

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA APLICADA

Métodos Iterativos para a Solução da Equação de Poisson

por

Valdirene da Rosa Rocho

Dissertação submetida como requisito parcial
para a obtenção do grau de
Mestre em Matemática Aplicada

Prof. Dagoberto Adriano Rizzotto Justo, Ph.D.
Orientador

Porto Alegre, 20 de abril de 2012.

CIP - CATALOGAÇÃO NA PUBLICAÇÃO

Rocho, Valdirene da Rosa

Métodos Iterativos para a Solução da Equação de Poisson / Valdirene da Rosa Rocho.—Porto Alegre: PPGMAp da UFRGS, 2012.

87 p.: il.

Dissertação (mestrado) —Universidade Federal do Rio Grande do Sul, Programa de Pós-Graduação em Matemática Aplicada, Porto Alegre, 2012.

Orientador: Justo, Ph.D., Dagoberto Adriano Rizzotto

Dissertação: Matemática Aplicada
Equação de Poisson, Condições de Contorno, Método diferenças finitas, Métodos iterativos, Equação de Navier-Stokes

Métodos Iterativos para a Solução da Equação de Poisson

por

Valdirene da Rosa Rocho

Dissertação submetida ao Programa de Pós-Graduação em Matemática Aplicada do Instituto de Matemática da Universidade Federal do Rio Grande do Sul, como requisito parcial para a obtenção do grau de

Mestre em Matemática Aplicada

Linha de Pesquisa: Análise Numérica e Computação Científica

Orientador: Prof. Dagoberto Adriano Rizzotto Justo, Ph.D.

Banca examinadora:

Profa. Dra. Sani de Carvalho Rutz da Silva
PPGECT/UTFPR

Prof. Dr. Rafael Rigão Souza
PPGMAT/UFRGS

Profa. Dra. Carolina Cardoso Manica
PPGMAp/UFRGS

Dissertação apresentada e aprovada em
20 de abril de 2012.

Prof. Dra. Maria Cristina Varriale
Coordenador

AGRADECIMENTOS

Eis um momento difícil: o agradecimento. A dificuldade não está somente em demonstrar a gratidão merecida àqueles que de alguma forma contribuíram para que mais uma etapa fosse superada, mas também na forma escolhida. É um tanto árdua esta tarefa, pois por mais que eu me esforce, que eu mencione alguns nomes importantes na elaboração desta pesquisa, por certo, alguns outros ficarão para trás. É esta provável falta que angustia, pois sou grata a todos.

Agradeço a Deus por iluminar-me sempre e conduzir-me em mais uma caminhada importante em minha vida; por dar-me a oportunidade de encontrar pessoas especiais neste caminho e pela determinação de eu nunca desistir de meus sonhos, mesmo tendo que ultrapassar alguns obstáculos... sem, no entanto, desviar de meus princípios.

À minha querida mãe Enoi, pela forma com que conduziu minhas atitudes e pensamentos, encorajando-me sempre para que conseguisse percorrer mais este caminho com sucesso, como em muitos outros que esteve ao meu lado. “Mãe, és mais que uma companheira, és uma amiga fiel, amável e dedicada. Superou comigo a distância, os momentos difíceis de desilusão, o medo, a saudade. Agradeço pela paciência, cumplicidade, incentivo, confiança e por acreditar que um dia eu conseguiria concluir o tão sonhado mestrado.”

Aos meus irmãos Ana e Lucas pelo carinho demonstrado e acreditar que eu seria capaz de alcançar este objetivo.

Agradeço também à prima Ana Maria e sua família, pela hospitalidade e carinho demonstrado em todo o período de mestrado e estada em seu lar.

De modo especial, agradeço ao professor Dr. Dagoberto Adriano Rizzato Justo, meu orientador. De fato, sem ele nada seria possível, o sonho não se

tornaria real. Sou imensamente grata por sua generosidade e por acreditar que seria possível quando, muitas vezes, nem ao menos eu acreditava. Sua confiança em mim, no meu trabalho, a forma como me guiou, provocou-me sérias reflexões quanto à existência do acaso.

Aos professores do Programa de Pós-Graduação em Matemática Aplicada, por suas colaborações que indiretamente e diretamente, influenciaram na construção de meu conhecimento. E também a todos os professores que tive ao longo de minha vida.

Aos meus colegas do Programa de Pós-Graduação em Matemática Aplicada, que compartilharam comigo alguns momentos de descontração e as dificuldades encontradas pelo caminho.

De modo especial, agradeço à colega Elisângela P. Francisqueti que não mediu esforços para ajudar-me nas dúvidas que surgiram ao longo desta caminhada. Muito obrigada por sua amizade.

À minha amiga Milena Titoni que sempre incentivou-me a não desistir deste objetivo.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), que me concedeu apoio financeiro, possibilitando um maior empenho à minha pesquisa.

Ao Programa de Pós-Graduação em Matemática Aplicada pela oportunidade.

De um modo geral, agradeço a todos os que, de uma forma ou de outra, contribuíram para que este trabalho fosse finalizado. Guardo por vocês uma gratidão que jamais poderá empalidecer, pois fazem parte da história de minha vida.

Sumário

LISTA DE FIGURAS	viii
LISTA DE TABELAS	x
RESUMO	xi
ABSTRACT	xii
1 INTRODUÇÃO	1
2 EQUAÇÃO DE POISSON	4
2.1 Modelo Contínuo	4
2.2 Modelo discreto unidimensional	5
2.2.1 Condições de contorno de Dirichlet	8
2.2.2 Condições de contorno de Neumann	9
2.3 O modelo discreto bidimensional	12
2.3.1 O problema de Dirichlet Bidimensional	14
2.3.2 O problema de Neumann Bidimensional	17
2.4 Autovalores da Matriz de Discretização	22
2.5 Existência de soluções	24
3 MÉTODOS ITERATIVOS CLÁSSICOS	27
3.1 Método de Jacobi	28
3.2 Método de Gauss-Seidel	30
3.3 Método das Sobre-relaxações Sucessivas - SOR	31
3.4 Convergência de Métodos Iterativos Lineares	33
3.5 Taxa de Convergência	38
3.6 Convergência do Método SOR	39

3.7 Problema de Poisson	40
3.7.1 Condição de Dirichlet Unidimensional	40
3.7.2 Condição de Dirichlet Bidimensional	41
3.7.3 Condição de Neumann Unidimensional	44
3.7.4 Condição de Neumann Bidimensional	48
4 RESULTADOS NUMÉRICOS	53
4.1 Existência de Solução	53
4.2 Raio Espectral e Taxa de Convergência	59
4.3 Método SOR e o parâmetro ω	61
4.3.1 Raio espectral, Taxa de convergência e $\omega_{\text{ótimo}}$	63
4.3.2 Comparação da Velocidade de Convergência	66
5 A EQUAÇÃO DE NAVIER-STOKES	68
5.1 Sistema de Equações	68
5.2 O Problema da Cavidade	71
6 CONCLUSÃO	78
APÊNDICE A	81
A.1 Algoritmo PRIME	81
A.1.1 Estrutura do Algoritmo	82
REFERÊNCIAS BIBLIOGRÁFICAS	85

Lista de Figuras

Figura 2.1	<i>Discretização do problema unidimensional</i>	6
Figura 2.2	<i>Discretização do problema bidimensional</i>	12
Figura 3.1	<i>Partição da matriz A</i>	29
Figura 4.1	<i>Condição (4.10) e termo fonte $f(x, y) = \sin x \cdot \sin y$ a cada iteração na malha cartesiana 21×21 e $tol = 10^{-10}$.</i>	55
Figura 4.2	<i>Condição (4.11) e termo fonte $f(x, y) = \sin x \cdot \sin y$ a cada iteração na malha cartesiana 21×21 e $tol = 10^{-10}$.</i>	56
Figura 4.3	<i>Diferença $\ \bar{p}^{k+1} - \bar{p}^k\$ quando $f(x, y) = x^2$ a cada iteração na malha cartesiana 21×21.</i>	57
Figura 4.4	<i>$\ \bar{p}^k\$ de $f(x, y) = x^2$ a cada iteração na malha cartesiana 21×21.</i>	57
Figura 4.5	<i>Diferença $\ \bar{p}^{k+1} - \bar{p}^k\$ quando $f(x, y) = x^2$ a cada iteração na malha cartesiana 21×21.</i>	58
Figura 4.6	<i>$\omega_{\text{ótimo}} \times \rho(G_{SOR})$ para malha cartesiana $h = \frac{1}{5}, \frac{1}{10}, \frac{1}{20}, \frac{1}{40}, \frac{1}{80}$</i>	62
Figura 4.7	<i>$\omega_{\text{ótimo}} \times \rho(G_{SOR})$ para malha cartesiana 21×21 do problema de Neumann</i>	63
Figura 5.1	<i>Solução do escoamento para o problema da cavidade em malha uniforme centrada 81×81. Gráfico dos vetores das componentes de velocidade.</i>	72
Figura 5.2	<i>Solução do escoamento para o problema da cavidade em malha uniforme deslocada 81×81. Gráfico dos vetores das componentes de velocidade.</i>	73
Figura 5.3	<i>Solução do escoamento para o problema da cavidade em malha uniforme centrada 81×81. Gráfico dos vetores das componentes de velocidade.</i>	74
Figura 5.4	<i>Solução do escoamento para o problema da cavidade em malha uniforme deslocada 81×81. Gráfico dos vetores das componentes de velocidade.</i>	74
Figura 5.5	<i>Solução do escoamento para o problema da cavidade em malha uniforme centrada 81×81. Gráfico dos vetores das componentes de velocidade.</i>	75

Figura 5.6	<i>Solução do escoamento para o problema da cavidade em malha uniforme deslocada 81×81. Gráfico dos vetores das componentes de velocidade.</i>	76
------------	--	----

Lista de Tabelas

Tabela 4.1	Experimento numérico para o problema de Dirichlet.	59
Tabela 4.2	Experimento numérico para o problema de Dirichlet. n_O é o número de iterações observado e n_T é o número de iterações teórico.	60
Tabela 4.3	Experimento numérico para equação de Poisson com condições de contorno do problema de Neumann.	61
Tabela 4.4	Experimento numérico para equação de Poisson com condições de contorno do problema de Neumann. n_O é o número de iterações observado.	61
Tabela 4.5	Resultados teóricos para ω , ρ , R e n_T com condições de contorno (4.1) e usando $h = \frac{1}{20}$	64
Tabela 4.6	Experimento numérico para equação de Poisson com condições de contorno de Dirichlet através do método iterativo SOR. n_O é o número de iterações observado e n_T é o número de iterações teórico.	64
Tabela 4.7	Experimento numérico para equação de Poisson com condições de contorno de Neumann através do método iterativo SOR. n_O é o número de iterações observado e n_T é o número de iterações teórico.	65
Tabela 4.8	Experimento numérico para equação de Poisson com condições de contorno de Dirichlet através dos métodos iterativos Jacobi, Gauss-Seidel e SOR.	66
Tabela 4.9	Experimento numérico para equação de Poisson com condições de contorno de Neumann através dos métodos iterativos Jacobi, Gauss-Seidel e SOR.	67

RESUMO

O objetivo deste trabalho é estudar o uso de métodos iterativos para a obtenção da solução da equação de Poisson com condições de contorno de Dirichlet e de Neumann para o caso uni e bidimensional em um retângulo $(0, 1) \times (0, 1)$.

Ao discretizar a equação de Poisson com o método de diferenças finitas obtém-se um sistema linear que pode ser resolvido através de um método iterativo. Para o problema de Neumann obtemos condições para que o problema tenha solução baseado na integral do termo fonte.

Fez-se um breve estudo referente aos métodos iterativos de Jacobi, Gauss-Seidel e SOR aplicados ao sistema obtido analisando o espectro da matriz de iteração dos respectivos métodos. A partir do maior autovalor em módulo podemos estudar a convergência dos métodos (quando os autovalores estão no disco unitário) e a taxa de convergência com a qual cada método convergirá.

Foi desenvolvido um código em linguagem Fortran e MATLAB para testar os resultados teóricos aplicados à solução do problema de Poisson num quadrado. Estudou-se também o método SOR e a obtenção do parâmetro ω ótimo.

Ainda neste trabalho destacamos também a aplicação dos resultados na solução do problema da cavidade. Utilizando o método PRIME, a partir da equação de Navier-Stokes podemos obter uma equação de Poisson para a pressão.

Dos problemas estudados montou-se os sistemas lineares, a partir destes pode-se verificar a existência de uma única ou infinitas soluções. E a partir da matriz de iteração de cada método iterativo pode-se determinar os autovalores e assim concluir quanto a convergência de cada método.

ABSTRACT

The aim of this work is to study the convergence of iterative methods for the solution of the Poisson equation with Dirichlet and Neumann boundary conditions in an interval $(0, 1)$ and, in the bi-dimensional case, in a rectangle $(0, 1) \times (0, 1)$.

The idea is to discretize the problem using the finite difference method and obtain a linear system that can be solved by an iterative method, such as Jacobi, Gauss-Seidel, and SOR.

We present eigenvalue formulas for the matrix of the linear system and for the matrices of the iterative methods. Such eigenvalues can be used to decide upon the convergence and rate of convergence of those methods.

For the iterative method to be convergent, the spectral radius of the iteration matrix should be less than one. We also obtain similar convergence criteria for semiconvergent and conditionally convergent methods that appear in the Neumann problem, where some eigenvalues have modulus equal to one.

We compare the Jacobi, Gauss-Seidel and SOR methods to obtain and optimal rate of convergence for the best choice of a parameter ω in the last one. In the end, we present an application of the results on the solution of a Poisson equation that appears in the solution of a Navier Stokes problem.

1 INTRODUÇÃO

A equação de Poisson é uma equação elíptica de derivadas parciais com uma ampla utilidade em Dinâmica de Fluidos, e em muitos casos é utilizada para modelar a pressão de um fluido. Para resolver esta equação pode-se utilizar vários métodos como, por exemplo, uma função de Green ou métodos numéricos. E ainda precisa-se de adequadas condições de contorno.

Neste trabalho objetivou-se resolver a equação de Poisson para o problema de Dirichlet dado por

$$\begin{aligned}\Delta p &= f & (x, y) \in \Omega \\ p(x, y) &= 0 & (x, y) \in \partial\Omega\end{aligned}\tag{1.1}$$

e o problema de Neumann

$$\begin{aligned}\Delta p &= f & (x, y) \in \Omega \\ \frac{\partial p}{\partial n}(x, y) &= 0 & (x, y) \in \partial\Omega\end{aligned}\tag{1.2}$$

para o caso uni e bidimensional. Na literatura de ([Young(1988)], [Fonseca(2009)], [Biezuner(2007)] entre outros) o problema de Dirichlet tem sido extensivamente estudado e é possível encontrar muitos trabalhos sobre este problema, pois este é um problema mais simples. Porém quando se trata do problema de Neumann as informações são limitadas (acredita-se que um dos fatores seja pela restrição de existência de solução).

No capítulo dois apresenta-se a discretização da equação de Poisson usando o método de diferenças finitas tanto para o caso unidimensional quanto o bidimensional.

A partir da equação discretizada obtém-se um sistema do tipo

$$A\vec{p} = \vec{b}\tag{1.3}$$

onde para cada condição de contorno determinada obtém-se uma matriz de discretização A , a qual difere pelos pontos do contorno. A partir desta pode-se determinar o espectro de cada uma dessas matrizes. Utilizando o produto de Kronecker pode-se relacionar as matrizes de discretização dos problemas em uma e duas dimensões. A partir dessa relação e de um teorema envolvendo o espectro e o produto de Kronecker, obtém-se fórmulas para os autovalores do problema bidimensional. O espectro da matriz A é útil, por exemplo, para determinar a existência de uma ou infinitas soluções do problema discretizado. Os problemas de Dirichlet encontrados sempre terão solução pois nenhuma das matrizes possui autovalor igual a zero (estes são amplamente encontrados na literatura). Os problemas de Neumann possuem um autovalor igual a zero, o que não garante solução para o problema. Se esta existir, não será única. Obtém-se nesse capítulo uma condição para que o sistema discreto tenha solução baseado na soma dos elementos do termo independente (seria o equivalente discreto da integral do termo fonte).

A idéia central dos métodos iterativos é transformar o sistema original (1.3), num sistema equivalente do tipo

$$\vec{x}^{k+1} = G\vec{x}^k + \vec{f} \quad (1.4)$$

onde uma sequência de estimativas \vec{x}^k pode ser obtida a partir de uma estimativa \vec{x}_0 . Sob certas condições, pode-se garantir a convergência dessa sequência a partir de qualquer condição inicial e estudar a taxa de convergência dos métodos.

No capítulo três apresenta-se vários métodos iterativos clássicos para a solução de (1.3) tais como Jacobi, Gauss-Seidel e SOR (Sobre-relaxação Sucessiva). Estes métodos são baseados na decomposição da matriz A de tal maneira a converter o sistema original numa iteração de ponto fixo linear como em (1.4). Para cada método obtém-se uma matriz de iteração G diferente e assim é possível definir quem são seus autovalores. Destaca-se ainda que a partir dos autovalores é possível fazer uma análise baseada na estimativa do raio espectral sobre a taxa de convergência, o número de iterações realizadas por cada método e sobretudo afirmar se o método é

convergente ou não. Apresenta-se as definições de métodos e sequências de matrizes convergente, semi-convergente e condicionalmente convergente, onde apenas a primeira é comumente utilizada no problema de Dirichlet. Apresenta-se também nesse capítulo fórmulas para os autovalores das matrizes de iteração e condições para que o problema de Neumann apresente convergência.

No quarto capítulo são descritos os resultados numéricos obtidos após as simulações dos problemas modelo, assim como sobre quais condições existe solução para os problemas em questão. Avaliou-se a taxa de convergência, quantidade de iterações realizadas até obter-se a convergência e o raio espectral da matriz de iteração de cada método, bem como a escolha do melhor coeficiente de relaxação ω para o método SOR.

No capítulo cinco é apresentado uma aplicação dos resultados obtidos no texto onde resolveu-se numericamente a equação de Navier-Stokes para escoamento em uma cavidade. Ao utilizar o método PRIME obtém-se uma equação de Poisson com condições de Neumann para a pressão do fluido o qual é estudada com diferentes métodos iterativos. O problema da cavidade é resolvido para diferentes números de Reynolds ($Re = 10, 100$ e 1000).

No último capítulo apresentam-se as considerações finais e perspectivas de trabalhos futuros.

2 EQUAÇÃO DE POISSON

Nesta seção apresenta-se o modelo contínuo da equação de Poisson, seguido pelos modelos discretos em uma e duas dimensões. Para cada um deles será apresentado o modelo contínuo com condições de contorno de Dirichlet e de Neumann.

2.1 Modelo Contínuo

A equação de Poisson é uma equação de derivadas parciais elíptica com uma ampla utilidade em escoamento de fluidos incompressíveis. Fisicamente, esta equação pode representar a distribuição de calor num estado permanente e na área de fluidos é utilizada para modelar a pressão.

Seja Ω um conjunto conexo em \mathbb{R}^d , $d = 1, 2, 3$ e $\partial\Omega$ a fronteira desse conjunto. De acordo com [Strikwerda(2004)] a equação de Poisson pode ser escrita como

$$\Delta p = f \quad \text{em } \Omega \quad (2.1)$$

onde $p : \mathcal{C}^2 \rightarrow \mathbb{R}$ é a incógnita (a pressão, por exemplo), e a função $f : \mathcal{C}^0 \rightarrow \mathbb{R}$ é um termo fonte.

Para resolver a equação (2.1) é necessário incluir condições de contorno para o problema que podem ser, por exemplo,

a) Condições de Dirichlet: a função é fixada no contorno, ou seja,

$$p = g \quad \text{em } \partial\Omega. \quad (2.2)$$

b) Condições de Neumann: a taxa de variação de p é fixa no contorno, ou seja,

$$\frac{\partial p}{\partial n} = g \quad \text{em } \partial\Omega. \quad (2.3)$$

c) Condições de Robin: é uma condição de contorno mista, ou seja,

$$p + \beta \frac{\partial p}{\partial n} = g \quad \text{em } \partial\Omega. \quad (2.4)$$

d) Condições periódicas: Num domínio quadrado, por exemplo, têm-se que p no contorno norte será igual a p no contorno sul (e o mesmo para a direção leste-oeste).

Neste trabalho estudar-se-á as condições a e b, dando atenção especial em b pois não são facilmente encontradas na literatura.

Segundo [Strikwerda(2004)] e [van Linde(1974)], ao utilizar-se as condições de contorno de Neumann para resolver-se a equação (2.1) é necessário que a hipótese de compatibilidade

$$\int \int_{\Omega} f \, d\Omega = \int_{\partial\Omega} g \, dS \quad (2.5)$$

seja satisfeita para existir solução para o problema. Deste modo, [Hackbusch(1992)] ressalta que existindo uma solução p , então $p + c$, com c uma constante arbitrária, é também uma solução. Caso a hipótese (2.5) não seja satisfeita, então não existe solução para o problema.

A necessidade da condição (2.5) pode ser facilmente provada. A partir da equação (2.1) e do teorema do divergente é necessário que

$$\int \int_{\Omega} f \, d\Omega = \int \int_{\Omega} \Delta p \, d\Omega = \int_{\partial\Omega} \vec{n} \cdot \vec{\nabla} p \, dS = \int_{\partial\Omega} \frac{\partial p}{\partial n} \, dS = \int_{\partial\Omega} g \, dS \quad (2.6)$$

onde \vec{n} é o vetor unitário normal a fronteira $\partial\Omega$.

