

MARTA HELENA DE OLIVEIRA

**MÉTODOS NUMÉRICOS NÃO  
OSCILATÓRIOS APLICADOS ÀS LEIS DE  
CONSERVAÇÃO HIPERBÓLICAS  
UNIDIMENSIONAIS**



UNIVERSIDADE FEDERAL DE UBERLÂNDIA  
FACULDADE DE MATEMÁTICA  
2010

MARTA HELENA DE OLIVEIRA

# MÉTODOS NUMÉRICOS NÃO OSCILATÓRIOS APLICADOS ÀS LEIS DE CONSERVAÇÃO HIPERBÓLICAS UNIDIMENSIONAIS

**Dissertação** apresentada ao Programa de Pós-Graduação em Matemática da Universidade Federal de Uberlândia, como parte dos requisitos para obtenção do título de **MESTRE EM MATEMÁTICA**.

**Área de Concentração:** Matemática.  
**Linha de Pesquisa:** Análise Numérica.

**Orientador:** Prof. Dr. César Guilherme de Almeida.

UBERLÂNDIA - MG  
2010

Dados Internacionais de Catalogação na Publicação (CIP)  
Sistema de Bibliotecas da UFU, MG , Brasil

---

O48m Oliveira, Marta Helena de, 1977-  
Métodos numéricos não oscilatórios aplicados às leis de conservação hiperbólicas unidimensionais [manuscrito] / Marta Helena de Oliveira. - 2010.  
130 f. : il.

Orientador: César Guilherme de Almeida.

Dissertação (mestrado) – Universidade Federal de Uberlândia, Programa de Pós-Graduação em Matemática.  
Inclui bibliografia.

1. Análise numérica - Teses. I. Almeida, César Guilherme de. II. Universidade Federal de Uberlândia. Programa de Pós-Graduação em Matemática. III. Título.

CDU: 519.6

---



UNIVERSIDADE FEDERAL DE UBERLÂNDIA  
FACULDADE DE MATEMÁTICA  
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA

Av. João Naves de Ávila, 2121, Bloco 1F, Sala 1F 152  
Campus Santa Mônica, Uberlândia - MG, CEP 38400-902

**ALUNA:** Marta Helena de Oliveira

**NÚMERO DE MATRÍCULA:** 93818.

**ÁREA DE CONCENTRAÇÃO:** Matemática.

**LINHA DE PESQUISA:** Análise Numérica.

**PÓS-GRADUAÇÃO EM MATEMÁTICA:** Nível Mestrado.

**TÍTULO DA DISSERTAÇÃO:** MÉTODOS NUMÉRICOS NÃO OSCILATÓRIOS APLICADOS ÀS LEIS DE CONSERVAÇÃO HIPERBÓLICAS UNIDIMENSIONAIS

**ORIENTADOR:** Prof. Dr. César Guilherme de Almeida.

A dissertação foi **APROVADA**, em reunião pública, realizada na Sala Multiuso da Faculdade de Matemática, Bloco 1F, Campus Santa Mônica, em 01 de abril de 2010, às 8 horas, com a seguinte Banca Examinadora:

Prof. Dr. César Guilherme de Almeida  
UFU - Universidade Federal de Uberlândia

Profa. Dra. Alice de Jesus Kozakevicius  
UFSM - Universidade Federal de Santa Maria

Profa. Dra. Simone Sousa Ribeiro  
UFF - Universidade Federal Fluminense

Profa. Dra. Ana Maria Amarillo Bertone  
UFU - Universidade Federal de Uberlândia

Uberlândia, 01 de abril de 2010.

# Dedicatória

*Este trabalho é dedicado aos meus pais*

*José Severino e Áurea Maria.*

# Agradecimentos

Todo grande sonho nasce amparado. Cada parte de sua realização têm o suor, a dedicação e a conquista de quem ousou sonhar junto.

Acima de tudo agradeço a Deus e a Nossa Senhora por todas as graças concebidas.

Aos meus pais, José e Áurea, pela confiança, pelo apoio, pelos incentivos e pelo amor incondicional dedicados durante esta jornada.

