração

End

 else

Begin

$$Q := 2r^{p/2}/(1 + r^p); B := \text{DELNP}/\text{DELNPI};$$

 if $B \geq 1.0$, then

Begin

$$T = T + 1;$$

 if $IE > 0$, then Go to Próxima Iteração

 else Go to Teste de esc. de Parâmetros;

End

 else

 if $B > Q$, then

Begin

$$X := \left[\left(\frac{1+r^p}{2} \right) \left(B + \sqrt{B^2 - Q^2} \right) \right]^{1/p};$$

$$M'_E = \frac{1}{2} \left[M_E + m_E + \left(\frac{2 - M_E - m_E}{1+r} \right) \left(\frac{X^2 + r}{X} \right) \right];$$

End

 else $M'_E = M_E$;

End

Teste de Escolha de Parâmetros

if p é par , **then continue**; **else Go to** Próxima Iteração;

if $p \geq p^*$, **then**

Begin

if $IE \geq 0$, **then continue**; **else Go to** Atualize m_E ;

if $B > (Q)^F$, **then continue**; **else Go to** Próxima Iteração;

if $T = 0$, **then**

Begin

$p := -1$; **Go to** Próxima Iteração;

End

else continue;

if $R > 1$ **and** $T \geq 5$, **then Go to** Atualize m_E ;

else Go to Próxima Iteração;

End

else Go to Próxima Iteração;

Atualize m_E

$m_E' := 0.0$; $m_E'' := 0.0$; $DR = 0$;

if $R > 1.0$, **then**

Begin

$U := \left[1/\sqrt{r} \right] \left[(1+r^p)/(1+r^{p-1}) \right] R$;

$m_E'' := \frac{1}{2} \left[M_E + m_E - \frac{1}{2} (M_E - m_E) ((U^2 + 1)/U) \right]$;

if $|RO - R| < 0.1$, **then** $DR = 1$; **else continue**

End

else continue

if $B > 1.0$, **then**

Begin

$Y := \left[(B/Q) + \sqrt{(B/Q)^2 - 1} \right]^{1/p}$;

$$m'_E := \frac{1}{2} \left[M_E + m_E - \frac{1}{2} (M_E - m_E) ((Y^2 + 1) / Y) \right];$$

End

else

Begin

if $IE = 0$, **then Go to** Próxima Iteração; **else continue**

End

if $IE = 0$, **then**

Begin

if $DR = 0$, **then Go to** Próxima Iteração;

else

Begin

$$m_E := \min[1.1m_E, 1.1m''_E, m_E];$$

$$M_E := 0.1; p := -1;$$

End

End

else

Begin

if $DR = 0$ **and** $n < ILIM$, **then Go to** Próxima Iteração;

else

Begin

$$m_E := \min[1.1 m'_E, 1.1 m''_E, m_E, -1.0];$$

$$M'_E := 0.1; IE := 0; p := -1; s := 0;$$

End

End

Go to Próxima Iteração;

Capítulo IV

Método dos Gradientes Conjugados

4-1 Introdução

O método dos Gradientes Conjugados empregado para solução de sistemas de equações algébricas lineares foi apresentado nos início da década de 50 por [Hestenes 56] e [Stiefel 52]. Embora seja um método iterativo, converge para a solução verdadeira, na ausência de erros de arredondamento, em um número finito conhecido de iterações [Young 81].

4-2 Método dos Gradientes Conjugados

O método dos gradientes conjugados (método GC) para resolver o sistema linear $A\mathbf{u} = \mathbf{b}$, dado por (2-1.1) será descrito. Supõe-se que A seja uma matriz de ordem n , simétrica positiva definida.

O método GC pode ser considerado como uma modificação do método "steepest descent". Para deduzir o método "steepest descent", considere a forma quadrática:

$$Q(\mathbf{u}) = \frac{1}{2} \mathbf{u}^T A \mathbf{u} - \mathbf{b}^T \mathbf{u}. \quad (4-2.1)$$

Pode-se mostrar que o problema de minimizar (4-2.1) é equivalente a resolver (2-1.1). Supondo que $\bar{\mathbf{u}}$ seja a solução de (2-1.1) e considerando $\mathbf{u} = \bar{\mathbf{u}} + \mathbf{v}$, tem-se

$$\begin{aligned} Q(\bar{\mathbf{u}} + \mathbf{v}) &= \frac{1}{2} (\bar{\mathbf{u}} + \mathbf{v})^T A (\bar{\mathbf{u}} + \mathbf{v}) - \mathbf{b}^T (\bar{\mathbf{u}} + \mathbf{v}) \\ &= \frac{1}{2} \bar{\mathbf{u}}^T A \bar{\mathbf{u}} + \frac{1}{2} \mathbf{v}^T A \bar{\mathbf{u}} + \frac{1}{2} \mathbf{v}^T A \mathbf{v} - \mathbf{b}^T \bar{\mathbf{u}} - \mathbf{b}^T \mathbf{v} \\ &= Q(\bar{\mathbf{u}}) + \frac{1}{2} \mathbf{v}^T A \mathbf{v} + \mathbf{v}^T (A \bar{\mathbf{u}} - \mathbf{b}) = Q(\bar{\mathbf{u}}) + \frac{1}{2} \mathbf{v}^T A \mathbf{v} \end{aligned} \quad (4-2.2)$$

Se A for simétrica positiva definida, $\mathbf{v}^T A \mathbf{v} > 0$, a menos que $\mathbf{v} = \mathbf{0}$. Portanto, se $\mathbf{w} \neq \bar{\mathbf{u}}$, então $Q(\mathbf{w}) > Q(\bar{\mathbf{u}})$. Assim, $\bar{\mathbf{u}}$ é o único mínimo absoluto de $Q(\mathbf{u})$.

No método de "steepest descent", obtém-se uma aproximação $\mathbf{u}^{(m+1)}$ a partir de $\mathbf{u}^{(m)}$ na forma

$$\mathbf{u}^{(m+1)} = \mathbf{u}^{(m)} + \lambda_m \mathbf{p}^{(m)}, \quad (4-2.3)$$

em que, $-\mathbf{p}^{(m)}$ é o vetor gradiente de Q no ponto $\mathbf{u}^{(m)}$ e λ_m é uma constante. O vetor $\mathbf{p}^{(m)}$ é definido por

$$-\mathbf{p}^{(m)} = \begin{pmatrix} \frac{\partial Q(\mathbf{u}^{(m)})}{\partial u_1} \\ \frac{\partial Q(\mathbf{u}^{(m)})}{\partial u_2} \\ \vdots \\ \frac{\partial Q(\mathbf{u}^{(m)})}{\partial u_{n-1}} \\ \frac{\partial Q(\mathbf{u}^{(m)})}{\partial u_n} \end{pmatrix} = \mathbf{A}\mathbf{u}^{(m)} - \mathbf{b} = \mathbf{r}^{(m)}. \quad (4-2.4)$$

Escolhe-se λ_m para minimizar $Q(\mathbf{u}^{(m)} + \lambda_m \mathbf{p}^{(m)})$. Derivando $Q(\mathbf{u}^{(m)} + \lambda_m \mathbf{p}^{(m)})$ em relação a λ , levando em conta que $\mathbf{b}^T \mathbf{p} = \mathbf{p}^T \mathbf{b}$ e $\mathbf{u}^T \mathbf{A} \mathbf{p} = \mathbf{p}^T \mathbf{A} \mathbf{u}$, obtém-se

$$\lambda_m = -\frac{(\mathbf{p}^{(m)})^T \mathbf{r}^{(m)}}{(\mathbf{p}^{(m)})^T \mathbf{A} \mathbf{p}^{(m)}} = \frac{(\mathbf{r}^{(m)})^T \mathbf{r}^{(m)}}{(\mathbf{r}^{(m)})^T \mathbf{A} \mathbf{r}^{(m)}}. \quad (4-2.5)$$

Se $-\mathbf{p}^{(m)}$ for o gradiente de Q , $\mathbf{r}^{(m)} = \mathbf{p}^{(m)}$.