2.2 Modelo discreto unidimensional

Para a discretização das equações usou-se o método diferenças finitas, e ao invés de trabalhar-se com o domínio contínuo Ω considerou-se apenas um conjunto

finito de pontos de Ω . No caso unidimensional, sendo $\Omega = (0, 1)$, pode-se dividir esse intervalo em $n + 1$ intervalos iguais de tamanho $h = \frac{1}{n+1}$ conforme a figura 2.1.

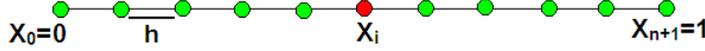


Figura 2.1: *Discretização do problema unidimensional*

Desta forma $x_0 < x_1 < \dots < x_{i-1} < x_i < x_{i+1} < \dots < x_{n+1}$ com $n + 2$ pontos igualmente espaçados e observe que $x_i = x_{i-1} + h$ para $1 \leq i \leq n$.

De posse desse conjunto, o qual denominado malha uniforme de espaçamento h , utilizou-se a expansão em Série de Taylor para a aproximação das equações diferenciais parciais, pois a ideia deste método é fazer a substituição das derivadas por diferenças finitas.

Supondo que f seja uma função contínua no intervalo $[a, b]$ e que possua derivadas n -ésima contínuas neste intervalo para todo $x \in [a, b]$. Expandindo em Série de Taylor, onde $\Delta x = x - x_0$.

$$f(x) = f(x_0) + (\Delta x)f_x(x) + \frac{(\Delta x)^2}{2!}f_{xx}(x) + \frac{(\Delta x)^3}{3!}f_{xxx}(x) + \dots \quad (2.7)$$

Determina-se a derivada de uma função f em um determinado ponto x_i expandindo em Série de Taylor neste mesmo ponto tem-se

$$f(x_i + \Delta x) = f(x_i) + (\Delta x)f_x(x_i) + \frac{(\Delta x)^2}{2!}f_{xx}(x_i) + \frac{(\Delta x)^3}{3!}f_{xxx}(x_i) + \dots \quad (2.8)$$

Portanto, isola-se a primeira derivada e re-arranjando os termos obtém-se

$$f_x(x_i) = \frac{f(x_i + \Delta x) - f(x_i)}{\Delta x} + \left[-\frac{\Delta x}{2!}f_{xx}(x_i) - \frac{(\Delta x)^2}{3!}f_{xxx}(x_i) - \dots \right] \quad (2.9)$$

haja visto que, a expressão (2.9) indica a derivada primeira

$$f_x(x_i) = \frac{f(x_i + \Delta x) - f(x_i)}{\Delta x} \quad (2.10)$$

os demais termos são denominados erro de truncamento local (ETL)

$$ETL = \left[-\frac{\Delta x}{2!} f_{xx}(x_i) - \frac{(\Delta x)^2}{3!} f_{xxx}(x_i) - \dots \right] \quad (2.11)$$

O erro de truncamento local deve tender a zero quando o número de pontos da malha tender ao infinito [Fortuna(2000)]. Aproxima-se as derivadas por meio de diferenças finitas e usou-se o método diferenças para frente a expressão (2.10) torna-se

$$f_x(x_i) = \frac{f_{i+1} - f_i}{\Delta x} + \mathcal{O}(\Delta x) \quad (2.12)$$

Do mesmo modo pode-se obter uma aproximação de diferenças finitas utilizando diferenças para trás.

Para as derivadas de segunda ordem é necessário as seguintes expansões em Série de Taylor

$$f(x_i + \Delta x) = f(x_i) + \Delta x f_x(x_i) + \frac{(\Delta x)^2}{2!} f_{xx}(x_i) + \mathcal{O}(\Delta x)^3 \quad (2.13)$$

$$f(x_i - \Delta x) = f(x_i) - \Delta x f_x(x_i) + \frac{(\Delta x)^2}{2!} f_{xx}(x_i) + \mathcal{O}(\Delta x)^3 \quad (2.14)$$

Combinou-se as equações (2.13) e (2.14) e reorganizando os termos obtém-se a derivada primeira e a equação torna-se

$$f_x(x_i) = \frac{f_{i+1} - f_{i-1}}{2\Delta x} + \mathcal{O}(\Delta x)^2 \quad (2.15)$$

De modo análogo obtém-se a equação da derivada segunda conforme a equação

$$f_x(x_i) = \frac{f_{i+1} - 2f_i + f_{i-1}}{(\Delta x)^2} + \mathcal{O}(\Delta x)^2 \quad (2.16)$$

Entretanto, estas aproximações apresentadas são denominadas diferenças centrais.

Do nosso problema original (2.1) a derivada de segunda ordem para o caso unidimensional é dada por

$$\left[\frac{\partial^2 p}{\partial x^2} \right]_i \approx \frac{p_{i+1} - 2p_i + p_{i-1}}{(\Delta x)^2} \quad (2.17)$$

onde $\Delta x = h$ sendo este o espaçamento na malha uniforme e p_i é a aproximação de $p(x_i)$.

Portanto discretizando-se o problema contínuo (2.1) têm-se

$$\frac{p_{i+1} - 2p_i + p_{i-1}}{h^2} = f_i \quad (2.18)$$

para $1 \leq i \leq n$ que fornece o sistema de equações lineares

$$p_0 - 2p_1 + p_2 = h^2 f_1 \quad (2.19)$$

$$p_1 - 2p_2 + p_3 = h^2 f_2 \quad (2.20)$$

$$\vdots \quad (2.21)$$

$$p_{n-2} - 2p_{n-1} + p_n = h^2 f_{n-1} \quad (2.22)$$

$$p_{n-1} - 2p_n + p_{n+1} = h^2 f_n \quad (2.23)$$

com n equações e $n + 2$ incógnitas.

Para resolver este problema aplicam-se condições de contorno discretas.

2.2.1 Condições de contorno de Dirichlet

A discretização das condições de Dirichlet são dadas por

$$p_0 = \alpha \quad \text{e} \quad p_{n+1} = \beta. \quad (2.24)$$

Ao colocar o sistema (2.19) na forma matricial obtém-se $A\vec{p} = \vec{b}$. Considerando que a matriz A contém os pontos do contorno, esta é de ordem

$(n + 2) \times (n + 2)$, ou seja,

$$A = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 1 & -2 & 1 & \\ & \ddots & \ddots & \ddots \\ & & 1 & -2 & 1 \\ 0 & \dots & & 0 & 1 \end{bmatrix}, \quad (2.25)$$

$$\vec{p} = \begin{bmatrix} p_0 \\ p_1 \\ \vdots \\ p_n \\ p_{n+1} \end{bmatrix} \quad \text{e} \quad \vec{b} = \begin{bmatrix} \alpha \\ h^2 f_1 \\ \vdots \\ h^2 f_n \\ \beta \end{bmatrix}. \quad (2.26)$$

Como a solução de p_0 e p_{n+1} já está determinada por (2.24), elimina-se do problema substituindo na equação 1 e n . Assim o sistema assume a forma matricial, $A_D \vec{p} = \vec{b}$ onde

$$A_D = \begin{bmatrix} -2 & 1 & \dots & 0 \\ 1 & -2 & 1 & \\ & \ddots & \ddots & \ddots \\ & & 1 & -2 & 1 \\ 0 & \dots & & 1 & -2 \end{bmatrix}, \quad \vec{p} = \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_{n-1} \\ p_n \end{bmatrix} \quad \text{e} \quad \vec{b} = \begin{bmatrix} h^2 f_1 - \alpha \\ h^2 f_2 \\ \vdots \\ h^2 f_{n-1} \\ h^2 f_n - \beta \end{bmatrix}. \quad (2.27)$$

Assim, para a solução do problema deve-se encontrar um vetor $\vec{p} = (p_1, \dots, p_n)$ que satisfaz a equação matricial (2.27).

2.2.2 Condições de contorno de Neumann

Ao resolver a equação (2.1) com condições de contorno de Neumann discretizamos-a por

$$\frac{p_1 - p_0}{h} = 0, \quad \text{o que implica que } p_0 = p_1, \quad (2.28)$$

e

$$\frac{p_{n+1} - p_n}{h} = 0, \quad \text{o que implica que } p_{n+1} = p_n. \quad (2.29)$$

A partir de (2.1) obtém-se um sistema de equações lineares, idêntico ao sistema (2.19). Considerando as condições de contorno (2.28) e (2.29) este sistema admite a forma matricial

$$\begin{bmatrix} -1 & 1 & & \dots & 0 \\ 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & & 1 & -2 & 1 \\ 0 & \dots & & 1 & -1 \end{bmatrix} \cdot \begin{bmatrix} p_0 \\ p_1 \\ \vdots \\ p_n \\ p_{n+1} \end{bmatrix} = \begin{bmatrix} 0 \\ h^2 f_1 \\ \vdots \\ h^2 f_n \\ 0 \end{bmatrix}. \quad (2.30)$$

Observe que a última linha poderia ser escrita como

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

porém, dessa forma o sistema não seria simétrico.

Note também que pode-se eliminar as incógnitas p_0 e p_{n+1} do problema. Fez-se isso somando a linha 1 e a linha 2 (semelhante para outro extremo) de tal forma a obter o sistema $A_N \vec{p} = \vec{b}$, onde

$$A_N = \begin{bmatrix} -1 & 1 & & \dots & 0 \\ 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & & 1 & -2 & 1 \\ 0 & \dots & & 1 & -1 \end{bmatrix} \quad (2.31)$$

é uma matriz de ordem $n \times n$,

$$\vec{p} = \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_{n-1} \\ p_n \end{bmatrix} \quad \text{e} \quad \vec{b} = \begin{bmatrix} h^2 f_1 \\ h^2 f_2 \\ \vdots \\ h^2 f_{n-1} \\ h^2 f_n \end{bmatrix}. \quad (2.32)$$

Dessa forma, o problema de encontrar uma aproximação para a função $p(x)$ que satisfaz a equação (2.1) se reduz em encontrar um vetor $\vec{p} = (p_1, \dots, p_n)$ que satisfaz a equação matricial (2.30).

Para as condições de contorno de Neumann note que mesmo que seja considerado apenas os nós interiores da malha a matriz de discretização é idêntica a matriz A de (2.30), porém a sua ordem é reduzida, ou seja, A_N passa ser de ordem $n \times n$.

Uma versão modificada para a representação das condições de contorno de Neumann pode ser feita definindo

$$\frac{2p_1 - 2p_0}{h} = 0, \quad \text{o que implica que} \quad 2p_0 = 2p_1, \quad (2.33)$$

e

$$\frac{2p_{n+1} - 2p_n}{h} = 0, \quad \text{o que implica que} \quad 2p_{n+1} = 2p_n. \quad (2.34)$$

A partir dessas equações montou-se um sistema $A_M \vec{p} = \vec{b}$ semelhante a (2.31) e (2.32) com matrizes

$$A_M = \begin{bmatrix} -2 & 2 & & \dots & 0 \\ 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & & 1 & -2 & 1 \\ 0 & \dots & & 2 & -2 \end{bmatrix} \quad (2.35)$$

onde A_M é uma matriz de ordem $n \times n$,

$$\vec{p} = \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_{n-1} \\ p_n \end{bmatrix} \quad \text{e} \quad \vec{b} = \begin{bmatrix} h^2 f_1 \\ h^2 f_2 \\ \vdots \\ h^2 f_{n-1} \\ h^2 f_n \end{bmatrix}. \quad (2.36)$$

2.3 O modelo discreto bidimensional

O produto de Kronecker será útil para descrever problemas lineares providos de duas ou mais dimensões porque permite representações mais compactas.

Definição 2.3.1. *Se $A \in \mathbb{R}^{m \times n}$ e $B \in \mathbb{R}^{p \times q}$, então o produto de Kronecker de A e B é definido como a matriz*

$$A \otimes B = \begin{bmatrix} a_{11}B & \dots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \dots & a_{mn}B \end{bmatrix} \in \mathbb{R}^{mp \times nq}. \quad (2.37)$$

Definição 2.3.2. *Se $A \in \mathbb{R}^{n \times n}$ e $B \in \mathbb{R}^{m \times m}$. Então a soma de Kronecker de A e B , definida por $A \oplus B$ é uma matriz $mn \times mn$ onde $A \oplus B = (I_m \otimes A) + (B \otimes I_n)$. Note que, em geral, $A \oplus B \neq B \oplus A$.*

Em um modelo bidimensional discreto ao invés de trabalhar com o domínio contínuo considerou-se apenas um conjunto finito de pontos em Ω . Considere um domínio quadrado onde $\Omega = (0, 1) \times (0, 1)$. O domínio pode ser discretizado conforme a figura 2.2.

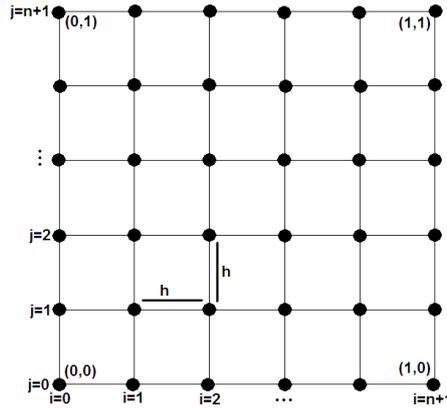


Figura 2.2: *Discretização do problema bidimensional*

Note que o espaçamento horizontal é dado por $\Delta x = \frac{1}{n+1}$ e o espaçamento vertical por $\Delta y = \frac{1}{n+1}$. Neste trabalho considerou-se $\Delta x = \Delta y = h$, ou seja,

espaçamento uniforme. Assim a malha bidimensional uniforme será definida como

$$\begin{aligned} x_{i,j} &= ih, & \text{para } 0 \leq i, j \leq n+1 \\ y_{i,j} &= jh, & \text{para } 0 \leq i, j \leq n+1 \end{aligned} \quad (2.38)$$

e $p_{i,j}$ será a aproximação para $p(x_{i,j}, y_{i,j})$.

Em seguida, aproximou-se as derivadas parciais de segunda ordem para o caso bidimensional através do método diferenças finitas centradas e estas tornam-se

$$\left[\frac{\partial^2 p}{\partial x^2} \right]_{i,j} \approx \frac{p_{i+1,j} - 2p_{i,j} + p_{i-1,j}}{(\Delta x)^2}, \quad (2.39)$$

$$\left[\frac{\partial^2 p}{\partial y^2} \right]_{i,j} \approx \frac{p_{i,j+1} - 2p_{i,j} + p_{i,j-1}}{(\Delta y)^2}. \quad (2.40)$$

Deste modo aproximou-se Δp de modo que se obtenha a fórmula de cinco pontos na qual nos fornece a seguinte expressão

$$\Delta p = p_{xx} + p_{yy} \approx \frac{1}{h^2} (p_{i-1,j} + p_{i,j-1} - 4p_{i,j} + p_{i+1,j} + p_{i,j+1}). \quad (2.41)$$

Portanto a equação de Poisson (2.1) discretizada para o ponto (x_{ij}, y_{ij}) torna-se

$$\frac{p_{i+1,j} + p_{i-1,j} - 4p_{i,j} + p_{i,j+1} + p_{i,j-1}}{h^2} = f_{i,j}. \quad (2.42)$$

O sistema de equações lineares pode ser determinado a partir de (2.42), para tal usou-se a ordem lexicográfica. Nesta ordem, os pontos da malha são percorridos linha por linha, da esquerda para a direita, de baixo para cima, ou seja,

$$p_{0,1}, p_{0,2}, \dots, p_{0,n}, p_{1,0}, p_{1,1}, \dots, p_{1,n}, \dots, \dots, p_{n,1}, p_{n,2}, \dots, p_{n,n}. \quad (2.43)$$

A escolha de uma ordenação para os pontos da malha é importante na resolução de sistemas lineares, pois de acordo com [Biezuner(2007)] e [Fonseca(2009)] como existem várias ordenações possíveis, existem várias matrizes associadas.

Assim, utilizando (2.42) e a ordenação lexicográfica e obtém-se o seguinte sistema

$$\begin{aligned}
p_{0,1} + p_{1,0} - 4p_{1,1} + p_{2,1} + p_{1,2} &= h^2 f_{1,1} \\
p_{0,2} + p_{1,1} - 4p_{1,2} + p_{2,2} + p_{1,3} &= h^2 f_{1,2} \\
&\vdots \\
p_{i-1,j} + p_{i,j-1} - 4p_{i,j} + p_{i+1,j} + p_{i,j+1} &= h^2 f_{i,j} \\
&\vdots \\
p_{n-1,n} + p_{n,n-1} - 4p_{n,n} + p_{n+1,n} + p_{n+1,n} &= h^2 f_{n,n}.
\end{aligned} \tag{2.44}$$

Para que seja possível resolver o problema é necessário transformá-lo num sistema de equações lineares $A\vec{p} = \vec{b}$ cuja solução será uma aproximação da solução \vec{p} de (2.1) assim deve-se determinar condições de contorno adequadas para a resolução deste.

2.3.1 O problema de Dirichlet Bidimensional

Para resolver o problema de Dirichlet para a equação de Poisson (2.1) no quadrado unitário $\Omega = (0, 1) \times (0, 1)$ usou-se as seguintes condições de contorno

$$p = g \quad \text{em } \partial\Omega. \tag{2.45}$$

Discretizando (2.45) obtém-se que $p_{i,j} = g_{i,j}$ em todo o contorno. Desta forma os contornos sul e norte são dados por

$$\begin{aligned}
p_{i,0} &= g_{i,0} \quad \text{para } 0 \leq i \leq n+1 \\
p_{i,n+1} &= g_{i,n+1}
\end{aligned} \tag{2.46}$$

e os contornos leste e oeste são

$$\begin{aligned}
p_{0,j} &= g_{0,j} \quad \text{para } 0 \leq j \leq n+1 \\
p_{n+1,j} &= g_{n+1,j}.
\end{aligned} \tag{2.47}$$

A matriz do sistema (2.44) juntamente com (2.46) e (2.47) é de ordem $(n+2)^2 \times (n+2)^2$. Considerando todos os pontos da malha (inclusive os pontos do contorno), a matriz do sistema (2.44) com condições de Dirichlet dados por (2.45) é dada por

$$A = \begin{bmatrix} I & & \dots & 0 \\ E & D & E & \\ & \ddots & \ddots & \ddots \\ & & E & D & E \\ 0 & \dots & & & I \end{bmatrix} \quad (2.48)$$

onde I é a matriz identidade de tamanho $(n+2) \times (n+2)$. E as matrizes D e E ambas com dimensões $(n+2) \times (n+2)$ são

$$D = \begin{bmatrix} 1 & & \dots & 0 \\ 1 & -4 & 1 & \\ & \ddots & \ddots & \ddots \\ & & 1 & -4 & 1 \\ 0 & \dots & & & 1 \end{bmatrix} \quad \text{e} \quad E = \begin{bmatrix} 0 & & \dots & 0 \\ & 1 & & \\ & & \ddots & \\ & & & 1 \\ 0 & \dots & & 0 \end{bmatrix}. \quad (2.49)$$

Eliminando as equações para os pontos no contorno a matriz A passa a ser da seguinte forma

$$L_D = A = \begin{bmatrix} B & I & & \dots & 0 \\ I & B & I & & \\ & \ddots & \ddots & \ddots & \\ & & I & B & I \\ 0 & \dots & & I & B \end{bmatrix} \quad (2.50)$$

onde I é a matriz identidade $n \times n$ e B é a matriz $n \times n$ dada por

$$B = \begin{bmatrix} -4 & 1 & & \dots & 0 \\ 1 & -4 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -4 & 1 \\ 0 & \dots & & 1 & -4 \end{bmatrix}. \quad (2.51)$$

A matriz L_D é a matriz que geralmente aparece na literatura como em [Strikwerda(2004)], [Young(1988)], [Datta(1994)] entre outros. Assim, o sistema linear é de $n^2 \times n^2$ equações com o mesmo número de incógnitas.

Utilizando o produto de Kronecker a matriz L_D (do caso bidimensional) é reescrita usando a matriz A_D em (2.27) (do caso unidimensional). Tomando

$$I \otimes A_D = \begin{bmatrix} 1A_D & \dots & 0 \\ & \ddots & \\ 0 & \dots & 1A_D \end{bmatrix} \quad (2.52)$$

e

$$A_D \otimes I = \begin{bmatrix} -2I & 1I & \dots & 0 \\ 1I & -2I & 1I & \\ & \ddots & \ddots & \ddots \\ & & 1I & -2I & 1I \\ 0 & \dots & & 1I & -2I \end{bmatrix} \quad (2.53)$$

Somando (2.52) e (2.53) e usando a definição 2.3.2 sabe-se que

$$A_D \oplus A_D = (I \otimes A_D) + (A_D \otimes I) \quad (2.54)$$

ou,

$$L_D = A_D \oplus A_D = \begin{bmatrix} A_D - 2I & I & \dots & 0 \\ I & A_D - 2I & I & \\ & \ddots & \ddots & \ddots \\ & & I & A_D - 2I & I \\ 0 & \dots & & I & A_D - 2I \end{bmatrix} \quad (2.55)$$

onde o elemento $A_D - 2I$ da matriz (2.55) é dado por

$$A_D - 2I = \begin{bmatrix} -4 & 1 & \dots & 0 \\ 1 & -4 & 1 & \\ & \ddots & \ddots & \ddots \\ & & 1 & -4 & 1 \\ 0 & \dots & & 1 & -4 \end{bmatrix}. \quad (2.56)$$

Desta forma têm-se que $A_D - 2I = B$, por (2.51) e (2.56), ou seja, a matriz em duas dimensões com condições de Dirichlet pode ser escrita a partir da matriz unidimensional como $L_D = A_D \oplus A_D$.

