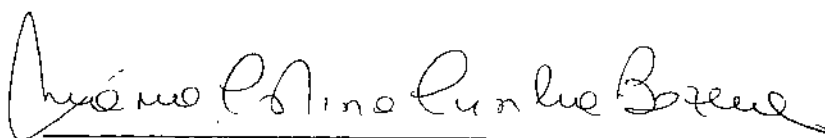


O MÉTODO DA COLOCAÇÃO EM ESPAÇOS POLINOMIAIS POR
PARTES: ASPECTOS GERAIS E APLICAÇÃO A UM
PROBLEMA PARABÓLICO NÃO-LINEAR

Este exemplar corresponde a redação final da tese devidamente corrigida e defendida pelo Sr. EDSON RODRIGUES CARVALHO e aprovada pela Comissão Julgadora.

Campinas, 10 de fevereiro de 1987



Prof^a. Dr^a. MARIA CRISTINA CUNHA BEZERRA
Orientadora

Dissertação apresentada ao Instituto de Matemática, Estatística e Ciência da Computação da Universidade Estadual de Campinas, como requisito parcial para a obtenção do título de Mestre em Matemática Aplicada, área: Análise Aplicada.

Í N D I C E

0.	NOTAÇÃO GERAL	
1.	INTRODUÇÃO	1
2.	O MÉTODO DA COLOCAÇÃO	3
2.1	Introdução	3
2.2	Uma Aplicação Simples	5
2.3	Colocação Ortogonal	11
2.4	Uma Relação entre os métodos de Colocação e Galerkin ..	17
3.	APLICAÇÃO A UMA EQUAÇÃO DIFERENCIAL DO TIPO PARABÓLICO	21
3.1	Introdução	21
3.2	Notação e Algumas Desigualdades	27
3.3	Alguns Resultados da Teoria da Aproximação	35
3.4	Uma Formulação de Galerkin	38
3.5	Análise de Convergência	42
3.6	O Método da Colocação no Tempo Discreto	49
4.	ASPECTOS COMPUTACIONAIS	53
4.1	Introdução	53
4.2	Discretização no Espaço	53
4.3	Discretização na Variável Temporal	58
4.4	A Matriz de Colocação	59
4.5	O Vetor de Colocação	61
4.6	Exemplos Numéricos	62
4.7	Detalhes do Programa	68
5.	CONCLUSÃO	84

Ã,

Fátima e ao Klausius,

os desencantos-uma constatação da vida acadêmica-,
e as certezas-um forte reflexo dos laços de família

AGRADECIMENTOS

- À Prof^a.Dr^a. Maria Cristina Cunha Bezerra, pela orientação criteriosa e segura com que me conduziu na realização deste trabalho.
- Aos colegas do Curso de Matemática do CEUL/UFMS que, assumindo minhas atividades na Instituição, permitiram a minha estada na UNICAMP.
- Aos Professores Petrônio Pulino e Antonio Carlos Moretti, pela solicitude e experiência no esclarecimento de dúvidas relativas ao aspecto computacional.
- Aos companheiros do Curso de Matemática Aplicada -eles e eu- um conjunto em busca de soluções de problemas que não o foram, dada a nossa união.
- À CAPES pelo necessário apoio.

0. NOTAÇÃO GERAL

$$I = [0, 1]$$

$$I_j = [x_{j-1}, x_j], \quad j = 1, 2, \dots, n$$

$$H_3 = \{v(x) \in C^1(I) \mid v \text{ é um polinômio cúbico sobre cada } I_j, j=1, \dots, n\}$$

$$\xi_{jk} = \frac{x_{j-1} + x_j}{2} + (-1)^k \frac{h_j}{2\sqrt{3}}, \quad j = 1, 2, \dots, n; \quad k = 1, 2, \text{ pontos de colocação (nós Gaussianos).}$$

$$L^2(E): \text{ espaço das funções } u \text{ de quadrados integráveis sobre } E$$

$$|\alpha| = \sum_{i=1}^n \alpha_i, \quad \text{para } \alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)$$

$$H^m(E) = \{u \in L^2(E); \quad \alpha, |\alpha| \leq m, D^\alpha u \in L^2(E)\}, \quad m \geq 0$$

$$\|u\|_{H^m(E)}^2 = \sum_{|\alpha| \leq m} \|D^\alpha u\|_{L^2(E)}^2$$

$$(u, v) = \int_0^1 u v dx = \sum_{j=1}^n (u, v)_j = \sum_{j=1}^n \int_{x_{j-1}}^{x_j} u v dx, \quad \text{produto escalar}$$

$$\text{de } L^2 \text{ sobre } I \text{ e } I_j.$$

$$\langle u, v \rangle_j = \frac{1}{2} [u(\xi_{j1})v(\xi_{j1}) + u(\xi_{j2})v(\xi_{j2})] h_j, \quad |u|_j^2 = \langle u, u \rangle_j$$

$$\langle u, v \rangle = \sum_{j=1}^n \langle u, v \rangle_j, \quad |u|^2 = \langle u, u \rangle, \quad \text{produto interno discreto.}$$

$$L^2(0, \tau; H^m(I)): \text{ espaço vetorial das funções } u(x, t) \text{ tais que para todo } t \in (0, T), u(x, t) \in H^m(I), \text{ cuja norma}$$

em $H^m(I)$ é uma função pertencente a $L^2(0, T)$.

$L^\infty(0, T; H^m(I))$: espaço vetorial das funções $u(x, t)$ tais que para todo $t \in (0, T)$, $u(x, t) \in H^m(I)$, cuja norma em $H^m(I)$ é uma função pertencente a $L^\infty(0, T)$.

$$\|u\|_{L^\infty(0, T; H^6(I))} = \sup_{0 < t < T} \|u(t)\|_{H^6(I)} < \infty$$

$$\|u_t\|_{L^2(0, T; H^6(I))}^2 = \int_0^T \|u_t(t)\|_{H^6(I)}^2 dt < \infty$$

$$H_3^0 = \{u \in H_3 \mid u(0) = u(1) = 0\}$$

$$H_0^1(I) = \{u \in H^1(I) \mid u(0) = u(1) = 0\}$$

$$\|u\|_{H_0^1}^2 = \int_0^1 (u')^2 dx$$

H_3 -interpolante de f : polinômio de Hermite cúbico por partes que interpola f e f' nos nós $x_j, j = 1, \dots, n$.

1. INTRODUÇÃO

Nesta dissertação estudaremos um método de colocação destinado a resolver problemas de valor inicial ou de fronteira para equações diferenciais, tendo por base funções polinomiais por partes. A razão de se considerar colocação com funções polinomiais por partes é que a utilização destas funções oferece vantagens significativas. A teoria é menos complexa e mais poderosa, pois, além de envolver cálculos com matrizes que são mais fáceis de construir, permite melhores resultados de convergência e demonstrações mais simples do que usando, por exemplo, polinômios ou funções trigonométricas [6].

O método da colocação que utilizaremos aqui basicamente envolve construir a solução aproximada como uma combinação linear de um conjunto conveniente de funções; os coeficientes desta aproximação são determinados exigindo que a combinação satisfaça a equação diferencial em certos pontos, chamados pontos de colocação. Uma escolha possível seriam os nós naturais; outra escolha, que enfatizaremos em nosso trabalho, são os nós Gaussianos da partição do domínio. Visto sob esta última forma, o método é bastante conhecido pelos químicos; de fato, o método da colocação tem sido utilizado de modo extensivo para resolver problemas de valor inicial e de fronteira, em uma e duas dimensões, que surgem em reatores dinâmicos e outros sistemas [3].

Na primeira parte deste trabalho nos propomos a desenvolver noções gerais sobre o método da colocação, envolvendo aplicações simples, visando fortalecer a idéia básica do mesmo; também, uma relação com o método de Galerkin será estabelecida e alguns resultados de convergência serão incluídos.

Na segunda parte desta dissertação, o nosso objetivo será aplicar o método da colocação a um problema parabólico não linear; faremos isto de maneira detalhada, incluindo uma análise da convergência e alguns esquemas de discretização na variável temporal visando a implementação futura.

Finalmente, analisaremos questões diretamente ligadas ao aspecto computacional, isto é, escolha da base, como obter a matriz de colocação (por elemento e global), solução do sistema linear, exemplos numéricos específicos para ilustrar a eficiência do método e sugestões para a implementação na forma de algumas subrotinas que serão detalhadas no final desta terceira parte.

2. O MÉTODO DA COLOCAÇÃO

2.1 INTRODUÇÃO: Nos últimos anos, estudiosos têm demonstrado que o método da colocação é uma técnica eficiente para aproximar a solução de problemas, envolvendo equações diferenciais. A idéia básica do método pode ser explicada brevemente. Vamos supor que estamos procurando uma solução para um problema que envolva a seguinte equação:

$$L(u) = f$$

onde L é um operador diferencial. Inicialmente escolhemos um espaço de dimensão finita P , chamado o espaço das funções de forma, e uma base $\{z_1, z_2, \dots, z_N\}$ de P . O método da colocação consiste em encontrar uma aproximação para u da forma:

$$(2.1.1) \quad U_N = \sum_{j=1}^N a_j z_j,$$

onde os coeficientes desconhecidos a_j , $j = 1, 2, \dots, N$, são determinados de tal modo que:

$$L U_N(x_i) = f(x_i), \quad 1 \leq i \leq N$$

onde $x_1, x_2, \dots, x_N$ são N pontos distintos pertencentes a um domínio D para os quais os termos de (2.1.1) fazem sentido.

Se esta função U_N existe dizemos que ela "coloca" $f(x)$

nos pontos $x_1, x_2, \dots, x_N$ e a chamaremos de aproximação por colocação para u no espaço P .

Em problemas de evolução os coeficientes a_j são funções da variável temporal, isto é, $a_j = a_j(t)$ ao passo que os elementos da base são funções da variável espacial, isto é,

$$z_j = z_j(x), \quad j = 1, 2, \dots, N.$$

A idéia de procurar uma solução aproximada como uma combinação linear de funções linearmente independentes para resolver problemas envolvendo equações diferenciais não é nova. Ela foi proposta pela primeira vez por Galerkin em 1915. O que é novo e tem tornado o método da colocação mais eficiente é a escolha do espaço P e a construção de bases apropriadas. Um exemplo importante, por suas aplicações não só no método em questão como também em outros métodos numéricos de solução de equações diferenciais e aproximação de funções, é o espaço P das funções polinomiais por partes. Neste caso as funções da base são escolhidas com suporte compacto, isto é, elas se anulam fora de um subintervalo "pequeno". Tal base é chamada uma base local. O método da colocação com um espaço de funções polinomiais por partes, uma base local e escolha conveniente de nós produz uma aproximação cuja convergência para uma solução u suficientemente suave é de ordem superior. Além disto, esta associação dá origem a um sistema de equações (linear quando o operador L o for) cuja matriz é esparsa. Esta característica tem vantagens

computacionais; na literatura especializada recente podemos constatar a importância e eficiência dos métodos numéricos destinados a explorar esta esparsidade.

2.2 UMA APLICAÇÃO SIMPLES: Consideremos o seguinte problema de valor de contorno

$$(2.2.1) \quad \begin{cases} Lu = u'' - \sigma(x)u(x) = f(x), & x \in [0, 1] \\ u(0) = u(1) = 0 \end{cases}$$

$\sigma(x) > 0$ e contínua em $[0, 1]$.

Seja P o espaço gerado pelas funções

$$B_i(x) = \frac{1}{h^3} \begin{cases} (x-x_{i-2})^3, & x \in [x_{i-2}, x_{i-1}], \\ h^3 + 3h^2(x-x_{i-1}) + 3h(x-x_{i-1})^2 - 3(x-x_{i-1})^3, & x \in [x_{i-1}, x_i], \\ h^3 + 3h^2(x_{i+1}-x) + 3h(x_{i+1}-x)^2 - 3(x_{i+1}-x)^3, & x \in [x_i, x_{i+1}], \\ (x_{i+2}-x)^3, & x \in [x_{i+1}, x_{i+2}], \\ 0, & \text{se } x \notin [x_{i-2}, x_{i+2}], \end{cases}$$

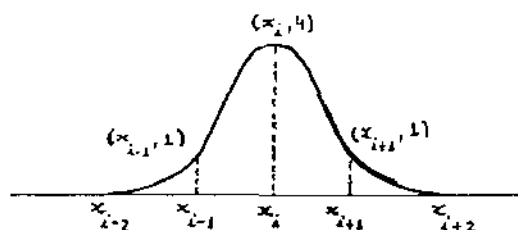
para $i = 0, 1, \dots, N$, onde $h = \frac{1}{N}$ e $x_0, x_1, \dots, x_N$ são $N+1$ pontos em $[0, 1]$ tais que $0 = x_0 < x_1 < \dots < x_N = 1$. Na definição acima usamos também quatro pontos adicionais satisfazendo:

$$x_{-2} < x_{-1} < x_0 \quad \text{e} \quad x_{N+2} > x_{N+1} > x_N .$$

Observemos que:

$$B_i(x_j) = \begin{cases} 4 & \text{se } j = i , \\ 1 & \text{se } j = i-1 \text{ ou } j = i+1 , \\ 0 & \text{se } j = i+2 \text{ ou } j = i-2 . \end{cases}$$

O gráfico de $B_i(x)$ é da forma:



Além disso, estas funções são duas vezes continuamente diferenciáveis apresentando os seguintes valores:

TABELA I

	x_{i-2}	x_{i-1}	x_i	x_{i+1}	x_{i+2}
$B_i(x)$	0	1	4	1	0
$B'_i(x)$	0	$\frac{3}{h}$	0	$-\frac{3}{h}$	0
$B''_i(x)$	0	$\frac{6}{h^2}$	$-\frac{12}{h^2}$	$\frac{6}{h^2}$	0

Funções como estas são conhecidas como funções splines e

desempenham um importante papel em muitos problemas de aproximação e interpolação. Em particular, a base que acabamos de definir é conhecida como B-splines cúbicos.

Retornando ao problema inicial, tomemos como base para P os B-splines acima introduzidos. Assim,

$$(2.2.2) \quad \begin{cases} LU_N(x_i) = f(x_i) & , \quad 0 \leq i \leq N \\ U_N(0) = U_N(1) = 0 & , \end{cases}$$

define um sistema de equações cuja matriz é chamada matriz de colocação e denotada por $C_N = (c_{ij})$, $-1 \leq i, j \leq N+1$.

Em C_N a primeira e a última linha surge dos coeficientes das condições de contorno $U_N(0) = 0$ e $U_N(1) = 0$, respectivamente, enquanto as $N+1$ equações interiores vêm de exigirmos que:

$$c_{ij} = LB_j(x_i) = B_j''(x_i) - \sigma(x_i)B_j(x_i), \quad -1 \leq j \leq N+1, \quad 0 \leq i \leq N.$$

Mais especificamente,

$$(2.2.3) \quad \begin{cases} U_N(0) = \sum_{j=-1}^{N+1} a_j B_j(0) = a_{-1} + 4a_0 + a_1 = 0 \\ U_N(1) = \sum_{j=-1}^{N+1} a_j B_j(1) = a_{N-1} + 4a_N + a_{N+1} = 0 \end{cases},$$

onde os valores de $B_j(x_i)$ são dados pela tabela I.