Aos meus irmãos Ailton, Meirinês e Inês por todo apoio e por todo incentivo.

As minhas amigas Raquel, Vanessa e Viviane pela presença, mesmo que na distância.

As minhas amigas Patrícia Barreto e Marline Ilha da Silva pelo apoio e pela colaboração.

Aos amigos do mestrado Milena, Alessandra, Carolina, Marcelo Lopes e Danilo por todas as soluções sanadas e pelo convívio.

Ao Fred pelos carinhos, pelas angústias, pelas alegrias e acima de tudo pela compreensão.

A todos os funcionários da FAMAT pela hospitalidade.

A todos docentes do Programa de Mestrado-FAMAT que muito contribuíram para a realização deste trabalho, especialmente Ana Maria Amarillo Bertone, Santos Alberto Enriquez Remigio e Alessandro Alves Santana.

Ao programa de Pós-Graduação em Matemática pelo apoio financeiro e ao professor Edson Augustini cordialidade.

Aos professores Ana Maria Amarillo Bertone, Simone Souza Ribeiro por participarem da banca examinadora e, de mesma forma, agradeço aos professores suplentes Alessandro Alves Santana e Alex André Schmidt.

À professora Alice de Jesus Kozakevicius pela orientação nos momentos precisos.

Ao meu orientador, César Guilherme de Almeida pela excelente e competente orientação, bem como pela paciência, pela tolerância e pela compreensão.

A todos aqueles que contribuíram, direta ou indiretamente, e participaram desta vitória meus eternos agradecimentos.

OLIVEIRA, M. H. *MÉTODOS NUMÉRICOS ESSENCIALMENTE NÃO OSCILATÓRIOS APLICADOS ÀS LEIS DE CONSERVAÇÃO HIPERBÓLICAS UNIDIMENSIONAIS* 2010. 130 p. Dissertação de Mestrado, Universidade Federal de Uberlândia, Uberlândia-MG.

## Resumo

A solução de uma lei de conservação pode desenvolver descontinuidades do tipo choque ou ondas de rarefação, mesmo que a condição inicial seja suave. Assim, é desejável o desenvolvimento de esquemas numéricos capazes de reproduzir esses comportamentos. Além de representar corretamente as descontinuidades, os esquemas possuem a tarefa de obter a solução correta conhecida como solução de entropia. O objetivo dessa dissertação é o estudo de métodos numéricos não-oscilatórios para aproximar soluções de leis de conservação hiperbólicas escalares unidimensionais. Para alcançar tal objetivo é preciso estudar alguns esquemas numéricos especiais, tais como esquemas ‘upwind’, esquemas TVD, esquemas conservativos e esquemas monótonos. Para critério de comparação entre os métodos numéricos será utilizada a solução clássica de equações conhecidas da literatura (Equação de Burgers e Equação de Advecção). Para o cálculo das soluções analíticas será empregada a teoria envolvendo as equações características. A aproximação numérica da lei de conservação se divide em duas etapas: a aproximação no espaço e a aproximação no tempo. Para a aproximação no espaço, serão utilizados os esquemas ENO (*Essentially Non-Oscillatory* - essencialmente não oscilatório) e WENO (*Weighted Essentially Non-Oscillatory* - essencialmente não oscilatório ponderado); para a aproximação no tempo, será utilizado o método numérico Runge-Kutta TVD (*Total Variation Dimishing*).

*Palavras-chave:* Leis de conservação, métodos numéricos não-oscilatórios, ENO, WENO, Runge-Kutta TVD

OLIVEIRA, M. H. *APLICACION OF NUMERICAL METHODS ESSENTIALLY NON OSCILLATORY CONSERVATION LAWS ONE DIMENSIONAL* 2010. 130 p. M. Sc. Dissertation, Federal University of Uberlândia, Uberlândia-MG.