2.3.2 O problema de Neumann Bidimensional

O problema de Neumann para a equação de Poisson definido sobre um quadrado unitário $\Omega = (0, 1) \times (0, 1)$, sob as condições de contorno

$$\frac{\partial p}{\partial n} = 0 \quad \text{sobre } \partial\Omega. \quad (2.57)$$

Discretizando (2.57) nos contornos sul e norte como

$$\begin{aligned} p_{i,0} &= p_{i,1} & \text{para } 0 \leq i \leq n+1 \\ p_{i,n+1} &= p_{i,n} \end{aligned} \quad (2.58)$$

e nos contornos leste e oeste como

$$\begin{aligned} p_{0,j} &= p_{1,j} & \text{para } 0 \leq j \leq n+1 \\ p_{n+1,j} &= p_{n,j}. \end{aligned} \quad (2.59)$$

Combinando as equações (2.44), (2.58) e (2.59) determinou-se a matriz de discretização L_N .

Assim, quando considera-se os pontos do contorno a matriz de discretização A torna-se

$$A = \begin{bmatrix} -I & E & \dots & 0 \\ E & D & E & \\ & \ddots & \ddots & \ddots \\ & & E & D & E \\ 0 & \dots & E & -I \end{bmatrix} \quad (2.60)$$

com

$$D = \begin{bmatrix} -1 & 1 & \dots & 0 \\ 1 & -4 & 1 & \\ & \ddots & \ddots & \ddots \\ & & 1 & -4 & 1 \\ 0 & \dots & & -1 & 1 \end{bmatrix} \quad \text{e} \quad E = \begin{bmatrix} 0 & & \dots & 0 \\ & 1 & & \\ & & \ddots & \\ & & & 1 \\ 0 & \dots & & 0 \end{bmatrix} \quad (2.61)$$

onde A é de ordem $(n+2)^2 \times (n+2)^2$ e, D e E são de ordem $(n+2) \times (n+2)$.

Costuma-se também representar A através da notação estêncil

$$\begin{bmatrix} 1 \\ 1 & -4 & 1 \\ 1 \end{bmatrix} \quad (2.62)$$

a qual consiste em dispor os coeficientes de $p_{i-1,j}$, $p_{i,j-1}$, $p_{i,j}$, $p_{i+1,j}$, $p_{i,j+1}$ da fórmula dos cinco pontos em (2.41) sendo que estes devem-se estar posicionado na mesma posição dos pontos $x_{i-1,j}$, $x_{i,j-1}$, $x_{i,j}$, $x_{i+1,j}$, $x_{i,j+1}$ na malha cartesiana.

Assim, os estênceis localizados próximo a fronteira (esquerda, direita, superior e inferior), são respectivamente

$$\begin{bmatrix} 1 \\ 0 & -3 & 1 \\ 1 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ 1 & -3 & 0 \\ 1 \end{bmatrix}, \quad (2.63)$$

$$\begin{bmatrix} 0 \\ 1 & -3 & 1 \\ 1 \end{bmatrix} \quad \text{e} \quad \begin{bmatrix} 1 \\ 1 & -3 & 1 \\ 0 \end{bmatrix}. \quad (2.64)$$

Note que para os pontos localizados no canto da malha, substituindo os valores da fronteira obtém-se

$$\begin{bmatrix} 1 \\ 0 & -2 & 1 \\ 0 \end{bmatrix} \quad \text{e} \quad \begin{bmatrix} 1 \\ 1 & -2 & 0 \\ 0 \end{bmatrix}. \quad (2.65)$$

Após a definição de todos os estênceis montou-se L_N na sua forma matricial, onde a mesma tem dimensão $n^2 \times n^2$ e é composta por blocos de dimensão $n \times n$ como

$$L_N = A = \begin{bmatrix} B & I & \dots & 0 \\ I & C & I & \\ & \ddots & \ddots & \ddots \\ & & I & C & I \\ 0 & \dots & & I & B \end{bmatrix}, \quad (2.66)$$

onde I é a matriz identidade,

$$B = \begin{bmatrix} -2 & 1 & \dots & 0 \\ 1 & -3 & 1 & \\ & \ddots & \ddots & \ddots \\ & & 1 & -3 & 1 \\ 0 & \dots & & 1 & -2 \end{bmatrix} \quad \text{e} \quad C = \begin{bmatrix} -3 & 1 & \dots & 0 \\ 1 & -4 & 1 & \\ & \ddots & \ddots & \ddots \\ & & 1 & -4 & 1 \\ 0 & \dots & & 1 & -3 \end{bmatrix}. \quad (2.67)$$

O sistema $L_N \vec{p} = \vec{b}$ pode ser reescrito utilizando o produto de Kronecker. Para isso, tome A_N como em (2.31) e

$$I \otimes A_N = \begin{bmatrix} 1A_N & \dots & 0 \\ & \ddots & \\ 0 & \dots & 1A_N \end{bmatrix} \quad (2.68)$$

e

$$A_N \otimes I = \begin{bmatrix} -1I & 1I & \dots & 0 \\ 1I & -2I & 1I & \\ & \ddots & \ddots & \ddots \\ & & 1I & -2I & 1I \\ 0 & \dots & & 1I & -1I \end{bmatrix}. \quad (2.69)$$

Considerando a definição 2.3.2, sabe-se que

$$L_N = (I \otimes A_N) + (A_N \otimes I) = \begin{bmatrix} A_N - I & I & \dots & 0 \\ I & A_N - 2I & I & \\ & \ddots & \ddots & \ddots \\ & & I & A_N - 2I & I \\ 0 & \dots & & I & A_N - I \end{bmatrix} \quad (2.70)$$

sendo que

$$A_N - I = \begin{bmatrix} -2 & 1 & \dots & 0 \\ 1 & -3 & 1 & \\ & \ddots & \ddots & \ddots \\ & & 1 & -3 & 1 \\ 0 & \dots & & 1 & -2 \end{bmatrix} \quad (2.71)$$

e

$$A_N - 2I = \begin{bmatrix} -3 & 1 & \dots & 0 \\ 1 & -4 & 1 & \\ & \ddots & \ddots & \ddots \\ & & 1 & -4 & 1 \\ 0 & \dots & & 1 & -3 \end{bmatrix}. \quad (2.72)$$

Logo, $A_N - I = B$ e $A_N - 2I = C$ como em (2.67), respectivamente.

Portanto $L_N = A_N \oplus A_N$.

Para o problema de Neumann a qual modificou-se as condições de contorno, observe que

$$A_M = \begin{bmatrix} -2 & 2 & \dots & 0 \\ 1 & -2 & 1 & \\ & \ddots & \ddots & \ddots \\ & & 1 & -2 & 1 \\ 0 & \dots & & 2 & -2 \end{bmatrix} \quad (2.73)$$

é de ordem $n \times n$ no problema unidimensional.

A partir de (2.73) e através das definições 2.3.1 e 2.3.2 determinou-se a matriz L_M do caso bidimensional (veja [Plemmons(1976)]). Deste modo,

$$A_M \otimes I = \begin{bmatrix} -2I & 2I & \dots & 0 \\ I & -2I & I & \\ & \ddots & \ddots & \ddots \\ & & I & -2I & I \\ 0 & \dots & & 2I & -2I \end{bmatrix} \quad (2.74)$$

e $I \otimes A_M$ é dada por (2.68). Somando obtém-se

$$(I \otimes A_M) + (A_M \otimes I) = \begin{bmatrix} A_M - 2I & 2I & \dots & 0 \\ I & A_M - 2I & I & \\ & \ddots & \ddots & \ddots \\ & & I & A_M - 2I & I \\ 0 & \dots & & 2I & A_M - 2I \end{bmatrix} \quad (2.75)$$

onde

$$A_M - 2I = \begin{bmatrix} -4 & 2 & \dots & 0 \\ 1 & -4 & 1 & \\ & \ddots & \ddots & \ddots \\ & & 1 & -4 & 1 \\ 0 & \dots & & 2 & -4 \end{bmatrix}. \quad (2.76)$$

Pode-se definir então que $L_M = A_M \oplus A_M$ como a matriz de discretização do problema de Neumann para o caso bidimensional.

2.4 Autovalores da Matriz de Discretização

Existem muitos modelos matemáticos que envolvem matrizes tridiagonais semelhantes a matriz

$$T = \begin{bmatrix} -\alpha + b & c & \dots & 0 \\ a & b & c & \\ & \ddots & \ddots & \ddots \\ & & a & b & c \\ 0 & \dots & a & -\beta + b \end{bmatrix}. \quad (2.77)$$

Os autovalores de T são dados por uma fórmula geral que, de acordo com [Yueh(2005)], quando, $\alpha = \beta = 0$ são dados por

$$\lambda_k = b + 2\sqrt{ac} \cos\left(\frac{k\pi}{n+1}\right) \quad \text{para } 1 \leq k \leq n \quad (2.78)$$

e quando $\alpha = \beta = -\sqrt{ac} \neq 0$ estes são

$$\lambda_k = b + 2\sqrt{ac} \cos\left[\frac{(k-1)\pi}{n}\right] \quad \text{para } 1 \leq k \leq n. \quad (2.79)$$

Teorema 2.4.1. *Os autovalores de A_D são*

$$\lambda_k = -2 + 2 \cos\left(\frac{k\pi}{n+1}\right), \quad \text{para } 1 \leq k \leq n. \quad (2.80)$$

Demonstração. Seja $a = c = 1$, $b = -2$ e $\alpha = \beta = 0$, então $T = A_D$ e os autovalores de A_D são dados pela fórmula (2.80) segundo [Yueh(2005)]. \square

Teorema 2.4.2. *Os autovalores de A_N são*

$$\lambda_k = -2 + 2 \cos\left[\frac{(k-1)\pi}{n}\right], \quad \text{para } 1 \leq k \leq n. \quad (2.81)$$

Demonstração. Segundo [Yueh(2005)], com $a = c = 1$, $b = -2$ e $\alpha = \beta = -\sqrt{ac}$ obtém-se os autovalores em (2.81). \square

O teorema 2.4.3 revela que os autovalores da matriz de discretização bidimensional são justamente dados pela soma dos autovalores da matriz de discretização unidimensional.

Teorema 2.4.3. *Se $A \in \mathbb{R}^{n \times n}$ tem autovalores λ_i e $B \in \mathbb{R}^{m \times m}$ tem autovalores μ_j , então a soma de kronecker $A \oplus B = (I_m \otimes A) + (B \otimes I_n)$ tem mn autovalores*

$$\lambda_1 + \mu_1, \dots, \lambda_1 + \mu_m, \lambda_2 + \mu_1, \dots, \lambda_2 + \mu_m, \dots, \lambda_n + \mu_m \quad (2.82)$$

Usou-se as propriedades do produto de Kronecker e a partir da matriz de discretização do problema de Dirichlet para o caso unidimensional obtém-se a matriz do caso bidimensional, ou seja, a matriz (2.50).

Teorema 2.4.4. *Os autovalores de $L_D = A_D \oplus A_D$ são*

$$\lambda_{i,j} = -2 + 2 \cos \left(\frac{i\pi}{n+1} \right) - 2 + 2 \cos \left(\frac{j\pi}{n+1} \right) \quad \text{para } 1 \leq i, j \leq n. \quad (2.83)$$

Demonstração. Através do teorema 2.4.1 e 2.4.3 segue a prova. \square

Teorema 2.4.5. *Os autovalores de $L_N = A_N \oplus A_N$ são*

$$\lambda_{i,j} = -2 + 2 \cos \left[\frac{(i-1)\pi}{n} \right] - 2 + 2 \cos \left[\frac{(j-1)\pi}{n} \right] \quad \text{para } 1 \leq i, j \leq n. \quad (2.84)$$

Demonstração. Através do teorema 2.4.2 e 2.4.3 segue a prova. \square

Os autovalores de A_M em (2.73) são dados por

$$\lambda_k = -2 + 2 \cos \left[\frac{(k-1)\pi}{n-1} \right] \quad \text{para } 1 \leq k \leq n. \quad (2.85)$$

Assim, através do teorema 2.4.3 determinou-se os autovalores de $L_M = A_M \oplus A_M$ em (2.75) e estes são dados por

$$\lambda_{i,j} = -2 + 2 \cos \left[\frac{(i-1)\pi}{n-1} \right] - 2 + 2 \cos \left[\frac{(j-1)\pi}{n-1} \right] \quad \text{para } 1 \leq i, j \leq n. \quad (2.86)$$

2.5 Existência de soluções

Nessa seção apresenta-se sob quais condições os sistemas lineares obtidos nesse capítulo tem solução. O problema de Dirichlet é mais fácil de estudar pois não apresenta autovalor igual a zero.

Teorema 2.5.1. *Seja a matriz A_D dada por (2.27). O sistema $A_D \vec{p} = \vec{b}$ sempre tem solução.*

Demonstração. Pelo teorema 2.4.1, a matriz A_D não possui autovalor igual a zero, portanto é inversível e o sistema sempre tem solução. Observe que o menor autovalor em módulo é obtido quando $k = 1$ em (2.80), ou seja,

$$\lambda_1 = -2 + 2 \cos \left(\frac{\pi}{n+1} \right). \quad (2.87)$$

□

Teorema 2.5.2. *Seja $L_D = A_D \oplus A_D$ dada por (2.55). O sistema $L_D \vec{p} = \vec{b}$ sempre tem solução.*

Demonstração. Como no teorema anterior, e pelo teorema 2.4.2, a matriz L_D não possui autovalor igual a zero, portanto é inversível e o sistema sempre tem solução.

□

Para o problema de Neumann, a matriz A_N será singular pois tem um autovalor $\lambda = 0$ quando $k = 1$ em (2.81) (da mesma forma para L_N quando $i = j = 1$ em (2.84)).

O autovetor associado a $\lambda = 0$ é dado pelo vetor constante $\vec{1} = [1 \ 1 \dots \ 1 \ 1]$ que pode ser facilmente verificado com $A_N \cdot \vec{1} = 0 \cdot \vec{1}$, ou

$$\begin{bmatrix} -1 & 1 & & \dots & 0 \\ 1 & -2 & 1 & & \\ & & \ddots & \ddots & \ddots \\ & & & 1 & -2 & 1 \\ 0 & \dots & & 1 & -1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 1 \end{bmatrix} = 0 \cdot \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 1 \end{bmatrix}. \quad (2.88)$$

Isso implica que se o sistema $A_N \vec{p} = \vec{b}$ tem solução \vec{p} , então $\vec{p} + \alpha \vec{1}$ também será solução desse sistema, ou seja,

$$A_N(\vec{p} + \alpha \vec{1}) = A_N \vec{p} + \alpha A_N \vec{1} = \vec{b} + 0 = \vec{b}. \quad (2.89)$$

Teorema 2.5.3. *Seja A_N e \vec{b} dados por (2.31) e (2.32). O sistema*

$$A_N \vec{p} = \vec{b} \quad (2.90)$$

terá solução se e somente se

$$\sum_{i=1}^n b_i = 0. \quad (2.91)$$

Demonstração. A prova segue como em [Strikwerda(2004), p. 366-367]. \square

É razoável assumir que a dimensão do espaço nulo é 1 (já que isto é verdade para o caso contínuo). Como o posto do espaço coluna é igual ao posto do espaço linha, então existe um vetor $\vec{y} \neq \vec{0}$ tal que $\vec{y}^T A$ seja $\vec{0}$. Esse vetor representa a restrição que \vec{b} no sistema deve satisfazer para que a solução exista. Tem-se então que

$$0 = (\vec{y}^T A) \vec{p} = \vec{y}^T (A \vec{p}) = \vec{y}^T \vec{b} \quad (2.92)$$

ou seja, $\vec{y}^T \vec{b} = 0$ então existe uma solução \vec{p} para o sistema (2.90). Se A é simétrica (que é o nosso caso) então $\vec{y} = \vec{1}$ que pertence ao espaço nulo de A . Assim, a condição para que o sistema tenha solução é que

$$0 = \vec{1}^T \vec{b} = \sum_{i=1}^n b_i. \quad (2.93)$$

Note que para \vec{b} dado em (2.32) esta condição torna-se

$$\sum_{i=1}^n h^2 f_i = 0 \quad (2.94)$$

que pode ser interpretada como o equivalente discreto da condição de integrabilidade que é a equação (2.5) (multiplicada por h), ou seja,

$$\int_{\Omega} f \, d\Omega = 0. \quad (2.95)$$

Teorema 2.5.4. *Seja L_N dada por (2.66) o sistema*

$$L_N \vec{p} = \vec{b} \quad (2.96)$$

tem solução se e somente se

$$\sum_{i=1}^{n^2} b_i = 0. \quad (2.97)$$

Demonstração. A prova é semelhante ao caso 1D. Basta verificar que L_N é simétrica (veja (2.70)) e que $L_N \vec{1} = \vec{0}$, ou seja, $\vec{1}$ é um autovetor de L_N associado ao autovalor $\lambda = 0$. □

3 MÉTODOS ITERATIVOS CLÁSSICOS

Descreve-se neste capítulo métodos iterativos clássicos para a solução de sistemas lineares tais como o método de Jacobi, Gauss-Seidel e SOR. Apresenta-se também resultados sobre a convergência de tais métodos seguindo como referência os trabalhos [Saad(2000)], [Kelley(1995)] e [Young(1988)].

Os métodos iterativos são utilizados para encontrar a solução de um sistema de equações lineares descritas como

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

com n equações e n incógnitas.

Para o nosso objeto de estudo tal sistema surge da discretização das equações diferenciais parciais, onde A é uma matriz quadrada de ordem n , \vec{x} e \vec{b} são vetores de n elementos. Em geral, a matriz A que obtém-se é uma matriz grande e esparsa, pois esta é obtida por meio de um esquema de diferenças finitas.

Uma simples aproximação para uma solução iterativa de um sistema linear é reescrever (3.1) como uma iteração de ponto fixo linear [Kelley(1995)] de modo a obter uma equação iterativa como

$$\vec{x}^{k+1} = G\vec{x}^k + \vec{f} \quad (3.2)$$

que convirja para a solução do mesmo, onde $G \in \mathbb{R}^{n \times n}$ é a matriz de iteração do método iterativo (3.2) e $\vec{f} \in \mathbb{R}^n$.

O método iterativo (3.2) está relacionado ao sistema linear

$$(I - G)\vec{x} = \vec{f}. \quad (3.3)$$

Sendo que o conjunto solução $\mathcal{S}(A, \vec{b})$ de (3.1) está relacionado com o conjunto solução de $\mathcal{S}(I - G, \vec{f})$ de (3.3). Assim, se a sequência $\vec{x}_0, \vec{x}_1, \vec{x}_2, \dots$ definida

por (3.2) converge para o limite \hat{x} , então $\hat{x} \in \mathcal{S}(I - G, \vec{f})$. Portanto, a sequência $\vec{x}_0, \vec{x}_1, \vec{x}_2, \dots$ converge para a solução de (3.1) (detalhes, veja [Young(1988)]).

Através de um método iterativo obtém-se uma sequência de aproximações sucessivas $\vec{x}_0, \vec{x}_1, \vec{x}_2, \vec{x}_3, \dots$ para a solução do sistema original (3.1). Uma vez que \vec{x}^k esteja suficientemente próximo da solução exata de acordo com uma margem de tolerância aceita, para-se o processo de iteração e aceita-se \vec{x}^k como a solução aproximada adequada para o problema. Por exemplo, o critério de parada pode ser estabelecido através de uma cota de tolerância, ou seja, $\|x^{k+1} - x^k\| < \epsilon$.

3.1 Método de Jacobi

Existem várias maneiras de encontrar uma matriz G que satisfaça (3.2). Um método que decompõe a matriz A como a soma de duas ou mais matrizes como

$$A = M - N \quad (3.4)$$

é chamado método *splitting*. Nesse caso têm-se

$$A\vec{x} = \vec{b} \quad (3.5)$$

$$(M - N)\vec{x} = \vec{b} \quad (3.6)$$

$$M\vec{x} = N\vec{x} + \vec{b} \quad (3.7)$$

$$\vec{x} = M^{-1}N\vec{x} + M^{-1}\vec{b}. \quad (3.8)$$

Nesse tipo de método, $G = M^{-1}N$ onde M é não-singular e $\vec{f} = M^{-1}\vec{b}$.

O método de Jacobi por ser um método *splitting* nos permite decompor a matriz A como

$$A = M - N = D - (E + F) \quad (3.9)$$

onde D é a diagonal de A , $-E$ é a parte triangular estritamente inferior de A e $-F$ é a parte triangular estritamente superior de A , conforme ilustramos na figura 3.1 (nesse caso $M = D$ e $N = E + F$).