Para encontrar as outras linhas da matriz C_N utilizamos no

vamente a tabela I a fim de avaliar cada c_{ij} nos nós da partição. A seguir, substituímos estes valores nas equações de colocação (2.2.2) e multiplicamos por h^2 para obter as $N+1$ equações lineares

$$(2.2.4) \quad (h^2\sigma_i - 6)a_{i-1} + (4h^2\sigma_i + 12)a_i + (h^2\sigma_i - 6)a_{i+1} = -h^2 f(x_i), \quad 0 \leq i \leq N.$$

Observemos que (2.2.4) é um sistema linear com $N+1$ equações e $N+3$ incógnitas. Eliminando a_{-1} da primeira equação de (2.2.4):

$$(h^2\sigma_0 - 6)a_{-1} + (4h^2\sigma_0 + 12)a_0 + (h^2\sigma_0 - 6)a_1 = -h^2 f(x_0)$$

e da primeira equação de (2.2.3):

$$a_{-1} + 4a_0 + a_1 = 0,$$

encontramos

$$(2.2.5) \quad 36a_0 = -h^2 f(x_0).$$

Analogamente, eliminando a_{N+1} da última equação de (2.2.4) e de (2.2.3), obtemos

$$(2.2.6) \quad 36a_N = -h^2 f(x_N).$$

Juntando (2.2.5) e (2.2.6) com as $(N-1)$ restantes equações de (2.2.4), teremos o sistema de $N+1$ equações lineares

$$(2.2.7) \quad A_N y = b$$

a $N+1$ incógnitas $y = (a_0, a_1, \dots, a_N)^t$ com o termo independente dado por:

$$b = -h^2 (f(x_0), f(x_1), \dots, f(x_N))^t.$$

É imediato que se $\sigma(x) > 0$ a matriz dos coeficientes

$$A_N = \begin{bmatrix} 36 & 0 & 0 & 0 \\ h^2 \sigma_1^{-6} & 4h^2 \sigma_1 + 12 & h^2 \sigma_1^{-6} & 0 \\ 0 & h^2 \sigma_1^{-6} & 4h^2 \sigma_1 + 12 & h^2 \sigma_1^{-6} \\ 0 & 0 & 0 & 36 \end{bmatrix}$$

é estritamente diagonalmente dominante e portanto não singular pelo teorema de Gershgorin [1].

Logo, o método da colocação aplicado a (2.2.1), usando uma base de splines cúbicos tem uma única solução $U_N(x)$ dada por:

$$U_N(x) = a_{-1}B_{-1}(x) + a_0B_0(x) + \dots + a_NB_N(x) + a_{N+1}B_{N+1}(x).$$

É possível mostrar que o erro da aproximação na norma do máximo é da ordem de h^2 [3].

Considerando a mesma norma e a mesma base, o método dos elementos finitos fornece uma ordem de convergência de h^4 . Deste ponto de vista, o método de Rayleigh-Ritz é mais preciso. Embora os resultados teóricos favoreçam o método dos elementos finitos, na prática experiências computacionais indicam igual preci-

são. Alguns fatos justificam estas vantagens do método da colocação. Em primeiro lugar, o cálculo dos elementos da matriz de colocação envolve simples avaliações de funções nos nós, ao passo que, no método dos elementos finitos, este cálculo exige avaliações de integrais que são usualmente aproximadas numericamente. Em segundo lugar, a matriz de colocação é tridiagonal enquanto a matriz que surge pela aplicação do método de Rayleigh-Ritz é de banda sete.

Uma forma de tornar o método da colocação competitivo em termos teóricos com o método dos elementos finitos é escolher como base funções splines de grau superior. Isto nem sempre é uma boa prática se considerarmos a complexidade computacional envolvida. Entretanto, para problemas do tipo que estamos tratando aqui, a utilização de colocação com splines quínticos fornece uma ordem de convergência de h^4 onde a matriz de colocação é de banda cinco e os cálculos envolvidos são ainda muito mais simples do que aqueles originados pelo método dos elementos finitos com spline cúbico.

Uma teoria geral sobre existência e análise do erro para o método da colocação aplicado a equações diferenciais ordinárias de ordem m , tendo como base funções polinomiais por partes, é desenvolvida em Russel e Shampine (1972) e Prenter (1975). O método de prova que adotaram utiliza a noção de função de Green associada com o operador diferencial.

2.3 COLOCAÇÃO-ORTOGONAL: Outra forma de acelerar a razão de convergência do método da colocação a fim de compará-lo favoravelmente ao método dos elementos finitos é colocar nos nós Gaussianos da partição.

Sob este aspecto, o método é conhecido na literatura especializada como método da colocação ortogonal. Do ponto de vista experimental sua utilização principalmente na Engenharia Química é muito ampla. Por exemplo, para problemas relacionados com reações químicas apenas o método da colocação é factível, visto que outros métodos requerem avaliações de complicadas integrais envolvendo funções exponenciais [2].

Existência, unicidade e estimativa do erro para colocação ortogonal em uma dimensão usando splines foi apresentado pela primeira vez por C. de Boor e B. Swartz (1973)[13].

A fim de mostrar uma aplicação simples do método, consideremos novamente o problema de valor de contorno de segunda ordem:

$$(2.3.1) \quad \begin{cases} Lu(x) = u''(x) - \sigma(x)u(x) = f(x) & , \quad 0 \leq x \leq 1 \\ u(0) = u(1) = 0 & , \end{cases}$$

onde $\sigma(x) > 0$ é contínua em $[0, 1]$. Seja

$$\pi: 0 = x_0 < x_1 < \dots < x_N = 1$$

uma partição de $[0, 1]$, com h_1 sendo o tamanho do subintervalo

lo i , $i = 1, 2, \dots, N$.

Consideremos uma das fórmulas de integração numérica bastante útil, a regra de quadratura Gaussiana para dois pontos

$$\frac{h_i}{2} [g(\bar{x}_{i1}) + g(\bar{x}_{i2})] \approx \int_{x_{i-1}}^{x_i} g(x) dx$$

onde

$$(2.3.2) \quad \begin{cases} \bar{x}_{i1} = x_{i-1} + \frac{h_i}{2} \left[1 - \frac{1}{\sqrt{3}} \right] \\ \bar{x}_{i2} = x_{i-1} + \frac{h_i}{2} \left[1 + \frac{1}{\sqrt{3}} \right] \\ h_i = x_i - x_{i-1}, \quad i = 1, 2, \dots, N. \end{cases}$$

Observemos que os elementos $\bar{x}_{ik}$, $i=1, 2, \dots, N$, $k=1, 2$ definem uma nova partição

$$\bar{\pi}: \bar{x}_{11} < \bar{x}_{12} < \bar{x}_{21} < \bar{x}_{22} < \dots < \bar{x}_{N1} < \bar{x}_{N2} = \bar{x}_1 < \bar{x}_2 < \dots < \bar{x}_{2N},$$

cujos nós serão chamados nós Gaussianos de π .

Nosso objetivo é colocar $f(x)$ nestes $2N$ nós Gaussianos.

Isto exige $2N$ funções splines linearmente independentes $\{z_j(x): 1 \leq j \leq 2N\}$ para as quais a matriz $[Lz(\bar{x}_i)] = c_{ij}$,

$1 \leq i, j \leq 2N$ é não singular.

A escolha natural é o conjunto dos polinômios de Hermite cúbico por partes com nós em π satisfazendo às condições de contorno.

Em particular, consideremos a seguinte base:

$$\phi_i(x) = \begin{cases} \frac{-2(x-x_{i-1})^3}{h^3} + \frac{3(x-x_{i-1})^2}{h^2}, & x_{i-1} \leq x \leq x_i \\ 1 + \frac{2(x-x_i)^3}{h^3} - \frac{3(x-x_i)^2}{h^2}, & x_i \leq x \leq x_{i+1} \\ 0, & x \notin [x_{i-1}, x_{i+1}] \end{cases}$$

$$1 \leq i \leq N-1,$$

$$\psi_i(x) = \begin{cases} \frac{(x-x_{i-1})^2(x-x_i)}{h^2}, & x_{i-1} \leq x \leq x_i \\ \frac{(x-x_i)(x_{i+1}-x)^2}{h^2}, & x_i \leq x \leq x_{i+1} \\ 0, & x \notin [x_{i-1}, x_{i+1}] \end{cases},$$

$$0 \leq i \leq N.$$

Notemos que de fato cada uma destas funções satisfaz às condições de contorno

$$\phi_i(0) = \phi_i(1) = \psi_i(0) = \psi_i(1) = 0$$

Vamos reenumerá-las. Seja

$$z_i(x) = \begin{cases} \frac{\psi_{\frac{i-1}{2}}}{2} & , \quad i = 1, 3, 5, \dots, 2N+1 \\ \frac{\phi_{\frac{i}{2}}}{2} & , \quad i = 2, 4, 6, \dots, 2N-2. \end{cases}$$

Então,

$$\{\psi_0, \phi_1, \psi_1, \phi_2, \psi_2, \dots, \phi_{N-1}, \psi_{N-1}, \psi_N\} = \{z_1, z_2, \dots, z_{2N}\}.$$

Consideremos P o espaço gerado pelas $\{z_i\}_{i=1}^{2N}$.

Buscamos $U_N(x) = a_1 z_1(x) + a_2 z_2(x) + \dots + a_{2N} z_{2N}(x)$ resolvendo as equações de colocação

$$LU_N(\bar{x}_i) = \sum_{j=1}^{2N} a_j Lz_j(\bar{x}_i) = f(\bar{x}_i), \quad 1 \leq i \leq 2N, \quad ,$$

onde $\bar{x}_1 < \bar{x}_2 < \dots < \bar{x}_{2N}$ são os nós Gaussianos da partição.

É possível demonstrar que a ordem de convergência da solução aproximada U_N é de h^4 [3].

O método pode ser generalizado para equações diferenciais ordinárias de ordem m . Neste caso, devemos introduzir m nós Gaussianos associados às m raízes do polinômio ortogonal de grau m em cada subintervalo $[x_{i-1}, x_i]$ da partição inicial de $[0, 1]$. Usamos então, uma base de polinômios Hermite por partes pertencentes a $C^{m-1}[0, 1]$ que são de grau $2m-1$ em cada subintervalo e satisfaçam às condições de contorno associadas com o operador diferencial.

Se $f^{(m)}$ é contínua por partes em $[0, 1]$ pode ser mostrado que a ordem de convergência é h^{2m} .

O método da colocação ortogonal também pode ser aplicado com sucesso a equações diferenciais parciais, produzindo uma ordem de convergência superior e competindo, desta forma, favoravelmente, com elementos finitos.

Prenter and Russell [4] demonstraram, sem utilizar a noção de funções de Green ou mesmo análise da matriz, que o método da colocação ortogonal com Hermite bicúbico numa grade retangular fornece uma esquema de quarta ordem para aproximar a solução $u(x,y)$ da equação diferencial parcial elítica

$$Lu = -D_x [P(x,y) D_x u(x,y)] - D_y [g(x,y) D_y u(x,y)] + C(x,y) u(x,y) = f(x,y)$$

com $u(x,y) = 0$ em ∂R , R um retângulo.

Neste caso, os nós Gaussianos em cada subretângulo $[x_{i-1}, x_i] \times [y_{j-1}, y_j]$ são os pontos

$$(\bar{x}_{i1}, \bar{y}_{j1}) , (\bar{x}_{i1}, \bar{y}_{j2}) , (\bar{x}_{i2}, \bar{y}_{j1}) , (\bar{x}_{i2}, \bar{y}_{j2}) ,$$

onde $\bar{x}_{ik}$ e $\bar{y}_{jk}$ são definidos como em (2.3.2) .

Mais recentemente (1984) Dyksen, Houstis, Lynch and Rice [5] estudaram a performance dos métodos de colocação e Galerkin tomando Hermite bicúbico como base, considerando o mesmo problema acima. Os sistemas lineares que surgem pela aplicação destes métodos foram resolvidos por métodos diretos. As medidas de comparação foram o tempo de computação para atingir uma precisão de três dígitos significativos e a memória utilizada. Os resultados a que chegaram, após uma análise de dezoito problemas foram:

1. A quantidade de memória requerida é a mesma;
2. Colocação requer menos tempo de computação para atingir três dígitos de precisão nos pontos da grade. A conclusão final é que no contexto considerado e com as implementações que fizeram dos métodos, colocação é sensivelmente melhor que Galerkin para os aspectos 1. e 2. citados acima.

Em problemas que dependem do tempo o método da colocação é aplicado no domínio das variáveis espaciais e dá origem a um problema de valor inicial para um sistema de equações diferenciais ordinárias. A implementação de tal problema é feita por um método de discretização da variável temporal, conhecido como colocação no tempo discreto e a aproximação resultante é chama-

da de aproximação por colocação no tempo discreto. No próximo capítulo analisaremos um problema deste tipo.

2.4 UMA RELAÇÃO ENTRE COLOCAÇÃO E GALERKIN: Uma forma de relacionar Galerkin e Colocação é através de esquemas de quadratura numérica.

Em particular, consideremos um espaço P_N de dimensão finita e uma base $\{z_1, z_2, \dots, z_N\}$ de P_N .

Vamos supor que estamos procurando uma solução para um problema que envolve a seguinte equação:

$$Lu = f.$$

O método de Galerkin consiste em encontrar uma aproximação para u da forma

$$U_N(x) = \sum_{j=1}^N a_j z_j(x)$$

cujos coeficientes $a_j, j=1, \dots, N$ são determinados resolvendo o sistema linear

$$(2.4.1) \quad \sum_{j=1}^N (Lz_j, z_i) a_j = (f, z_i), \quad 1 \leq i \leq N,$$

onde

$$(f, g) = \int_E f(x) g(x) dx$$

é o produto interno usual em $L^2(E)$, E um domínio qualquer em R^1 ou R^2 .

Se L é um operador diferencial usualmente o que se faz é integrar o lado esquerdo de (2.4.1) por partes substituindo a integral pela forma bilinear $a(z_j, z_i)$.

Vamos proceder, entretanto, de uma forma diferente. Ao invés de integrar por partes, usamos a quadratura numérica

$$\int_E g(y) dy \approx \sum_{i=1}^N g(y_i) w_i, \quad ,$$

onde $y_1, y_2, \dots, y_N$ são N pontos distintos de E e os w_i 's são constantes determinadas pelo esquema de aproximação que dá origem à regra.

Assim, temos:

$$\int_E Lz_j(y) z_i(y) dy \approx \sum_{k=1}^N Lz_j(y_k) z_i(y_k) w_k$$

e

$$\int_E f(y) z_i(y) dy \approx \sum_{k=1}^N z_i(y_k) w_k f(y_k).$$

Obtemos então, um sistema de equações

$$(2.4.2) \quad B_N D_N C_N^t \hat{a} = B_N D_N \hat{f} \quad ,$$

onde

$$B_N = (Z_i(y_k)) \quad ,$$

$$D_N = \text{diag}(w_k) \quad ,$$

$$C_N = (LZ_j(y_k)) \quad ,$$

$$\hat{f} = (f_1, \dots, f_N)^t, \quad f_i = f(y_k),$$

$$\hat{a} = (a_1, \dots, a_N)^t.$$

Deste modo, quando B_N é não singular, o sistema de Galerkin completamente discretizado é equivalente a

$$C_N^t \hat{a} = \hat{f}.$$

Seja $\hat{U}_N(x) = \sum_{j=1}^N \hat{a}_j Z_j(x)$ a solução para (2.4.2)

Se os erros da quadratura forem suficientemente pequenos, $B_N D_N C_N^t$ é não singular e o erro para esta solução $\hat{U}_N(x)$ de Galerkin discreto é praticamente o mesmo que para a solução por Galerkin $U_N(x)$.

Além disso, desde que as matrizes são quadradas, a não singularidade de $B_N D_N C_N^t$ implica a não singularidade de C_N^t , portanto de C_N .

Assim, $C_N^t \hat{a} = \hat{f}$ é equivalente a

$$L\hat{U}_N(y_k) = \hat{f}(y_k) ,$$

e portanto, $\hat{U}_N(y)$ é a solução por colocação.