Pode ser mostrado que

$$Q(\mathbf{u}^{(m+1)}) - Q(\mathbf{u}^{(m)}) = -\frac{((\mathbf{r}^{(m)})^T \mathbf{r}^{(m)})^2}{2((\mathbf{r}^{(m)})^T \mathbf{A} \mathbf{r}^{(m)})}. \quad (4-2.6)$$

Assim, a seqüência $Q(\mathbf{u}^{(0)})$, $Q(\mathbf{u}^{(1)})$, . . . é uma seqüência decrescente limitada por $Q(\bar{\mathbf{u}})$.

Desta forma, a seqüência $Q(\mathbf{u}^{(0)})$, $Q(\mathbf{u}^{(1)})$, . . . converge.

Um modificador para o método de "steepest descent" é apresentado na ausência de erros, convergindo em não mais que n iterações, em que n é a ordem da matriz A . Esse método é conhecido como método dos gradientes conjugados dado por

$\mathbf{u}^{(0)}$ é arbitrário,

$$\mathbf{u}^{(m+1)} = \mathbf{u}^{(m)} + \lambda_m \mathbf{p}^{(m)} \quad m = 0, 1, 2, \dots,$$

$$\mathbf{p}^{(m)} = \begin{cases} \mathbf{r}^{(m)}, & m = 0, \\ \mathbf{r}^{(m)} + \alpha_m \mathbf{p}^{(m-1)}, & m = 1, 2, \dots, \end{cases} \quad (4-2.7)$$

$$\alpha_m = -(\mathbf{r}^{(m)})^T A \mathbf{p}^{(m-1)} / (\mathbf{p}^{(m-1)})^T A \mathbf{p}^{(m-1)}, \quad m = 1, 2, \dots,$$

$$\mathbf{r}^{(m)} = \mathbf{b} - A \mathbf{u}^{(m)}, \quad m = 0, 1, 2, \dots,$$

$$\lambda_m = (\mathbf{p}^{(m)})^T \mathbf{r}^{(m)} / (\mathbf{p}^{(m)})^T A \mathbf{p}^{(m)}, \quad m = 0, 1, 2, \dots,$$

Assim, os resíduos $\mathbf{r}^{(0)}, \mathbf{r}^{(1)}, \dots$ e os vetores direção $\mathbf{p}^{(0)}, \mathbf{p}^{(1)}, \dots$ gerados por (4-2.7) satisfazem as relações

$$(\mathbf{r}^{(i)})^T \mathbf{r}^{(j)} = 0 \quad \text{para } i \neq j,$$

$$(\mathbf{p}^{(i)})^T A \mathbf{p}^{(j)} = 0 \quad \text{para } i \neq j, \quad (4-2.8)$$

$$(\mathbf{r}^{(i)})^T A \mathbf{p}^{(j)} = 0 \quad \text{para } i \neq j \text{ e } i \neq j + 1.$$

Os vetores resíduos $\mathbf{r}^{(0)}, \mathbf{r}^{(1)}, \dots$ são mutuamente ortogonais e os vetores direção $\mathbf{p}^{(0)}, \mathbf{p}^{(1)}, \dots$ são mutuamente A -conjugados. Da primeira relação em (4-2.8), resulta que $\mathbf{r}^{(s)} = 0$ para algum $s \leq n$. Dessa forma, o método dos gradientes conjugados, converge na ausência de erros de arredondamento, em não mais que n iterações.

4-3 Método dos Gradientes Conjugados Acelerado

O método dos gradientes conjugados pode ser modificado, a fim de se adotar o procedimento de aceleração polinomial aplicado aos métodos iterativos básicos.

Considere o método iterativo básico dado pela equação (2-1.2), $\mathbf{u}^{(m+1)} = G\mathbf{u}^{(m)} + \mathbf{k}$, em que para alguma matriz Q não singular,

$$G = I - Q^{-1}A \quad \text{e} \quad \mathbf{k} = Q^{-1}\mathbf{b}.$$

Assumindo que o método seja simetrizável, existe uma matriz de simetriação W não singular, tal que, $W(I - G)W^{-1}$ é SPD. Se A for SPD, a discussão inclui os métodos de Jacobi bem como qualquer outro método em que a matriz Q seja SPD. A discussão será restrita ao método de Jacobi em que a matriz W tem a forma $D^{1/2}$, em que D é uma matriz diagonal cujos elementos da diagonal principal são os elementos da diagonal principal da matriz A , pois o método de Jacobi é simetrizável tomando a matriz $W = D^{1/2}$ [Young 81].

De maneira análoga à obtenção da equação (3-4.9), é possível obter o procedimento de aceleração do método dos gradientes conjugados na forma

$$\mathbf{u}^{(m+1)} = \rho_{m+1} \left\{ \gamma_{m+1} \delta^{(m)} + \mathbf{u}^{(m)} \right\} + (1 - \rho_{m+1}) \mathbf{u}^{(m-1)}, \quad (4-3.1)$$

em que

$$\delta^{(m)} = G\mathbf{u}^{(m)} + \mathbf{k} - \mathbf{u}^{(m)},$$

$$\gamma_{m+1} = \left[-\frac{(W\delta^{(m)})^T W G \delta^{(m)}}{(W\delta^{(m)})^T W \delta^{(m)}} \right]^{-1}, \quad \rho_1 = 1,$$

$$\rho_{m+1} = \left[-\frac{\gamma_{m+1}}{\gamma_m} \frac{(W\delta^{(m)})^T W \delta^{(m)}}{(W\delta^{(m-1)})^T W \delta^{(m-1)}} \frac{1}{\rho_m} \right]^{-1}, \quad \text{se } m \geq 1 \quad (4-3.2)$$

Pára-se o processo iterativo e aceita-se $\mathbf{u}^{(m)}$ como uma aproximação satisfatória da solução verdadeira $\bar{\mathbf{u}}$ sempre que $\mathbf{u}^{(m)}$ satisfaça a desigualdade,

$$\frac{1}{1 - M_E^{(m)}} \frac{\|\delta^{(m)}\|_w}{\|\mathbf{u}^{(m)}\|_w} \leq \xi,$$

em que ξ é o grau de tolerância desejado, $M_E^{(m)}$ é uma estimativa para $M(G)$, o maior autovalor de G [Young 81].