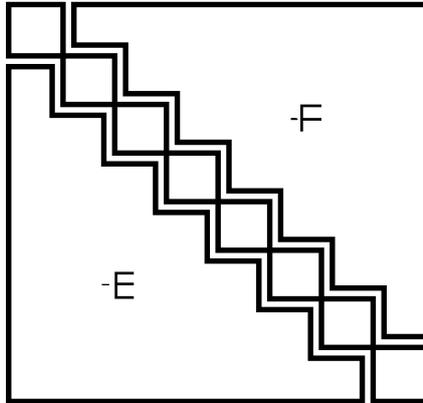


Figura 3.1: *Partição da matriz A*

Substituindo (3.9) em (3.5) obtemos

$$D\vec{x} = (E + F)\vec{x} + \vec{b}. \quad (3.10)$$

Deste modo, de acordo com [Saad(2000)] transformar a equação (3.10) numa equação de ponto fixo esta torna-se

$$D\vec{x}^{k+1} = (E + F)\vec{x}^k + \vec{b}, \quad (3.11)$$

ou ainda, isolando o termo \vec{x}^{k+1} ,

$$\vec{x}^{k+1} = D^{-1}(E + F)\vec{x}^k + D^{-1}\vec{b}. \quad (3.12)$$

Note que, para que a matriz D tenha inversa é necessário que $a_{ii} \neq 0$ para $1 \leq i \leq n$.

Compara-se (3.12) com a equação (3.2) e assim obtém-se

$$G_{Jac} = D^{-1}(E + F) \quad (3.13)$$

e

$$\vec{f} = D^{-1}\vec{b}. \quad (3.14)$$

3.2 Método de Gauss-Seidel

O método de Gauss-Seidel é também *splitting*, onde $M = D - E$ e $N = F$. Assim o método pode ser escrito como

$$(D - E)\vec{x}^{k+1} = F\vec{x}^k + \vec{b} \quad (3.15)$$

ou ainda,

$$\vec{x}^{k+1} = (D - E)^{-1}F\vec{x}^k + (D - E)^{-1}\vec{b}. \quad (3.16)$$

De modo análogo ao método de Jacobi, aplica-se o método de Gauss-Seidel uma condição necessária é que $a_{ii} \neq 0$ para $1 \leq i \leq n$.

Note que a matriz de iteração é

$$G_{GS} = (D - E)^{-1}F \quad (3.17)$$

e

$$\vec{f} = (D - E)^{-1}\vec{b}. \quad (3.18)$$

Também pode-se utilizar o método de outra forma na qual é conhecido como *backward* Gauss-Seidel que pode ser definido como

$$(D - F)\vec{x}^{k+1} = E\vec{x}^k + \vec{b} \quad (3.19)$$

ou,

$$\vec{x}^{k+1} = (D - F)^{-1}E\vec{x}^k + (D - F)^{-1}\vec{b}. \quad (3.20)$$

Nesse caso a matriz de iteração é dada por

$$G_{BGS} = (D - F)^{-1}E \quad (3.21)$$

e

$$\vec{f} = (D - F)^{-1}\vec{b}. \quad (3.22)$$

Um terceiro método pode ser obtido, aplica-se uma iteração do método de Gauss-Seidel pra frente seguida por uma iteração do método de Gauss-Seidel para trás. Esse método é chamado de Gauss-Seidel simétrico. Para este caso a matriz de iteração é definida como

$$G_{SGS} = G_{BGS}G_{GS} = (D - F)^{-1}E(D - E)^{-1}F. \quad (3.23)$$

3.3 Método das Sobre-relaxações Sucessivas - SOR

O método SOR é uma generalização do método de Gauss-Seidel, adiciona-se um parâmetro de relaxação ω e de fato constrói-se uma iteração sobre-relaxação sucessiva, ou seja, o método iterativo SOR. Uma das características do método iterativo SOR é extrapolar o valor de \vec{x}^{k+1} de tal forma que ele se aproxime mais rapidamente da solução numérica do sistema de equações [Fortuna(2000), Kelley(1995)].

Para a obtenção do coeficiente de relaxação do método SOR, multiplicou-se por ω a equação (3.1), obtendo

$$\omega A\vec{x} = \omega\vec{b} \quad (3.24)$$

e substituindo (3.9) em (3.24) têm-se

$$(\omega D - \omega E - \omega F)\vec{x} = \omega\vec{b}. \quad (3.25)$$

Pode-se reescrever o lado esquerdo de (3.25) obtendo

$$(\omega D - \omega E - \omega F) = (D - \omega E) - [\omega F + (1 - \omega)D], \quad (3.26)$$

assim

$$[(D - \omega E) - (\omega F + (1 - \omega)D)]\vec{x} = \omega\vec{b}. \quad (3.27)$$

De modo semelhante ao método de Jacobi e Gauss-Seidel obtém-se o método SOR

$$(D - \omega E)\vec{x}^{k+1} = [\omega F + (1 - \omega)D]\vec{x}^k + \omega\vec{b} \quad (3.28)$$

e, rearranjando os termos, têm-se

$$\vec{x}^{k+1} = G_{SOR}\vec{x}^k + (D - \omega E)^{-1}\omega\vec{b} \quad (3.29)$$

onde,

$$G_{SOR} = (D - \omega E)^{-1}[\omega F + (1 - \omega)D] \quad (3.30)$$

e

$$\vec{f} = \omega(D - \omega E)^{-1}\vec{b}. \quad (3.31)$$

Observe que, quando $\omega = 1$, o método iterativo SOR se reduz ao método de Gauss-Seidel.

Seguindo [Young(1988)] a matriz G_{SOR} (3.30) pode ser reescrita de outra maneira (multiplicando por D^{-1} no interior)

$$\begin{aligned} G_{SOR} &= (D^{-1}D - \omega D^{-1}E)^{-1}[\omega D^{-1}F + (1 - \omega)D^{-1}D] \\ &= (I - \omega D^{-1}E)^{-1}[\omega D^{-1}F + (1 - \omega)I] \end{aligned} \quad (3.32)$$

de modo análogo, reescreve-se a função \vec{f}

$$\vec{f} = \omega(D^{-1}D - \omega D^{-1}E)^{-1}D^{-1}\vec{b} \quad (3.33)$$

onde têm-se que

$$\vec{f} = \omega(I - \omega L)^{-1}D^{-1}\vec{b}. \quad (3.34)$$

Definindo as matrizes $D^{-1}E = L$ e $D^{-1}F = U$ obtém-se

$$G_{SOR} = (I - \omega L)^{-1}[\omega U + (1 - \omega)I] \quad (3.35)$$

e

$$\vec{f} = \omega(I - \omega L)^{-1}D^{-1}\vec{b}. \quad (3.36)$$

Substituindo (3.35) e (3.36) em (3.2) assim, a equação matricial para o método iterativo SOR torna-se

$$\vec{x}^{k+1} = (I - \omega L)^{-1}[\omega U + (1 - \omega)I]\vec{x}^k + (I - \omega L)^{-1}\omega D^{-1}\vec{b}. \quad (3.37)$$

A seguir apresenta-se o resultado geral da convergência da matriz de iteração gerada a partir dos métodos iterativos.

3.4 Convergência de Métodos Iterativos Lineares

Faz-se uma breve exposição de algumas definições e resultados importantes que se farão necessários posteriormente.

Definição 3.4.1. *Seja $T \in \mathbb{R}^{n \times n}$ uma matriz $n \times n$. Um vetor $\vec{v} \in \mathbb{R}^n$, $\vec{v} \neq \vec{0}$, é dito autovetor de T , se existir um $\lambda \in \mathbb{C}$ tal que $T\vec{v} = \lambda\vec{v}$. O escalar λ é denominado autovalor de T associado ao autovetor \vec{v} .*

Definição 3.4.2. *O raio espectral da matriz A denotado por $\rho(A)$ é definido como o máximo do módulo dos autovalores de A , ou seja,*

$$\rho(A) = \max_{\lambda \in \sigma(A)} |\lambda| \quad (3.38)$$

onde $\sigma(A)$ é o conjunto dos autovalores, também denominado de espectro de A .

Definição 3.4.3. *Se \vec{u} e \vec{v} são vetores no \mathbb{R}^n e se $\vec{u} \neq \vec{0}$, então*

$$\text{proj}_{\vec{u}}\vec{v} = \frac{\vec{v}^T \vec{u}}{\vec{u}^T \vec{u}} \vec{u} \quad (3.39)$$

As próximas três definições referem-se a convergência de sequências de matrizes.

Definição 3.4.4. *A sequência de matrizes A_1, A_2, A_3, \dots é convergente se $A_i \rightarrow O$ quando $i \rightarrow \infty$. Temos $\|A_i\| \rightarrow 0$ e O é a matriz nula.*

Definição 3.4.5. *A sequência de matrizes A_1, A_2, A_3, \dots é semi-convergente se $A_i \rightarrow A$ quando $i \rightarrow \infty$. Temos $\|A_i - A\| \rightarrow 0$.*

Definição 3.4.6. *Uma sequência de matrizes A_1, A_2, A_3, \dots é condicionalmente convergente se existe um subespaço $V \subseteq \mathbb{R}^n$ onde $\|A_i \vec{x} - A \vec{x}\| \rightarrow 0$, $\forall \vec{x} \in V$ quando $i \rightarrow \infty$.*

As três próximas definições referem-se a convergência de sequências de potências da matriz A , ou seja, A, A^2, A^3, \dots

Definição 3.4.7. *A matriz A é convergente se a sequência A, A^2, A^3, \dots é convergente.*

Definição 3.4.8. *A matriz A é semi-convergente se a sequência A, A^2, A^3, \dots é semi-convergente.*

Definição 3.4.9. *A matriz A é dita condicionalmente convergente se a sequência A, A^2, A^3, \dots é condicionalmente convergente.*

A partir das definições acima pode-se classificar um método iterativo quanto a sua convergência.

Se existe uma solução \vec{x}^* para (3.3), têm-se que

$$x^* = Gx^* + \vec{f}. \quad (3.40)$$

Subtraindo (3.40) de (3.2) obtém-se

$$\vec{x}^{k+1} - \vec{x}^* = G(\vec{x}^k - \vec{x}^*) = \dots = G^{k+1}(\vec{x}^0 - \vec{x}^*). \quad (3.41)$$

Pode-se definir o erro absoluto na iteração $k + 1$ como

$$\vec{\epsilon}^{k+1} = \vec{x}^{k+1} - \vec{x}^* \quad (3.42)$$

e interessados na convergência verifica-se quando $\vec{\epsilon}^k \rightarrow \vec{0}$ na equação

$$\vec{\epsilon}^{k+1} = G\vec{\epsilon}^k = \dots = G^{k+1}\vec{\epsilon}^0 \quad (3.43)$$

A definição de método iterativo convergente normalmente encontrada é a seguinte (veja referências [Saad(2000), Datta(1994), Young(1988)]).

Definição 3.4.10. *Um método iterativo é convergente se para todo \vec{x}_0 e \vec{f} a iteração (3.2) é convergente. Desta forma têm-se que $\vec{x}^{k+1} - \vec{x}^* \rightarrow \vec{0}$ quando $k \rightarrow \infty$.*

Lema 3.4.1. *Seja G uma matriz convergente. Então para todo \vec{x}_0 a equação (3.43) converge para $\vec{0}$.*

Demonstração. Se G é uma matriz convergente então $G^k \rightarrow O$, a matriz nula, o que implica que $\vec{e}^{k+1} \rightarrow \vec{0}$ quando $k \rightarrow \infty$. \square

Para determinar se uma matriz G é convergente pode-se verificar o seu raio espectral, $\rho(G)$.

Para que possamos falar nessa sequência convergente deve-se garantir que a equação matricial possui solução. [Datta(1994)] e [Young(1988)] afirmam que uma condição necessária e suficiente para que o método iterativo seja convergente é que o raio espectral de G seja estritamente menor do 1, isto é, $\rho(G) < 1$.

O teorema a seguir nos dá uma condição para garantir a convergência de métodos da forma (3.2).

Teorema 3.4.1. *A matriz G é convergente e o método iterativo (3.2) é convergente, se $\rho(G) < 1$.*

Demonstração. A prova completa pode ser vista em [Saad(2000), Young(1988)].

No caso mais simples assumindo que o maior autovalor em módulo tem multiplicidade algébrica igual a 1 reescreve-se o vetor inicial

$$\vec{x}_0 = c_1\vec{v}_1 + c_2\vec{v}_2 + \dots + c_n\vec{v}_n \quad (3.44)$$

onde \vec{v}_k são os autovalores de G , para $1 \leq k \leq n$.

Assim,

$$\begin{aligned} G\vec{x}_0 &= G(c_1\vec{v}_1 + c_2\vec{v}_2 + \dots + c_n\vec{v}_n) \\ &= c_1G\vec{v}_1 + c_2G\vec{v}_2 + \dots + c_nG\vec{v}_n \\ &= c_1\lambda_1\vec{v}_1 + c_2\lambda_2\vec{v}_2 + \dots + c_n\lambda_n\vec{v}_n \end{aligned} \quad (3.45)$$

e

$$G^k \vec{x}_0 = c_1 \lambda_1^k \vec{v}_1 + c_2 \lambda_2^k \vec{v}_2 + \dots + c_n \lambda_n^k \vec{v}_n \quad (3.46)$$

Caso $\rho(G) < 1$, temos que $\lambda_i^k \rightarrow 0$ para $1 \leq i \leq n$ quando $k \rightarrow \infty$. \square

A dificuldade aparece quando encontrou-se autovalores iguais a 1. Neste caso a matriz A de (3.1) é singular, e se \vec{b} está na imagem de A , então a sequência gerada por (3.2) converge se e somente se G é semi-convergente. Neste caso a solução de (3.1) para qual (3.2) converge depende do vetor inicial \vec{x}_0 (detalhes veja, [Neumann and Plemmons(1978), Berman and Plemmons(1979), Young(1988)]).

Definição 3.4.11. *Um método iterativo (3.2) é semi-convergente se a iteração converge para todo \vec{x}_0 . Note que o vetor \vec{x}^* para o qual o método converge depende do vetor \vec{x}_0 inicial.*

Para G ser semi-convergente, o limite

$$\lim_{i \rightarrow \infty} G^i \quad (3.47)$$

deve existir, embora não seja necessário a matriz ser nula. Ainda tem-se que G é semi-convergente se e somente se

1. $\rho(G) \leq 1$.
2. se $\rho(G) = 1$ então:
 - a) os blocos de Jordan associados a $\lambda = 1$ tem tamanho 1.
 - b) o único autovalor com $|\lambda| = 1$ é $\lambda = 1$.

Isso leva ao seguinte lema.

Lema 3.4.2. *Seja $G \in \mathbb{R}^{n \times n}$. Então G é semi-convergente se e somente se existe uma matriz P não-singular tal que*

$$G = P \begin{bmatrix} I & 0 \\ 0 & K \end{bmatrix} P^{-1} \quad (3.48)$$

onde I é a identidade e $\rho(K) < 1$.

Pode-se ainda permitir que os autovalores de G sejam iguais a -1 .

Veja o seguinte exemplo. Seja a matriz G dada por

$$G = \begin{bmatrix} -1 & \\ & K \end{bmatrix} \quad (3.49)$$

onde $\rho(K) < 1$. A sequência G^k tende a

$$G^k = \begin{bmatrix} (-1)^k & \\ & K^k \end{bmatrix}. \quad (3.50)$$

Como $\rho(K) < 1$ tem-se que K^k tende a O e o bloco $(-1)^k$ ficará oscilando entre -1 e 1 , originando uma sequência limitada cuja norma é zero.

Inicialmente elimina-se do vetor \vec{x}_0 a componente na direção do autovetor \vec{v}_1 associado a $\lambda_1 = -1$, tem-se que esta componente será sempre zero o que permitirá a convergência do método iterativo.

Assim, tem-se o seguinte lema.

Lema 3.4.3. *Seja $G \in \mathbb{R}^{n \times n}$. Então G é condicionalmente convergente se existe uma matriz P não-singular tal que*

$$G = P \begin{bmatrix} J & & \\ & -I & \\ & & K \end{bmatrix} P^{-1} \quad (3.51)$$

onde I e J são matrizes identidades não necessariamente do mesmo tamanho e $\rho(K) < 1$.

Definição 3.4.12. *Um método iterativo é condicionalmente convergente se a iteração (3.2) possui G condicionalmente convergente.*

Note que a iteração converge para vetores \vec{x}_0 que não possuam componentes nas direções dos autovetores associados a $\lambda = -1$. Se, além disso, \vec{x}_0 não possui componentes na direção dos autovetores associados a $\lambda = 1$, tem-se que a iteração (3.2) converge sempre para a mesma solução.

3.5 Taxa de Convergência

Para que exista convergência de um método iterativo da forma (3.2), depende-se do raio espectral de G . Além disso, o raio espectral influencia na velocidade de convergência do método. Todavia, quanto menor o raio espectral menor será a quantidade de iterações necessárias para que satisfaça a condição de convergência do método. Pode-se então definir a taxa de convergência assintótica de um método iterativo como (ver [Young(1988)])

$$R(G) = -\ln \rho(G) \quad (3.52)$$

Observe que, pela definição de convergência assintótica a taxa de convergência depende unicamente do raio espectral da matriz de iteração.

Se G for semi-convergente ou condicionalmente convergente o raio de convergência assintótica da iteração (3.2) é dado por

$$R(G) = -\ln \delta(G) \quad (3.53)$$

onde

$$\delta(G) = \max\{|\lambda| : \lambda \in \sigma(G), |\lambda| \neq 1\}. \quad (3.54)$$

Assim, $\delta(G)$ é o segundo maior autovalor em módulo de G . Deste modo, de acordo com [Berman and Plemmons(1979)] têm-se que $\delta(G) = \rho(K)$ onde K é dada por (3.48).

Através do raio espectral de G ainda existe a possibilidade de estimar teoricamente o número de iterações realizada por cada método. Desta forma, este é dado pela definição a seguir.

Definição 3.5.1. *Seja tol a tolerância, então o número de iterações n para que a solução satisfaça $\|x^k - x^*\| < tol$ é dado por*

$$n = \frac{\ln\left(\frac{1}{tol}\right)}{R(G)} \quad (3.55)$$

(ver [Young(1988)]).

3.6 Convergência do Método SOR

Em muitos casos, a convergência pode ser substancialmente acelerada. Isso significa que ao invés de fazer uma correção para a qual a equação é satisfeita exatamente, faz-se uma correção maior. No caso do método iterativo SOR, para este convergir deve-se seguir o seguinte teorema.

Teorema 3.6.1. *O método iterativo SOR não converge para uma aproximação inicial arbitrária se ω estiver fora do intervalo $(0, 2)$ (veja [Datta(1994)]).*

Ao fazer uma boa escolha para o coeficiente ω , melhora-se o desempenho do método SOR. Uma vez que, seguindo o teorema tem-se uma boa opção na escolha do ω ótimo.

Teorema 3.6.2. *Se o maior autovalor da matriz de iteração de Jacobi, G_{Jac} é real e $\rho(G) < 1$, então o ω ótimo que produz o menor raio espectral para G_{SOR} é dado por*

$$\omega_{\text{ótimo}} = \frac{2}{1 + \sqrt{1 - \rho(G_{Jac})^2}} \quad (3.56)$$

e, o raio espectral da G de SOR é dado por

$$\rho(G_{SOR}) = \omega_{\text{ótimo}} - 1 \quad (3.57)$$

(veja [Datta(1994)]).

3.7 Problema de Poisson

A seguir apresenta-se a matriz G de (3.2) gerada a partir do método iterativo de Jacobi para o problema de Poisson sob as condições de contorno de Dirichlet e de Neumann e seus respectivos autovalores.

3.7.1 Condição de Dirichlet Unidimensional

Em se tratando do problema de Dirichlet unidimensional, tem-se que A_D é a matriz dos coeficientes em (2.27) e a matriz de iteração é dada por $J_D = D^{-1}(E + F)$, portanto

$$D = \begin{bmatrix} -2 & & \dots & 0 \\ & -2 & & \\ & & \ddots & \\ & & & -2 \\ 0 & \dots & & -2 \end{bmatrix} \quad (3.58)$$

e

$$-(E + F) = \begin{bmatrix} 0 & 1 & & \dots & 0 \\ 1 & 0 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & 0 & 1 \\ 0 & \dots & & 1 & 0 \end{bmatrix}. \quad (3.59)$$

Como a matriz D é diagonal, sua inversa é dada pelo inverso dos seus elementos, então J_D pode ser escrita como

$$J_D = \frac{1}{2} \begin{bmatrix} 0 & 1 & & \dots & 0 \\ 1 & 0 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & 0 & 1 \\ 0 & \dots & & 1 & 0 \end{bmatrix}. \quad (3.60)$$

Segundo [Yueh(2005)], os autovalores da matriz J_D para o caso unidimensional são dados por

$$\lambda_k = \cos\left(\frac{k\pi}{n+1}\right) \quad \text{para } 1 \leq k \leq n. \quad (3.61)$$

Segue que, o maior autovalor em módulo ocorre quando $k = 1$, sendo este definido como o raio espectral de J_D ,

$$\rho(J_D) = \cos\left(\frac{\pi}{n+1}\right). \quad (3.62)$$

Assim como em [Young(1988)], a partir dos autovalores de J_D pode-se determinar os autovalores da matriz de iteração GS_D do método iterativo Gauss-Seidel, e estes são

$$\mu_k = \left[\cos\left(\frac{k\pi}{n+1}\right) \right]^2 \quad \text{para } 1 \leq k \leq \left\lfloor \frac{n}{2} \right\rfloor, \quad (3.63)$$

e quando $k > \lfloor \frac{n}{2} \rfloor$ temos que $\mu_k = 0$.