Logo, pelos argumentos acima temos a equivalência entre os métodos da colocação e Galerkin discreto.

3. APLICAÇÃO A UMA EQUAÇÃO DIFERENCIAL DO TIPO PARABÓLICO

3.1 INTRODUÇÃO: Nos últimos anos equações de difusão não lineares têm chamado a atenção da comunidade matemática, fato este que reflete o crescimento na amplitude e importância de suas aplicações. Tais equações surgem em modelos matemáticos de um grande número de fenômenos físicos, biológicos e sociais que ocorrem em áreas como dinâmica de população, teoria de reatores químicos, etc.

Como ilustração, consideremos um modelo de pesca no qual a distribuição espacial e a movimentação dos peixes são levados em consideração [10]. Imaginemos que a margem seja retilínea, chamemos x a distância do ponto em análise à margem e $u(x, t)$ a densidade da população de peixe em x no instante t . A equação diferencial parcial que representa um modelo de crescimento e difusão da população é:

$$(3.1.1) \quad \frac{\partial u}{\partial t} = \sigma^2 \frac{\partial^2 u}{\partial x^2} + F(x, u) \quad (0 \leq x \leq S) \quad ,$$

onde $F(x, u)$ é a razão de crescimento natural da população na distância x da margem, S denota o contorno do habitat da população e σ^2 é o coeficiente de difusão.

A equação (3.1.1) é uma equação diferencial não linear do tipo parabólico; equações como esta é que passaremos a analisar.

Consideremos a equação diferencial parabólica

$$(3.1.2) \quad c(x, t, u) \frac{\partial u}{\partial t} = a(x, t, u) \frac{\partial^2 u}{\partial x^2} + b(x, t, u, \frac{\partial u}{\partial x}), \quad 0 < x < 1, 0 < t \leq T,$$

sujeita à condição inicial,

$$(3.1.3) \quad u(x, 0) = f(x) \quad , \quad 0 < x < 1 \quad ,$$

e às condições de contorno,

$$(3.1.4) \quad u(0, t) = g_0(t) \quad , \quad u(1, t) = g_1(t) \quad , \quad 0 < t \leq T.$$

Vamos assumir a condição de estabilidade,

$$(3.1.5) \quad 0 < m \leq c(x, t, u) \leq M \quad , \quad m \leq a(x, t, u) \leq M < \infty \quad , \quad \text{para}$$

$$0 \leq x \leq 1, \quad 0 \leq t \leq T \quad \text{e} \quad -\infty < u < \infty .$$

Procuramos a solução numérica de (3.1.2)-(3.1.4) pelo método da colocação para o caso particular em que a solução aproximada é um polinômio de Hermite cúbico por partes na variável espacial x para cada tempo t . Mais precisamente, consideremos a partição do intervalo $I=[0, 1]: 0 = x_0 < x_1 < \dots < x_n = 1$, com $h_j = x_j - x_{j-1}$ e $I_j = [x_{j-1}, x_j]$.

Seja,

$H_3 = \{v=v(x) \in C^1(I)/v \text{ é um polinômio cúbico sobre cada } I_j, j = 1, \dots, n\}.$

Buscamos uma aplicação $U: [0, T] \rightarrow H_3$ tal que U seja uma aproximação para a solução u de (3.1.2)-(3.1.4), $0 \leq t \leq T$, no sentido que mais tarde definiremos.

Uma base para H_3 pode ser construída a partir das funções,

$$V(x) = \begin{cases} 1 - 3x^2 + 2x^3 & , \quad 0 \leq x \leq 1, \\ 1 - 3x^2 - 2x^3 & , \quad -1 \leq x \leq 0, \\ 0 & , \quad |x| > 1, \end{cases}$$

$$S(x) = \begin{cases} x(1-x)^2 & , \quad 0 \leq x \leq 1 \\ x(1+x)^2 & , \quad -1 \leq x \leq 0 \\ 0 & , \quad |x| > 1. \end{cases}$$

De fato, tomemos

$$v_j(x) = \begin{cases} V[(x-x_j)/h_{j+1}] & , \quad x \geq x_j \\ V[(x-x_j)/h_j] & , \quad x \leq x_j \end{cases}$$

$$s_j(x) = \begin{cases} h_{j+1} S[(x-x_j)/h_{j+1}] & , \quad x \geq x_j \\ h_j S[(x-x_j)/h_j] & , \quad x \leq x_j, \quad j=0,1,\dots,n \end{cases}$$

Então,

$$H_3 = \text{Span}\{V_0, S_0, \dots, V_n, S_n\} \text{ e a } \dim(H_3) = 2n+2.$$

Portanto, necessitamos de $2n+2$ relações para cada tempo t para especificar a solução aproximada $U(t)$. Duas destas condições podem ser obtidas das condições de contorno, isto é, os coeficientes de V_0 e V_n são dados por $g_0(t)$ e $g_1(t)$, respectivamente. O método da colocação exige que as relações restantes sejam obtidas fazendo com que a equação diferencial seja satisfeita para $2n$ pontos. Uma vez que existem n intervalos I_j , devemos selecionar dois pontos em cada um deles.

Como a escolha dos pontos de colocação é crítica na obtenção da ordem de convergência vamos escolhê-los como sendo as imagens afim das raízes do polinômio de Legendre de grau 2.

Sejam $-1/\sqrt{3}$ e $1/\sqrt{3}$ os zeros do polinômio de Legendre $P_2(x) = (3x^2 - 1)/2$ e consideremos a transformação afim $L(\xi) = A\xi + B$ definida em $[-1, 1]$ e com valores em $[x_{j-1}, x_j]$, satisfazendo, $L(-1) = x_{j-1}$ e $L(1) = x_j$. Então, temos:

$$L(-1) = -A + B = x_{j-1}$$

$$L(1) = A + B = x_j$$

$$\text{Portanto, } 2B = x_{j-1} + x_j, \quad B = (x_{j-1} + x_j)/2$$

$$A = x_j - (x_{j-1} + x_j)/2, \quad A = (x_j - x_{j-1})/2$$

Assim,
$$L(\xi) = \frac{x_{j-1} - x_{j-1}}{2} \cdot \xi + \frac{x_{j-1} + x_j}{2}, \text{ ou } L(\xi) = \frac{x_{j-1} + x_j}{2} + \frac{h_j}{2} \cdot \xi.$$

Logo, os pontos de colocação são dados por:

$$(3.1.7) \quad \xi_{jk} = \frac{x_{j-1} + x_j}{2} + (-1)^k \frac{h_j}{2\sqrt{3}}, \quad j=1, 2, \dots, n, k=1, 2.$$

Deste modo, o método da colocação fica especificado pelas equações (estamos suprimindo parcialmente a escrita das variáveis independentes x e t):

$$(3.1.8) \quad [c(U) \frac{\partial U}{\partial t} - a(U) \frac{\partial^2 U}{\partial x^2} - b(U, \frac{\partial U}{\partial x})](\xi_{jk}, t) = 0, \quad ,$$

$$j=1, 2, \dots, n, \quad k=1, 2, \quad U(0, t)=g_0(t), \quad U(1, t)=g_1(t), \quad \text{para}$$

$$0 < t \leq T.$$

É necessário, ainda, especificar condição inicial para U ; a maneira mais fácil de fazer isto se $f \in C^1(I)$ é considerar $U(x, 0)$ como o H_3 -interpolante de f , isto é $U(x, 0)$ e $\frac{\partial U}{\partial x}(x, 0)$ interpolarem f e f' nos nós naturais x_j , respectivamente.

Neste capítulo analisaremos a convergência da solução U de (3.1.8) para u , isto é, demonstraremos que existe uma constante C dependendo de u e de suas derivadas tal que

$$\|u - U\|_{L^\infty(0, T; L^\infty(I))} \leq C h^4, \quad h = \max_j h_j, \quad ,$$

se $U(x, 0)$ for escolhida como indicamos acima.

Observemos que em vista da hipótese (3.1.5) podemos dividir (3.1.8) pelo coeficiente $a(x, t, u)$ sem perda de generalidade, uma vez que a aritmética permanece inalterada. Portanto, doravante assumiremos que

$$a(x, t, u) \equiv 1$$

Também assumiremos que $g_0(t) = g_1(t) = 0$ pela modificação de b e c , o que é irrelevante computacionalmente, mas útil na análise que há de se seguir.

O espaço dos polinômios de Hermite cúbico por partes H_3 também pode ser utilizado no método de Galerkin da mesma forma que no método da colocação, produzindo uma precisão da mesma ordem, isto é, $O(h^4)$. Assim, alguma comparação entre os dois métodos poderia ser realizada. Praticamente, o método da colocação é sensivelmente mais rápido no computador do que Galerkin, considerando a mesma partição, pois, não existem quadraturas para serem calculadas no método da colocação e além disso, existem apenas quatro coeficientes não nulos em quaisquer das $2N$ equações geradas por (3.1.8), enquanto existem seis no caso de Galerkin. Portanto, resolver o sistema algébrico que resulta da discretização no tempo é mais simples por colocação. Entretanto, os argumentos que serão apresentados aqui exigem mais suavidade para u no método da colocação do que é exigido para o método de Galer-

kin a fim de obter a razão de convergência pretendida.

3.2 NOTAÇÃO E ALGUMAS DESIGUALDADES: Todas as funções a serem consideradas para o que segue são de valores reais. Seja E um aberto qualquer do R^n , o espaço Euclidiano n -dimensional ; $x = \{x_1, x_2, \dots, x_n\} \in E$. Designemos por $L^2(E)$ o espaço das funções u de quadrados integráveis sobre E , isto é,

$$(3.2.1) \quad \|u\|_{L^2(E)}^2 = \int_E |u|^2 dx < \infty$$

Associado à norma (3.2.1) temos o produto escalar em $L^2(E)$:

$$(u, v)_{L^2(E)} = \int_E u(x) v(x) dx.$$

Definimos o espaço de Sobolev $H^m(E)$ de ordem $m(m \geq 0)$ sobre E como o conjunto de funções $u \in L^2(E)$ tais que todas as derivadas parciais $D^\alpha u$ (no sentido de distribuição), com $|\alpha| \leq m$, pertencem ao espaço $L^2(E)$, isto é,

$$(3.2.2) \quad H^m(E) = \{u \in L^2(E) / D^\alpha u \in L^2(E), \forall \alpha, |\alpha| \leq m\}, \quad \text{onde}$$

$$D^\alpha = \frac{\partial^{\alpha_1 + \alpha_2 + \dots + \alpha_n}}{\partial x_1^{\alpha_1} \dots \partial x_n^{\alpha_n}}, \quad \alpha = \{\alpha_1, \alpha_2, \dots, \alpha_n\},$$

$$|\alpha| = \alpha_1 + \alpha_2 + \dots + \alpha_n$$

Nestes espaços podemos considerar

$$(3.2.3) \quad \|u\|_{H^m(E)} = \left(\sum_{|\alpha| \leq m} \|D^\alpha u\|_{L^2(E)}^2 \right)^{\frac{1}{2}}; \text{ munido}$$

do desta norma $H^m(E)$ é um espaço de Hilbert [7].

Em nosso caso, $E=(0, 1)$. Assim, indicaremos o produto escalar de L^2 sobre I e I_j como segue:

$$(3.2.4) \quad (u, v) = \int_0^1 uv dx = \sum_{j=1}^n (u, v)_j = \sum_{j=1}^n \int_{x_{j-1}}^{x_j} uv dx$$

É conveniente, ainda, definir o produto interno discreto:

$$(3.2.5) \quad \langle u, v \rangle_j = \frac{1}{2} [u(\xi_{j1})v(\xi_{j1}) + u(\xi_{j2})v(\xi_{j2})] h_j, \quad |u|_j^2 = \langle u, u \rangle_j$$

e

$$(3.2.6) \quad \langle u, v \rangle = \sum_{j=1}^n \langle u, v \rangle_j, \quad |u|^2 = \langle u, u \rangle.$$

Representamos por $L^2(0, T; H^m(I))$ o espaço vetorial das funções $u(x, t)$ tais que para todo $t \in (0, T)$ a função $u(x, t) \in H^m(I)$,

cuja norma em $H^m(I)$ é uma função pertencente a $L^2(0,T)$.

De modo análogo definimos $L^\infty(0,T;H^m(I))$ como espaço vetorial das funções $u(x,t)$ tais que para todo $t \in (0,T)$, $u(x,t) \in H^m(I)$, cuja norma em $H^m(I)$ é uma função pertencente a $L^\infty(0,T)$.

Então, podemos definir as seguintes normas sobre estes espaços:

$$(3.2.7) \quad \|u\|_{L^\infty(0,T;H^6(I))} = \sup_{0 < t < T} \|u(t)\|_{H^6(I)} < \infty$$

$$(3.2.8) \quad \|u_t\|_{L^2(0,T;H^6(I))}^2 = \int_0^T \|u_t(t)\|_{H^6(I)}^2 dt < \infty$$

No decorrer desta dissertação K denotará uma constante genérica com valores diferentes possíveis em diferentes contextos.

Os lemas que apresentaremos a seguir fornecem desigualdades que serão úteis para a análise do erro em parágrafo seguinte.

LEMA 3.2.1(desigualdade de Gronwall). Se as funções $f(t)$ e $g(t)$ são não negativas definidas em $[a,b]$, e se $K \geq 0$, então a desigualdade

$$(i) \quad g(t) \leq K + \int_a^t f(s)g(s)ds,$$

implica em

$$(ii) \quad g(t) \leq K e^{\int_a^t f(s) ds}$$

Demonstração

De (i) temos

$\frac{g(t)f(t)}{K + \int_a^t f(s)g(s)ds} \leq f(t)$. Integrando de a até t , obtemos

$$K + \int_a^t f(s)g(s)ds$$

$$\ln\left[K + \int_a^t f(s)g(s)ds\right] - \ln(K) \leq \int_a^t f(s)ds. \text{ Isto acarreta}$$

$$(iii) \quad K + \int_a^t f(s)g(s)ds \leq K e^{\int_a^t f(s)ds}$$

Portanto, de (i) e (iii) temos o resultado. ■

As demonstrações dos lemas seguintes são consequências da aplicação das definições do produto interno discreto (3.2.5),

(3.2.6) e da desigualdade de Schwartz.

LEMA 3.2.2. Para todo f e g em H_3 ,

$$-(f'', g) = (f', g') - f'g \Big|_0^1 + \frac{1}{1080} \sum_{j=1}^N f_j''' g_j''' h_j^5$$

onde f_j é o valor(constante) da terceira derivada de f em I_j .

Demonstração. É suficiente considerar apenas um intervalo de comprimento h_1 , uma vez que os termos da fronteira se anulam na soma sobre os intervalos.