### Abstract

The solution of a conservation law may develop discontinuities like shocks and rarefaction waves, even if the initial condition is a smooth function. Then, numerical schemes should be able to generate efficient approximations in order to reproduce the same behavior as the analytic solution. Besides, these schemes have to capture the physically correct solution or entropy solution. The goal of this master dissertation is to study non-oscillatory schemes applied to one-dimensional scalar hyperbolic conservation laws. In order to reach the is objective, it is necessary to understand some special methods, such as, upwind scheme, TVD schemes, conservative schemes and monotone schemes. The effectiveness of the methods will be verified through the comparison with the well-known classical solutions exhibited in literature: Advection Equation and Burgers' Equation. The characteristic equations will be employed for getting analytic solutions of conservation laws. We will derive numerical approximations for conservation laws using ENO (Essentially Non-Oscillatory) and WENO (Weighted Essentially Non-Oscillatory) schemes for space discretization, and Runge-Kutta TVD (Total Variation Dimishing) for time discretization.

*Key-words:* Conservation Laws, Numerical Methods non-oscillatory, ENO, WENO and Runge-Kutta TVD

# Lista de Figuras

|      |                                                                                                                                           |    |
|------|-------------------------------------------------------------------------------------------------------------------------------------------|----|
| 1.1  | Tubo de gás . . . . .                                                                                                                     | 5  |
| 1.2  | Curvas características da equação de advecção, $a > 0$ . . . . .                                                                          | 13 |
| 1.3  | Curvas características da equação de advecção, $a < 0$ . . . . .                                                                          | 13 |
| 1.4  | Condição inicial suave $N = 100$ , $CFL = 0.6$ , $t = 0$ . . . . .                                                                        | 15 |
| 1.5  | Solução antes da formação do choque: $t = 0.2$ ; $N = 100$ e $CFL = 0.6$ . . . . .                                                        | 15 |
| 1.6  | Solução na formação de choque: $t = \frac{1}{\pi}$ , $N = 100$ , $CFL = 0.6$ . . . . .                                                    | 15 |
| 1.7  | Região aberta $V \subset \mathbb{R} \times (0, +\infty)$ para condição de Rankine-Hugoniot. . . . .                                       | 17 |
| 1.8  | Caso particular para fronteira da condição de Rankine-Hugoniot . . . . .                                                                  | 18 |
| 1.9  | Condição inicial do Problema de Riemann . . . . .                                                                                         | 22 |
| 1.10 | Curva característica e solução da equação de advecção. . . . .                                                                            | 22 |
| 1.11 | Formação de choque através da colisão das curvas características. . . . .                                                                 | 24 |
| 1.12 | Formação de choque - equação 1.68 . . . . .                                                                                               | 25 |
| 1.13 | Leque de rarefação . . . . .                                                                                                              | 26 |
| 1.14 | Curvas características - Equação 1.71 . . . . .                                                                                           | 27 |
| 1.15 | Construção da solução que satisfaz a condição de entropia . . . . .                                                                       | 27 |
| 1.16 | Solução entrópica para equação de Burgers . . . . .                                                                                       | 28 |
| 1.17 | Solução não entrópica para equação de Burgers . . . . .                                                                                   | 28 |
| 2.1  | Diagrama para condição CFL da aproximação $U_j^{n+1}$ . . . . .                                                                           | 33 |
| 2.2  | Triângulo computacional da condição CFL para um esquema de diferenças finitas de três pontos, ilustrado no diagrama 2.1. . . . .          | 33 |
| 2.3  | Diagrama para o Método de Lax-Friedrichs . . . . .                                                                                        | 36 |
| 2.4  | Diagrama do método ‘upwind’ quando $a > 0$ . . . . .                                                                                      | 40 |
| 2.5  | Método de Lax-Friedrichs: no domínio $[0, 1]$ , $\Delta x = 0.01$ , $t = 0.5$ , $\Delta t = 0.02$ e $CFL = 0.2$ . . . . .                 | 41 |
| 2.6  | Método ‘Upwind’: no domínio $[0, 1]$ , $\Delta x = 0.01$ e $t = 0.5$ , $\Delta t = 0.02$ , $CFL = 0.2$ . . . . .                          | 41 |
| 2.7  | Ilustração do erro pontual do método de Lax-Friedrichs para $N = 100$ no intervalo $[0, 1]$ . . . . .                                     | 41 |
| 2.8  | Ilustração do erro pontual do método ‘upwind’ para $N = 100$ no intervalo $[0, 1]$ . . . . .                                              | 41 |
| 2.9  | ‘Upwind’ não conservativo: no domínio $[0, 1]$ com $\Delta t/\Delta x = 0.5$ , $\Delta x = 0.01$ ; $t = 1$ , $\Delta t = 0.005$ . . . . . | 45 |
| 2.10 | ‘Upwind’ conservativo: no domínio $[0, 1]$ com $\Delta t/\Delta x = 0.5$ , $\Delta x = 0.01$ ; $t = 1$ , $\Delta t = 0.005$ . . . . .     | 45 |