Portanto o máximo autovalor ocorre quando $k = 1$ e este é o raio espectral da matriz de iteração de Gauss-Seidel

$$\rho(GS_D) = \cos^2\left(\frac{\pi}{n+1}\right). \quad (3.64)$$

Uma vez que $\rho(GS_D) < \rho(J_D) < 1$, assim pelo teorema 3.4.1 pode-se garantir que os métodos são convergentes e que o método de Gauss-Seidel possui uma taxa de convergência mais rápida do que o Jacobi.

3.7.2 Condição de Dirichlet Bidimensional

Para o problema de Dirichlet bidimensional, montou-se a matriz de iteração de (3.2) a partir de (3.58) e (3.59) juntamente com as definições do produto de Kronecker. Assim, determinou-se a matriz de iteração J_{2D} do método de Jacobi para o caso bidimensional.

A matriz de Jacobi para o caso 2D relacionada com a matriz L_D (2.55) pode ser decomposta como

$$D = \begin{bmatrix} -4 & & \\ & \ddots & \\ & & -4 \end{bmatrix} \quad \text{e} \quad -(E + F) = L_D - D. \quad (3.65)$$

Novamente é fácil obter D^{-1} que possui $-\frac{1}{4}$ ao longo da diagonal e pode-se obter a matriz $J_{2D} = D^{-1}(E + F)$ tendo o mesmo padrão que L_D sem os elementos da diagonal com os valores $\frac{1}{4}$ nos termos diferentes de zero.

Utilizando o produto de Kronecker pode-se obter também que

$$J_{2D} = \frac{1}{2}(J_D \oplus J_D) = \frac{1}{2}(I \otimes J_D + J_D \otimes I) \quad (3.66)$$

onde,

$$\frac{1}{2}(I \otimes J_D) = \frac{1}{2} \begin{bmatrix} 1J_D & \dots & 0 \\ & \ddots & \\ 0 & \dots & 1J_D \end{bmatrix} \quad (3.67)$$

e

$$\frac{1}{2}(J_D \otimes I) = \frac{1}{4} \begin{bmatrix} 0 & I & & \\ I & 0 & I & \\ & \ddots & \ddots & \ddots \\ & & I & 0 & I \\ & & & I & 0 \end{bmatrix}. \quad (3.68)$$

Logo,

$$\frac{1}{2}(I \otimes J_D + J_D \otimes I) = \frac{1}{4} \begin{bmatrix} 2J_D & I & \dots & 0 \\ I & 2J_D & I & \\ & \ddots & \ddots & \ddots \\ & & I & 2J_D & I \\ 0 & \dots & & I & 2J_D \end{bmatrix}. \quad (3.69)$$

Desta forma, determinou-se a matriz J_{2D} do caso bidimensional. Mais do que isso, ainda pode-se obter os autovalores de J_{2D} em (3.69) a partir de J_D em (3.61), onde estes são dados por (usando o teorema 2.4.3)

$$\lambda_{i,j} = \frac{\cos\left(\frac{i\pi}{n+1}\right) + \cos\left(\frac{j\pi}{n+1}\right)}{2} \quad \text{para} \quad 1 \leq i, j \leq n. \quad (3.70)$$

Além disso, o maior autovalor ocorre quando $i = j = 1$, ou seja, o raio espectral de J_{2D} é

$$\rho(J_{2D}) = \frac{\cos\left(\frac{\pi}{n+1}\right) + \cos\left(\frac{\pi}{n+1}\right)}{2} \quad (3.71)$$

ou ainda,

$$\rho(J_{2D}) = \cos\left(\frac{\pi}{n+1}\right). \quad (3.72)$$

Considerando o método iterativo Gauss-Seidel, pode-se gerar uma matriz de iteração semelhante a (3.17) e nomeamos-a por GS_{2D} . Segundo [Young(1988)], pode-se verificar numericamente que seus autovalores são determinados a partir de (3.70), ou seja, $\mu = \lambda^2$, ou ainda

$$\mu_{i,j} = \left[\frac{\cos\left(\frac{i\pi}{n+1}\right) + \cos\left(\frac{j\pi}{n+1}\right)}{2} \right]^2 \quad \text{para} \quad 1 \leq i, j \leq \left\lfloor \frac{n}{2} \right\rfloor \quad (3.73)$$

e os autovalores restantes são iguais a zero.

Portanto o máximo autovalor ocorre quando $i = j = 1$,

$$\rho(GS_{2D}) = \cos^2\left(\frac{\pi}{n+1}\right) \quad (3.74)$$

e este é o raio espectral de GS_{2D} .

Como $\rho(GS_{2D}) < \rho(J_{2D}) < 1$, garantiu-se pelo teorema 3.4.1 que os respectivos métodos iterativos são convergentes.

3.7.3 Condição de Neumann Unidimensional

Considere o problema de Neumann unidimensional. Seja A_M dada por (2.35) e portanto como $J_M = D^{-1}(E + F)$ obtém-se que

$$D = \begin{bmatrix} -2 & & \dots & 0 \\ & -2 & & \\ & & \ddots & \\ & & & -2 \\ 0 & \dots & & -2 \end{bmatrix} \quad (3.75)$$

e

$$-(E + F) = \begin{bmatrix} 0 & 2 & & \dots & 0 \\ 1 & 0 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & & 1 & 0 & 1 \\ 0 & \dots & & 2 & 0 \end{bmatrix}. \quad (3.76)$$

Portanto J_M é

$$J_M = \frac{1}{2} \begin{bmatrix} 0 & 2 & & \dots & 0 \\ 1 & 0 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & & 1 & 0 & 1 \\ 0 & \dots & & 2 & 0 \end{bmatrix}. \quad (3.77)$$

Usando a matriz A_N para obter-se a matriz de Jacobi que é idêntica a J_M , ou seja, $J_N = J_M$.

Pode-se verificar numericamente que os autovalores desta matriz de acordo com [Neumann and Plemmons(1978)], são

$$\lambda_k = \cos \left[\frac{(k-1)\pi}{n-1} \right] \quad \text{para} \quad 1 \leq k \leq n. \quad (3.78)$$

Observe que de modo análogo ao problema de Dirichlet, obtém-se os autovalores da matriz de iteração do método Gauss-Seidel, e estes são

$$\mu_k = \left[\cos \left(\frac{k-1}{n-1} \right) \pi \right]^2 \quad \text{para} \quad 1 \leq k \leq \left\lfloor \frac{n}{2} \right\rfloor \quad (3.79)$$

e os demais autovalores são iguais a zero.

Quando $k = 1$ em (3.78) o autovalor $\lambda_1 = 1$ e quando $k = n$ obtém-se o autovalor $\lambda_n = -1$ (todos os outros autovalores satisfazem $|\lambda| < 1$). Sendo este em módulo o raio espectral da matriz de iteração para o método iterativo de Jacobi e também o de Gauss-Seidel, e seguindo-se rigorosamente o teorema 3.4.1, tem-se que a matriz de iteração do problema de Neumann unidimensional não seria convergente.

Portanto obtém-se a convergência do método iterativo ao resolver o problema de Neumann definiu-se algumas preliminares que são importantes na solução deste.

A única possibilidade então é que o método seja condicionalmente convergente conforme o lema 3.4.3. É fácil verificar que o autovetor \vec{v}_1 de J_M correspondente a $\lambda_1 = 1$ é

$$\vec{v}_1 = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}, \quad (3.80)$$

ou seja,

$$\frac{1}{2} \begin{bmatrix} 0 & 2 & & \dots & 0 \\ 1 & 0 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & 0 & 1 \\ 0 & \dots & & 2 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 1 \end{bmatrix} = 1 \cdot \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 1 \end{bmatrix} \quad (3.81)$$

o que implica que $J_M \vec{v}_1 = 1 \vec{v}_1$.

Para $\lambda_n = -1$ o autovetor \vec{v}_n é dado por

$$\vec{v}_n = \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \\ \vdots \end{bmatrix} \quad (3.82)$$

satisfazendo

$$\frac{1}{2} \begin{bmatrix} 0 & 2 & \dots & 0 \\ 1 & 0 & 1 & \\ & \ddots & \ddots & \ddots \\ & & 1 & 0 & 1 \\ 0 & \dots & & 2 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \\ \vdots \end{bmatrix} = -1 \cdot \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \\ \vdots \end{bmatrix} \quad (3.83)$$

o qual implica que $J_M \vec{v}_n = -1 \vec{v}_n$.

Dado um vetor inicial \vec{x}_0 arbitrário, segundo o teorema 3.4.1, eliminar as componentes na direção de \vec{v}_1 e \vec{v}_n .

Sendo

$$\vec{x} = \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix}^T \quad (3.84)$$

o vetor inicial, a projeção na direção de \vec{v}_1 é

$$\vec{p} = \frac{\vec{x}^T \vec{v}_1}{\vec{v}_1^T \vec{v}_1} \vec{v}_1 \quad (3.85)$$

assim,

$$\vec{p} = \frac{\begin{bmatrix} x_1 & \dots & x_n \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}}{\begin{bmatrix} 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}} \vec{v}_1 = \frac{\sum_{i=1}^n x_i}{n} \vec{v}_1 \quad (3.86)$$

Portanto determinou-se a projeção de \vec{x}_0 na direção de \vec{v}_n calculou-se

$$\vec{q} = \frac{\vec{x}_0^T \vec{v}_n}{\vec{v}_n^T \vec{v}_n} \vec{v}_n \quad (3.87)$$

onde,

$$\vec{q} = \frac{\begin{bmatrix} x_1 & \dots & x_n \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \\ \vdots \end{bmatrix}}{\begin{bmatrix} 1 & -1 & 1 & -1 & \dots \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \\ \vdots \end{bmatrix}} \vec{v}_n \quad (3.88)$$

e

$$\vec{q} = \frac{\sum_{i=1}^n (-1)^{i+1} x_i}{n} \vec{v}_n. \quad (3.89)$$

Portanto com base nos resultados apresentados anteriormente, para que o método convirja deve-se ter o vetor inicial dado por

$$\hat{x} = \vec{x}_0 - \alpha \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 1 \end{bmatrix} - \beta \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \\ \vdots \end{bmatrix} \quad (3.90)$$

onde,

$$\alpha = \frac{\sum_{i=1}^n x_i^0}{n} \quad \text{e} \quad \beta = \frac{1}{n} \sum_{i=1}^n (-1)^i x_i^0 \quad (3.91)$$

Assim, determinou-se \vec{x}_0 através da definição 3.4.12 e pelo lema 3.4.3 tem-se que os métodos iterativos Jacobi e Gauss-Seidel são condicionalmente convergentes. Uma vez que garantiu-se que o método iterativo é condicionalmente

convergente, determinou-se o seu raio espectral e este é dado pelo maior autovalor diferente de $|\lambda| = 1$.

Para $k = 2$ tem-se que, o pseudo raio espectral da matriz de iteração de Jacobi é dado por

$$\delta(J_M) = \cos\left(\frac{\pi}{n-1}\right) \quad (3.92)$$

e, no caso da matriz de iteração de Gauss-Seidel este é dado por

$$\delta(GS_M) = \cos^2\left(\frac{\pi}{n-1}\right). \quad (3.93)$$

3.7.4 Condição de Neumann Bidimensional

O caso bidimensional utilizando condições de contorno de Neumann será semelhante ao problema de Dirichlet. Portanto pode-se obter a matriz de iteração de Jacobi $J_{2M} = \frac{1}{2}(J_M \oplus J_M)$ utilizando o mesmo raciocínio.

Para determinar os autovalores de J_{2M} no caso bidimensional usou-se o teorema 2.4.3, já que sabe-se quem são os autovalores de J_M no caso unidimensional. Para determiná-los é necessário conhecer quem são os autovalores no caso unidimensional e estes são dados por (3.78). Assim os autovalores são fornecidos por

$$\lambda_{i,j} = \frac{\cos\left[\frac{(i-1)\pi}{n-1}\right] + \cos\left[\frac{(j-1)\pi}{n-1}\right]}{2} \quad \text{para } 1 \leq i, j \leq n. \quad (3.94)$$

Conhecendo os autovalores da matriz de iteração do método de Jacobi, segundo [Young(1988)] definiu-se os autovalores para o método de Gauss-Seidel que são

$$\mu_{i,j} = \left[\frac{\cos\left[\frac{(i-1)\pi}{n-1}\right] + \cos\left[\frac{(j-1)\pi}{n-1}\right]}{2} \right]^2 \quad \text{para } 1 \leq i, j \leq \left\lfloor \frac{n}{2} \right\rfloor \quad (3.95)$$

e os demais $\mu_{i,j} = 0$.

Quando tomou-se $i = j = 1$ tem-se que o maior autovalor da matriz de iteração dos métodos iterativos Jacobi e Gauss-Seidel é dado por $|\lambda| = |\mu| = 1$ e de acordo com teorema 3.4.1 não obtém-se convergência para a solução do problema através dos respectivos métodos.

Seja a matriz do método iterativo Jacobi (definiu-se em forma de blocos)

$$J_{2M} = \begin{bmatrix} E & F & \dots & 0 \\ H & E & H & \\ & \ddots & \ddots & \ddots \\ & & H & E & H \\ 0 & \dots & & F & E \end{bmatrix} \quad (3.96)$$

onde

$$E = \begin{bmatrix} 0 & \frac{1}{2} & \dots & 0 \\ \frac{1}{4} & 0 & \frac{1}{4} & \\ & \ddots & \ddots & \ddots \\ & & \frac{1}{4} & 0 & \frac{1}{4} \\ 0 & \dots & & \frac{1}{2} & 0 \end{bmatrix}, \quad F = \begin{bmatrix} \frac{1}{2} & \dots & 0 \\ & \frac{1}{2} & \\ & & \ddots \\ & & & \frac{1}{2} \\ 0 & \dots & & & \frac{1}{2} \end{bmatrix} \quad (3.97)$$

e

$$H = \begin{bmatrix} \frac{1}{4} & \dots & 0 \\ & \frac{1}{4} & \\ & & \ddots \\ & & & \frac{1}{4} \\ 0 & \dots & & & \frac{1}{4} \end{bmatrix}. \quad (3.98)$$

Os maiores autovalores em módulo de J_{2M} são $\lambda_1 = 1$ e $\lambda_{n^2} = -1$. Tomando $\lambda_1 = 1$, pode-se verificar que

$$\vec{v}_1 = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \quad (3.99)$$

é o autovetor associado a este autovalor, usando

$$\begin{bmatrix} E & F & \dots & 0 \\ H & E & H & \\ & \ddots & \ddots & \ddots \\ & & H & E & H \\ 0 & \dots & F & E \end{bmatrix} \cdot \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = 1 \cdot \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \quad (3.100)$$

Para $\lambda_{n^2} = -1$ a situação é um pouco mais complicada. O autovetor \vec{v}_{n^2} deve corresponder a matriz

$$\begin{bmatrix} 1 & -1 & 1 & -1 & \dots \\ -1 & 1 & -1 & 1 & \dots \\ 1 & -1 & 1 & -1 & \dots \\ -1 & 1 & -1 & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \quad (3.101)$$

(lida do elemento a_{11} para baixo seguindo coluna por coluna). Quando n for ímpar tem-se que

$$\vec{v}_{n^2} = \left[1 \quad -1 \quad 1 \quad -1 \quad 1 \quad -1 \quad \dots \quad 1 \quad -1 \right]^T \quad (3.102)$$

simplesmente alternando 1 e -1 . Porém, quando n é par pode-se haver repetições. Por exemplo, para $n = 4$ tem-se a seguinte matriz

$$\begin{bmatrix} 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \end{bmatrix} \quad (3.103)$$

e o vetor \vec{v}_{n^2} será

$$\vec{v}_{n^2} = \left[1 \quad -1 \quad 1 \quad -1 \quad -1 \quad 1 \quad -1 \quad 1 \quad 1 \quad -1 \quad 1 \quad -1 \quad -1 \quad 1 \quad -1 \quad 1 \right] \quad (3.104)$$

Note que para n ímpar (e par também) tem-se que

$$\begin{bmatrix} E & F & & \dots & 0 \\ H & E & H & & \\ & \ddots & \ddots & \ddots & \\ & & H & E & H \\ 0 & \dots & & F & E \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \\ \vdots \end{bmatrix} = -1 \cdot \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \\ \vdots \end{bmatrix}. \quad (3.105)$$

Calculando a projeção de \vec{x}_0 na direção de \vec{v}_1

$$\vec{p} = \frac{\vec{x}_0^T \vec{v}_1}{\vec{v}_1^T \vec{v}_1} \vec{v}_1 \quad (3.106)$$

assim,

$$\vec{p} = \frac{\begin{bmatrix} x_{11} & \dots & x_{nn} \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}}{\begin{bmatrix} 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}} \vec{v}_1 = \frac{\sum_{i,j=1}^n x_{i,j}}{n^2} \vec{v}_1 = \frac{1}{n^2} \left(\sum_{i,j=1}^n x_{i,j} \right) \vec{v}_1 \quad (3.107)$$

De modo análogo determina-se \vec{q} a projeção de \vec{x}_0 na direção de \vec{v}_n

$$\vec{q} = \frac{\vec{x}_0^T \vec{v}_n}{\vec{v}_n^T \vec{v}_n} \vec{v}_n \quad (3.108)$$

onde

$$\vec{q} = \frac{\begin{bmatrix} x_{11} & \dots & x_{nn} \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \\ \vdots \end{bmatrix}}{\begin{bmatrix} 1 & -1 & 1 & -1 & \dots \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \\ \vdots \end{bmatrix}} \vec{v}_{n^2} \quad (3.109)$$

e

$$\vec{q} = \frac{1}{n^2} \left(\sum_{i=1}^n \sum_{j=1}^n (-1)^{i+j} x_{ij} \right) \vec{v}_{n^2} \quad (3.110)$$

Necessariamente, o vetor inicial deve satisfazer a seguinte condição

$$\hat{x} = \vec{x}_0 - \alpha \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 1 \end{bmatrix} - \beta \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \\ \vdots \end{bmatrix} \quad (3.111)$$

onde,

$$\alpha = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n x_{ij} \quad \text{e} \quad \beta = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n (-1)^{i+j} x_{ij} \quad (3.112)$$

Assim, pela definição 3.4.12 e pelo lema 3.4.3 tem-se que os métodos iterativos Jacobi e Gauss-Seidel são condicionalmente convergentes. E assim, o raio espectral dos respectivos métodos é o maior autovalor diferente de 1. Ou seja, em (3.94) e (3.95) se $i = 1$ e $j = 2$, ou vice-versa, o pseudo raio espectral da matriz de iteração dos respectivos métodos para o problema de Neumann bidimensional.

4 RESULTADOS NUMÉRICOS

Neste capítulo apresenta-se os resultados obtidos após a implementação e simulação do algoritmo. A análise refere-se à convergência espectral da matriz de iteração dos métodos iterativos usados para resolver a equação de Poisson sob as condições de contorno de Dirichlet e Neumann.

4.1 Existência de Solução

Nesta etapa do trabalho descreve-se alguns aspectos referente a metodologia adotada na simulação da equação de Poisson (2.1) para as respectivas condições de contorno (2.45) e (2.57). Apresenta-se sua forma discretizada torna-se

Problema de Dirichlet

$$\begin{aligned} p_{i,1} = 0, \quad p_{i,ny} = 0, \quad \text{para } 1 \leq i \leq nx \\ p_{1,j} = 0, \quad p_{nx,j} = 0, \quad \text{para } 1 \leq j \leq ny \end{aligned} \quad (4.1)$$

Problema de Neumann

$$\begin{aligned} p_{i,1}^{n+1} = p_{i,2}^{n+1}, \quad p_{i,ny}^{n+1} = p_{i,ny-1}^{n+1}, \quad \text{para } 1 \leq i \leq nx \\ p_{1,j}^{n+1} = p_{2,j}^{n+1}, \quad p_{nx,j}^{n+1} = p_{nx-1,j}^{n+1}, \quad \text{para } 1 \leq j \leq ny \end{aligned} \quad (4.2)$$

Na realização dos testes foi considerada a seguinte condição inicial para o problema objeto deste estudo

$$p_{i,j}^0 = 0 \quad \text{para } 1 \leq i \leq nx, \quad 1 \leq j \leq ny. \quad (4.3)$$

A equação de Poisson (2.1) exige um termo fonte e este deve satisfazer a condição de que

$$\int_{\Omega} f dx = 0 \quad (4.4)$$

Dessa forma, garantindo que o sistema (2.44) aplicado as condições de contorno (2.45) ou (2.57), de fato possui solução, aplica-se métodos numéricos para obtenção de uma solução aproximada do problema.

Diante deste fato, utilizou-se a função

$$f(x, y) = \sin(2\pi x) \cdot \sin(2\pi y) \quad (4.5)$$

como sendo o termo fonte da equação (2.1) já que a condição (4.4) é satisfeita, ou seja,

$$\sum_{i=1}^n b_i = 0 = \sum_{i=2}^{nx-1} \sum_{i=2}^{ny-1} f_{ij} \quad (4.6)$$

É importante ressaltar que em $\sum_{i=2}^{nx-1} \sum_{i=2}^{ny-1} f_{ij}$ não considera-se os pontos do contorno, ou seja, são considerados apenas os pontos interiores na malha cartesiana.