Sejam

$$f(x) = \sum_{i=0}^3 a_i x^i \quad \text{e} \quad g(x) = \sum_{i=0}^3 b_i x^i. \quad \text{Então,}$$

$$-(f'', g)_1 = - \int_0^{h_1} f'' g \, dx = -(6a_3 x, b_3 x^3)_1 + \int_0^{h_1} 6a_3 b_3 x^4 \, dx.$$

Avaliando os dois últimos termos do lado direito da expressão acima obtemos:

$$-(6a_3 x, b_3 x^3)_1 + \int_0^{h_1} 6a_3 b_3 x^4 \, dx = \frac{h_1^5}{1080} f''' g''' ; \quad \text{assim,}$$

$$-\langle f'', g \rangle_1 = - \int_0^{h_1} f'' g \, dx + \frac{h_1^5}{1080} f''' g''' .$$

Uma integração por partes nos leva a

$$-\langle f'', g \rangle_1 = (f', g')_1 - f' g \Big|_0^{h_1} + \frac{h_1^5}{1080} f''' g''' .$$

Por homogeneidade temos o resultado. ■

LEMA 3.2.3. Para $f \in H_3$,

$$\max \left\langle \frac{(f''')^2}{\int_0^h (f')^2 dx} \right\rangle = \frac{720}{h^5} .$$

O lema acima nos permite fazer a seguinte estimativa:

$$\begin{aligned} (i) \quad \frac{1}{1080} \sum_{j=1}^N f_j''' g_j''' h_j^5 &\leq \frac{2}{3} \sum_{j=1}^N \|f_j'\|_{L^2(I_j)} \|g_j'\|_{L^2(I_j)} \\ &\leq \frac{2}{3} \left[\int_0^1 (f')^2 dx \right]^{\frac{1}{2}} \left[\int_0^1 (g')^2 dx \right]^{\frac{1}{2}} = \frac{2}{3} \|f\|_{H_0^1} \|g\|_{H_0^1} \end{aligned}$$

$f, g \in H_3$, onde

$$\|f\|_{H_0^1}^2 = \int_0^1 (f')^2 dx \quad \text{e} \quad H_0^1(I) = \{f \in H^1(I) \mid f(0) = f(1) = 0\}.$$

Seja H_3^0 o subespaço de H_3 consistindo das funções que se anulam em $x=0$ e $x=1$.

LEMA 3.2.4. Se $f \in H_3^0$, então

$$\|f\|_{H_0^1}^2 \leq - \langle f'', f \rangle \leq \frac{5}{3} \|f\|_{H_0^1}^2 .$$

Demonstração. É uma consequência imediata da aplicação do lema (3.2.2) e da estimativa (i). ■

LEMA 3.2.5. Para $f \in H_3$,

$$\langle f', g' \rangle = (f', g') - \frac{1}{720} \sum_{j=1}^N f_j''' g_j''' h_j^5$$

e

$$|f'|^2 \leq \|f\|_{H_0^1}^2 .$$

Demonstração. A primeira parte segue um argumento análogo àquele do lema 3.2.2.

A outra parte do lema é consequência de

$$\langle f', f' \rangle = |f'|^2 = \|f\|_{H_0^1}^2 - \frac{1}{720} \sum_{j=1}^N (f_j''')^2 h_j^5$$

e, portanto,

$$(ii) \quad |f'|^2 \leq \|f\|_{H_0^1}^2, \quad f \in H_3 \quad \blacksquare$$

LEMA 3.2.6. Para $f \in H^1$,

$$\|f\|_{H_0^1}^2 + |f|^2 \geq \frac{1}{4} \|f\|_{H^1}^2.$$

Demonstração. A relação

$$f(x) = f(\xi_{j1}) + \int_{\xi_{j1}}^x f'(\tau) d\tau \quad \text{implica em}$$

$$\|f\|_{L^2(I_j)}^2 \leq 4|f|_j^2 + 2h_j^2 \|f\|_{H_0^1(I_j)}^2. \quad \text{Por outro lado}$$

$$\|f\|_{H^1}^2 = \|f\|_{L^2}^2 + \|f\|_{H_0^1}^2 \leq 4|f|_j^2 + 2\|f\|_{H_0^1}^2 + \|f\|_{H_0^1}^2 \leq 4|f|_j^2 + 4\|f\|_{H_0^1}^2.$$

Logo,

$$\|f\|_{H_0^1}^2 + |f|^2 \geq \frac{1}{4} \|f\|_{H^1}^2 . \quad \blacksquare$$

3.3 ALGUNS RESULTADOS DA TEORIA DA APROXIMAÇÃO: A fim de estabelecer um limite do erro do método da colocação é necessário primeiro que limitemos a diferença entre a solução aproximada U e o interpolante cúbico de Hermite da solução exata u . Portanto, necessitaremos de uma avaliação do erro da interpolação.

Na obtenção de estimativas em problemas envolvendo uma variável, o teorema do núcleo de Peano é de importância fundamental. Nós o enunciaremos aqui considerando funcionais lineares da forma

$$L(u) = \sum_{i=0}^{r-1} \left\{ \int_a^b a_i(x) D^i u(x) dx + \sum_{j=1}^{N_i} b_{ij} D^i u(x_{ij}) \right\} ,$$

definidos para todo $u \in H^r(J)$ onde $J=(a,b)$. Estamos assumindo que as funções $a_i(x)$ são contínuas por partes sobre $[a,b]$ e os pontos $x_{ij} \in [a,b]$.

TEOREMA 3.3.1. Seja $L(p)=0$ para todo polinômio p de grau menor do que r . Então, para todo $u \in H^r(J)$, onde $J=(a,b)$,
 $-\infty < a < b < \infty$,

$$L(u) = \int_a^b D^r u(t) K(t) dt \quad , \quad \text{onde}$$

$$K(t) = \frac{1}{(r-1)!} L_x [(x-t)_+^{r-1}] \quad , \quad \text{e}$$

$$(x-t)_+^{r-1} = \begin{cases} (x-t)^{r-1} & , \quad x \geq t \\ 0 & , \quad x < t \end{cases}$$

A notação $L_x [(x-t)_+^{r-1}]$ significa que o funcional L é aplicado a $(x-t)_+^{r-1}$ como uma função de x .

A demonstração deste teorema é uma consequência da fórmula de Taylor com resto [8].

Para funções suficientemente suaves $u \in [0,1]$, defina o seguinte interpolante de u

$$T_3(u)(x) = u(0)V(x) + u'(0)\dot{S}(x) + u(1)V(x-1) + u'(1)\dot{S}(x-1) \quad ,$$

onde $\dot{S}(x)$ e $V(x)$ são definidas em (3.1.6).

Como consequência do teorema 3.3.1 o erro da interpolação tem a seguinte representação para funções $u \in H^{(4)}(I)$:

$$(3.3.1) \quad (u - T_3 u)(x) = \int_0^1 K_3(x,t) u^{(4)}(t) dt \quad , \quad \text{onde para ca}$$

da t fixado, o núcleo é dado por

$$K_3(x,t) = g_t(x) - T_3(g_t)(x) ,$$

$$g_t(x) = \begin{cases} \frac{1}{3!} (x-t)^3 & , \quad 0 \leq t \leq x \\ 0 & , \quad x \leq t \leq 1 . \end{cases}$$

Vamos definir as seguintes relações para funções $u \in H^6(0,h)$:

a. $T_{3,h}(u)(x) = T_3(w)(x/h)$, $0 \leq x \leq h$, onde $w(x) = u(hx)$

b. ${}_h\langle f, g \rangle = \frac{h}{2} \sum_{i=1}^2 fg(\frac{h}{2}(1+(-1)^i/\sqrt{3}))$, ${}_h|f|^2 = {}_h\langle f, f \rangle$.

Os dois lemas seguintes serão de importância para a análise do erro global e suas demonstrações dependem dos resultados anteriores.

LEMA 3.3.1. Sejam $u \in H^6(I)$ e $e = u - T_3 u$. Então existe uma constante K tal que

(i) ${}_1|e^{(r)}| \leq K \|u^{(4)}\|_{L^2(I)} , \quad r = 0, 1 ,$

(ii) ${}_1|e''| \leq K \|u^{(5)}\|_{L^2(I)} ,$

$$(iii) \quad |{}_1\langle e', 1 \rangle| \leq K \|u^{(5)}\|_{L^2(I)},$$

$$(iv) \quad |{}_1\langle e'', 1 \rangle| \leq K \|u^{(6)}\|_{L^2(I)}.$$

LEMA 3.3.2. Sejam $u \in H^6(0, h)$ e $e = u - T_{3,h}u$. Então, existe uma constante K independente de h tal que:

$$(i) \quad h |e^{(r)}| \leq K h^{4-r} \|u\|_{H^4(0, h)}, \quad r = 0, 1,$$

$$(ii) \quad h |e''| \leq K h^3 \|u\|_{H^5(0, h)},$$

$$(iii) \quad |{}_h\langle e', 1 \rangle| \leq K h^{\frac{9}{2}} \|u\|_{H^5(0, h)},$$

$$(iv) \quad |{}_h\langle e'', 1 \rangle| \leq K h^{\frac{9}{2}} \|u\|_{H^6(0, h)}.$$

3.4 UMA FORMULAÇÃO DE GALERKIN: Para demonstrar a convergência da solução aproximada U de (3.1.8) para a solução exata u pelo método da colocação precisamos estabelecer uma relação entre este e o método de Galerkin cuja convergência já tem sido estudada.

Consideremos o método da colocação especificado pelas equa

ções:

$$(3.4.1) \quad \{c(U)U_t - U_{xx} - b(U, U_x)\}(\xi_{jk}, t) = 0, \quad j=1, 2, \dots, N, \quad k=1, 2$$

$$U(0, t) = g_0(t), \quad U(1, t) = g_1(t), \quad 0 \leq t \leq T.$$

Nesta seção mostraremos que a solução (3.4.1) pode ser vista como a solução do seguinte esquema tipo Galerkin:

$$(3.4.2) \quad \langle c(U)U_t - U_{xx} - b(U, U_x), Z \rangle = 0, \quad Z \in H_3^0,$$

sendo a ortogonalidade relativa ao produto interno discreto (3.2.5)-(3.2.6).

Sejam $\{z_i\}_{i=1}^{2N}$ uma base para H_3^0 e

$$U(x, t) = \sum_{i=1}^{2N} b_i(t) z_i(x).$$

Substituindo a expressão para $U(x, t)$ no sistema (3.4.1) podemos reescrevê-lo como

$$(3.4.3) \quad Fb' + Gb = R,$$

onde

$$b = (b_1, b_2, \dots, b_{2N})^t, \quad F = F(b) = (f_{ij}), \quad G = (g_{ij}), \quad R = (r_i).$$

Se $\{\xi_i | i=1, 2, \dots, 2N\} = \{\xi_{ij} | i=1, \dots, N; j=1, 2\}$, então podemos expressar f_{ij} , g_{ij} e r_i como segue:

$$f_{ij} = c \left(\sum_{i=1}^{2N} b_i Z_i(\xi_i) \right) Z_j(\xi_i),$$

$$g_{ij} = -Z_j^i(\xi_i),$$

$$r_i = b \left(\sum_{\ell=1}^{2N} b_\ell Z_\ell(\xi_i), \sum_{\ell=1}^{2N} b_\ell Z_\ell'(\xi_i) \right).$$

De maneira semelhante, (3.4.2) pode ser expresso como

$$(3.4.4) \quad Cb' + Ab = S,$$

onde

$$C = C(b) = (c_{ij}), \quad A = (a_{ij}), \quad S = (s_i)$$

$$c_{ij} = \langle c \left(\sum_{\ell=1}^{2N} b_\ell Z_\ell \right) Z_j, Z_i \rangle,$$

$$a_{ij} = - \langle Z_j^i, Z_i \rangle,$$

$$s_i = \langle b \left(\sum_{\ell=1}^{2N} b_{\ell} z_{\ell}, \sum_{\ell=1}^{2N} b_{\ell} z'_{\ell} \right), z_i \rangle .$$

Se mostrarmos que a matriz $C(b)$ é não singular isto implicará que (3.4.4) tem no máximo uma solução e, portanto, pode ser resolvido localmente no tempo. Pela equivalência entre (3.4.1) e (3.4.2) segue que $F(b)$ também é não singular e deste modo (3.4.3) pode também ser resolvido localmente no tempo. De fato, temos:

LEMA 3.4.1. Para todo b , $C(b)$ é não singular.

Demonstração. Suponhamos que $a, b \in \mathbb{R}^{2N}$, $a \neq 0$, são tais que $C(b)a=0$. Se $U(x) = \sum_{i=1}^{2N} b_i z_i(x)$ e $W(x) = \sum_{i=1}^{2N} a_i z_i(x)$, então, por hipótese $\langle C(U)W, W \rangle = 0$. Portanto, isto implica que $W(\xi_{jk}) = 0$, $j = 1, \dots, N$, $k = 1, 2$, visto que $C > 0$ para qualquer argumento. Como W é um polinômio cúbico em $[x_0, x_1]$ e se anula para x_0, ξ_{11}, ξ_{12} , então, ou $W \equiv 0$ em $[x_0, x_1]$, ou $W(x_1)W'(x_1) > 0$. Se $W \equiv 0$ em qualquer $[x_i, x_{i+1}]$, então $W \equiv 0$ em I , pois em cada intervalo adjacente W teria uma raiz dupla no ponto final e duas raízes nos pontos de colocação. Assim, como $a \neq 0$, vemos que $W(x_1)W'(x_1) > 0$. Em $[x_1, x_2]$, W' tem raízes em (x_1, ξ_{21}) e (ξ_{21}, ξ_{22}) . Portanto, $W(x_2)W'(x_2) > 0$.

Procedendo desta forma temos que $W(x_N)W'(x_N) > 0$, o que é uma contradição, uma vez que $W(x_N) = 0$.

Logo, $C(b)$ é não singular. ■

3.5 ANÁLISE DA CONVERGÊNCIA: Nesta seção obteremos uma estimativa para o erro da solução aproximada U obtida em (3.4.1); vamos assumir sem perda de generalidade que $g_0(t)=g_1(t)=0$.

A limitação do erro que tentaremos obter será dada em termos da norma de $W-U$, onde W é uma aproximação para u convenientemente escolhida em H_3 [11].

Deste modo, reduzimos a questão de limitar o erro na solução aproximada para uma questão de teoria da aproximação.

Seja

$$\langle c(U) \frac{\partial U}{\partial t} - \frac{\partial^2 U}{\partial x^2} - b(U, \frac{\partial U}{\partial x}), z \rangle = 0, \quad z \in H_3^0.$$

TEOREMA 3.5.1. Seja u a solução de (3.1.2)-(3.1.4). Vamos assumir que

- (i) os coeficientes a, b e c na equação diferencial (3.1.2) tenham terceiras derivadas limitadas,
- (ii) $0 < m \leq c(x, t, u) \leq M, m \leq a(x, t, u) \leq M < \infty$, para

$$0 \leq x \leq 1, 0 \leq t \leq T, -\infty < u < \infty$$

- (iii) $u \in L^\infty(0, T; H^6)$ e $u_t \in L^2(0, T; H^6)$.

Seja, ainda, $U(x, 0)$ o H_3 -interpolante de $f(x)=u(x, 0)$.

Então, existe uma única solução U para o sistema (3.1.8) e U converge para u com um erro dado por

$$\|u-U\|_{L^\infty(0,T;L^\infty)} \leq K[\|u\|_{L^\infty(0,T;H^6)} + \|u_t\|_{L^2(0,T;H^6)}] h^4,$$

onde

$$h = \max_j h_j, \quad j=1, \dots, N$$

Demonstração. Suponhamos que $u \in C^3(I \times [0, T])$ e seja $W: [0, T] \rightarrow H_3$ o H_3 -interpolante de u para cada tempo t .

Inicialmente obteremos uma estimativa para $v = W - U$.

Sejam $\eta = u - W$ e

$$R = \frac{\partial^2 \eta}{\partial x^2}$$

Por um cálculo direto temos

$$\begin{aligned} \langle c(U) v_t - v_{xx}, Z \rangle &= \langle c_u^* W_t v - c(W) \eta_t - c_u^* u_t \eta, Z \rangle + \langle b_u^* \eta + b_u^{**} v + b_{u_x}^* v_x, Z \rangle + \\ &+ \langle R, Z \rangle + \langle b(W, u_x) - b(W, W_x), Z \rangle, \end{aligned}$$

onde

$$c_u^* = \frac{[c(U) - c(W)]}{U - W}, \quad ,$$

$$c_u^{**} = \frac{[c(u) - c(W)]}{u - W}, \quad ,$$

$$b_u^* = \frac{[b(u, u_x) - b(W, u_x)]}{u - W}, \quad ,$$

$$b_u^{**} = \frac{[b(W, W_x) - b(U, W_x)]}{W - U}, \quad ,$$

$$b_{u_x}^* = \frac{[b(U, W_x) - b(U, U_x)]}{W_x - U_x} \quad e$$

estamos admitindo que b e c sejam diferenciáveis com respeito a u e u_x .