Para o processo de aceleração de Chebyshev adaptativo descrito no capítulo anterior, a última estimativa para $M(G)$, obtida pelo processo adaptativo, é usada para o teste de parada. Para o caso da aceleração do método dos gradientes conjugados, não são necessárias estimativas para $M(G)$ no decorrer do processo iterativo. Entretanto, pode-se obter uma estimativa para $M(G)$ usando

$$M_E^{(m)} = M(T_m),$$

onde T_m é a matriz tridiagonal dada por

$$T_m = \begin{bmatrix} 1 - \gamma_1^{-1} & \frac{\rho_2 - 1}{\rho_2 \gamma_2} & & & & & 0 \\ \gamma_1^{-1} & 1 - \gamma_2^{-1} & \frac{\rho_3 - 1}{\rho_3 \gamma_3} & & & & \\ & (\rho_2 \gamma_2)^{-1} & 1 - \gamma_3^{-1} & & & & \\ & & & \ddots & & & \\ & & & & \ddots & & \\ 0 & & & & & (\rho_{m-1} \gamma_{m-1})^{-1} & \frac{\rho_m - 1}{\rho_m \gamma_m} \\ & & & & & & 1 - \gamma_m^{-1} \end{bmatrix}.$$

Para computar $M(T_m)$, usa-se o método da bissecção, aplicado à função $\Delta_m(\lambda)$, definida por

$$\Delta_m(\lambda) \equiv \det(T_m - \lambda I).$$

A avaliação de $\Delta_m(\lambda)$ para um λ dado é obtida usando a relação de recorrência

$$\Delta_{m+1}(\lambda) = \left[-\frac{1-\gamma_{m+1}}{\gamma_{m+1}} - \gamma \right] \Delta_m(\lambda) + \frac{1-\rho_{m+1}}{\rho_{m+1}\gamma_{m+1}\rho_m\gamma_m} \Delta_{m-1}(\lambda).$$

Desde que $\Delta_{m-1}(\lambda)$ e $\Delta_m(\lambda)$ possuam zeros comuns e que todos os zeros de $\Delta_m(\lambda)$ sejam menores que a unidade, existe precisamente um zero de $\Delta_m(\lambda)$ entre $M(T_{m-1})$ e 1 [Young 81].

Assim,

$$\Delta_m(M(T_{m-1}))\Delta_m(1) < 0.$$

Essa condição é suficiente para garantir a convergência.

Na prática, entretanto, pode ocorrer que o valor computado para $M(T_{m-1})$ não seja correto, podendo ocorrer $\Delta_m(M(T_{m-1}))\Delta_m(1) > 0$. Neste caso, usa-se $M(T_m) = M(T_{m-1})$. Isso será suficiente para ser usado como teste de parada.

Na implementação foi utilizado outro teste de parada que será descrito no capítulo 5.

4-4 Procedimento computacional

Baseado nas discussões acima, um programa informal é apresentado pelo **Algoritmo 4-4.1** [Young 81]. Os dados necessários são:

ξ grau de aproximação desejado

\mathbf{u} aproximação inicial para o vetor $\mathbf{u}^{(0)}$

Em geral, durante a iteração, \mathbf{u} representa $\mathbf{u}^{(m)}$, \mathbf{u}_ϕ representa $\mathbf{u}^{(m-1)}$ e \mathbf{un} representa $\mathbf{u}^{(m+1)}$; analogamente, γ e γ_ϕ , δ , δ_ϕ e δ_N . Também, θ representa $G\delta^{(m)}$, ψ representa $WG\delta^{(m)}$, e τ representa $W\delta^{(m)}$.

Algoritmo 4-4.1

Entrada : (ξ, \mathbf{u})

Condições iniciais:

$n := 0; \gamma := 1.0; \mathbf{u}_\phi := \mathbf{0}$ (vetor) ; $\delta_\phi := \mathbf{0}$ (vetor);

$\delta := G\mathbf{u} + \mathbf{k} - \mathbf{u};$

$\boldsymbol{\tau} := W\delta; DELNP :=; \boldsymbol{\tau}^T \boldsymbol{\tau}$

Proxima iteração:

$n := n + 1; \gamma_\phi := \gamma;$

Calculo da nova iteração:

$\boldsymbol{\theta} := G\delta;$

$\boldsymbol{\psi} := W\boldsymbol{\theta}; DEN := \boldsymbol{\tau}^T \boldsymbol{\psi}; \gamma := \left[1 - \frac{DEN}{DELNP} \right]^{-1};$

if $n = 1$, then $\rho := 1.0$; else $\rho := \left[1 - \left(\frac{\gamma}{\gamma_\phi} \right) \left(\frac{1}{\rho} \right) \left(\frac{DELNP}{DELNO} \right) \right]^{-1};$

$\mathbf{un} := \rho(\gamma\delta + \mathbf{u}) + (1-\rho)\mathbf{u}_\phi$; $YUN := \|\mathbf{un}\|_\eta$;

$DELNO := DELNP;$

$\delta\mathbf{n} := \rho[\gamma\boldsymbol{\theta} + (1-\gamma)\delta] + (1-\rho)\delta_\phi$; $DELNE := \|\delta\mathbf{n}\|_\beta$;

$\boldsymbol{\tau} := W\delta\mathbf{n}; DELNP := \boldsymbol{\tau}^T \boldsymbol{\tau}$

$\mathbf{u}_\phi := \mathbf{u}; \delta_\phi := \delta;$

$\mathbf{u} := \mathbf{u}\mathbf{n}; \delta := \delta\mathbf{n};$

Teste de parada:

Compute M_E (se necessário) para a matriz T_m usando a bisseção.

if $\frac{DELNE}{YUN} \leq \xi(1 - M_E)$, then imprima a saída e stop

else continue

go to proxima iteração.

Capítulo V

Resultados e discussões

5-1 Introdução

Neste capítulo serão apresentadas as descrições das matrizes dos coeficientes dos sistemas usados em teste, um plano de testes, os resultados dos testes realizados e uma avaliação comparativa dos métodos. Nas análises tomar-se-ão como base o número de iterações, a ordem da matriz, o raio espectral da matriz de iteração, a taxa de convergência, a tolerância requerida, o tempo de execução e a qualidade da solução encontrada.

5-2 Descrição das matrizes

No sistema $Ax = b$ usado nos testes, a matriz A dos coeficientes é do tipo banda simétrica ou não-simétrica. A simétrica tem bandas superior e inferior de largura 1 (exemplo 5.1). A não-simétrica tem banda superior de largura 2 e banda inferior de largura 1 (exemplo 5.2). As bandas superiores e inferiores são compostas de elementos “-1” e os elementos da diagonal principal são escolhidos para que torne a matriz A diagonal dominante ou estritamente diagonal dominante. O vetor independente b é escolhido de modo que todos os componentes do vetor

solução sejam iguais e conhecidos. Os testes foram realizados com matrizes de ordens 50, 100, 300 e 500.

5-2.1 Matriz simétrica

Em cada linha da matriz simétrica, o elemento da diagonal principal é escolhido para que a matriz A seja diagonal dominante ou estritamente diagonal dominante.

Exemplo 5.1:

$$A = \begin{bmatrix} d_{1,1} & -1 & 0 & 0 & 0 & 0 & \cdot & \cdot & 0 & 0 \\ -1 & d_{2,2} & -1 & 0 & 0 & 0 & \cdot & \cdot & 0 & 0 \\ 0 & -1 & d_{3,3} & -1 & 0 & 0 & \cdot & \cdot & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & \cdot & 0 & -1 & d_{n-2,n-2} & -1 & 0 \\ 0 & 0 & \cdot & \cdot & \cdot & 0 & 0 & -1 & d_{n-1,n-1} & -1 \\ 0 & 0 & \cdot & \cdot & \cdot & 0 & 0 & 0 & -1 & d_{n,n} \end{bmatrix}, \quad (5-2.1)$$

em que $d_{i,i}$, $1 \leq i \leq n$, é escolhido de forma que torne a matriz A diagonal dominante ou estritamente diagonal dominante.

5-2.2 Matriz não-simétrica

Em cada linha da matriz não-simétrica, o elemento da diagonal principal é escolhido para que a matriz A seja diagonal dominante ou estritamente diagonal dominante.