Considere também como um outro termo fonte a função

$$g(x, y) = \sin x \cdot \sin y \quad (4.7)$$

certamente no domínio $[0, 1] \times [0, 1]$ a

$$\int_0^1 \int_0^1 g(x, y) dx dy \neq 0 \quad (4.8)$$

o que implica que o problema (2.1) não terá solução pois não satisfaz a condição de integrabilidade.

Garantindo que o termo fonte satisfaz (4.4) numericamente. Pode-se então calcular

$$S = \sum_{i=2}^{nx-1} \sum_{i=2}^{ny-1} f_{ij} \quad \text{e} \quad R = \sum_{i=1}^{nx} \sum_{i=1}^{ny} f_{ij} \quad (4.9)$$

e garantir que a integral da f seja zero no interior do domínio usando (caso a)

$$f_{ij} = f_{ij} - \frac{S}{(nx-2)(ny-2)} \quad (4.10)$$

ou que a integral seja zero em todo o domínio usando (caso b)

$$f_{ij} = f_{ij} - \frac{R}{nx \times ny} \quad (4.11)$$

O gráfico 4.1 refere-se ao resultado da simulação da equação de Poisson (2.1) com o termo fonte (4.7) juntamente com a condição (4.10). Neste caso a condição de integrabilidade (4.4) é satisfeita, isso implica em (4.6), muito mais que isso pode-se afirmar que o método iterativo é convergente para $tol = 10^{-10}$.

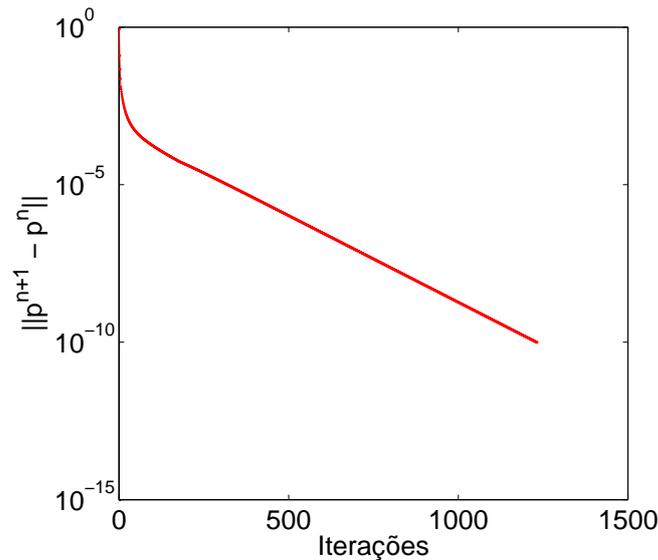


Figura 4.1: Condição (4.10) e termo fonte $f(x, y) = \sin x \cdot \sin y$ a cada iteração na malha cartesiana 21×21 e $tol = 10^{-10}$.

Quando usou-se as condições (4.11) e (4.7) a integral é igual a zero em todo o domínio, porém não é igual a zero no interior do mesmo, o que não satisfaz o critério de convergência dos métodos iterativos para a tolerância determinada. As primeiras iterações do método são apresentadas na figura 4.2 onde notou-se que o método não é convergente.

Quando usou-se a função fonte $f(x, y) = x^2$, a condição de existência de solução (2.94) não é satisfeita, ou seja, o método iterativo diverge pois $\sum f_{i,j} \neq 0$.

Uma peça importante usada para determinar a convergência do método iterativo é a diferença entre a solução na iteração k e a iteração $k + 1$. Uma vez que não se conhece de fato a solução exata para se medir a “distância” entre o iterado

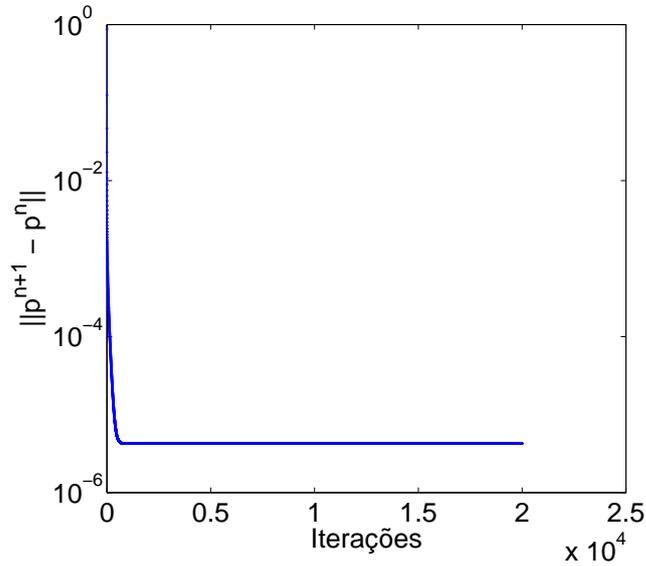


Figura 4.2: Condição (4.11) e termo fonte $f(x, y) = \sin x \cdot \sin y$ a cada iteração na malha cartesiana 21×21 e $tol = 10^{-10}$.

\vec{p}^k e a solução exata \vec{p} utilizou-se

$$\|\vec{p}^{k+1} - \vec{p}^k\|. \quad (4.12)$$

Portanto itera-se até que esta quantidade seja menor que uma determinada tolerância, dada por tol .

Por fim, observe que para $f(x, y) = x^2$ não obtém-se a convergência. Utilizou-se $tol = 10^{-10}$ para plotar $\|\vec{p}^{k+1} - \vec{p}^k\|$ e obteve-se o gráfico 4.3 o qual quando de fato o método converge este deve tender a zero, mas não está pois $f(x, y) = x^2$ não satisfaz a condição de integrabilidade.

Ao ilustra-se a norma de p , $\|p^k\|$ a medida que k cresce notou-se que esta quantidade está sempre crescendo, mesmo que lentamente, conforme a figura 4.4. Isto é devido a $\sum f \neq 0$.

Utiliza-se a condição

$$\vec{p}_{i,j}^{k+1} = \vec{p}_{i,j}^{k+1} - \vec{p}_{1,2}^{k+1}, \quad 1 \leq i, j \leq nx \quad (4.13)$$

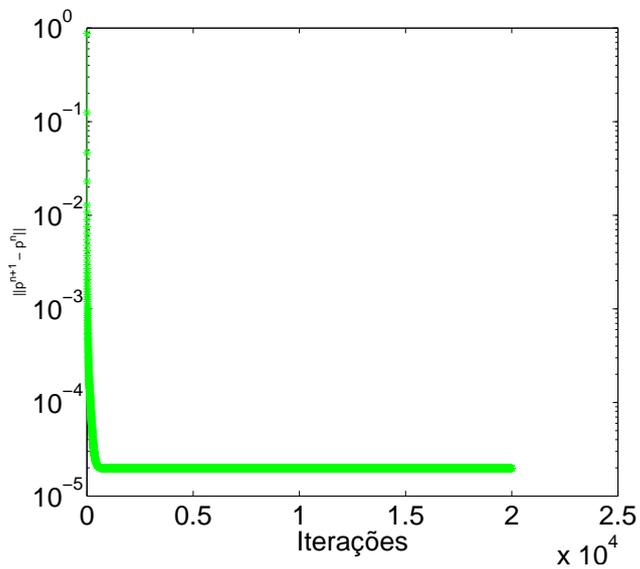


Figura 4.3: Diferença $\|\bar{p}^{k+1} - \bar{p}^k\|$ quando $f(x, y) = x^2$ a cada iteração na malha cartesiana 21×21 .

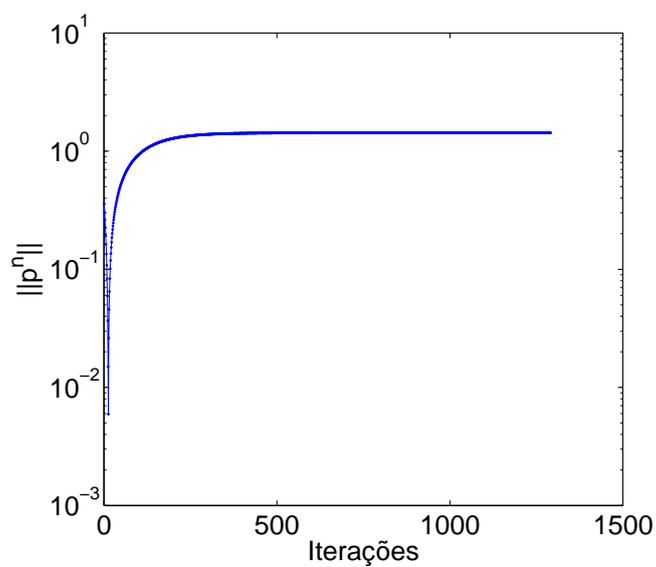


Figura 4.4: $\|\bar{p}^k\|$ de $f(x, y) = x^2$ a cada iteração na malha cartesiana 21×21 .

juntamente com as condições de Neumann 4.2. Esta condição faz uma “correção” na solução a cada iteração, deslocando as componentes do vetor solução \vec{p} , de tal forma que a solução em $x_{1,2}$ seja igual a zero, ou seja, $p_{1,2} = 0$. Poderia ser usado qualquer outro ponto da malha fixando este a zero. De acordo com a literatura [Strikwerda(2004), Pozrikidis(2001)] não se pode concluir que esta é uma solução do problema pois não satisfaz (4.4). Entretanto este é um artifício numérico que garante a convergência das iterações se (4.4) não puder ser satisfeita exatamente.

Utilizando (4.13) juntamente com a função $f(x, y) = x^2$ ilustrou-se $\|\vec{p}^{k+1} - \vec{p}^k\|$ e obteve-se o gráfico 4.5 onde pode ser visto que existe convergência do método iterativo. Neste caso foi usado $tol = 10^{-12}$.

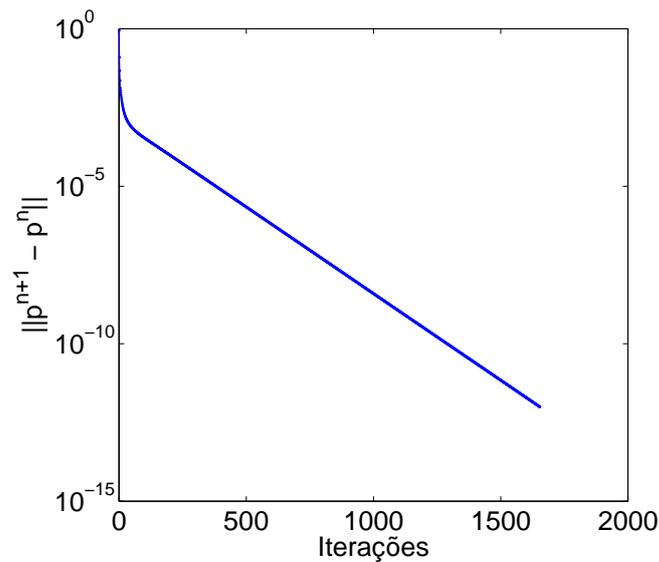


Figura 4.5: Diferença $\|\vec{p}^{k+1} - \vec{p}^k\|$ quando $f(x, y) = x^2$ a cada iteração na malha cartesiana 21×21 .

4.2 Raio Espectral e Taxa de Convergência

Apresenta-se alguns resultados de experimentos numéricos envolvendo a equação de Poisson (2.1).

Durante o desenvolvimento desta pesquisa, surgiram algumas perguntas. Tais como, por exemplo, sobre quais condições de contorno o método iterativo é convergente? Assim, partiu-se para as simulações a fim de procurar resposta para tal questão, ou algo relevante servindo como alguma pista.

Primeiramente, tratou-se de resolver (2.1) com as condições de contorno de Dirichlet (4.1) e uma determinada tolerância ($tol = 10^{-6}$) e assim comparou-se nossos resultados numéricos com resultados teóricos, ou seja, informações relevantes encontradas na literatura, obteve-se resultados que concordam com valores encontrados na bibliografia de [Young(1988), Neumann and Plemmons(1978), Laub(2005)]. E estes são de grande valia na comparação e validação dos resultados. Tais afirmações apresenta-se na tabela 4.1.

Tabela 4.1: Experimento numérico para o problema de Dirichlet.

h	$\rho(J_{2D})$	$\rho(GS_{2D})$	$R(J_{2D})$	$R(GS_{2D})$
$\frac{1}{5}$	0,8090	0,6545	0,2119	0,4238
$\frac{1}{9}$	0,9510	0,9045	0,0502	0,1003
$\frac{1}{10}$	0,9876	0,9755	0,0124	0,0248
$\frac{1}{20}$	0,9969	0,9938	0,0031	0,0062
$\frac{1}{40}$	0,9992	0,9984	0,0008	0,0016
$\frac{1}{80}$				

Através do algoritmo implementado e da matriz de iteração gerada a partir deste, determinou-se o raio espectral das matrizes dos métodos de Jacobi e Gauss-Seidel o qual apresenta-se na tabela 4.1 para vários tamanho de malha. Comparando o resultado do experimento numérico com as equações (3.72) e (3.74) verificou-se que eles são iguais. Assim, para o problema de Dirichlet os que os métodos iterativos são convergentes, pois apresentam raio espectral menor que 1.

A taxa de convergência de $R(J_{2D})$ e $R(GS_{2D})$ foi determinada por (3.52) o qual esta depende do raio espectral. Na tabela 4.2, n_O é o número de iterações observado na simulação numérica a qual utilizou-se $tol = 10^{-6}$. Note que o número de iterações teórico e observado aumenta rapidamente ao diminuir h .

Tabela 4.2: Experimento numérico para o problema de Dirichlet. n_O é o número de iterações observado e n_T é o número de iterações teórico.

h	$n_O(J_{2D})$	$n_O(GS_{2D})$	$n_T(J_{2D})$	$n_T(GS_{2D})$
$\frac{1}{5}$	57	29	66	33
$\frac{1}{10}$	181	73	276	138
$\frac{1}{20}$	501	120	1116	558
$\frac{1}{40}$	1112	256	4475	2238
$\frac{1}{80}$	1574	724	17275	8638

Contudo a estimativa que n_T é proporcional a $\frac{1}{h^2} \approx nx^2$ enquanto que n_O é proporcional a $\frac{1}{h^{\frac{3}{2}}} \approx nx^{\frac{3}{2}}$. Quanto a quantidade de iterações realizadas pelos métodos iterativos utilizados para resolver o sistema $A\vec{p} = \vec{b}$, podemos estimar a quantidade de iterações realizadas por cada método (n_T) através de (3.55).

Para este problema modelo (condições de contorno de Dirichlet), comparou-se os resultados com dados apresentados por ([Young(1988), p. 132]) onde o autor faz referência aos valores obtidos em seu experimento numérico após as simulações. É importante ressaltar que quanto ao raio espectral conseguimos os mesmos resultados que Young, a diferença ocorre somente quanto ao número de iterações.

Com isto nas simulações, resolveu-se (2.1) com as condições de contorno de Neumann 4.2 e analisou-se do ponto de vista numérico os autovalores das matrizes obtidas quanto ao critério para obter-se a convergência. Com base nos resultados apresenta-se as tabelas 4.3 e 4.4. Observe que no método de Gauss-Seidel a convergência ocorre de modo mais acelerado que o método de Jacobi.

Para o problema de Neumann usou-se (3.94) e (3.95) e tanto para o método iterativo de Jacobi quanto para o de Gauss-Seidel tem-se que o raio espectral é igual a 1, e assim o método não seria convergente. Entretanto usando a definição

Tabela 4.3: Experimento numérico para equação de Poisson com condições de contorno do problema de Neumann.

h	$\delta(J_{2N})$	$\delta(GS_{2N})$	$R(J_{2N})$	$R(GS_{2N})$
$\frac{1}{5}$	0,8536	0,7286	0,1582	0,3166
$\frac{1}{10}$	0,9698	0,9405	0,0306	0,0613
$\frac{1}{20}$	0,9931	0,9862	0,0069	0,0138
$\frac{1}{40}$	0,9983	0,9966	0,0017	0,0034
$\frac{1}{80}$	0,9996	0,9992	0,0004	0,0008

Tabela 4.4: Experimento numérico para equação de Poisson com condições de contorno do problema de Neumann. n_O é o número de iterações observado.

h	$n_O(J_{2N})$	$n_O(GS_{2N})$
$\frac{1}{5}$	62	51
$\frac{1}{10}$	217	160
$\frac{1}{20}$	571	424
$\frac{1}{40}$	1550	1014
$\frac{1}{80}$	4315	2702

(3.54) de $\delta(G)$ calculou-se numericamente $\delta(J_{2N})$ e $\delta(GS_{2N})$. Esses valores são usados segundo [Neumann and Plemmons(1978), Pozrikidis(2001)] para estimar a taxa de convergência R , de acordo com o lema 3.4.2.

4.3 Método SOR e o parâmetro ω

Visando acelerar a taxa de convergência na solução do sistema de equações (2.44) implementou-se o método iterativo SOR que emprega um parâmetro de relaxação ω . É possível escolher ω , de tal forma que a taxa de convergência seja sensivelmente superior àquela obtida nos métodos de Jacobi e Gauss-Seidel.

Para o problema de Dirichlet uni e bidimensional para encontrar o melhor coeficiente de relaxação $\omega_{\text{ótimo}}$ utilizou-se a fórmula dada em ([Young(1988),

p. 173]) que é dada por

$$\rho(G_{SOR}) = \begin{cases} \left[\frac{\omega\rho(J) + \sqrt{\omega^2\rho(J)^2 - 4(\omega-1)}}{2} \right]^2, & \text{se } 0 < \omega \leq \omega_{\text{ótimo}}; \\ \omega - 1, & \text{se } \omega_{\text{ótimo}} \leq \omega < 2. \end{cases} \quad (4.14)$$

onde $\rho(J)$ é o raio espectral da matriz de Jacobi.

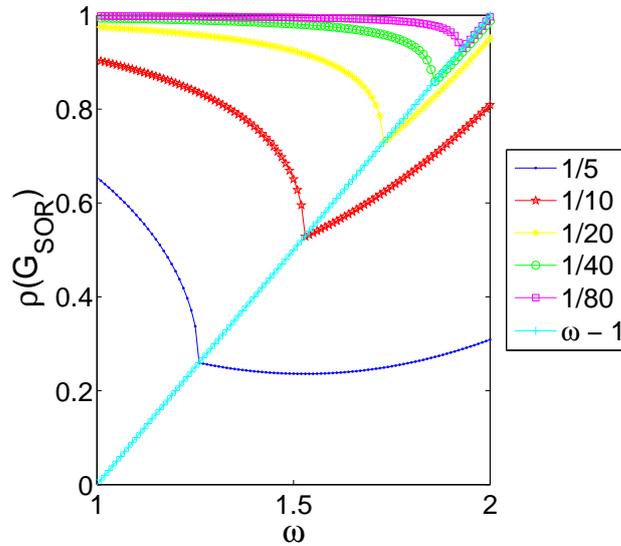


Figura 4.6: $\omega_{\text{ótimo}} \times \rho(G_{SOR})$ para malha cartesiana $h = \frac{1}{5}, \frac{1}{10}, \frac{1}{20}, \frac{1}{40}, \frac{1}{80}$

Note que usando (4.14), encontra-se o ponto mínimo da curva para cada tamanho de malha usada nas simulações. O ponto em que acontece a intersecção da curva com a reta é exatamente onde temos o menor raio espectral de G_{SOR} para o $\omega_{\text{ótimo}}$ conforme a figura 4.6.

É importante lembrar que (4.14) só é válida quando considera-se os pontos interiores da malha cartesiana e também para a malha centrada.

Todavia o problema de Neumann, não encontra-se na literatura uma fórmula na qual pode-se estimar o melhor coeficiente de relaxação. Assim, como sabe-se que o método SOR converge somente quando $0 < \omega < 2$, de acordo com o teorema 3.6.1 partiu-se para o teste em torno deste intervalo a qual consiste em

encontrar o melhor coeficiente de relaxação em uma malha cartesiana de dimensões 21×21 e $tol = 10^{-6}$. Montou-se a matriz do método iterativo SOR para vários valores de ω e calculou-se o raio espectral para cada uma dessas matrizes, plotando os valores na figura 4.7. Note que a curva é semelhante as curvas da figura 4.6.

Para $h = \frac{1}{20}$ o valor do $\omega_{\text{ótimo}}$ está próximo de $\omega_{\text{ótimo}} \approx 1.8$.

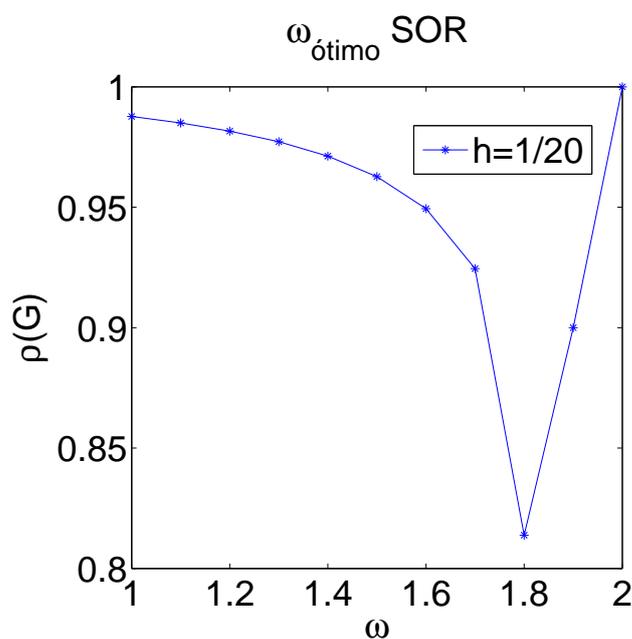


Figura 4.7: $\omega_{\text{ótimo}} \times \rho(G_{SOR})$ para malha cartesiana 21×21 do problema de Neumann

4.3.1 Raio espectral, Taxa de convergência e $\omega_{\text{ótimo}}$

Inicialmente apresenta-se os resultados numéricos referente ao problema de Dirichlet bidimensional e posteriormente para o problema de Neumann bidimensional em uma malha cartesiana 21×21 e $tol = 10^{-6}$.