Escolhendo $Z = v_t$ e supondo que as derivadas de b e c com relação a u e u_x são limitadas, obtemos:

$$\begin{aligned} \langle m v_t - v_{xx}, v_t \rangle &= m \langle v_t, v_t \rangle - \langle v_{xx}, v_t \rangle = m |v_t|^2 - \langle v_{xx}, v_t \rangle \leq K |v|^2 + |v_x|^2 + |\eta|^2 + |\eta_t|^2 + \\ &+ \frac{m}{3} |v_t|^2 + \langle R, v_t \rangle + \langle b(W, u_x) - b(W, W_x), v_t \rangle. \end{aligned}$$

Acrescentando à expressão anterior a desigualdade

$$\frac{1}{2} \frac{d}{dt} |v|^2 \leq \frac{m}{6} |v_t|^2 + \frac{3}{2m} |v|^2$$

temos:

$$(3.5.1) \quad \frac{m}{2} |v_t|^2 + \frac{1}{2} \frac{d}{dt} |v|^2 - \langle v_{xx}, v_t \rangle \leq K [|v|^2 + |v_x|^2 + |\eta|^2 + |\eta_t|^2] + \\ + \langle R, v_t \rangle + \langle b(W, u_x) - b(W, W_x), v_t \rangle .$$

O lema 3.2.2 implica que

$$-\langle v_{xx}, v_t \rangle = \frac{1}{2} \frac{d}{dt} \langle v_{xx}, v \rangle .$$

Integrando (3.5.1) no intervalo de tempo $(0, t)$, obtemos

$$(3.5.2) \quad \frac{m}{2} \int_0^t |v_t|^2 d\tau - \langle v_{xx}, v \rangle(t) + \frac{1}{2} |v|^2(t) \leq K [|v|^2(0) + \int_0^t |v|^2 + |v_x|^2 dx + \\ + \int_0^t (|\eta|^2 + |\eta_t|^2) d\tau] - \frac{1}{2} \langle v_{xx}, v \rangle(0) + \int_0^t \langle R, v_t \rangle d\tau + \int_0^t \langle b(W, u_x) - b(W, W_x), v_t \rangle d\tau .$$

Aplicações dos lemas 3.3.2, e 3.2.4 e da desigualdade

$$(3.5.3) \quad ab \leq \varepsilon a^2 + \frac{1}{4\varepsilon} b^2, \text{ para todo } a \text{ e } b \text{ e } \varepsilon > 0 ,$$

implicam que

$$\left| \int_0^t \langle R, v_t \rangle d\tau \right| \leq \frac{1}{16} [-\langle v_{xx}, v \rangle + |v|^2] + \frac{1}{16} \int_0^t [-\langle v_x, v \rangle + |v|^2] d\tau +$$

$$+ K \sum_{j=1}^N [\|u\|_{H^6(I_j)}^2 + \int_0^t \|u_t\|_{H^6(I_j)}^2 d\tau] h_j^8 .$$

A última integral do lado direito de (3.5.2) pode ser escrita como

$$\int_0^t \langle b(W, u_x) - b(W, W_x), v_t \rangle d\tau = \langle b(W, u_x) - b(W, W_x), v \rangle \Big|_0^t -$$

$$- \int_0^t \left\langle \frac{\partial}{\partial t} [\langle b(W, u_x) - b(W, W_x) \rangle], v \right\rangle d\tau .$$

Utilizando o lema 3.3.2, o teorema do valor médio para derivadas e novamente a desigualdade (3.5.3) com ε conveniente, para avaliar o lado direito da expressão acima, obtemos:

$$\int_0^t \langle b(W, u_x) - b(W, W_x), v_t \rangle d\tau \leq \frac{1}{16} [-\langle v_{xx}, v \rangle + |v|^2] - \frac{1}{16} \int_0^t [-\langle v_{xx}, v \rangle + |v|^2] d\tau +$$

$$+K \sum_{j=1}^N [\|u\|_{H^5(I_j)}^2 + \int_0^t [\|u\|_{H^5(I_j)}^2 + \|u_t\|_{H^5(I_j)}^2] d\tau] h_j^8.$$

Portanto, (3.5.2) pode ser reescrito como

$$\begin{aligned} \frac{m}{2} \int_0^t |v_t|^2 d\tau - \langle v_{xx}, v \rangle(t) + \frac{1}{2} |v|^2(t) \leq & K [|v|^2(0) + \int_0^t [|v|^2 + |v_x|^2] dx + \\ & + \int_0^t [|n|^2 + |n_t|^2] d\tau - \langle v_{xx}, v \rangle(0) + |v|^2 - \langle v_{xx}, v \rangle + \sum_{j=1}^N \{ \|u\|_{H^6(I_j)}^2 + \\ & + \int_0^t \|u_t\|_{H^6(I_j)}^2 d\tau \} h_j^8]. \end{aligned}$$

Segue de (3.2.7), (3.2.8), segunda parte do lema 3.2.5 e lema 3.2.1 que

$$\begin{aligned} \int_0^t |v_t|^2 d\tau + \max [\|v\|_{H_0^1}^2 + |v|^2] \leq & K [\|v(0)\|_{H_0^1}^2 + |v(0)|^2 + \\ & + \sum_{j=1}^N h_j^8 \{ \|u\|_{L^\infty(0,T;H^6(I_j))}^2 + \|u_t\|_{L^2(0,T;H^6(I_j))}^2 \}]. \end{aligned}$$

O lema (3.2.6) implica que

$$\int_0^T |v_t|^2 d\tau + \|v\|_{L^\infty(0,T;H^1)}^2 \leq K[\|v(0)\|_{H^1}^2 + \sum_{j=1}^N h_j^8 \{\|u\|_{L^\infty(0,T;H^6(I_j))}^2 + \|u_t\|_{L^2(0,T;H^6(I_j))}^2\}]$$

Em particular,

$$\|v\|_{L^\infty(0,T;L^\infty)}^2 \leq K[\|v(0)\|_{H^1}^2 + \sum_{j=1}^N h_j^8 \{\|u\|_{L^\infty(0,T;H^6(I_j))}^2 + \|u_t\|_{L^2(0,T;H^6(I_j))}^2\}].$$

Obtida a estimativa para v , lançaremos mão do seguinte resultado estabelecido na teoria da aproximação[9]:

$$\|\eta\|_{L^\infty(0,T;L^\infty)}^2 \leq K \sum_{j=1}^N h_j^8 \|u\|_{L^\infty(0,T;H^5(I_j))}^2. \text{ De fato, como}$$

$\eta + v = u - U$, temos

$$\|u-U\|_{L^\infty(0,T;L^\infty)}^2 \leq K[\|v(0)\|_{H^1}^2 + \sum_{j=1}^N h_j^8 \{\|u\|_{L^\infty(0,T;H^6(I_j))}^2 + \|u_t\|_{L^2(0,T;H^6(I_j))}^2\}].$$

Escolhendo $U(x,0)$ como o H_3 -interpolante de $u(x,0)$, obtemos

$$\|u-U\|_{L^\infty(0,T;L^\infty)}^2 \leq K[\|u\|_{L^\infty(0,T;H^6)}^2 + \|u_t\|_{L^2(0,T;H^6)}^2] h^4,$$

$$h = \max h_j, \quad j = 1, \dots, N$$

3.6 O MÉTODO DA COLOCAÇÃO NO TEMPO DISCRETO: A fim de obtermos soluções aproximadas para o sistema de equações diferenciais ordinárias definido por (3.1.8) é necessário discretizar a variável t .

Nesta seção introduziremos alguns esquemas de aproximação que podem ser utilizados para este fim.

Seja,

$$t_m = m \Delta t, \text{ onde } \Delta t = T/M, M \text{ um inteiro positivo.}$$

O primeiro esquema que apresentamos aqui é o método da colocação-Crank-Nicolson obtido da seguinte forma:

Buscamos uma aplicação $U: \{t_0, t_1, \dots, t_M\} \rightarrow H_3$ tal que

$$(3.6.1) \quad \begin{aligned} & \text{(i)} \quad U^0 - f \text{ é pequena} \\ & \text{(ii)} \quad \left\{ c(U^{m+\frac{1}{2}})_t U^m - a(U^{m+\frac{1}{2}})_{xx} - b(U^{m+\frac{1}{2}}, U_x^{m+\frac{1}{2}}) \right\} (\xi_{ij}) = 0, \end{aligned}$$

$$i = 1, \dots, N, j = 1, 2, m = 0, 1, \dots, M-1$$

$$\text{(iii)} \quad U^m(0) = g_0(t_m), U^m(1) = g_1(t_m), m = 0, 1, \dots, M$$

onde

$$U^m = U^m(x) = U(x, t_m),$$

$$U^{m+\frac{1}{2}} = \frac{U^{m+1} + U^m}{2},$$

$$d_t U^m = (U^{m+1} - U^m) / \Delta t.$$

A interpretação de $c(U^{m+\frac{1}{2}})$, bem como dos outros coeficientes é

$$c(U^{m+\frac{1}{2}})(\xi_{ij}) = c(\xi_{ij}, t_{m+\frac{1}{2}}, U^{m+\frac{1}{2}}(\xi_{ij})).$$

É possível mostrar que este método é de segunda ordem no tempo [12].

A menos que a equação diferencial (3.1.2) seja linear, (3.6.1) requer a solução, para cada tempo, de um sistema de equações algébricas não linear. A não linearidade pode ser contornada usando, em vez de (3.6.1), o método preditor-corretor que descrevemos a seguir.

Com a notação anterior consideremos o esquema de aproximação:

$$(3.6.2) \quad \{c(U^m) d_t \tilde{U}^m - a(U^m) \tilde{U}_{xx}^m - b(U^m, U_x^m)\}(\xi_{ij}) = 0$$

$$(3.6.3) \{c(\hat{U}^m) d_t U^m - a(\hat{U}^m) U_{xx}^{m+1/2} - b(\hat{U}^m, \hat{U}_x^m)\} (\xi_{ij}) = 0, \quad m \geq 0,$$

$$i = 1, \dots, N, \quad j = 1, 2,$$

onde

$$d_t \hat{U}^m = \frac{\hat{U}^{m+1} - U^m}{\Delta t},$$

$$\hat{U}^m = \frac{\hat{U}^{m+1} + U^m}{2},$$

$$\hat{U}^{m+1} = \sum_{j=1}^{2N} \hat{\alpha}_j^{m+1} z_j(\xi_{ik}).$$

No esquema acima (3.6.2) é chamado o preditor e (3.6.3) é o corretor.

Observemos que (3.6.2)-(3.6.3) requer a solução, para cada tempo, de dois sistemas algébricos lineares. Este método é também de segunda ordem no tempo [12].

Outro procedimento de aproximação para (3.1.8) e que requer apenas a solução de um conjunto de equações lineares para cada tempo é o chamado método de Crank-Nicolson extrapolado.

Nele, extrapolamos U^m para predizer U^{m+1} e então corrigimos. Mais especificamente, tomemos o esquema:

$$\{c(\hat{U}^m) \hat{d}_t U^m - a(\hat{U}^m) U_{xx}^{m+\frac{1}{2}, \theta} - b(\hat{U}^m, \hat{U}_x^m)\}(\xi_{ij}) = 0$$

$$i = 1, \dots, N, \quad j = 1, 2, \quad m = 1, 2, \dots, M-1,$$

onde

$$U^{m+\frac{1}{2}, \theta} = U^{m+\frac{1}{2}} + \theta [U^{m+1} - 2U^m + U^{m-1}], \quad \theta \geq 0,$$

$$\hat{U}^m = (3U^m - U^{m-1})/2.$$

Tal equação de diferenças requer um procedimento inicial diferente dos métodos anunciados previamente. Neste caso, temos que conhecer a solução para dois passos iniciais no tempo. A condição inicial pode ser utilizada para determinar a primeira aproximação U^0 . O vetor U^1 pode ser determinado por algum outro método; um conveniente pode ser obtido utilizando a própria equação (3.6.1) com os coeficientes sendo avaliados em U^0 , isto é,

$$\{c(U^0) \hat{d}_t U^0 - a(U^0) U_{xx}^{1, \theta} - b(U^0, U_x^0)\}(\xi_{ij}) = 0$$

$$i = 1, \dots, N, \quad j = 1, 2.$$

4. ASPECTOS COMPUTACIONAIS

4.1 INTRODUÇÃO: Neste capítulo apresentaremos alguns aspectos práticos da implementação do problema (3.1.2)-(3.1.4) bem como resultados numéricos de testes realizados. Os cálculos foram realizados no VAX/VMS v4.4, com a linguagem estruturada FORTRAN 77, usando precisão dupla. Uma rotina de eliminação Gaussiana com aproveitamento da banda foi utilizada para resolver o sistema obtido pela aplicação do método em questão. A discretização na variável temporal foi feita pelo método de Crank-Nicolson modificado que é um esquema de três passos no tempo.

4.2 DISCRETIZAÇÃO NO ESPAÇO: Por conveniência vamos reduzir o problema (3.1.2)-(3.1.4) a um problema de contorno homogêneo, isto é, $u(0,t)=u(1,t)=0$. Para isto, definimos a função

$$(4.2.1) \quad w(x,t) = u(x,t) - v(x,t) \quad ,$$

onde $v(x,t)$ para cada t fixado é uma função de x dada por

$$v(x,t) = [g_1(t) - g_0(t)]x + g_0(t)$$

Assim, resolver (3.1.2)-(3.1.4) é equivalente a encontrar a solução do problema

$$(4.2.2) \quad \begin{cases} \bar{c}(w) w_t - \bar{a}(w) w_{xx} = \bar{b}(w, w_x) \\ w(x, 0) = \bar{f}(x) \\ w(0, t) = w(1, t) = 0 \end{cases} ,$$

onde w é definida como em (4.2.1). No problema acima estamos suprimindo parcialmente a escrita das variáveis x e t em $\bar{c}$, $\bar{a}$ e $\bar{b}$.

Buscamos uma função $W(x, t) \in H_3^0$ que melhor aproxime a solução $w(x, t)$ de (4.2.2). Para cada t fixado podemos expressar W como

$$W(x, t) = \sum_{j=1}^{2N} \alpha_j(t) Z_j(x) ,$$

onde as funções $Z_j(x)$ são definidas como em (2.3.3).

Seja $\pi: 0 = x_0 < x_1 < \dots < x_N = 1$ uma partição de $[0, 1]$.

Como temos $2N$ funções linearmente independentes necessitamos de $2N$ pontos distintos para compatibilização do número de incógnitas e equações; vamos escolhê-los como sendo os nós Gaussianos de π , isto é ,

$$\bar{\pi}: 0 = x_0 < \bar{x}_{11} < \bar{x}_{12} < x_1 < \bar{x}_{21} < \bar{x}_{22} < x_2 < \dots < x_{N-1} < \bar{x}_{N1} < \bar{x}_{N2} < x_N = 1,$$

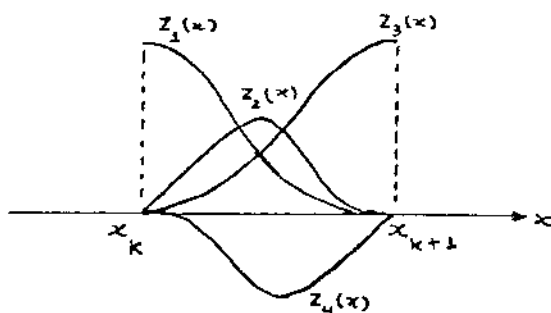
onde $\bar{x}_{ik}$, $i = 1, 2, \dots, N$; $K = 1, 2$ são definidos em (2.3.2).