Exemplo 5.2:

$$A = \begin{bmatrix} d_{1,1} & -1 & -1 & 0 & 0 & 0 & \cdot & \cdot & 0 & 0 \\ -1 & d_{2,2} & -1 & -1 & 0 & 0 & \cdot & \cdot & 0 & 0 \\ 0 & -1 & d_{3,3} & -1 & -1 & 0 & \cdot & \cdot & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & \cdot & 0 & -1 & d_{n-2,n-2} & -1 & -1 \\ 0 & 0 & \cdot & \cdot & \cdot & 0 & 0 & -1 & d_{n-1,n-1} & -1 \\ 0 & 0 & \cdot & \cdot & \cdot & 0 & 0 & 0 & -1 & d_{n,n} \end{bmatrix}, \quad (5-2.2)$$

em que $d_{i,i}$, $1 \leq i \leq n$, é escolhido de forma que torne a matriz A diagonal dominante ou estritamente diagonal dominante.

A escolha das matrizes com essas estruturas particulares se deve ao fato de todos os métodos serem aplicáveis a essas matrizes. Só é possível fazer algumas comparações práticas de todos os métodos usando matrizes dos coeficientes que satisfaçam as condições de convergência.

5-3 Plano de testes

Os testes de execução deverão servir para

- Estudar a variação do número de iterações com o raio espectral da matriz de iteração,
- Estudar a variação do número de iterações com o número de equações,
- Estudar a variação do tempo de execução com o raio espectral,
- Comparar os métodos tendo como base o estudo acima para que um método possa ser usado.

5-4 Observações sobre implementação

Na implementação dos métodos estudados decidiu-se:

1. privilegiar o processamento de matrizes por colunas;
2. minimizar o erro cometido em somatórios utilizando o algoritmo de Kahan;
3. adotar critérios de interrupção das iterações que fossem econômicos e ao mesmo tempo eficientes, evitando que ocorram interrupções prematuras ou errôneas;
4. permitir a escolha da norma necessária nos critérios de interrupção das iterações;
5. somente executar operações aritméticas com elementos não nulos da matriz dos coeficientes;
6. não adotar método algum para economizar espaço de memória para armazenar a matriz dos coeficientes e nem a matriz de iteração; essas matrizes são inteiramente armazenadas; e
7. utilizar, sempre que possível, uma estimativa do raio espectral da matriz de iteração.

5-4.1 Processamento de matrizes por colunas

Na implementação dos algoritmos foi utilizada a linguagem de programação Fortran. Nas operações entre matrizes e vetores optou-se pelo processamento por coluna sempre que possível, já que em Fortran uma matriz é armazenada por coluna; isso significa que elementos consecutivos de uma coluna ocupam posições consecutivas de memória e os elementos consecutivos de uma linha ficam a uma distância igual a um múltiplo do número de linhas que for declarado para o conjunto que armazena a matriz. Num sistema de computação que usa paginação os elementos $a_{i,j}$ e $a_{i+1,j}$ podem estar em páginas diferentes.

5-4.2 Algoritmo de Kahan

Para minimizar o efeito dos dígitos perdidos numa seqüência de adições devido ao ajuste de potência da parcela menor, em todos os somatórios foi usado o algoritmo de Kahan, que é eficiente e melhora significativamente o grau de exatidão de um somatório. Esse algoritmo é apresentado a seguir.

Algoritmo de Kahan

```
S = x1
C = 0
for j = 2 until n
    Y = xj - C
    T = S + Y
    C = (T - S) - Y
    S = T
endfor
```

A variável C captura os dígitos transbordados no ajuste da potência de Y na adição $S + Y$ e procura agregar na parcela seguinte ($Y = x_j - C$).

5-4.3 Critérios de interrupção das iterações

Para interromper as iterações são usados os seguintes critérios:

1. Inicialmente testa a variação absoluta: $\|\mathbf{u}^{(m+1)} - \mathbf{u}^{(m)}\| < \epsilon_1$.

Se $\epsilon_1 = 10^{-p}$ e $\|\mathbf{u}^{(m+1)} - \mathbf{u}^{(m)}\| < \epsilon_1$, então todos os componentes de $\mathbf{u}^{(m+1)}$ e $\mathbf{u}^{(m)}$ têm os primeiros p dígitos coincidentes.

2. Teste de variação relativa:
$$\frac{\|\mathbf{u}^{(m+1)} - \mathbf{u}^{(m)}\|}{\|\mathbf{u}^{(m)}\|} \tag{5-4.3.1}$$

Quando $\|\mathbf{u}^{(m+1)}\|$ e $\|\mathbf{u}^{(m)}\|$ forem pequenos os seus componentes serão pequenos e $\|\mathbf{u}^{(m+1)} - \mathbf{u}^{(m)}\|$ também será pequeno, então o Teste1 pode interromper prematuramente as iterações.

$$\frac{\|\mathbf{u}^{(m+1)} - \mathbf{u}^{(m)}\|}{\|\mathbf{u}^{(m)}\|} < \varepsilon_2$$

3. Se a solução $\bar{\mathbf{u}} = \mathbf{0}$ ou muito próximo de $\mathbf{0}$, então $\|\mathbf{u}^{(m)}\|$ será muito pequeno e pode causar overflow. Só aplicar o Teste2 se $1 + \|\mathbf{u}^{(m)}\| > 1$, ou seja, se $\|\mathbf{u}^{(m)}\|$ for maior que o ε da máquina.

4. Teste do resíduo

Pode haver uma divergência lenta. Neste caso $\|\mathbf{u}^{(m+1)} - \mathbf{u}^{(m)}\|$ será pequeno, isto é, a variação na solução de uma iteração para outra será pequena. Se $\|\mathbf{u}^{(m)}\|$ for finito, (5-4.3.1) será pequeno e o Teste3 é automaticamente satisfeito. Com uma falsa convergência. Para evitar que isso ocorra é necessário ainda fazer um teste do resíduo:

$$\|\mathbf{b} - A\mathbf{u}^{(m+1)}\| < \varepsilon_3$$

em que $\varepsilon_1 = \varepsilon_3$ e $\varepsilon_2 = \sqrt{\varepsilon_1}$.

Baseado nas condições acima a implementação foi construída da seguinte forma:

```

indicador de parada falso
if Teste1 then
    if Teste3 then
        if Teste2 then
            if Teste4 then
                indicador de parada verdadeiro
            endif
        endif
    endif
else
    if Teste4 then
        indicador de parada verdadeiro
    endif
endif

```

Se a convergência, de acordo com critérios acima, não for alcançado em um número pré-estabelecido de iterações, a execução do programa é interrompida.

5-4.4 Escolha da norma a ser usada

Observando-se o custo operacional de computação das normas Euclidiana, um e infinita, a norma infinita seria a escolhida. Testando os métodos com essas normas, foi observado que a norma um apresentou melhor resultado para a solução procurada. Comparando a norma um

e a norma infinita, obtém-se uma explicação. Se $\delta = \mathbf{u}^{(m+1)} - \mathbf{u}^{(m)}$, então $\sum_{i=1}^n |\delta_i| \geq \max_i |\delta_i|$ e

$$\sum_{i=1}^n |u_i^{(m)}| \geq \max_i |u_i^{(m)}|, \quad 1 \leq i \leq n, \text{ implica}$$

$$\left| \sum_{i=1}^n |\delta_i| - \sum_{i=1}^n |u_i^{(m)}| \right| \leq \left| \max_i |\delta_i| - \max_i |u_i^{(m)}| \right|, \quad 1 \leq i \leq n.$$

5-4.5 Raio espectral

Não havendo dúvidas, uma estimativa do raio espectral pode ser obtida calculando-se a norma da matriz uma vez que satisfaz a relação $S(B) \leq \|B\|$ em que B é a matriz de iteração de Jacobi. Para os métodos em que há necessidade de conhecer os autovalores dessa matriz, foram obtidos através das rotinas GENRV (matrizes reais gerais) e SYMRV (matrizes reais simétricas), da biblioteca SEDAN disponível no IBM 4381 da UFPB Campus II.