De acordo com a literatura de [Young(1988), Datta(1994)] deveríamos encontrar os resultados numéricos conforme a tabela 4.5.

Tabela 4.5: Resultados teóricos para ω , ρ , R e n_T com condições de contorno (4.1) e usando $h = \frac{1}{20}$.

ω	$\rho(G_{SOR})$	$R(G_{SOR})$	n_T
1,7286	0,7286	0,3166	44

Na tabela 4.6 a partir do coeficiente de relaxação ω apresenta-se o raio espectral ρ da matriz de SOR, a taxa de convergência R , o número de iterações n_O realizadas pelo método até obter a convergência e uma estimativa referente a quantidade de iterações n_T que o método deveria realizar para obter a convergência. O objetivo em construir a tabela é encontrar o melhor coeficiente de relaxação ω , que nos fornece o menor raio espectral para a matriz de SOR. Mostra-se apenas o intervalo (1.70, 1.74) pois é onde temos o ponto mínimo.

Tabela 4.6: Experimento numérico para equação de Poisson com condições de contorno de Dirichlet através do método iterativo SOR. n_O é o número de iterações observado e n_T é o número de iterações teórico.

ω	$\rho(G_{SOR})$	$R(G_{SOR})$	n_O	n_T
1,700	0,8262	0,1810	52	77
1,705	0,8191	0,1997	53	70
1,710	0,8108	0,2097	53	66
1,715	0,8011	0,2219	53	63
1,720	0,7888	0,2372	54	59
1,725	0,7715	0,2594	55	54
1,7286	0,7486	0,2896	55	48
1,730	0,7300	0,3147	55	44
1,735	0,7350	0,3079	56	45
1,740	0,7400	0,3011	56	46

Entretanto o teste realizado com as condições de contorno (4.2) montou-se a tabela 4.7 mostrando para diversos valores de ω o pseudo raio espectral δ , a taxa de convergência R e número de iterações referente ao método iterativo SOR. A busca em torno deste coeficiente $\omega_{\text{ótimo}}$ é para uma malha de dimensão 21×21 , assim como o menor raio espectral de G_{SOR} .

Dentro do intervalo $(1.80, 1.84)$ quando $\omega = 1,805$ (veja a tabela 4.7) conseguiu-se o menor raio espectral da matriz de iteração SOR e para este coeficiente temos a melhor taxa de convergência entre os coeficientes testados. Entretanto o número de iterações realizadas não é o melhor possível. Esta discrepância é explicada em ([Young(1988), p. 226]).

Tabela 4.7: Experimento numérico para equação de Poisson com condições de contorno de Neumann através do método iterativo SOR. n_O é o número de iterações observado e n_T é o número de iterações teórico.

ω	$\rho(G_{SOR})$	$R(G_{SOR})$	n_O
1,800	0,8138	0,2060	113
1,805	0,8050	0,2169	112
1,810	0,8100	0,2107	110
1,815	0,8150	0,2045	108
1,820	0,8200	0,1984	107
1,825	0,8250	0,1923	105
1,830	0,8300	0,1863	104
1,835	0,8350	0,1803	102
1,840	0,8400	0,1743	100

4.3.2 Comparação da Velocidade de Convergência

Usou-se os métodos iterativos Jacobi, Gauss-Seidel e SOR para resolver o sistema linear $A\vec{p} = \vec{b}$ onde A é a matriz de discretização obtida a partir da fórmula dos cinco pontos do laplaciano no quadrado unitário $\Omega = (0, 1) \times (0, 1)$ e \vec{b} é estabelecido pelas condições de contorno de Dirichlet ou Neumann (nosso objeto de estudo).

Apresenta-se resultados referente ao número de iterações necessárias para o problema de Dirichlet e de Neumann convergir de acordo com a margem de tolerância 10^{-6} , para três refinamentos possíveis da malha 6×6 , 11×11 e 21×21 (correspondentes a matrizes de dimensões $n = 16$, 81 e 361 , respectivamente), de acordo com cada método e para diferentes valores de ω no caso do método SOR é apresentado nas tabelas 4.8 e 4.9.

Para o Problema de Dirichlet note que na tabela o método de Gauss-Seidel é cerca de duas vezes mais rápido para convergir que o método de Jacobi e que dependendo da escolha de ω , o método SOR pode ser muito mais rápido que o método de Gauss-Seidel para a malha mais refinada. Subrelaxamento $\omega = 0.8$ não ajuda e para $\omega = 2$ o método SOR é divergente.

Tabela 4.8: Experimento numérico para equação de Poisson com condições de contorno de Dirichlet através dos métodos iterativos Jacobi, Gauss-Seidel e SOR.

	$h = 0,2$	$h = 0,1$	$h = 0,05$
Jacobi	57	181	501
SOR ($\omega = 0,8$)	42	98	147
Gauss-Seidel	29	73	120
SOR ($\omega = 1,4$)	20	38	83
SOR ($\omega = 1,6$)	31	33	62
SOR ($\omega = 1,7$)	43	45	52
SOR ($\omega = 1,8$)	65	65	77
SOR ($\omega = 1,9$)	134	132	136
SOR ($\omega = 2,0$)	∞	∞	∞

Para o problema de Neumann, o método de Gauss-Seidel é mais rápido para convergir que o método de Jacobi e que dependendo da escolha de ω , o método SOR pode ser mais rápido que o método de Gauss-Seidel para a malha mais refinada, mas no máximo que conseguimos é que ele seja duas vezes mais rápido, conforme a tabela 4.9.

Tabela 4.9: Experimento numérico para equação de Poisson com condições de contorno de Neumann através dos métodos iterativos Jacobi, Gauss-Seidel e SOR.

	$h = 0,2$	$h = 0,1$	$h = 0,05$
Jacobi	62	217	571
SOR ($\omega = 0,8$)	68	215	557
Gauss-Seidel	51	160	424
SOR ($\omega = 1,4$)	24	90	244
SOR ($\omega = 1,6$)	26	64	176
SOR ($\omega = 1,7$)	28	49	145
SOR ($\omega = 1,8$)	29	46	113
SOR ($\omega = 1,9$)	37	82	98
SOR ($\omega = 2,0$)	50	387	2491

5 A EQUAÇÃO DE NAVIER-STOKES

Neste capítulo mostra-se os resultados da simulação de fluidos através da equação de Navier-Stokes em uma cavidade quadrada. Para a discretização das equações empregou-se o método de diferenças finitas estudado por [Chorin(1997)], [Johnston and Liu(2002)] entre outros. Apresenta-se também as metodologias adotadas na resolução e discretização das equações que segue o método PRIME. A visualização dos resultados é realizada através do *software* Visual [Justo(1998)].

5.1 Sistema de Equações

Em coordenadas cartesianas bidimensionais para fluxos laminares de fluidos incompressíveis e isotérmicos a equação de Navier-Stokes é representada pela equação da continuidade que reflete o princípio físico da conservação de massa, e pelas equações do movimento nas direções x e y . A formulação diferencial da equação de Navier-Stokes na forma adimensional é dada por

$$\frac{\partial u}{\partial t} + \frac{\partial uu}{\partial x} + \frac{\partial uv}{\partial y} = -\frac{\partial p}{\partial x} + \frac{1}{Re} \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right) \quad (5.1)$$

$$\frac{\partial v}{\partial t} + \frac{\partial uv}{\partial x} + \frac{\partial vv}{\partial y} = -\frac{\partial p}{\partial y} + \frac{1}{Re} \left(\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} \right) \quad (5.2)$$

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0 \quad (5.3)$$

onde, u e v representam as velocidades nas direções dos eixos coordenados x e y , p a pressão, Re representa o número de Reynolds que está relacionado com as forças viscosas e com as forças de inércia. Segundo [Hughes and Brighton(1967)] quando o número de Re é pequeno, ou seja, $Re \ll 1$ as forças viscosas dominam, e quando $Re \gg 1$ de inércia predominam.

Tomando uma região limitada Ω , ou seja, definiu-se esta região por $0 \leq x, y \leq 1$ e $\partial\Omega$ a fronteira desta região. Quanto ao tipo de condições de contorno

[Fortuna(2000)] ressalta que isso depende muito do problema que está sendo resolvido, pois a escolha incorreta destas condições pode levar a conclusões incorretas e até mesmo afetar na estabilidade do método a ser utilizado na solução do problema.

Conforme [Chorin(1968)], usando o método PRIME (descrito no Apêndice A.1) surge uma equação de Poisson para a pressão da forma

$$\frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} = f \quad (5.4)$$

onde

$$f = \nabla \cdot \vec{u} = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \quad (5.5)$$

é o termo fonte.

As condições de contorno de Neumann, segundo [Johnston and Liu(2004), Orszag et al.(1986)Orszag, Israeli, and Deville], são as mais apropriadas para o termo da pressão. Estas são definidas como

$$\frac{\partial p}{\partial x} = 0, \quad y = 0, \quad \text{ou} \quad y = 1 \quad \text{e} \quad 0 \leq x \leq 1 \quad (5.6)$$

$$\frac{\partial p}{\partial y} = 0, \quad x = 0, \quad \text{ou} \quad x = 1 \quad \text{e} \quad 0 \leq y \leq 1 \quad (5.7)$$

Embora sejam determinadas condições de Neumann para o termo da pressão, ainda é necessário que o termo fonte na equação de Poisson (5.4) satisfaça a condição (4.4) para $t > 0$. Por exemplo, se o problema tiver condições de contorno de parede para a velocidade ($u = 0$ e $v = 0$ em todo o contorno) e condições de Neumann para a pressão, fisicamente isto equivale a não termos nenhum fluido entrando no sistema (pelos contornos). Nesse caso se a condição acima não for satisfeita, teremos um termo fonte (ou sumidouro) no interior do fluido.

Para um fluido incompressível, garantir que

$$\int \int f dx dy = \int \int \vec{\nabla} \cdot \vec{u} dx dy = 0 \quad (5.8)$$

equivale a dizer que o fluido deva ser *divergence-free*, caso contrário teremos fluido sendo inserido no interior do domínio a cada passo de tempo, ou seja, uma condição puramente numérica e errônea.

Segundo [Chorin(1968)] garantindo que a cada passo de tempo o fluido seja *divergence-free* para que seja sempre *divergence-free*. Se esta condição não for satisfeita, fluido é inserido lentamente ao domínio (uma fonte de erro numérico). Assim o ideal é iniciar a simulação com um campo de velocidade que satisfaça a equação da continuidade. Nem sempre a solução é conhecida a priori, o que torna difícil a resolução deste problema.

Nos problemas estudados, iniciou-se com $u = 0$ e $v = 0$ e uma descontinuidade no contorno gerando uma solução que não é *divergence-free* no tempo inicial. Isso introduz um termo fonte numérico errôneo. Para amenizar esse problema, usou-se o fato que a solução da equação de Poisson com condições de Neumann não é única a menos de uma constante. Assim, subtraindo a cada iteração do método iterativo da pressão uma constante da solução de modo que a solução fique fixa em um ponto em $p = 0$. Esta condição é utilizada em [Pozrikidis(2001)]. Numericamente isto equivale a usar a condição numérica¹

$$p^{n+1} = p^{n+1} - p_{1,2}^{n+1} \quad (5.9)$$

Para [Pozrikidis(2001)] isso equivale dizer que (5.9) é uma regularização a cada iteração e ao ser implementado simplesmente estamos deslocando todos os componentes do vetor solução p pela mesma quantidade após cada iteração. Note que a explicação acima pode ser entendida como “fixe $p = 0$ ” em algum ponto do domínio, neste caso em $x_{1,2}$.

¹Para a malha deslocada não é necessário usar a restrição (5.9) o método converge diretamente usando apenas as condições de contorno de Neumann para a equação de Poisson, o que não acontece com a malha centrada.

5.2 O Problema da Cavidade

Resolvendo a equação de Navier-Stokes sob o sistema de equações gerado é necessário de condições iniciais que são os valores de entrada e estes são de suma importância para o escoamento de um fluido. As condições iniciais para o problema proposto devem assumir valores suaves para a convergência ocorrer rapidamente. Estão elas apresentadas a seguir:

$$u(x, y, 0) = 1 \quad \text{para } (x, y) \in \Omega \quad (5.10)$$

$$v(x, y, 0) = 0 \quad \text{para } (x, y) \in \Omega \quad (5.11)$$

$$p(x, y, 0) = 0 \quad \text{para } (x, y) \in \Omega \quad (5.12)$$

Para este problema temos apenas condições de Dirichlet, onde o escoamento é induzido pelo movimento de deslizamento na parede do topo da cavidade. As condições de contorno nas paredes estacionárias estão sob os contornos leste, oeste e sul e são aplicadas condições *inflow* somente no contorno norte. Sendo que as condições de contorno são definidas para $t > 0$ e portanto são determinadas por

$$u(x, 0, t) = 0, \quad v(x, 0, t) = 0 \quad \text{para } 0 < x < 1 \quad (5.13)$$

$$u(x, 1, t) = 1, \quad v(x, 1, t) = 0 \quad \text{para } 0 < x < 1 \quad (5.14)$$

$$u(0, y, t) = 0, \quad v(0, y, t) = 0 \quad \text{para } 0 < y < 1 \quad (5.15)$$

$$u(1, y, t) = 0, \quad v(1, y, t) = 0 \quad \text{para } 0 < y < 1 \quad (5.16)$$

Os resultados são casos de simulações referente ao problema da cavidade onde utilizou-se as condições de contorno para pressão (5.6) e (5.7) e para as componentes de velocidade (5.13), (5.14), (5.15) e 5.16.

Para comparar gráficos de linhas de corrente para os escoamentos propostos definiu-se ($Re = 10$, $Re = 100$ e $Re = 1000$) desta forma pode-se comparar

os perfis de velocidade, para os vários números de Reynolds. Ainda que, aumentando o número de Reynolds há o deslocamento do centro do vórtice com pequenas perturbações. Para ambos os testes realizou-se as simulações deste problema em malha centrada e deslocada para $t_{final} = 1$, isto equivale a 640000 iterações.

Ilustra-se o comportamento do fluido através de 5.1 e 5.2, para $Re = 10$ na malha uniforme centrada e deslocada. É possível identificar que o perfil de velocidade é parabólico, e o fluido apresenta um vórtice maior na região central e dois menores nos cantos inferiores da malha com pequenas perturbações, pois é nestes pontos (próximo a parede) em que contêm os valores extremos das pressões. Pode-se observar ainda que nos pontos onde acontece a colisão do fluido com a parede a pressão é maior.

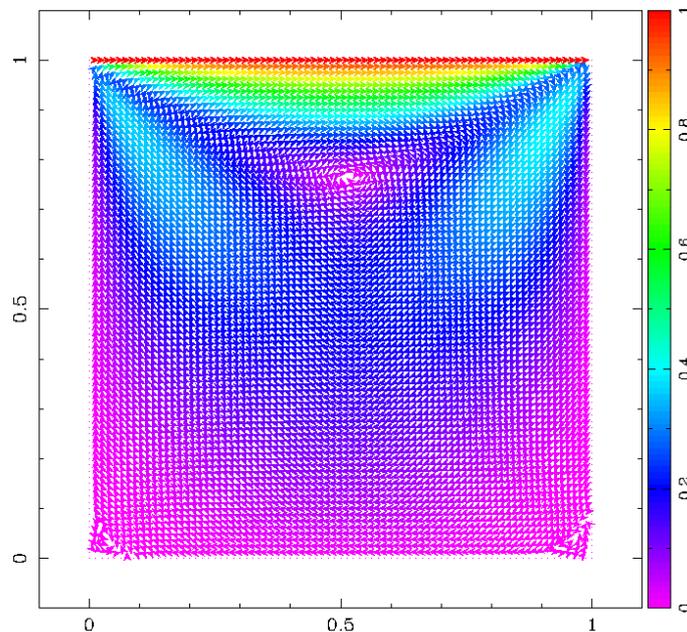


Figura 5.1: *Solução do escoamento para o problema da cavidade em malha uniforme centrada 81×81 . Gráfico dos vetores das componentes de velocidade.*

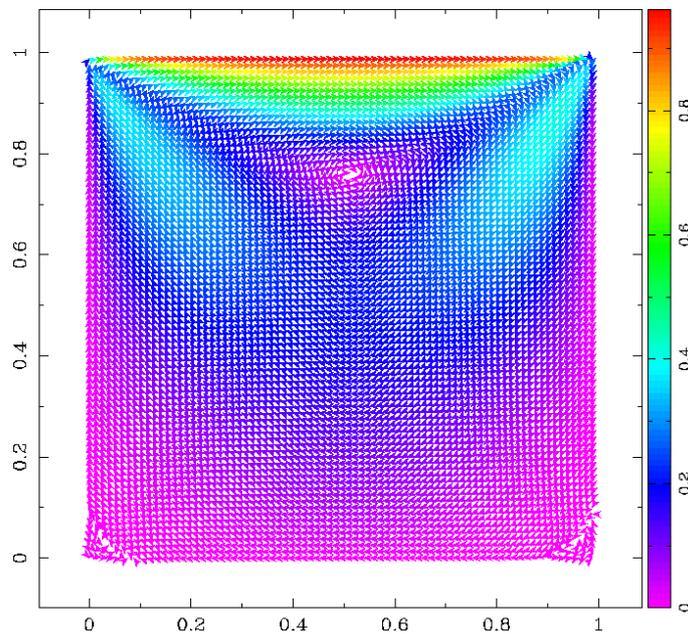


Figura 5.2: *Solução do escoamento para o problema da cavidade em malha uniforme deslocada 81×81 . Gráfico dos vetores das componentes de velocidade.*

Nas figuras 5.3 e 5.4 apresenta-se o comportamento do fluido no interior da cavidade quadrada para o número de Reynolds $Re = 100$. Os vórtices dependem principalmente do número de Reynolds, o fluido apresenta um vórtice maior na região central e dois menores nos cantos inferiores.

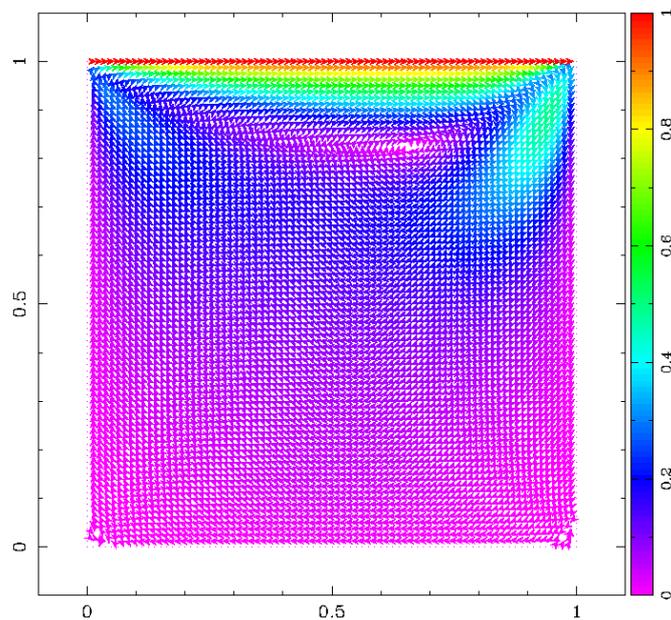


Figura 5.3: Solução do escoamento para o problema da cavidade em malha uniforme centrada 81×81 . Gráfico dos vetores das componentes de velocidade.

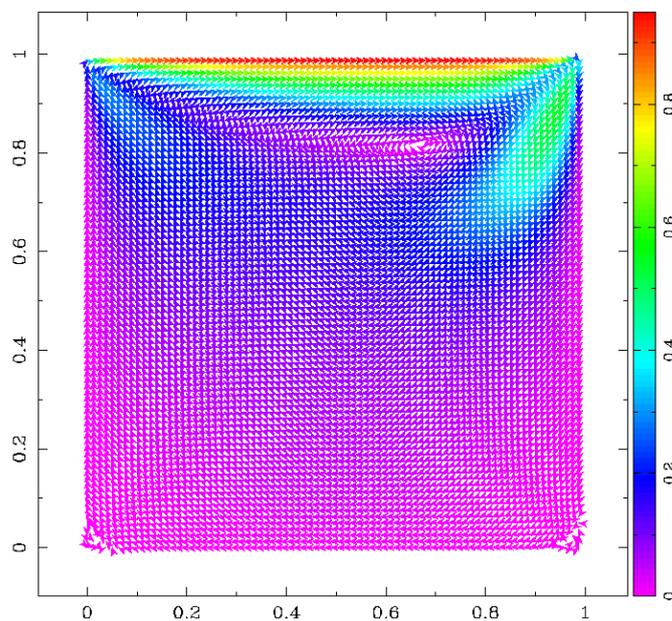


Figura 5.4: Solução do escoamento para o problema da cavidade em malha uniforme deslocada 81×81 . Gráfico dos vetores das componentes de velocidade.