Fundamentalmente o método da colocação exige que a equação diferencial seja satisfeita nos pontos de colocação, ou seja, queremos uma aplicação diferenciável $W: [0, T] \rightarrow H_3^0$ tal que

$$(4.2.3) \quad \left\{ \begin{array}{l} \{ \bar{c}(W)W_t - \bar{a}(W)W_{xx} - \bar{b}(W, W_x) \}(\bar{x}_{ij}, t) = 0, \\ i = 1, \dots, N, j = 1, 2, \\ 0 < t \leq T, \\ \bar{W}(x, 0) - \bar{f}(x) \text{ seja pequena}, \\ \bar{W}(0, t) = \bar{W}(1, t) = 0, \quad 0 < t \leq T \end{array} \right.$$

Por questão prática não trabalharemos diretamente com as funções $Z_j(x)$; adotaremos o seguinte procedimento:

(i) Enumeração local das funções da base



Assim, em cada elemento $I_k = [x_k, x_{k+1}]$ podemos escrever a função $W(x, t)$ como

$$W(x,t) /_{I_k} = W^k(x,t) = \sum_{j=1}^4 \alpha_j^k(t) z_j(x) \quad , \quad k = 2, 3, \dots, N-1 \quad ,$$

$$W^1(x,t) = \sum_{j=2}^4 \alpha_j^1(t) z_j(x) \quad , \quad \text{no 1º elemento} \quad ,$$

$$W^N(x,t) = \sum_{\substack{j=1 \\ j \neq 3}}^4 \alpha_j^N(t) z_j(x) \quad , \quad \text{no n-ésimo elemento.}$$

Utilizando esta enumeração local das funções da base podemos impor a condição de interpolação em cada um dos elementos obtendo assim duas equações:

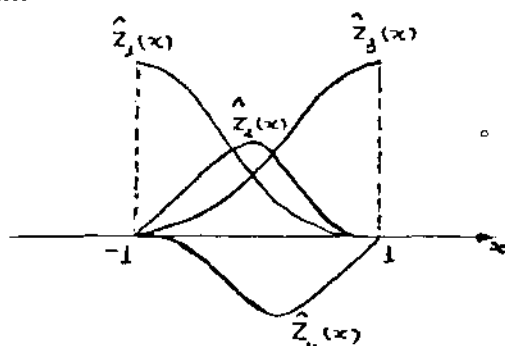
$$\begin{cases} \{ \bar{c}(W^k) W_t^k - \bar{a}(W^k) W_{xx}^k - \bar{b}(W^k, W_x^k) \} (\bar{x}_{k1}, t) = 0 \\ \{ \bar{c}(W^k) W_t^k - \bar{a}(W^k) W_{xx}^k - \bar{b}(W^k, W_x^k) \} (\bar{x}_{k2}, t) = 0 \end{cases} \quad ,$$

$$k = 1, 2, \dots, N-1, N.$$

Portanto, teremos um sistema global de equações diferenciais ordinárias em t formado por $2N$ equações a $2N$ incógnitas.

(ii) Elemento padrão.

Vamos definir as funções de base local no intervalo $[-1, 1]$ da forma



onde

$$Z_1(x) = (2+x)(1-x)^2 \cdot 0,25$$

$$Z_2(x) = (1+x)(1-x)^2 \cdot 0,25$$

$$Z_3(x) = (2-x)(1+x)^2 \cdot 0,25$$

$$Z_4(x) = -(1-x)(1+x)^2 \cdot 0,25$$

Para trabalhar no elemento padrão $[-1,1]$ devemos fazer uma mudança de variável definida pela seguinte transformação afim:

$$T: [x_k, x_{k+1}] \rightarrow [-1, 1] \quad ,$$

$$x \rightarrow T(x) = Ax + B \quad ,$$

onde

$T(x_k) = -1$ e $T(x_{k+1}) = 1$, isto é, T é expressa por

$$T(x) = \frac{2}{h_k} x - \frac{x_k + x_{k+1}}{h_k} .$$

Para $\bar{x} \in (x_k, x_{k+1})$ temos de

$$z_j(\bar{x}) = z_j(T(\bar{x}))$$

que

$$z'_j(\bar{x}) = z'_j(T(\bar{x})) \cdot T'(\bar{x}) = z'_j(T(\bar{x})) \frac{2}{h_k},$$

$$z''_j(\bar{x}) = z''_j(T(\bar{x})) \frac{4}{h_k^2}.$$

Observemos que os fatores $\frac{2}{h_k}$ e $\frac{4}{h_k^2}$ provenientes da mudança

de variável, devem ser considerados quando necessitarmos das derivadas das funções de base local.

4.3 DISCRETIZAÇÃO NA VARIÁVEL TEMPORAL: A fim de que cálculos práticos possam ser executados vamos discretizar (4.2.3) no tempo utilizando o esquema de Crank-Nicolson extrapolado definido em (3.6). Assim, na notação utilizada no capítulo 3 temos:

$$(4.3.1) \quad \{ \bar{c}(\bar{w}^m) d_t \bar{w}^m - \bar{a}(\bar{w}^m) \bar{w}_{xx}^{m+\frac{1}{2}, \theta} - \bar{b}(\bar{w}^m, \bar{w}_x^m) \} (\xi_{ij}) = 0$$

$$i = 1, \dots, N, \quad j = 1, 2, \quad m = 1, 2, \dots, M-1,$$

onde

$$d_t W^m = \frac{W^{m+1} - W^m}{\Delta t}$$

$$W^{m+\frac{1}{2}, \theta} = W^{m+\frac{1}{2}} + \theta [W^{m+1} - 2W^m + W^{m-1}] \quad , \quad \theta \geq 0 \quad ,$$

$$\hat{W}^m = (3W^m - W^{m-1})/2 \quad ,$$

e os coeficientes são avaliados para $t = (m+\frac{1}{2})\Delta t = t_{m+\frac{1}{2}}$.

Para iniciar este procedimento tomamos W^0 como sendo o H_3 -interpolante de w^0 e então calculamos W^1 utilizando a mesma equação definida em (4.3.1) substituindo $\hat{W}^m$ por W^0 com m e θ iguais a zero.

4.4 A MATRIZ DE COLOCAÇÃO: A equação (4.3.1) pode ser reescrita em cada elemento na forma

$$A^k_{\alpha_j^{m+1}} = B^k_{\alpha_j^m} + C^k_{\alpha_j^{m-1}} + D^k \quad ,$$

$$k = 1, 2, \dots, N \quad ; \quad m = 1, 2, \dots, M-1$$

onde

$$(4.4.1) \quad A^k = \sum_{j=1}^4 [\bar{c}(\tilde{W}^m) \cdot \bar{z}_j(\bar{x}_{ki}) - \bar{a}(\tilde{W}^m) \Delta t \cdot 0,5 \bar{z}_j''(\bar{x}_{ki}) - \bar{a}(\tilde{W}^m) \Delta t \cdot \theta \cdot \bar{z}_j''(\bar{x}_{ki})] ,$$

$$k = 1, \dots, N; \quad i = 1, 2 \quad , \quad \theta \geq 0,$$

Observemos que no primeiro elemento eliminamos a primeira coluna da matriz $K^1(I,J)$, enquanto que no n -ésimo elemento permutamos a quarta com a terceira coluna e depois eliminamos esta última. A rotina MCOL apresentada no final deste capítulo constrói a matriz de colocação por elemento.

A matriz de colocação global será construída de tal modo que somente as entradas diferentes de zero sejam armazenadas. Assim, teremos uma matriz de ordem $5 \times 2N$ definida como segue:

- 1ª linha: diagonal principal da matriz A
- 2ª linha: primeira diagonal inferior da matriz A
- 3ª linha: segunda diagonal inferior da matriz A
- 4ª linha: primeira diagonal superior da matriz A
- 5ª linha: segunda diagonal superior da matriz A

O armazenamento da matriz A nesta estrutura será feito pela subrotina MVG como detalharemos no final do capítulo.

4.5 O VETOR DE COLOCAÇÃO: Em cada elemento $[x_k, x_{k+1}]$ construímos um vetor auxiliar $M^k(I)$ através do algoritmo:

Algoritmo 1.

```

DO I = 1, 2
  Mk(I) = 0,      (k=1, 2, ..., N)
  IL = 2.k-2      (k=1, 2, ..., N)
  DO J = 1, 4
    Mk(I) = Mk(I) + K1(I,J) . αm(IL) + K2(I,J) αm-1(IL) (m=1, 2, ..., M-1; k=1, ..., N)
    IL = IL+1
  Mk(I) = Mk(I) + Dk      (k = 1, 2, ..., N)

```

A seguir, com o algoritmo abaixo, simplesmente montamos passo a passo o vetor de colocação global VG.

Algoritmo 2

IB = 0

DO J = 1, 2

IB = IB+1

VG(IB) = $M^k(J)$ (k = 1, 2, ..., N)

As subrotinas que descrevem estas situações serão detalhadas na última seção deste capítulo.

4.6 EXEMPLOS NUMÉRICOS: Com a finalidade de testar o programa implementado bem como a comprovação numérica dos resultados teóricos estabelecidos neste trabalho e em [12], três exemplos serão apresentados a seguir. Os resultados numéricos correspondentes a cada exemplo estão dispostos em tabelas.

Na notação da tabela o erro é definido por:

$$\|u-U\|_{\infty} = \max_t \max_x |u(x,t) - U(x,t)|, \quad t \in [0,T], \quad x \in [0,1],$$

Por conveniência os valores máximos são escolhidos apenas nos pontos da malha.

Nestas tabelas as colunas relativas aos valores de α e β serão detalhadas posteriormente.

Ao parâmetro θ que aparece na discretização da variável temporal foi atribuído o valor 0,0625. Também, estaremos utilizando a representação $1,4(-5)$ para $1,4 \times 10^{-5}$.

Exemplo 1

$$\begin{cases} u_t - uu_{xx} = \cos(x+t) + u^2 \\ u(x,0) = \sin(x), \quad 0 \leq x \leq 1 \\ u(0,t) = \sin(t), \quad 0 \leq t \leq 1 \\ u(1,t) = \sin(1+t) \end{cases}$$

Solução exata: $U(x,t) = \sin(x+t)$

Tabela 1

$\Delta x \backslash \Delta t$	$\frac{1}{5}$		$\frac{1}{10}$		$\frac{1}{20}$		$\frac{1}{40}$	
	$\ u-U\ _{\infty}$	α	$\ u-U\ _{\infty}$	α	$\ u-U\ _{\infty}$	α	$\ u-U\ _{\infty}$	α
$\frac{1}{100}$	1,4831(-5)	-	1,547(-5)	-	1,547(-5)	-	1,547(-5)	-
$\frac{1}{200}$	3,670(-6)	2,01	3,867(-6)	2,00	3,870(-6)	1,99	3,870(-6)	1,99
$\frac{1}{400}$	8,954(-7)	2,03	9,650(-7)	2,00	9,676(-7)	2,00	9,678(-7)	2,00
$\frac{1}{600}$	4,602(-7)	1,6	4,274(-7)	2,01	4,300(-7)	1,99	4,301(-7)	1,99

Exemplo 2

$$\begin{cases} u_t - uu_{xx} = u + u^2 \\ u(x,0) = \sin(x) \\ u(0,t) = 0 \\ u(1,t) = e^t \sin(1) \end{cases}$$

Solução exata: $u(x,t) = e^t \text{sen}(x)$.

Tabela 2

$\Delta x \backslash \Delta t$	$\frac{1}{5}$		$\frac{1}{10}$		$\frac{1}{20}$		$\frac{1}{40}$	
	$\ u-U\ _\infty$	α	$\ u-U\ _\infty$	α	$\ u-U\ _\infty$	α	$\ u-U\ _\infty$	α
$\frac{1}{100}$	4,884(-6)	-	4,805(-6)	-	4,600(-6)	-	4,828(-6)	-
$\frac{1}{200}$	1,322(-6)	1,9	1,209(-6)	1,9	1,159(-6)	1,9	1,208(-6)	1,9
$\frac{1}{400}$	4,303(-7)	1,6	3,087(-7)	1,9	2,910(-7)	1,9	3,024(-7)	1,9
$\frac{1}{600}$	2,649(-7)	1,2	1,418(-7)	1,9	1,349(-7)	1,9	1,344(-7)	2,0

Exemplo 3

$$\begin{cases} u_t - u^2 u_{xx} = x - x^4 + 12tx^2 u^2 \\ u(x,0) = 0 \\ u(0,t) = u(1,t) = 0 \end{cases}$$

Solução exata: $u(x,t) = x \cdot t(1-x^3)$

Tabela 3

$\Delta x \backslash \Delta t$	$\frac{1}{100}$		$\frac{1}{200}$		$\frac{1}{400}$		$\frac{1}{600}$	
	$\ u-U\ _\infty$	β	$\ u-U\ _\infty$	β	$\ u-U\ _\infty$	β	$\ u-U\ _\infty$	β
$\frac{1}{5}$	4,632(-5)	-	4,632(-5)	-	4,632(-5)	-	4,632(-5)	-
$\frac{1}{10}$	2,897(-6)	3,99	2,897(-6)	3,99	2,897(-6)	3,99	2,897(-6)	3,99
$\frac{1}{20}$	1,810(-7)	4,00	1,810(-7)	4,00	1,810(-7)	4,00	1,810(-7)	4,00
$\frac{1}{40}$	1,131(-8)	4,04	1,131(-8)	4,04	1,131(-8)	4,04	1,131(-8)	4,04

A seguir apresentaremos alguns argumentos no sentido de comprovar que as ordens de convergência em (Δt) e (Δx) são respectivamente $\alpha = 2$ e $\beta = 4$, como estabelece a estimativa teórica [12]:

$$\max_{1 \leq r \leq M} \| (u-U)^r \|_{L^\infty(I)} \leq K [(\Delta t)^2 + C(\Delta x)^4] ,$$

onde as constantes K e C não dependem das partições em x e em t , M é um inteiro que corresponde ao número de partições do intervalo $[0, T]$ e r define cada nó desta partição. De fato, suponhamos que assintoticamente o erro, nas mesmas normas utilizadas neste trabalho, fosse dado por:

$$(4.6.1) \quad E(\Delta x, \Delta t) \cong k [(\Delta t)^\alpha + C(\Delta x)^\beta]$$

Nesta expressão, fixando uma das partições e reduzindo a outra à metade teríamos:

$$(4.6.2) \quad E(\Delta x, \frac{\Delta t}{2}) \cong k [\frac{(\Delta t)^\alpha}{2^\alpha} + C(\Delta x)^\beta]$$

$$(4.6.3) \quad E(\frac{\Delta x}{2}, \Delta t) \cong K [(\Delta t)^{\tilde{\alpha}} + \frac{C(\Delta x)^\beta}{2^\beta}]$$

Analisemos inicialmente a tabela 1. Observando os valores apresentados segundo as linhas e colunas respectivamente, bem como as expressões (4.6.1), (4.6.2), e (4.6.3) verificamos que:

(i) fixando Δt e reduzindo Δx à metade (segundo as linhas) os valores permanecem praticamente inalterados. De (4.6.1) e (4.6.3) deduzimos que o erro correspondente à contribuição de Δx , isto é, $K(\Delta x)^\beta$ é desprezível com respeito ao termo $K(\Delta t)^\alpha$. Assim, neste exemplo $E(\Delta x, \Delta t) \cong K(\Delta t)^\alpha$.