5-5 Resultados de testes

Nesta seção serão apresentados os resultados de testes de execução dos programas que implementam os métodos vistos.

O Quadro 1 mostra para $\epsilon = 10^{-6}$ e a matriz 100×100 dos coeficientes, uma relação entre o raio espectral e o número de iterações. Na matriz simétrica (5-2.1), fazendo algumas mudanças na diagonal principal da matriz A foram obtidos os seguintes raios espectrais da matriz de iteração de Jacobi.

$d_{i,i} = 2.0$, $1 \leq i \leq n$	raio espectral igual a 0,999516
$d_{i,i} = 2.02$, $1 \leq i \leq n$	raio espectral igual a 0,989602
$d_{i,i} = 2.1$, $1 \leq i \leq n$	raio espectral igual a 0,951920
$d_{i,i} = 2.2$, $1 \leq i \leq n$	raio espectral igual a 0,908651
$d_{i,i} = 2.5$, $1 \leq i \leq n$	raio espectral igual a 0,799613

O Quadro 2 mostra a variação do tempo de execução para $\epsilon = 10^{-6}$ e a matriz 100×100 dos coeficientes com o raio espectral da matriz de iteração de Jacobi. Os mesmos raios espectrais usados no Quadro 1 foram usados.

O Quadro 3 mostra uma relação entre o raio espectral da matriz de iteração de Jacobi, em torno de 0,99, e o número de iterações. São apresentados os números de iterações para sistemas de ordens 50, 100, 300 e 500 para tolerâncias 10^{-2} , 10^{-4} e 10^{-6} . O raio espectral é obtido, para cada ordem, da matriz simétrica (5-2.1) fazendo $d_{i,i} = 2.0$ para $1 \leq i \leq n$. O valor do raio cresce ligeiramente com a ordem: 0,9981103 para 50, 0,9999516 para 100, 0,999945 para 300 e 0,999980 para 500.

O Quadro 4 mostra uma relação entre o resíduo e o número de equações para uma tolerância 10^{-6} . O raio espectral da matriz de iteração de Jacobi é obtido para cada ordem da matriz simétrica (5-2.1) fazendo $d_{i,i} = 2.0$ para $1 \leq i \leq n$, o qual cresce ligeiramente com a ordem: 0,9981103 para 50, 0,9999516 para 100, 0,999945 para 300 e 0,999980 para 500.

O Quadro 5 mostra uma relação entre o raio espectral da matriz de iteração de Jacobi, em torno de 0,95, e o número de iterações. São apresentados os números de iterações para sistemas de ordens 50, 100, 300 e 500 para tolerâncias 10^{-2} , 10^{-4} e 10^{-6} . O raio espectral é obtido, para cada ordem, da matriz simétrica (5-2.1) fazendo $d_{i,i} = 2.1$ para $1 \leq i \leq n$. O valor do raio cresce ligeiramente com a ordem: 0,950574 para 50, 0,951920 para 100, 0,952329 para 300 e 0,952362 para 500.

O Quadro 6 mostra uma relação entre o resíduo e o número de equações para uma tolerância 10^{-6} . O raio espectral da matriz de iteração de Jacobi é obtido para cada ordem da matriz simétrica (5-2.1) fazendo $d_{i,i} = 2.1$ para $1 \leq i \leq n$, o qual cresce ligeiramente com a ordem: 0,950574 para 50, 0,951920 para 100, 0,952329 para 300 e 0,952362 para 500.

O Quadro 7 mostra uma relação entre o raio espectral da matriz de iteração de Jacobi, em torno de 0,66, e o número de iterações. São apresentados os números de iterações para sistemas de ordens 50, 100, 300 e 500 para tolerâncias 10^{-2} , 10^{-4} e 10^{-6} . O raio espectral é obtido, para cada ordem, da matriz simétrica (5-2.1) fazendo $d_{i,i} = 3.0$ para $1 \leq i \leq n$. O valor do raio cresce ligeiramente com a ordem: 0,665402 para 50, 0,666344 para 100, 0,666630 para 300 e 0,666653 para 500.

O Quadro 8 mostra uma relação entre o resíduo e o número de equações para uma tolerância 10^{-6} . O raio espectral da matriz de iteração de Jacobi é obtido para cada ordem da matriz simétrica (5-2.1) fazendo $d_{i,i} = 3.0$ para $1 \leq i \leq n$, o qual cresce ligeiramente com a ordem: 0,665402 para 50, 0,666344 para 100, 0,666630 para 300 e 0,666653 para 500.

O Quadro 9 mostra uma relação entre o raio espectral da matriz de iteração de Jacobi, em torno de 0,86, e o número de iterações. São apresentados os números de iterações para sistemas de ordens 50, 100, 300 e 500 para tolerâncias 10^{-2} , 10^{-4} e 10^{-6} . O raio espectral é obtido, para cada ordem, da matriz não-simétrica (5-2.2) fazendo $d_{i,i} = 3.0$ para $1 \leq i \leq n$. O valor do

raio cresce ligeiramente com a ordem: 0,867791 para 50, 0,869612 para 100, 0,8701687 para 300 e 0,870214 para 500.

O Quadro 10 mostra uma relação entre o resíduo e o número de equações para uma tolerância 10^{-6} . O raio espectral da matriz de iteração de Jacobi é obtido para cada ordem da matriz não-simétrica (5-2.2) fazendo $d_{i,i} = 3.0$ para $1 \leq i \leq n$, o qual cresce ligeiramente com a ordem: 0,867791 para 50, 0,869612 para 100, 0,8701687 para 300 e 0,870214 para 500.

O Quadro 11 mostra uma relação entre o raio espectral da matriz de iteração de Jacobi, em torno de 0,65, e o número de iterações. São apresentados os números de iterações para sistemas de ordens 50, 100, 300 e 500 para tolerâncias 10^{-2} , 10^{-4} e 10^{-6} . O raio espectral é obtido, para cada ordem, da matriz não-simétrica (5-2.2) fazendo $d_{i,i} = 4.0$ para $1 \leq i \leq n$. O valor do raio cresce ligeiramente com a ordem: 0,650843 para 50, 0,652200 para 100, 0,652626 para 300 e 0,652660 para 500.

O Quadro 12 mostra uma relação entre o resíduo e o número de equações para uma tolerância 10^{-6} . O raio espectral da matriz de iteração de Jacobi é obtido para cada ordem da matriz não-simétrica (5-2.2) fazendo $d_{i,i} = 4.0$ para $1 \leq i \leq n$, o qual cresce ligeiramente com a ordem: 0,650843 para 50, 0,652200 para 100, 0,652626 para 300 e 0,652660 para 500.

O Quadro 13 mostra uma relação entre o Raio espectral $S(B)$ e a Norma $\|B\|$ da matriz de iteração de Jacobi. Na matriz (5-2.1), fazendo $d_{i,i}$ igual a 2.0, 2.02, 2.1, 2.2 e 2.5 para $1 \leq i \leq 100$, o raio espectral e norma são respectivamente 0,999516 e 1,000000, 0,989201 e 0,990099, 0,951920 e 0,952381, 0,908651 e 0,909090, 0,799613 e 0,800000.