Procedendo da mesma maneira como nos casos anteriores mas agora com $Re = 1000$ os gráficos 5.5 e 5.6 representam o comportamento do fluido na cavidade quadrada. É possível identificar o vórtice principal e os dois secundários na parte inferior da cavidade.

Ainda aumentando o número de Reynolds há um deslocamento do vórtice principal para o canto direito superior da cavidade.

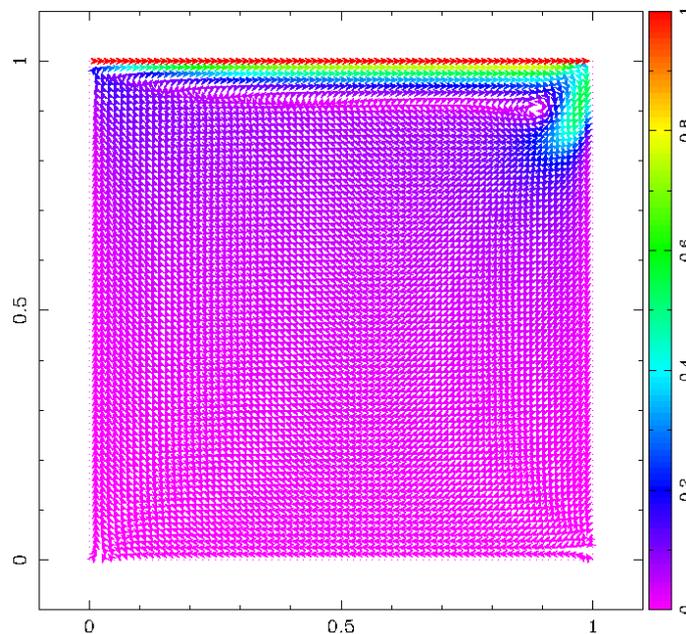


Figura 5.5: *Solução do escoamento para o problema da cavidade em malha uniforme centrada 81×81 . Gráfico dos vetores das componentes de velocidade.*

Realizou-se um teste com o intuito de responder a seguinte questão: Em que situação vale a pena usar uma malha deslocada ou centrada?

O problema da cavidade bidimensional foi iterado utilizando o método iterativo Gauss-Seidel até que obtivéssemos a solução permanente. Temos o resultado para 40000 iterações e tempo de execução 16.33s para a malha cartesiana deslocada e de 15.18s para a malha cartesiana centrada. Ainda fez-se o teste com o método iterativo SOR ($\omega = 1.8$) o tempo de execução para malha deslocada foi de

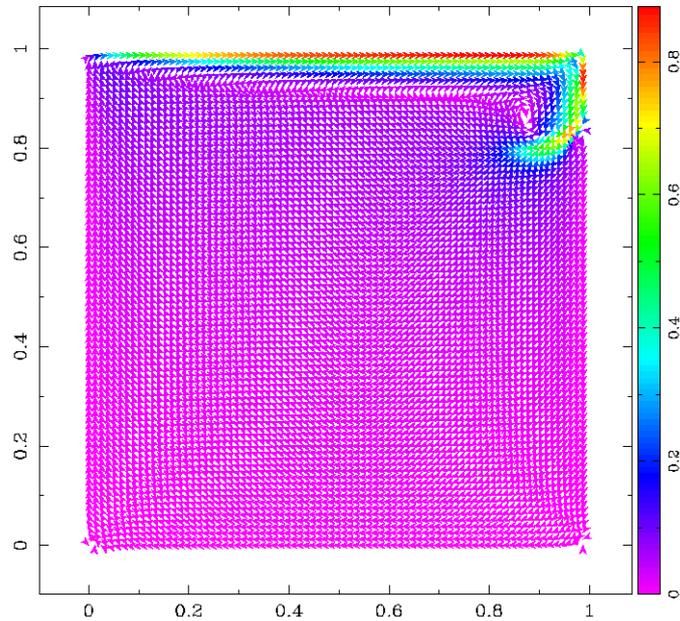


Figura 5.6: *Solução do escoamento para o problema da cavidade em malha uniforme deslocada 81×81 . Gráfico dos vetores das componentes de velocidade.*

16.38s e centrada de 15.58s. Para uma tolerância de $tol = 10^{-6}$ e malha cartesiana de 21×21 .

Justamente isto nos mostra que estamos obtendo a mesma solução embora estejamos com estruturas de malhas diferentes (centrada e deslocada). É claro que não podemos concluir algo definitivo, pois para o tamanho de malha testado não obteve-se vantagem quanto ao tempo de execução do algoritmo.

É importante ressaltar que o problema da cavidade bidimensional é uma aplicação de escoamento que tem sido extensivamente estudado. É possível encontrar muitos trabalhos sobre este problema na literatura [Francisquetti(2010)], [Griebel et al.(1997)Griebel, Dornseifer, and Neunhoffer], entre outros. Em nosso trabalho o objetivo em estudar este problema deu-se pelo fato de validar quanto à precisão, eficiência numérica e principalmente as condições de contorno para o termo da pressão na qual é dada pela equação de Poisson. Para tal equação usou-se condições de contorno de Neumann.

Apresentou-se nos capítulos anteriores as condições para o qual obtém-se uma aproximação da solução desta equação, e que é necessário que a condição de integrabilidade (4.4) seja satisfeita, caso contrário o problema não tem solução e não irá convergir. Isso também se dá para o problema da cavidade a qual temos como termo fonte (5.5).

6 CONCLUSÃO

O presente trabalho teve por objetivo estudar a análise espectral da equação de Poisson, em um domínio uni e bidimensional. Além disso, objetivou-se estudar métodos iterativos na qual seria utilizado para calcular a solução aproximada do sistema linear obtido a partir da discretização da equação de Poisson.

Primeiramente buscamos entender as matrizes de discretização a partir da equação de Poisson para os problemas modelos, e principalmente quando estávamos tratando do problema de Neumann, pois para este pouquíssimas informações encontramos na literatura. A matriz gerada a partir de cada um dos problemas por ser uma matriz esparsa e tridiagonal permite com facilidade encontrar os autovalores de cada uma das matrizes, a qual seguiu-se a literatura de [Yueh(2005)]. Assim, com o intuito de verificar os resultados obtidos, partiu-se para o estudo dos autovalores das matrizes de discretização A , assim como a obtenção das fórmulas dos autovalores para um caso geral.

Após a obtenção do sistema $A\vec{p} = \vec{b}$ investigamos os métodos iterativos básicos para a resolução de sistemas lineares. Assim fez-se um breve estudo referente aos métodos iterativos Jacobi, Gauss-Seidel e SOR para resolver o sistema linear gerado a partir da equação de Poisson. Esta equação exige a utilização de condições de contorno adequadas, a escolha destas não é única e a convergência do método depende delas.

Começamos então com o problema de Dirichlet, para este não encontramos obstáculos pelo caminho pois com as simulações através de testes de precisão e com relevantes informações encontradas na literatura nos levou a concluir que o sistema linear obtido a partir deste método nos fornece a existência de uma única solução. Além disso obteve-se uma fórmula geral para os autovalores de cada ma-

triz de iteração referente aos métodos Jacobi e Gauss-Seidel, e a partir desta é que concluiu-se quanto a convergência de cada método iterativo.

Ao tratar do problema de Neumann analisou-se as matrizes de iteração a que se refere aos métodos iterativos e obteve-se que o maior autovalor é exatamente 1, e de acordo com algumas literaturas [Young(1988), Datta(1994), Saad(2000)] deveríamos ter que o raio espectral fosse estritamente menor do que 1 para o método iterativo ser convergente. Por isso fez necessário um estudo mais avançado sobre este problema.

A literatura de [Pozrikidis(2001)] e [Neumann and Plemmons(1978)] apresenta teoremas e definições na qual conseguiu-se encontrar uma fórmula geral para os autovalores e então concluí que os métodos usados para encontrar a solução do problema de Neumann é condicionalmente convergente, pois encontrou-se que o autovalor $\lambda = 1$ é único, e assim seria considerado o segundo maior autovalor como sendo o pseudo raio espectral da matriz de iteração.

Uma das referências importantes aos métodos iterativos é que estes nos indicam melhorias referente as taxas de convergência de um método para o outro. Quanto aos métodos em si, pode-se comprovar através de testes e também concluir que o método iterativo SOR foi o que apresentou melhores resultados, com a vantagem de ter um menor esforço computacional. Para este método podemos sensivelmente melhorá-lo na procura do coeficiente de relaxação na qual tem por objetivo acelerar a convergência.

Baseados nos objetivos que nortearam esta pesquisa concluí-se e pode-se comprovar através das simulações e de teoremas encontrados na literatura [Datta(1994), Saad(2000), Young(1988)] a eficiência do algoritmo PRIME em relação a convergência espectral da equação de Poisson, sendo assim o algoritmo implementado supriu nossas expectativas para as condições de contorno definidas para a equação de Poisson.

Assim como, o uso desta no termo da pressão na equação de Navier-Stokes já que conseguiu-se concordâncias razoáveis com dados pesquisados na bibliografia.

Uma condição para a existência de solução do problema de Poisson que aparece na solução da equação de Navier Stokes é necessário que a integral de $\int f = \nabla \cdot \vec{u} = 0$ o que implica que a integral do divergente da velocidade seja zero o que é chamado de *divergence-free*.

Ainda, estudou-se a equação de Navier-Stokes para o problema da cavidade quadrada em Dinâmica de Fluidos Computacional, onde fez-se vários testes com o intuito de verificar a eficiência do código computacional, assim como o comportamento da solução dos problemas em estudo.

Com o objetivo de validar os resultados das simulações numéricas fez-se vários testes para diversos número de Reynolds a fim de verificar o comportamento do fluido que foram apresentados forma de gráficos para a velocidade. Além disso, foram apresentadas as superfícies de corrente que permitem construir a representação do movimento do fluido na cavidade, na qual percebe-se a formação de vórtices a partir dos cantos da cavidade e na região central.

Sugere-se para trabalhos futuros, num primeiro momento estudar a convergência espectral da equação de Poisson para o caso 3D que é direta utilizando o produto de Kronecker com três matrizes. E num segundo momento estudar a convergência desta equação sobre outros métodos iterativos e também outras condições de contorno, por exemplo, Robin ou periódica.

Apêndice A

Neste apêndice apresentamos o algoritmo **PRIME**.

A.1 Algoritmo PRIME

A solução da equação de Navier-Stokes é encontrada com a utilização do algoritmo **PRIME** – (**PR**essão **Imp**lícita, **M**omento **Exp**lícito) conforme [Griebel et al.(1997)Griebel, Dornseifer, and Neunhoffer].

Existem vários métodos para aproximarmos as derivadas nas equações governantes no escoamento de um fluido. Escolheu-se o método de Euler para a aproximação de derivada temporal de primeira ordem.

$$\left[\frac{\partial u}{\partial t} \right]_{i,j}^{n+1} = \frac{u_{i,j}^{n+1} - u_{i,j}^n}{\Delta t} \quad (\text{A.1})$$

Desta forma fazendo a substituição em (5.1) e (5.2) respectivamente, obtém-se

$$u^{n+1} = u^n + \Delta t \left[\frac{1}{Re} \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right) - \frac{\partial uu}{\partial x} + \frac{\partial uv}{\partial y} - \frac{\partial p}{\partial x} \right] \quad (\text{A.2})$$

$$v^{n+1} = v^n + \Delta t \left[\frac{1}{Re} \left(\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} \right) - \frac{\partial uv}{\partial x} + \frac{\partial vv}{\partial y} - \frac{\partial p}{\partial y} \right] \quad (\text{A.3})$$

A partir de (A.2) e (A.3) obtém-se as velocidades intermediárias \hat{u}^n e \hat{v}^n sem o termo da pressão, pois a pressão é tratada de forma implícita, e assim as equações ficam da seguinte forma

$$\hat{u}^n = u^n + \Delta t \left[\frac{1}{Re} \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right) - \frac{\partial uu}{\partial x} - \frac{\partial uv}{\partial y} \right] \quad (\text{A.4})$$

$$\hat{v}^n = v^n + \Delta t \left[\frac{1}{Re} \left(\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} \right) - \frac{\partial uv}{\partial x} - \frac{\partial vv}{\partial y} \right] \quad (\text{A.5})$$

Podemos reescrever as equações (A.2) e (A.3) de modo que se obtenha

$$u^{n+1} = \hat{u}^n - \partial t \frac{\partial p^{n+1}}{\partial x} \quad (\text{A.6})$$

$$v^{n+1} = \hat{v}^n - \partial t \frac{\partial p^{n+1}}{\partial y} \quad (\text{A.7})$$

Haja visto que é necessário obtermos uma equação para calcularmos a pressão, na sequência fez-se a substituição de (A.6) e (A.7) na equação (5.3) e obteve-se

$$0 = \frac{\partial u^{n+1}}{\partial x} + \frac{\partial v^{n+1}}{\partial y} = \frac{\partial \hat{u}^n}{\partial x} - \partial t \frac{\partial^2 p^{n+1}}{\partial x^2} + \frac{\partial \hat{v}^n}{\partial y} - \partial t \frac{\partial^2 p^{n+1}}{\partial y^2} \quad (\text{A.8})$$

Após a reorganização dos termos, temos a equação de Poisson para a pressão p^{n+1} no tempo t^{n+1}

$$\frac{\partial^2 p^{n+1}}{\partial x^2} + \frac{\partial^2 p^{n+1}}{\partial y^2} = \frac{1}{\partial t} \left(\frac{\partial \hat{u}^n}{\partial x} + \frac{\partial \hat{v}^n}{\partial y} \right) \quad (\text{A.9})$$

Deste modo quando estamos usando a malha centrada tem-se velocidade e pressão acopladas, representadas por (A.9). As velocidades são tratadas de forma explícita, sendo que as mesmas devem ser atualizadas após o cálculo implícito da pressão.

A.1.1 Estrutura do Algoritmo

Descreve-se a estrutura do algoritmo do método PRIME.

1. $t=0$
2. Determinar condições iniciais para u, v e p
3. Enquanto $t < t_{final}$
4. Calcule as velocidades intermediárias \hat{u} e \hat{v} usando (A.4) e (A.5)

5. Aplica condições de contorno para \hat{u} e \hat{v}
6. Calcule a função f
7. Enquanto $it < it_{max}$
8. Calcula pressão (A.9)
9. Aplica condições de contorno para a pressão
10. Calcule u^{n+1} e v^{n+1} usando (A.6) e (A.7)
11. Aplica condições de contorno para u^{n+1} e v^{n+1}
12. $t = t + \Delta t$

A estabilidade numérica depende do método numérico e ainda é obtida através da limitação do passo de tempo, ou seja, uma escolha apropriada para Δt . Existe bastante literatura a respeito das propriedades de estabilidade de métodos numéricos, onde [Guy and Fogelson(2005)] ressalta que um dos pioneiros neste estudo foi [Chorin(1968)] na década de 60.

O método numérico PRIME utilizado nesse trabalho segue os resultados apresentados em [Johnston and Liu(2002), Francisquetti(2010)]. A estabilidade desse método numérico depende portanto de:

1. Estabilidade difusiva do fluxo:

$$\frac{\Delta t}{(Re\Delta x)^2} \leq \left(\frac{1}{2}\right)^d \quad (\text{A.10})$$

onde Δx é o espaçamento entre as células na malha computacional e d é a dimensão do domínio, o que implica que

$$\Delta t \leq \frac{Re}{2} \left[\frac{1}{(\Delta x)^2} + \frac{1}{(\Delta y)^2} \right]^{-1} \quad (\text{A.11})$$

2. A estabilidade convectiva do fluxo:

$$\|\mathbf{u}\|_{L^\infty} \frac{\Delta t}{\Delta x} = CFL \leq 1 \quad (\text{A.12})$$

Segundo [Justo(2001)], devemos escolher Δt de modo que o método seja estável e não ocorra oscilações na solução. Portanto as seguintes condições são de suma importância para se determinar o valor de Δt , deste modo temos

$$|u_{max}|\Delta t < \Delta x \quad \text{e} \quad |v_{max}|\Delta t < \Delta y \quad (\text{A.13})$$

Assim, para a estabilidade numérica devemos ter que

$$\Delta t \leq \tau \cdot \min \left(\frac{Re}{2} \left[\frac{1}{(\Delta x)^2} + \frac{1}{(\Delta y)^2} \right]^{-1}, \frac{\Delta x}{|u_{max}|}, \frac{\Delta y}{|v_{max}|} \right) \quad (\text{A.14})$$

onde $0 < \tau \leq 1$, é chamado fator de segurança.

Referências Bibliográficas

- [Berman and Plemmons(1979)] A. Berman and R. J. Plemmons. *Nonnegative Matrices in the Mathematical Sciences*. Computer Science and Applied Mathematics, 1979.
- [Biezuner(2007)] R. J. Biezuner. *Métodos Numéricos para Equações Diferenciais Parciais Elípticas*. Departamento de Matemática - UFMG, 2007.
- [Chorin(1968)] A. J. Chorin. Numerical solution of the navier-stokes equations. *Journal Mathematics of Computation*, 22(104):745–762, 1968. ISSN 0025-5718.
- [Chorin(1997)] A. J. Chorin. A numerical method for solving incompressible viscous flow problems. *Journal of Computational Physics*, (135):118–125, 1997.
- [Datta(1994)] B. N. Datta. *Numerical Linear Algebra and Applications*. Books/Cole Publishing Company, 1994.
- [Fonseca(2009)] B. M. C. Fonseca. *O Ensino de Multigrid em Álgebra Linear Numérica*. Departamento de Matemática - UFMG, 2009.
- [Fortuna(2000)] A. O. Fortuna. *Técnicas computacionais para dinâmica dos fluidos: conceitos básicos e aplicações*. Edusp, 2000.
- [Francisquetti(2010)] E. P. Francisquetti. Estudo de quadtree para uso em dinâmica de fluidos computacional. Master's thesis, UFRGS, 2010.
- [Griebel et al.(1997)Griebel, Dornseifer, and Neunhoffer] M. Griebel, T. Dornseifer, and T. Neunhoffer. *Numerical Simulation in Fluid Dynamics: A practical Introduction*. SIAM, 1997. ISBN 0898713986.

- [Guy and Fogelson(2005)] R. D. Guy and A. L. Fogelson. Stability of approximate projection methods on cell-centered grids. *Journal of Computation Physics*, (203):517–538, 2005.
- [Hackbusch(1992)] W. Hackbusch. *Elliptic Differential Equations: Theory and Numerical Treatment*. Springer, 1992.
- [Hughes and Brighton(1967)] W. F. Hughes and J. A. Brighton. *Dinâmica dos Fluidos*. 1967.
- [Johnston and Liu(2002)] H. Johnston and J. Liu. Finite difference schemes for incompressible flow based on local pressure boundary conditions. *Journal of Computation Physics*, (180):120–154, 2002.
- [Johnston and Liu(2004)] H. Johnston and J. Liu. Accurate, stable and efficient navier-stokes solvers based on explicit treatment of the pressure term. *Journal of Computation Physics*, (199):221–259, 2004.
- [Justo(1998)] D. A. R. Justo. *Visual*. LICC, 1998.
- [Justo(2001)] D. A. R. Justo. Geração de malhas, condições de contorno e discretização de operadores para dinâmica de fluidos computacional. Master's thesis, UFRGS, 2001.
- [Kelley(1995)] C. T. Kelley. *Iterative Methods for Linear and Nonlinear Equations*. Society for Industrial Mathematics, 1995. ISBN 0898713528.
- [Laub(2005)] A. J. Laub. *Matrix Analysis for Scientists and Engineers*. Siam, 2005.
- [Neumann and Plemmons(1978)] M. Neumann and R. J. Plemmons. Convergent nonnegative matrices and iterative methods for consistent linear systems. *Journal Numerical Mathematics*, pages 265–279, 1978.

- [Orszag et al.(1986)Orszag, Israeli, and Deville] S. A. Orszag, M. Israeli, and M. O. Deville. Boundary conditions for incompressible flows. *Journal of Scientific Computing*, 1(1):75–111, 1986.
- [Plemmons(1976)] R. J. Plemmons. Regular splittings and the discrete poisson-neumann problem. *Numer. Math 25*, pages 153–161, 1976.
- [Pozrikidis(2001)] C. Pozrikidis. A note on the regularization of the discrete poisson-neumann problem. *Journal of Computational Physics*, pages 917–923, 2001.
- [Saad(2000)] Y. Saad. *Iterative Methods for Sparse Linear Systems*. Society for Industrial Mathematics, 2 edition, 2000. ISBN 0898715342.
- [Strikwerda(2004)] J. C. Strikwerda. *Finite Difference Schemes and Partial Differential Equations*. Siam, 2 edition, 2004.
- [van Linde(1974)] H. J. van Linde. High-order finite-difference methods for poisson’s equation. *Journal Mathematics of Computation*, pages 369–391, 1974.
- [Young(1988)] D. M. Young. *Iterative Solution of Large Linear Systems*. Academic Press, Inc, 1988.
- [Yueh(2005)] W. C. Yueh. Eigenvalues of several tridiagonal matrices. *Applied Mathematics E-Notes*, pages 66–74, 2005.