(ii) fixando Δx , reduzindo Δt à metade (segundo agora as colunas) e tendo em vista (4.6.2), (4.6.1) e (i) temos:

$$\frac{E(\frac{\Delta t}{2}, \Delta x)}{E(\Delta t, \Delta x)} \cong \frac{1}{2^\alpha} \quad \text{e, portanto, } \alpha \cong 2 \quad \text{como indica, na média, a coluna correspondente a } \alpha \text{ desta tabela.}$$

Neste exemplo não conseguimos calcular o valor de β uma vez que os erros provenientes da discretização em x (segundo as linhas) são praticamente os mesmos.

A análise da tabela 2 é a mesma procedida no exemplo anterior: concluímos que $\alpha \cong 2$ e não conseguimos calcular o valor de β .

Consideremos agora a tabela 3. Observe que invertemos propositalmente a apresentação da tabela de modo a introduzir uma coluna correspondente a β .

(i) Seguindo as linhas, isto é, fixando Δx e reduzindo Δt à metade os valores não se alteram. Concluímos então que o termo $K(\Delta t)^\alpha$ de (4.6.1) é desprezível com respeito a $KC(\Delta x)^\beta$. Assim, neste exemplo

$$E(\Delta x, \Delta t) \cong KC(\Delta x)^\beta$$

(ii) Seguindo agora as colunas e observando que

$$\frac{E(\frac{\Delta x}{2}, \Delta t)}{E(\Delta x, \Delta t)} \cong \frac{1}{2^\beta}$$

obtemos o valor de $\beta \cong 4$, o que está de acordo com os resultados teóricos [12].

Embora reconheçamos que o número de exemplos seja reduzido ousamos concluir que, a despeito das peculiaridades que cada problema possa apresentar, a constante C pode ser um fator decisivo no cálculo da aproximação. Além disso, as estimativas no trabalho teórico [12] são ótimas no sentido de que as ordens de convergência lá estabelecidas correspondem aos resultados numéricos.

4.7 DETALHES DO PROGRAMA

Como indicamos previamente apresentaremos nesta seção algumas subrotinas que utilizamos na implementação do problema(4.2.2); antes de detalhar cada uma delas listaremos as variáveis mais utilizadas.

NE	Número de elementos na variável espacial x
NT	Número de elementos na variável temporal t
HX	Tamanho de cada subintervalo da partição de x
HT	Tamanho de cada subintervalo da partição de t
L1,L2	Extremos do domínio da variável espacial
T1,T2	Extremos do domínio da variável temporal
G(2)	Zeros do polinômio de Legendre de grau dois
M(2)	Vetor de colocação por elemento
X(NE+1)	Vetor que define os nós naturais na variável x
T(NT+1)	Vetor que define os nós naturais na variável t
T5 (NT+1)	Vetor que define os nós intermediários referentes à partição em t
XG(2,NE)	Nós Gaussianos da partição
BASE(4)	Funções da base definidas no elemento padrão
DERIV(4)	Derivadas segundas das funções da base no elemento padrão
K(2,4)	Matriz de colocação por elemento

MG(5,2.NE) Matriz de colocação global
 VG(2.NE) Vetor de colocação global
 ALFA(2.NE) Vetor solução do sistema linear
 U(NE+1,NT+1) Solução aproximada nos nós naturais da partição
 da variável x.

4.7.1 Subrotina IAUX

Esta subrotina calcula matrizes auxiliares que representam as funções de base local e suas derivadas avaliadas nos zeros do polinômio de Legendre de grau dois.

Observemos que estamos multiplicando as derivadas segundas das funções da base pelo fator $4/(Hx)^2$, devido à transformação linear que realizamos.

```

C -----
C      SUBROUTINE IAUX(G,HX,X1,X2)
C -----
C      MATRIZES AUXILIARES
      REAL*8 Y,G(2),BASE(4),DERIV(4),HX,X3,X1(2,4),X2(2,4)
      X3=4.D0/(HX*HX)
      DO 5 I=1,2
        Y=G(I)
        CALL SHAPE(Y,BASE,DERIV)
      DO 10 J=1,4
        X1(I,J)=BASE(J)
        X2(I,J)=X3*DERIV(J)
10     CONTINUE
5      CONTINUE
      RETURN
      END
C -----

```

4.7.2 Subrotina NN

Esta subrotina calcula os nós naturais da partição da variável espacial.

```

C -----
C SUBROUTINE NN(L1,L2,NE,HX,X)
C -----
C CALCULA OS NOS NATURAIS DE [L1,L2]
C REAL*8 L1,L2,NE1,X(1),HX
C NE1=DFLOAT(NE)
C HX=(L2-L1)/NE1
C X(1)=L1
C DO 5 IE=2,NE+1
C   J=IE-1
C   X(IE)=X(J)+HX
5 CONTINUE
RETURN
END

```

4.7.3 Subrotina NG

Calcula os nós Gaussianos da partição da variável espacial através de uma mudança de variável definida pela transformação afim analisada em (ii) da seção 4.2.

```

C -----
C SUBROUTINE NG(NE,X,HX,XG,G)
C -----
C CALCULA OS NOS GAUSSIANOS DE [L1,L2]
C REAL*8 X(1),XG(2,1),G(2),HX,EI,ES
C G(1)=-DSQRT(3.D0)/3.D0
C G(2)=-G(1)
C DO 10 IE=2,NE+1
C   J=IE-1
C   EI=X(J)
C   ES=X(J+1)
C   DO 10 K=1,2
C     XG(K,J)=((EI+ES)+HX*G(K))/2.D0
10 CONTINUE
RETURN
END

```

4.7.4. Subrotina TEMPO

Calcula os nós naturais, os nós intermediários e a amplitude de cada subintervalo da partição da variável temporal.

```

C -----
C SUBROUTINE TEMPO(T1,T2,IT,NT,HT,T,T5)
C -----
C CALCULA OS NOS INTERMEDIARIOS DA PARTICAO DE [T1,T2]
C E OS NOS NATURAIS DA MESMA PARTICAO
C REAL*8 NT2,HT,T1,T2,T(1),T5(1)
C NT2=DFLOAT(NT)
C HT=(T2-T1)/NT2
C IF (IT.NE.1) THEN
C   T(IT)=HT/2.00+(IT-2)*HT
C   T5(IT)=T(IT)+HT/2.00
C ELSE
C   T(1)=T1
C   T5(1)=T1
C END IF
C RETURN
C END

```

4.7.5. Subrotina SHAPE

Avalia as funções de base local bem como suas derivadas segundas nos pontos de Gauss definidos no elemento padrão.

```

C -----
C SUBROUTINE SHAPE(Y,BASE,DERIV)
C -----
C FUNCOES DA BASE DEFINIDAS NO ELEMENTO PADRAO E SUAS
C DERIVADAS SEGUNDAS AVALIADAS NOS NOS GAUSSIANS
C REAL*8 Y,BASE(4),DERIV(4)
C BASE(1)=(2+Y)*(1-Y)*(1-Y)*.2500
C BASE(2)=(1+Y)*(1-Y)*(1-Y)*.2500
C BASE(3)=(2-Y)*(1+Y)*(1+Y)*.2500
C BASE(4)=-(1-Y)*(1+Y)*(1+Y)*.2500
C DERIV(1)=1.500*Y
C DERIV(2)=2.00*(3.00*Y-1)/4.00
C DERIV(3)=-1.500*Y
C DERIV(4)=2.00*(3.00*Y+1)/4.00
C RETURN
C END

```

4.7.6. Subrotina SI

Esta subrotina fornece a solução aproximada do problema nos nós naturais da partição da variável espacial no tempo $T(1)=0$. Isto é feito avaliando simplesmente a condição inicial $F(x)$ e sua derivada $D(x)$ nestes mesmos nós. Para este caso, o erro entre a solução aproximada e a solução exata é zero uma vez que estamos fazendo a avaliação exatamente nos pontos de interpolação.

```

C -----
C SUBROUTINE SI(IE,NE,X,HX,ALFA)
C -----
C SOLUCAO INICIAL NOS NOS NATURAIS PARA IT=1
C IMPLICIT REAL*8 (A-H,O-Z)
C REAL*8 X(1),ALFA(1),HX,P3
C P3=HX*.5D0
C IF (IE.EQ.1) THEN
C   ALFA(1)=P3*D(X(1))
C   ALFA(2)=F(X(IE+1))
C ELSE IF (IE.EQ.NE) THEN
C   J=2*IE
C   K=2*IE-1
C   ALFA(K)=P3*D(X(IE))
C   ALFA(J)=P3*D(X(IE+1))
C ELSE
C   J=2*IE
C   K=2*IE-1
C   ALFA(K)=P3*D(X(IE))
C   ALFA(J)=F(X(IE+1))
C END IF
C RETURN
C END

```

4.7.7. Subrotina JAUX

Calcula, a cada passo no tempo, a solução aproximada nos nós Gaussianos da partição.

```

C -----
C SUBROUTINE JAUX(IE,NE,X1,ALFA,U2)
C -----
C VETOR QUE CALCULA A SOLUCAO APROXIMADA NO NOS GAUSSIANS
REAL*8 U2(2,1),X1(2,4),ALFA(1)
-- DO 15 I=1,2
U2(I,IE)=0.D0
I1=2*IE-2
-- IF (IE.LI.NE) THEN
-- DO J=1,4
U2(I,IE)=U2(I,IE)+X1(I,J)*ALFA(I1)
I1=I1+1
-- -- END DO
-- -- ELSE
-- DO 30 J=1,4
-- IF (J.EQ.3) GO TO 30
U2(I,IE)=U2(I,IE)+X1(I,J)*ALFA(I1)
I1=I1+1
30 CONTINUE
-- -- END IF
15 CONTINUE
RETURN
-- -- END

```

4.7.8. Subrotina LAUX

Armazena o vetor ALFA(solução do sistema linear) em BETA.

```

C -----
C SUBROUTINE LAUX(NE,ALFA,BETA)
C -----
C ARMAZENA ALFA EM BETA
REAL*8 ALFA(1),BETA(1)
DO IE=1,2*NE
BETA(IE)=0.D0
BETA(IE)=ALFA(IE)
END DO
RETURN
END

```

4.7.9. CRN

Calcula o fator de correção do método de Crank-Nicolson extrapolado utilizando a fórmula:

$$U^m = (3U^m - U^{m-1})/2.$$

```

C -----
C SUBROUTINE CRN(IE,U1,U2,UTIL)
C -----
C CRANK-NICOLSON EXTRAPOLADO
  REAL*8 U1(2,1),U2(2,1),UTIL(2,1)
  DO I=1,2
    UTIL(I,IE)=0.D0
    UTIL(I,IE)=1.5D0*U2(I,IE)-.5D0*U1(I,IE)
  END DO
  RETURN
  END

```

4.7.10. Subrotina KAUX

Esta subrotina transfere os valores atuais da matriz U2(I, IE) para U1(I,IE) a fim de que esta possa ser utilizada como parâmetro em outras subrotinas.

```

C -----
C SUBROUTINE KAUX(IE,U2,U1)
C -----
C ARMAZENA U2 EM U1
  REAL*8 U2(2,1),U1(2,1)
  DO I=1,2
    U1(I,IE)=0.D0
    U1(I,IE)=U2(I,IE)
  END DO
  RETURN
  END

```

4.7.11. Subrotina MCOL

Nesta subrotina calculamos a matriz de colocação por elemento na forma como foi definida em (4.4.1). Observemos que na presente aplicação esta matriz é de ordem 2×4 e seus elementos são armazenados temporariamente em $K(2,4)$.

```

C -----
C SUBROUTINE MCOL(XG,IE,NE,IT,T,U1,X1,X2,HT,TETA,UTIL,K,K1,S,K2)
C -----
C MATRIZ DE COLOCACAO POR ELEMENTO
  IMPLICIT REAL*8 (A-H,O-Z)
  REAL*8 S(2),XG(2,1),Y,T(1),U1(2,1),UTIL(2,1),X1(2,4)
  REAL*8 X2(2,4),K(2,4),K1(2,4),HT,TETA,X4,X5,X6,X7,X8
  REAL*8 K2(2,4),W
  DO 10 I=1,2
    S(I)=XG(I,IE)
    Y=S(I)
    DO 15 J=1,4
      IF (IT.EQ.2) THEN
        X4=C(Y,T(IT),U1(I,IE))*X1(I,J)
        X5=.5D0*HT*A(Y,T(IT),U1(I,IE))*X2(I,J)
        K(I,J)=X4-X5
        K1(I,J)=X4+X5
      ELSE
        X6=HT*TETA*A(Y,T(IT),UTIL(I,IE))*X2(I,J)
        X7=C(Y,T(IT),UTIL(I,IE))*X1(I,J)
        X8=.5D0*HT*A(Y,T(IT),UTIL(I,IE))*X2(I,J)
        K(I,J)=X7-X8-X6
        K1(I,J)=X7+X8-2*X6
        K2(I,J)=X6
      END IF
    CONTINUE
  CONTINUE
  IF (IE.EQ.NE) THEN
    DO 20 I=1,2
      W=K(I,3)
      K(I,3)=K(I,4)
      K(I,4)=W
    CONTINUE
  END IF
  RETURN
END
15
10
20

```

4.7.12. Subrotina VCOL

Como explicamos na seção 4.5 o vetor de colocação por elemento é obtido por esta subrotina de acordo com o esquema fornecido pelo algoritmo 1. Observemos que tal vetor é constituído de duas componentes e estas são armazenadas temporariamente em $M(2)$.

```

C      SUBROUTINE VCOL(IE,IT,NE,HT,T,K1,ALFA,S,P,BETA,K2,M)
C      -----
C      VETOR DE COLOCACAO POR ELEMENTO
C      -----
      IMPLICIT REAL*8 (A-H,O-Z)
      REAL*8 T(1),S(2),P(2,1),BETA(1)
      REAL*8 M(2),K1(2,4),ALFA(1),K2(2,4),HT
      DO 44 I=1,2
        M(I)=0.D0
        I1=2*IE-2
        IF (IE.LT.NE) THEN
          DO J=1,4
            IF (IT.EQ.2) THEN
              M(I)=M(I)+K1(I,J)*ALFA(I1)
            ELSE
              M(I)=M(I)+K1(I,J)*ALFA(I1)+K2(I,J)*BETA(I1)
            END IF
            I1=I1+1
          END DO
          M(I)=M(I)+HT*B(S(I),T(IT),P(I,IE))
        ELSE
          DO 40 J=1,4
            IF (J.EQ.3) GO TO 40
            IF (IT.EQ.2) THEN
              M(I)=M(I)+K1(I,J)*ALFA(I1)
            ELSE
              M(I)=M(I)+K1(I,J)*ALFA(I1)+K2(I,J)*BETA(I1)
            END IF
            I1=I1+1
          CONTINUE
          M(I)=M(I)+ HT*B(S(I),T(IT),P(I,IE))
        END IF
      CONTINUE
      RETURN
      END
40
44