Quadro 1 Raio espectral \times Número de iterações (matriz 5-2.1)

Método		Raio espectral				
		0,999516	0,989620	0,951920	0,908651	0,799613
Nº de	Jacobi	21.882	1.310	310	167	76
	Gauss-Sidel	11.658	691	166	91	44
I t e	Gradientes Conjugados	51	51	51	39	25
	SOR	314	100	54	40	27
r a	Chebyshev adaptativa	632	167	73	53	38
	Chebyshev não-adaptativa	524	120	57	42	28
ç õ e s	Aceleração dos gradientes conjugados	51	51	51	39	25

Matriz de ordem 100 \times 100.**Quadro 2** Raio espectral \times Tempo de execução em segundos (matriz 5-2.1)

Método		Raio espectral				
		0,999516	0,986020	0,951920	0,908651	0,799613
Jacobi		266,80	13,66	3,19	1,73	0,79
Gauss-Sidel		126,16	7,57	1,77	0,98	0,47
Gradientes Conjugados		1,11	1,15	1,14	0,87	0,56
SOR		11,77	9,31	8,72	8,56	8,42
Chebyshev adaptativa		9,24	2,46	1,08	0,76	0,54
Chebyshev não- adaptativa		14,79	9,78	8,96	8,62	8,44
Aceleração dos gradientes conjugados		1,16	1,13	1,19	0,90	0,57

Matriz de ordem 100 \times 100.

Quadro 3

Nº de equações × Nº de iterações
(matriz SPD diagonal dominante (5-2.1))

Tolerância		10 ⁻²				10 ⁻⁴				10 ⁻⁶			
Nº de Equações		50	100	300	500	50	100	300	500	50	100	300	500
Nº de I t e r a ç õ e s	Jacobi	1086	2846	6346	6366	3512	12364	89769	222783	5937	21882	174317	457016
	Gauss-Seidel	727	2140	8978	12410	1939	6899	51248	129020	3152	11658	93522	246137
	Gradientes Conjugados	26	51	151	251	26	51	151	251	26	51	151	251
	SOR	77	153	454	756	118	233	694	1154	153	314	934	1554
	Chebyshev adaptativa	163	330	931	1341	238	481	1495	4078	314	632	2067	6815
	Chebyshev não adaptativa	110	217	646	1076	191	381	1140	1898	263	524	1570	2615
	Aceleração dos gradientes conjugados	26	51	151	251	26	51	151	251	26	51	151	251

Raio espectral varia ligeiramente com a ordem: 0,9981103 para 50, 0,9999516 para 100, 0,999945 para 300 e 0,999980 para 500.

Quadro 4

Nº de Equações × Resíduo
matriz SPD diagonal dominante - Tolerância 10⁻⁶
Raio espectral em torno de 0,99

Nº de Equações	50	100	300	500
Jacobi	0.199561D-05	0.199834D-05	0.199998D-05	0.999994D-05
Gauss-Sidel	0.997324D-06	0.999315D-06	0.999920D-06	0.999959D-06
Gradientes Conjugados	0.688338D-13	0.497213D-12	0.680013D-11	0.229903D-10
SOR	0.145671D-06	0.473112D-07	0.236972D-07	0.183541D-07
Chebyshev adaptativa	0.609838D-07	0.311051D-07	0.134292D-07	0.269916D-06
Chebyshev não-adaptativa	0.504417D-06	0.422585D-06	0.340120D-06	0.323631D-06
Aceleração dos gradientes conjugados	0.336259D-12	0.157017D-11	0.219209D-10	0.759735D-10

Raio espectral varia ligeiramente com a ordem: 0,9981103 para 50, 0,9999516 para 100, 0,999945 para 300 e 0,999980 para 500.

Quadro 5

N° de equações × N° de iterações
(matriz SPD estritamente diagonal dominante (5-2.1))

Tolerância		10^{-2}				10^{-4}				10^{-6}			
N° de Equações		50	100	300	500	50	100	300	500	50	100	300	500
N° de I t e r a ç õ e s	Jacobi	107	124	149	160	197	217	243	254	288	310	337	348
	Gauss-Seidel	62	71	84	89	108	119	132	138	154	166	180	186
	Gradientes Conjugados	23	25	26	26	26	39	41	41	26	51	55	55
	SOR	23	26	30	32	36	40	44	46	47	54	59	61
	Chebyshev adaptativa	57	33	39	42	58	53	59	63	66	73	81	80
	Chebyshev não adaptativa	25	28	32	34	40	42	46	48	55	57	61	63
	Aceleração dos gradientes conjugados	23	25	26	26	26	39	41	41	26	51	55	55

Raio espectral varia ligeiramente com a ordem: 0,950574 para 50, 0,951920 para 100, 0,952329 para 300 e 0,952362 para 500.

Quadro 6

N° de Equações × Resíduo
matriz SPD estritamente diagonal dominante
Tolerância 10^{-6} - Raio espectral em torno de 0,95

N° de Equações	50	100	300	500
Jacobi	0.196361D-05	0.199202D-05	0.197085D-05	0.199584D-05
Gauss-Sidel	0.946739D-06	0.954751D-06	0.954289D-06	0.937264D-06
Gradientes Conjugados	0.652256D-14	0.119765D-13	0.373722D-06	0.385723D-06
SOR	0.712009D-06	0.280652D-06	0.224327D-06	0.209083D-06
Chebyshev adaptativa	0.289912D-06	0.311051D-07	0.222127D-06	0.227961D-06
Chebyshev não adaptativa	0.270005D-06	0.331923D-06	0.223258D-06	0.206796D-06
Aceleração dos gradientes conjugados	0.101599D-12	0.447087D-12	0.373722D-06	0.385723D-06

Raio espectral varia ligeiramente com a ordem: 0,950574 para 50, 0,951920 para 100, 0,952329 para 300 e 0,952362 para 500.

Quadro 7

Nº de equações × Nº de iterações
(matriz SPD estritamente diagonal dominante (5-2.1))

Tolerância		10 ⁻²				10 ⁻⁴				10 ⁻⁶			
Nº de Equações		50	100	300	500	50	100	300	500	50	100	300	500
Nº de I t e r a ç õ e s	Jacobi	20	21	24	25	31	33	36	37	42	44	47	48
	Gauss-Seidel	13	14	15	16	19	20	22	23	26	27	29	29
	Gradientes Conjugados	9	9	9	9	14	14	14	14	18	18	19	19
	SOR	10	10	12	12	14	15	16	17	19	20	21	22
	Chebyshev adaptativa	20	21	21	21	29	30	33	34	38	38	42	42
	Chebyshev não adaptativa	10	11	12	13	15	16	17	18	20	21	22	20
	Aceleração dos gradientes conjugados	9	9	9	9	14	14	14	14	18	18	19	19

Raio espectral varia ligeiramente com a ordem: 0,665402 para 50, 0,666344 para 100, 0,666630 para 300 e 0,666653 para 500.

Quadro 8

Nº de Equações × Resíduo
matriz SPD estritamente diagonal dominante
Tolerância 10⁻⁶ - Raio espectral em torno de 0,66

Nº de Equações	50	100	300	500
Jacobi	0.164163D-05	0.161840D-05	0.153622D-05	0.172949D-05
Gauss-Sidel	0.536540D-06	0.637067D-06	0.529930D-06	0.902459D-06
Gradientes Conjugados	0.546648D-06	0.661738D-06	0.265577D-06	0.267506D-06
SOR	0.388007D-06	0.361272D-06	0.469528D-06	0.306162D-06
Chebyshev adaptativa	0.172688D-05	0.360201D-05	0.155268D-05	0.266320D-05
Chebyshev não-adaptativa	0.317754D-06	0.256466D-06	0.339019D-06	0.590743D-06
Aceleração dos gradientes conjugados	0.546648D-06	0.661738D-06	0.265578D-06	0.267507D-06

Raio espectral varia ligeiramente com a ordem: 0,665402 para 50, 0,666344 para 100, 0,666630 para 300 e 0,666653 para 500.