|      |                                                                                                                                                                                     |    |
|------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 2.11 | Ilustração do erro pontual do método ‘upwind’ não conservativo no intervalo $[0, 1]$ .                                                                                              | 45 |
| 2.12 | Ilustração do erro pontual para método ‘upwind’ conservativo no intervalo $[0, 1]$ .                                                                                                | 45 |
| 3.1  | Aproximação ENO: fluxo ‘splitting’ Lax-Friedrichs . . . . .                                                                                                                         | 61 |
| 3.2  | Malha numérica apresentando os primeiros e os últimos pontos, da malhas numéricas inteira e metade, para ilustrar a condição de fronteira do método de diferenças finitas. . . . .  | 62 |
| 3.3  | Aproximações à esquerda e à direita das fronteiras das células. . . . .                                                                                                             | 66 |
| 3.4  | Ilustração da diferença entre o índice $i$ e o índice $j$ . . . . .                                                                                                                 | 66 |
| 3.5  | Interpolação de grau 3 com $r = 0$ e $\Delta x = 0.01$ , no intervalo $[0, 1]$ . . . . .                                                                                            | 71 |
| 3.6  | Interpolação de grau 3 com $r = 1$ e $\Delta x = 0.01$ , no intervalo $[0, 1]$ . . . . .                                                                                            | 71 |
| 3.7  | Interpolação de grau 3 com $r = 2$ e $\Delta x = 0.01$ , no intervalo $[0, 1]$ . . . . .                                                                                            | 71 |
| 3.8  | Interpolação de grau 3 com $r = 3$ e $\Delta x = 0.01$ , no intervalo $[0, 1]$ . . . . .                                                                                            | 71 |
| 3.9  | Aproximação ENO de grau 3 com $\Delta x = 0.01$ , no intervalo $[0, 1]$ . . . . .                                                                                                   | 77 |
| 3.10 | Runge-Kutta TVD de 3ª ordem para EDO’s . . . . .                                                                                                                                    | 87 |
| 3.11 | Ilustração do erro pontual do método de Runge-Kutta de 3ª ordem para $N = 10$ no intervalo $[0, 1.2]$ . . . . .                                                                     | 88 |
| 4.1  | Aproximação do exemplo 4.1: Runge-Kutta de 3ª ordem no domínio $[-0.1, 1]$ com $\Delta x = 0.0055$ , $t = 0.5$ , $\Delta t = 0.00275$ e $CFL = \Delta t / \Delta x = 0.5$ . . . . . | 90 |
| 4.2  | Aproximação do exemplo 4.1: Runge-Kutta de 2ª ordem no domínio $[-0.1, 1]$ com $\Delta x = 0.0055$ , $t = 0.5$ , $\Delta t = 0.00275$ e $CFL = \Delta t / \Delta x = 0.5$ . . . . . | 90 |
| 4.3  | Ilustração do erro pontual para a aproximação ilustrada na figura 4.1. . . . .                                                                                                      | 90 |
| 4.4  | Ilustração do erro pontual para a aproximação ilustrada na figura 4.2. . . . .                                                                                                      | 90 |
| 4.5  | Aproximação do exemplo 4.2: Runge-Kutta de 3ª ordem no domínio $[-0.1, 1.6]$ , $\Delta x = 0.0085$ e $t = 1$ , $\Delta t = 0.00425$ , $CFL = 0.6$ . . . . .                         | 92 |
| 4.6  | Aproximação do exemplo 4.2: Runge-Kutta de 2ª ordem no domínio $[-0.1, 1.6]$ , $\Delta x = 0.0085$ e $t = 1$ , $\Delta t = 0.00425$ , $CFL = 0.6$ . . . . .                         | 92 |
| 4.7  | Erro pontual da aproximação ilustrada na figura 4.5. . . . .                                                                                                                        | 93 |
| 4.8  | Erro pontual da aproximação ilustrada na figura 4.6. . . . .                                                                                                                        | 93 |
| 4.9  | Aproximação Runge-Kutta não TVD de 2ª ordem: $[-1, 1.6]$ , $\Delta x = 0.026$ e $t = 1$ , $\Delta t = 0.01$ , $CFL = 0.3846$ . . . . .                                              | 94 |
| 4.10 | Aproximação Runge-Kutta de terceira ordem para rarefação : $[-0.2, 1.5]$ , $\Delta x = 0.017$ e $t = 0$ , $CFL = 0.5$ . . . . .                                                     | 95 |
| 4.11 | Solução do exemplo 4.3: no domínio $[-0.2, 1.5]$ , $\Delta x = 0.017$ e $t = 0.5$ , $\Delta t = 0.0085$ , $CFL = 0.5$ . . . . .                                                     | 95 |
| 4.12 | Aproximação Runge-Kutta de terceira ordem para rarefação: $[-0.2, 1.5]$ , $\Delta x = 0.017$ e $t = 1$ , $\Delta t = 0.0085$ , $CFL = 0.5$ . . . . .                                | 95 |
| 4.13 | Comparação da aproximação numérica, formando o leque de rarefação para $t = 0, 0.5, 1, 5$ : $[-0.2, 1.5]$ , $\Delta x = 0.017$ , $\Delta t = 0.0085$ , $CFL = 0.5$ . . . . .        | 95 |
| 4.14 | Erro pontual da aproximação ilustrada na figura 4.10. . . . .                                                                                                                       | 95 |