```

4.7.13. Subrotina MVG

Esta subrotina constrói o vetor e a matriz de colocação global. O vetor de colocação global é obtido segundo o algoritmo 2 apresentado na seção 4.5 e seus elementos são armazenados em VG(2,NE). A matriz de colocação global é montada de tal forma que se aproveite sua estrutura de banda (ver seção 4.4) e tem seus elementos diferentes de zero guardados em MG(5,2,NE).

```

C -----
C SUBROUTINE MVG(J1,J2,J3,J4,J5,IB,IE,NE,M,K,VG,MG)
C -----
C CALCULA A MATRIZ E O VETOR GLOBAL
C REAL*8 VG(1),MG(5,1),M(2),K(2,4)
C MONTAGEM DO VETOR, DE COLOCACAO GLOBAL
C DO J=1,2
C   IB=IB+1
C   VG(IB)=M(J)
C END DO
C MONTAGEM DA PRIMEIRA LINHA
C DO J=2,3
C   I=J-1
C   J1=J1+1
C   MG(1,J1)=K(I,J)
C END DO
C MONTAGEM DA SEGUNDA LINHA
C IF (IE.EQ.1) THEN
C   J2=J2+1
C   MG(2,J2)=K(2,2)
C ELSE
99 DO J=1,2
C   J2=J2+1
C   MG(2,J2)=K(J,J)
C END DO
C MONTAGEM DA TERCEIRA LINHA
C J3=J3+2
C MG(3,J3)=K(2,1)
C END IF
C MONTAGEM DA QUARTA LINHA
C DO J=3,4
C   I=J-2
C   J4=J4+1
C   MG(4,J4)=K(I,J)
C END DO
C MONTAGEM DA QUINTA LINHA
C J5=J5+2
C MG(5,J5)=K(1,4)
C RETURN
C END

```

4.7.14. Subrotina ELG

Esta subrotina resolve o sistema linear utilizando um método de eliminação Gaussiana que não destrói a característica de banda da matriz de colocação.

```

C      -----
C      SUBROUTINE ELG(NE,MG,VG,ALFA)
C      -----
C      RESOLVE O SISTEMA LINEAR
      REAL*8 Q,VG(1),MG(5,1),ALFA(1),S
      NP1=2*NE
      DO 55 I=1,NP1-1
        Q=-MG(2,I)/MG(1,I)
        J=I+1
        MG(1,J)=MG(1,J)+Q*MG(4,I)
        VG(J)=VG(J)+Q*VG(I)
        IF (I.LT.(NP1-1)) THEN
          MG(4,J)=MG(4,J)+Q*MG(5,I)
          Q=-MG(3,I)/MG(1,I)
          K=I+2
          MG(2,J)=MG(2,J)+Q*MG(4,I)
          MG(1,K)=MG(1,K)+Q*MG(5,I)
          VG(K)=VG(K)+Q*VG(I)
        END IF
55     CONTINUE
C      RETRO-SUBSTITUICAO
      ALFA(NP1)=VG(NP1)/MG(1,NP1)
      J=NP1-1
      ALFA(J)=(VG(J)-ALFA(NP1)*MG(4,J))/MG(1,J)
      J=J-1
      DO 65 K=J,1,-1
        L=K+1
        I=K+2
        S=ALFA(L)*MG(4,K)+ALFA(I)*MG(5,K)
        ALFA(K)=(VG(K)-S)/MG(1,K)
65     CONTINUE
      RETURN
      END

```

4.7.15. Subrotina FINAL

Esta subrotina chamada no final do programa principal calcula o maior erro a partir dos erros máximos avaliados em cada tempo pela subrotina SOL.

```

C -----
C SUBROUTINE FINAL(ERMAX,NT,MAXIMO)
C -----
REAL*8 ERMAS(1),MAXIMO
MAXIMO=ERMAX(1)
DO IT=2,NT+1
  IF (MAXIMO.LT.ERMAX(IT)) MAXIMO=ERMAX(IT)
END DO
WRITE(70,*)'ERRO GLOBAL=',MAXIMO
RETURN
END

```

4.7.16. Subrotina SOL

Apresenta a solução aproximada e a solução exata avaliadas nos nós naturais da partição, bem como o erro absoluto entre estas soluções.

```

C -----
C SUBROUTINE SOL(NE,HX,NT,HT,IT,TETA,X,T5,ALFA,U,SE,ERR,ERMAX)
C -----
C VETOR SOLUCAO NOS NOS GAUSSIANOS DA PARTICAO DE X
IMPLICIT REAL*8 (A-H,O-Z)
REAL*8 X(1),SE(33,1),DIF(33,514),HT,HX
REAL*8 W(33,514),Z1(33,514),U(33,1),ALFA(1),T5(1)
REAL*8 ERR(33,1),TETA,ERMAX(1)
WRITE(70,*)'N.ELEMENTOS EM X=',NE,'PARTICAO EM X=',HX
WRITE(70,*)'N.ELEMENTOS EM T=',NT,'PARTICAO EM T=',HT
WRITE(70,*)'DISCRETIZACAO NO TEMPO=',T5(IT)
WRITE(70,*)'PARAMETRO TETA=',TETA
IF (IT.EQ.1) THEN
  W(1,IT)=F(0.D0)
  W(NE+1,IT)=F(1.D0)
ELSE
  W(1,IT)=0.D0
  W(NE+1,IT)=0.D0
END IF
Z1(1,IT)=V(X(1),T5(IT))
Z1(NE+1,IT)=V(X(NE+1),T5(IT))
U(1,IT)=W(1,IT)+Z1(1,IT)

```

```

U(NE+1,IT)=W(NE+1,IT)+Z1(NE+1,IT)
SE(1,IT)=DSIN(X(1)+T5(IT))
SE (NE+1,IT)=DSIN(X(NE+1)+T5(IT))
DIF(1,IT)=SE(1,IT)-U(1,IT)
DIF(NE+1,IT)=SE(NE+1,IT)-U(NE+1,IT)
ERR(1,IT)=DABS(DIF(1,IT))
ERR(NE+1,IT)=DABS(DIF(NE+1,IT))
-- DO 75 I=1,NE-1
J=I+1
II=2*I
W(J,IT)=ALFA(II)
Z1(J,IT)=V(X(J),T5(IT))
U(J,IT)=W(J,IT)+Z1(J,IT)
SE(J,IT)=DSIN(X(J)+T5(IT))
DIF(J,IT)=SE(J,IT)-U(J,IT)
ERR(J,IT)=DABS(DIF(J,IT))
75 CONTINUE
ERMAX(1)=ERR(1,IT)
-- DO IE=2,NE+1
-- IF (ERMAX(IT).LT.ERR(IE,IT)) ERMAX(IT)=ERR(IE,IT)
-- END DO
-- IF (IT.GE.NI-1) THEN
WRITE(70,120)
120 FORMAT(2X,'SOL.APROXIMADA',12X,'SOL.EXATA',14X,'ERRO',/)
-- DO 80 I=1,NE+1
WRITE(70,180) U(I,IT),SE(I,IT),ERR(I,IT)
180 FORMAT(2X,E16.9,12X,E16.9,12X,E16.9,/)
80 CONTINUE
-- END IF
WRITE(70,*) 'ERMAX(',IT,')=',ERMAX(IT)
RETURN
END

```

4.7.17. Subprograma FUNCTION

As funções definidas nos subprogramas seguintes se referem ao problema 1 apresentado como exemplo numérico na seção 4.6

```

C -----
REAL*8 FUNCTION F(X)
C -----
REAL*8 X
F=DSIN(X)-DSIN(1.D0)*X
RETURN
END

```

```

C -----
C REAL*8 FUNCTION D(X)
C -----
REAL*8 X
D=DCOS(X)-DSIN(1.D0)
RETURN
END

C -----
C REAL*8 FUNCTION A(X,T,U)
C -----
REAL*8 X,T,U
A=U+(DSIN(1.D0+T)-DSIN(T))*X+DSIN(T)
RETURN
END

C -----
C REAL*8 FUNCTION B(X,T,U)
C -----
IMPLICIT REAL*8 (A-H,O-Z)
Y1=(U+(DSIN(1.D0+T)-DSIN(T))*X+DSIN(T))*2
Y2=DCOS(X+T)
Y3=-(DCOS(1.D0+T)-DCOS(T))*X-DCOS(T)
B=Y1+Y2+Y3
RETURN
END

C -----
C REAL*8 FUNCTION C(X,T,U)
C -----
REAL*8 X,T,U
C=1.D0
RETURN
END

C -----
C REAL*8 FUNCTION V(X,T)
C -----
REAL*8 X,T
V=(DSIN(1.D0+T)-DSIN(T))*X+DSIN(T)
RETURN
END

```

4.7.18. Subrotina ZERO

Esta subrotina cada vez que é chamada zera os parâmetros J1, J2, J3, J4 e IB e atribui o valor -1 ao parâmetro J5, que serão utilizados como contadores na subrotina MVG.

```

C      -----
C      SUBROUTINE ZERO(J1,J2,J3,J4,J5,IB)
C      -----
C      VARIÁVEIS UTILIZADAS EM MVG
      J1=0
      J2=0
      J3=0
      J4=0
      J5=-1
      IB=0
      RETURN
      END

```

4.7.19. Programa PRINCIPAL

```

C      PROGRAMA PRINCIPAL
C
C      PROBLEMA DE VALOR INICIAL E DE FRONTEIRA
C
C      EQUACAO DIFERENCIAL PARABOLICA NAO LINEAR
C
C      DOMINIO DA VARIÁVEL ESPACIAL: 0,1
C
C      DISCRETIZACAO EM X: METODO DA COLOCACAO
C
C      DISCRETIZACAO EM T: CRANK-NICOLSON MODIFICADO
C
C      BASE UTILIZADA: HERMITE CUBICO POR PARTES
C
      IMPLICIT REAL*8 (A-H,O-Z)
      REAL*8 MG(5,65),VG(65),ALFA(65),X(33),XG(2,65),X1(2,4)
      REAL*8 X2(2,4),T(514),U(33,514),SE(33,514),ERR(33,514)
      REAL*8 UTIL(2,65),BASE(4),DERIV(4),W(33,514),Z1(33,514)
      REAL*8 M(2),G(2),TS(514),K(2,4),K1(2,4),L1,L2,DIF(33,514)
      REAL*8 HX,T1,T2,HT,TETA,U1(2,65),U2(2,65),P(2,65)
      REAL*8 BETA(65),K2(2,4),S(2),ERMAX(514),MAXIMO

```

```

C -----
C ENTRADA DE DADOS
C -----
      TYPE 5
5  FORMAT(3(/),T5,' TEMPO INICIAL,FINAL E N. ELEMENTOS:',%)
      ACCEPT * ,T1,T2,NT
      TYPE * ,T1,T2,NT
      TYPE 15
15  FORMAT(3(/),T5,'ENTRE COM EXTREMOS E N. DE ELEMENTOS:',%)
      ACCEPT * , L1,L2,NE
      TYPE * ,L1,L2,NE
      TYPE 25
25  FORMAT(3(/),T5,'ENTRE COM O PARAMETRO TETA,TETA)=0:',%)
      ACCEPT 30,TETA
30  FORMAT(G)
      CALL NN(L1,L2,NE,HX,X)
      CALL NG(NE,X,HX,XG,G)
      CALL IAUX(G,HX,X1,X2)
C -----
C PRIMEIRO PASSO (SOLUCAO EM T5(1)=0.)
C -----
      IT=1
      CALL TEMPO(T1,T2,IT,NT,HT,T,T5)
      DO IE=1,NE
      CALL SI(IE,NE,X,HX,ALFA)
      END DO
      CALL SOL(NE,HX,NT,HT,IT,TETA,X,T5,ALFA,U,SE,ERR,ERMAX)
C -----
C SEGUNDO PASSO (SOLUCAO EM T5(2)=HT)
C -----
      IT=2
      CALL TEMPO(T1,T2,IT,NT,HT,T,T5)
      CALL ZERO(J1,J2,J3,J4,J5,IB)
      DO IE=1,NE
      CALL JAUX(IE,NE,X1,ALFA,U2)
      CALL KAUX(IE,U2,U1)
      CALL MCOL(XG,IE,NE,IT,T,U1,X1,X2,HT,TETA,UTIL,K,K1,S,K2)
      CALL VCOL(IE,IT,NE,HT,T,K1,ALFA,S,U1,BETA,K2,M)
C -----
      CALL MVG(J1,J2,J3,J4,J5,IB,IE,NE,M,K,VG,MG)
      END DO
      CALL LAUX(NE,ALFA,BETA)
      CALL ELG(NE,MG,VG,ALFA)
      CALL SOL(NE,HX,NT,HT,IT,TETA,X,T5,ALFA,U,SE,ERR,ERMAX)

```

5. CONCLUSÃO

O método da colocação tal como visto neste trabalho se apresenta como uma ferramenta eficaz para encontrar a solução numérica de equações diferenciais. De fato, com uma escolha apropriada da base e de nós, o método apresenta características interessantes como, por exemplo, matrizes de banda com um pequeno número de banda; tal peculiaridade é particularmente desejável, principalmente, quando lidamos com equações diferenciais parciais, uma vez que, neste caso, o sistema tende a tornar-se muito grande. Além disso, colocando nos nós Gaussianos da partição, é possível acelerar sua razão de convergência tornando-o competitivo, em termos teóricos, com outros métodos.

Outro fator fundamental que contribui para a eficiência do método da colocação é aquele referente ao aspecto computacional. Realmente, como constatamos com a implementação que fizemos do problema parabólico não linear, o método apresenta grande eficiência computacional, bem como uma implementação muito simples.

Em contrapartida, observamos que de maneira geral, teorias sobre existência de solução e análise do erro pelo método da colocação para equações diferenciais parciais não são fáceis de se obter e demonstrar.

Assim, nossa conclusão é que o método da colocação, quando aplicável, é com relação a outros métodos, sensivelmente mais fácil e mais eficiente em termos computacionais.

REFERENCIAS

- [1] VARGA, Richard, S. Matrix Iterative Analysis. Prentice-Hall, Inc. Englewood cliffs, N. J., 1962.
- [2] FERGUSON, Noble, B. and FINLAYSON, Bruce, A. (1970). Transient Chemical Reaction Analysis by Orthogonal Collocation. The Chemical Engineering Journal, vol. 1, pp. 327-336.
- [3] PRENTER, P, M. Splines and Variational Methods. A Wiley-Interscience Publication, 1975.
- [4] PRENTER, P, M. and RUSSEL, R, D. (1976). Orthogonal collocation for Elliptic Partial Differential Equations. Siam J. Numer. Anal., vol. 13, nº 6 , pp. 923-939.
- [5] DYKSEN, W, R, HOUSTIS, E, N, LYNCH, R, E. and RICE, J , R. (1984). The performance of the collocation and Galerkin Methods with Hermite Bi-Cubics. SIAM J. Numer. Anal., vol. 21, nº 4, pp. 695-715
- [6] RUSSEL, R, D. and SHAMPINE, L, F. (1972). A collocation Method for Boundary Value Problems. Numer. Math., vol. 19, pp. 1-28
- [7] LIONS, J, L. and MAGENES, E. Problèmes aux limites non homogènes et applications. Vol. 1, Travaux et Recherches Mathématiques, nº 17, Dunod, Paris , 1968.

- [8] DAVIS, P, J. Interpolation and Approximation. Blaisdell Publishing Company, 1963.
- [9] WENDROFF, B. First Principles of Numerical Analysis, Addison -Wesley, Reading, Mass., 1969.
- [10] CLARK, C, W. Mathematical Bioeconomics. John Wiley & Sons, N. Y., 1976. pp. 325-326.
- [11] DOUGLAS, J, J. and DUPONT, T.(1973). A finite Element Collocation Method for Quasilinear Parabolic Equations. Mathematics of Computation, vol. 27 , n° 121, pp.17-28.
- [12] DOUGLAS, J, J. and DUPONT, T.(1974). Collocation Methods for Parabolic Equations in a Single Space Variable. Lecture Notes in Mathematics, 385.
- [13] De BOOR, C. and B. SWARTZ(1973). Collocation at Gaussian Points. SIAM J. Numer. Anal., Vol. 10, n° 4, pp. 582-606.