Quadro 9

Nº de equações × Nº de iterações
(matriz não-simétrica diagonal dominante (5-2.2))

Tolerância		10 ⁻²				10 ⁻⁴				10 ⁻⁶			
Nº de Equações		50	100	300	500	50	100	300	500	50	100	300	500
Nº de I t e r a ç õ e s	Jacobi	114	204	541	897	154	246	615	957	190	297	675	1028
	Gauss-Seidel	69	126	343	555	86	147	374	593	101	165	399	623
	Gradientes Conjugados	336	712	***	***	410	779	***	***	495	856	***	***
	SOR	46	88	249	408	52	95	258	420	58	102	267	430
	Chebyshev adaptativa	***	***	***	***	***	***	***	***	***	***	***	***
	Chebyshev não adaptativa	***	***	***	***	***	***	***	***	***	***	***	***
	Aceleração dos gradientes conjugados	***	***	***	***	***	***	***	***	***	***	***	***

(***) Método não converge.

Raio espectral varia ligeiramente com a ordem: 0,867791 para 50, 0,869612 para 100, 0,8701687 para 300 e 0,870214 para 500.

Quadro 10

Nº de Equações × Resíduo
matriz não-simétrica diagonal dominante
Tolerância 10⁻⁶ - Raio espectral em torno de 0,86

Nº de Equações	50	100	300	500
Jacobi	0.254614D-05	0.243609D-05	0.254939D-05	0.274711D-05
Gauss-Sidel	0.127742D-05	0.130518D-05	0.160914D-05	0.161669D-05
Gradientes Conjugados	0.218983D-05	0.218983D-05	***	***
SOR	0.384380D-06	0.301790D-06	0.748022D-06	0.588308D-06
Chebyshev adaptativa	***	***	***	***
Chebyshev não-adaptativa	***	***	***	***
Aceleração dos gradientes conjugados	***	***	***	***

(***) Método não converge.

Raio espectral varia ligeiramente com a ordem: 0,867791 para 50, 0,869612 para 100, 0,8701687 para 300 e 0,870214 para 500.

Quadro 11

Nº de equações × Nº de iterações
(matriz não-simétrica estritamente diagonal dominante (5-2.2))

Tolerância		10 ⁻²				10 ⁻⁴				10 ⁻⁶			
Nº de Equações		50	100	300	500	50	100	300	500	50	100	300	500
Nº de iterações	Jacobi	25	28	32	34	40	44	48	50	54	59	64	66
	Gauss-Seidel	19	21	24	25	29	32	35	37	38	43	46	48
	Gradientes Conjugados	24	24	23	23	45	45	45	45	57	67	67	66
	SOR	16	17	20	21	24	26	29	30	31	35	38	39
	Chebyshev adaptativa	***	***	***	***	***	***	***	***	***	***	***	***
	Chebyshev não adaptativa	***	***	***	***	***	***	***	***	***	***	***	***
	Aceleração dos gradientes conjugados	***	***	***	***	***	***	***	***	***	***	***	***

(***) Método não converge.

Raio espectral varia ligeiramente com a ordem: 0,650843 para 50, 0,652200 para 100, 0,652626 para 300 e 0,652660 para 500.

Quadro 12

Nº de Equações × Resíduo
matriz não-simétrica estritamente diagonal dominante
Tolerância 10⁻⁶ - Raio espectral em torno de 0,65

Nº de Equações	50	100	300	500
Jacobi	0.262846D-05	0.260458D-05	0.260258D-05	0.259159D-05
Gauss-Sidel	0.150593D-05	0.138481D-05	0.197046D-05	0.157327D-05
Gradientes Conjugados	0.133780D-05	0.216918D-05	0.189690D-05	0.219697D-05
SOR	0.796831D-06	0.990358D-06	0.106187D-05	0.115700D-05
Chebyshev adaptativa	***	***	***	***
Chebyshev não-adaptativa	***	***	***	***
Aceleração dos gradientes conjugados	***	***	***	***

(***) Método não converge.

Raio espectral varia ligeiramente com a ordem: 0,650843 para 50, 0,652200 para 100, 0,652626 para 300 e 0,652660 para 500.

Quadro 13

Raio espectral $S(B) \times$ Norma $\|B\|$

na matriz (5-2.1), fazendo $d_{i,i}$ igual a 2.0, 2.02, 2.1, 2.2 e 2.5 para $1 \leq i \leq n$, a matriz de iteração de Jacobi possui respectivamente Raio espectral e Norma

$S(B)$	0,999516	0,989201	0,951920	0,908651	0,799613
$\ B\ $	1,000000	0,990099	0,952381	0,909090	0,800000

Matriz de ordem 100×100 .

5-6 Observações sobre os resultados

A partir dos testes realizados foi observado que a convergência dos métodos iterativos está relacionado diretamente com o raio espectral da matriz de iteração de Jacobi. O Quadro 1 mostra a relação entre o número de iterações de cada método com o raio espectral da matriz de iteração de Jacobi.

O número de iterações (Quadro 1) em todos os métodos, exceto o método dos Gradientes Conjugados e sua versão acelerada, aumentam com o raio espectral. Por outro lado, nos métodos básicos, quando o raio espectral for superior a 0,951920 o número de iterações cresce muito.

Quanto ao tempo de execução (Quadro 2), todos os métodos têm o mesmo comportamento descrito no parágrafo anterior, exceto os métodos SOR e Aceleração de Chebyshev não-adaptativa que apresentaram maior tempo de execução devido aos cálculos adicionais para obter o maior e menor autovalores, necessário nesses métodos.

O Quadro 3 mostra a variação do número de iterações em todos os métodos com o número de equações e com tolerância de erro usadas, com o raio espectral da matriz de iteração de Jacobi em torno de 0,99. O quadro mostra que nos métodos básicos o crescimento do número de iterações com o número de equações é muito maior.

Os Quadros 4 e 5 mostram que se o raio espectral da matriz de iteração de Jacobi for inferior a 0,95, o número de iterações necessário para satisfazer os critérios de parada é sempre inferior ao número de equações independentemente da tolerância requerida.

Quando a matriz A for não-simétrica, os métodos de aceleração de Chebyshev adaptativa e não-adaptativa e o método de dos Gradientes Conjugados acelerado não convergem. Para a matriz não-simétrica (5-2.2) usada em teste de ordens 50, 100, 300 e 500 respectivamente, fazendo $d_{i,i} = 3.0$ para $1 \leq i \leq n$, o raio espectral da matriz de iteração de Jacobi for superior a 0,86. Nessa situação os métodos básicos convergem com um número de iterações superior ao número de equações como mostra o Quadro 6, o método SOR, converge com um número de iterações bem inferior ao dos métodos básicos. Se não for conhecido previamente o raio espectral da matriz de iteração, o método SOR não é indicado, uma vez que o tempo de execução para obtê-lo é superior ao tempo que os métodos básicos levam para encontrar a solução. O método dos Gradientes Conjugados quando o raio espectral é dessa ordem não converge para um sistema com um número superior a 140 equações, porque a convergência só é garantida se a matriz dos coeficientes for simétrica positiva definida (condição suficiente).

Se o raio espectral da matriz de iteração do método de Jacobi for inferior a 0,86 e a matriz for SPD (5-2.2) o Quadro 7 mostra que o método dos Gradientes Conjugados converge para todas as ordens das matrizes testadas. Também os métodos básicos e SOR convergem.

Capítulo VI

Conclusões

Baseado nos estudos realizados foi verificado que as características da matriz dos coeficientes associada ao sistema define os métodos a serem aplicados.

I – Estrutura da matriz.

- Não-simétrica diagonal dominante.