|      |                                                                                                               |     |
|------|---------------------------------------------------------------------------------------------------------------|-----|
| 4.15 | Erro pontual da aproximação ilustrada na figura 4.12. . . . .                                                 | 95  |
| A.1  | Possibilidades de estênceis para o interpolador ENO de ordem 3. . . . .                                       | 98  |
| A.2  | Estêncil do polinômio interpolador de Lagrange para cálculo das constantes $d_r$ ;<br>$r = 0, 1, 2$ . . . . . | 100 |

# Lista de Tabelas

|     |                                                                                                                      |     |
|-----|----------------------------------------------------------------------------------------------------------------------|-----|
| 2.1 | Tabela de erros para equação de advecção, na norma $L_1$ , com $N = 100$ . . . . .                                   | 42  |
| 2.2 | Tabela de erro para equação de Burgers, na norma $L_1$ , com $N = 100$ . . . . .                                     | 46  |
| 3.1 | Constantes $c_{ri}$ referente a Eq. (3.46) . . . . .                                                                 | 72  |
| 3.2 | Constantes $\alpha_{ik}$ para Runge-Kutta de 3ª ordem . . . . .                                                      | 86  |
| 3.3 | Constantes $\beta_{ik}$ para Runge-Kutta de 3ª ordem . . . . .                                                       | 86  |
| 3.4 | Constantes $\lambda_i$ para Runge-Kutta de 3ª ordem . . . . .                                                        | 87  |
| 4.1 | Tabela de erros para equação de advecção, na norma $L_1$ , com ENO de 4ª ordem.                                      | 91  |
| 4.2 | Tabela de erros para equação de Burgers, na norma $L_1$ , com ENO de 4ª ordem.                                       | 92  |
| 4.3 | Tabela de erros na norma $L_1$ , para a onda de rarefação, com ENO de 4ª ordem<br>e Runge-Kutta de 3ª ordem. . . . . | 94  |
| A.1 | Constantes $d_r$ referentes a Eq. (A.5) . . . . .                                                                    | 101 |