Métodos que convergem:

de Jacobi,

de Gauss-Seidel,

de Sobre-Relaxação Sucessiva (SOR).

O método SOR é recomendado se o raio espectral da matriz de iteração de Jacobi for superior a 0,86. Se o raio espectral for inferior a 0,86, os métodos básicos são recomendados, pela sua simplicidade (não precisa do valor do raio espectral).

- Simétrica positiva definida diagonal dominante.

Todos os métodos convergem.

O método dos gradientes conjugados é o melhor de todos para tolerâncias até 10^{-12} em precisão dupla. Para uma margem de tolerância próximo ao ϵ da máquina recomenda-se a utilização do método dos gradientes conjugados acelerado. Como o método dos gradientes conjugados não requer nenhum parâmetro especial, é o recomendado quando aplicável. Se o raio espectral da matriz de iteração de Jacobi for inferior a 0,95, todos os métodos se comportam de maneira uniforme e é indiferente usar qualquer método.

II – Velocidade de convergência

A velocidade de convergência dos métodos iterativos está diretamente ligada ao raio espectral da matriz de iteração. Sendo a matriz de iteração de Jacobi obtida com um menor número de operações, optou-se pelo cálculo do seu raio espectral, pois fornece uma boa idéia qualitativa sobre a velocidade de convergência dos outros métodos. Quanto menor o raio espectral da matriz de iteração de Jacobi, maior é a velocidade de convergência de todos os outros métodos aplicáveis.

- Não-simétrica diagonal dominante.

Se o raio espectral da matriz de iteração de Jacobi for inferior a 0,86, os métodos básicos bem como os outros métodos aplicáveis convergem rapidamente.

- Simétrica positiva definida diagonal dominante.

Se o raio espectral da matriz de iteração de Jacobi for inferior a 0,95, os métodos básicos e todos os outros, com exceção do método dos gradientes conjugados e seus derivados, convergem rapidamente. O método dos gradientes tem velocidade de convergência independente do raio espectral.

Se a matriz dos coeficientes associada ao sistema for SPD estritamente diagonal dominante, o raio espectral da matriz de iteração de Jacobi será inferior a 0,95, é indiferente usar qualquer método.

III – Norma

A norma usada nos critérios de interrupção tem influência sobre:

- a) tempo de execução de um programa que implementa um método e
- b) a qualidade dos resultados. Embora a norma infinita seja a mais econômica dentre as normas mais comuns, a qualidade dos resultados obtidos medida pela norma do resíduo $\|b - Au\|$ é inferior a aquela que se consegue com a norma 1.

A norma euclidiana, embora mais dispendiosa de calcular que a norma 1 forneceu resultados com qualidade inferior comparados com os fornecidos pela norma 1.

IV – Raio Espectral

Em vez de calcular os autovalores da matriz de iteração para obter o seu raio espectral que é dispendioso, uma estimativa $S(B) \leq \|B\|$ da matriz de iteração de Jacobi é suficiente em muitos casos para se ter uma idéia da velocidade de convergência.

Trabalhos futuros

Este trabalho concentrou sua atenção na aceleração de métodos iterativos para solução de sistemas lineares.

Como trabalho adicional é sugerido:

1. estudar métodos básicos não simetrizáveis;
2. estudar procedimentos não polinomiais de aceleração dos métodos iterativos básicos;

3. generalizar o estudo do método dos gradientes conjugados para direções conjugadas;
4. implementar os métodos vistos com economia do espaço de memória armazenando somente elementos não nulos da matriz dos coeficientes do sistema de equações lineares.

Referências bibliográficas

- [Albrecht 73] Albrecht, P., *Análise Numérica*, Livros Técnicos e Científicos Editora S. A., Rio de Janeiro, 1973.
- [Golub 89] Golub, Gene H., and Van Loan, Charles F, *Matrix Computation*, Johns Hopkins University Press, Baltimore, 1989, 2ª edição.
- [Varga 62] Varga, R. S., *Matrix Iterative Analysis*, Prentice Hall, Inc., Englewood Cliffs, New Jersey, 1962.
- [Young 71] Young, D. M., *Iterative Solution of Large Linear Systems*, Academic Press, New York and London, 1971.
- [Young 72] Young, D. M., and Gregory, R. T., *A Survey of Numerical Mathematics*, Vol. II, Addison-Wesley Publishing Company, Reading, MA, 1972.
- [Young 78] Young, D. M., and Kincaid David R., *Survey of Iterative Methods*, Center for Numerical Analysis, the University of Texas at Austin, Abril 1978, (relatório técnico).
- [Young 81] Young, D. M., e Hageman Louis A., *Applied Iterative Methods*, Academic Press, New York and London, 1981.
- [Reid 71] Reid, J. K., *On the Method of Conjugate Gradients for the Solution of Large Sparse Systems of Linear Equations*, Proc. Conf. Large Sparse Sets of Linear Equations. pp. 231 - 254. Academic Press, New York, 1971.
- [Reid 72] Reid, J.K., *The Use of Conjugate Gradients for Systems of Linear Equations Possessing Property A*, SIAM J. Num. Anal. **9**, 325-332, 1972.
- [Ortega 73] Ortega, J. M., *Numerical Analysis, A Second Course*, Academic Press, New York and London, 1973.
- [Rutishauser 75] Rutishauser, H., *The Jacobi Method for Real Symmetric Matrices*, Numer Math. **9**, 1-10, ver também HACA, pp. 202-11, 1966.
- [Gill 74] Gill, P. e W. Murray, *Numerical Methods for Constrained Optimization*, Academic Press. London and New York, 1974.
- [Conte 72] Conte, S. D., *Elementary Numerical Analysis — An Algorithmic Approach*, Second Edition, McGraw-Hill Book Company, 1972.

- [Young 80] Young, D. M., Hayes, L., and Jea, K. C., *Conjugate Gradient Acceleration of Iterative Methods*, Part I, The Symmetrizable case. Rep. CNA - 162, Center for Numerical Analysis, University of Texas at Austin, 1980.
- [Reid 66] Reid, J. K., *A method for finding the optimum successive overrelaxation factor*, *Comput. J.* **9**, pp. 200-204, 1966.
- [Hagman 77] Hageman, L. A., Luk F., and Young, D. M., *On the acceleration of Iterative Methods*, Preliminary Report. Rep. CNA -129, Center for Numerical Analysis, University of Texas at Austin, 1977.
- [Hestenes 56] Hestenes, M. R., *The Conjugate Gradient Method for Solving Linear Systems*, In *Numerical Analysis*, J. Curtiss, Vol. VI, McGraw-Hill, New York, 1956.
- [Stiefel 52] Stiefel, E. R., and Hestenes, M. R., *Methods of Conjugate Gradient for Solving Linear Systems*, *Nat. Bur. Std. J. Res.* **49**, pp. 409-436, 1952.
- [Daniel 67] Daniel, J. W., *The Conjugate Gradient Method for Linear and Nonlinear Operator Equations*. *SIAM J. Numer. Anal.* **4**, pp. 10 - 26, 1976.
- [Ralston 65] Ralston, A. , *A First Course in Numerical Analysis*, McGraw-Hill Book Company, New York, 1965.
- [Strang 76] Strang, G., *Linear Algebra and its Applications*, Academic Press, New York, 1976.
- [Householder 75] Householder, Alston S., *The Theory of Matrices in Numerical Analysis*, Dover Publications, Inc. New York, 1975.
- [Soares 83] Soares, M. N. Almeida, *Estudos e implementação de métodos iterativos para solução de sistemas de equações lineares integrados a uma biblioteca numérica*, Dissertação de Mestrado do Departamento de Sistemas e Computação, UFPB, Campus II, 1983.