# Lista de Símbolos

|                                 |                                                                       |
|---------------------------------|-----------------------------------------------------------------------|
| $x$                             | variável no espaço;                                                   |
| $t$                             | variável no tempo;                                                    |
| $\Delta x$                      | espaçamento no eixo $x$ ;                                             |
| $\Delta t$                      | espaçamento no eixo $t$ ;                                             |
| $f(u)$                          | função de fluxo da EDP;                                               |
| $F$                             | função de fluxo numérica;                                             |
| $\frac{\partial u}{\partial x}$ | derivada parcial de $u$ em relação a $x$ ;                            |
| $\frac{\partial u}{\partial t}$ | derivada parcial de $u$ em relação a $t$ ;                            |
| $u_0$                           | condição inicial da lei de conservação;                               |
| $u$                             | solução da lei de conservação;                                        |
| $x(t)$                          | curvas características;                                               |
| $a$                             | coeficiente constante da equação de advecção;                         |
| $C_0^1$                         | espaço das funções continuamente diferenciáveis com suporte compacto; |
| $\Phi$                          | função teste de classe $C_0^1$ ;                                      |
| $\vec{n}$                       | vetor normal;                                                         |
| $u_l$                           | parte da condição inicial do problema de Riemann;                     |
| $u_r$                           | parte da condição inicial do problema de Riemann;                     |
| $\nabla G$                      | divergente do campo $G$ ;                                             |
| $s$                             | velocidade com a qual viaja a descontinuidade da lei de conservação;  |
| $\theta$                        | coeficiente angular das retas características;                        |
| $x_j$                           | pontos discretos da malha numérica no eixo $x$ ;                      |
| $t_n$                           | pontos discretos da malha numérica no eixo $t$ ;                      |
| $I_j$                           | célula;                                                               |
| $x_{j+\frac{1}{2}}$             | extremo superior da célula $I_j$ ;                                    |
| $\bar{u}_j$                     | aproximação média da célula $I_j$ ;                                   |
| $U_j^n$                         | aproximação numérica da função $u$ da célula $j$ no tempo $n$ ;       |
| $U_{\Delta t}$                  | função constante por partes;                                          |
| $\alpha_{CFL}$                  | número CFL;                                                           |
| $D$                             | domínio de dependência da EDP;                                        |
| $D_{\Delta t}$                  | domínio de dependência numérico;                                      |
| $\bar{E}_j^n$                   | erro pontual;                                                         |
| $E_{\Delta t}$                  | função erro;                                                          |
| $\ \cdot\ _1$                   | norma $L_1$ ;                                                         |
| $\ \cdot\ _2$                   | norma $L_2$ ;                                                         |
| $\ U^n\ _1$                     | norma $L_1$ para dados discretos;                                     |
| $\ \cdot\ _{1,T}$               | norma- $T$ no intervalo de tempo $[0, T]$ ;                           |

|                       |                                                                                         |
|-----------------------|-----------------------------------------------------------------------------------------|
| $L_{\Delta t}$        | operador erro de truncamento local;                                                     |
| $L$                   | constante Lipschitz;                                                                    |
| $W$                   | conjunto das soluções fracas;                                                           |
| $\Delta x_j$          | tamanho da célula $I_j$ ;                                                               |
| $v$                   | função suave;                                                                           |
| $p$                   | polinômio para aproximar $v$ ;                                                          |
| $S$                   | estêncil escrito em termos das células $I_j$ ;                                          |
| $\overline{S}$        | estêncil escrito em termos dos pontos $x_j$ ;                                           |
| $k$                   | ordem de precisão do polinômio $p$ ;                                                    |
| $v_{j+\frac{1}{2}}^e$ | aproximação na fronteira da célula $I_j = [x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}]$ ;     |
| $v_{j+\frac{1}{2}}^d$ | aproximação na fronteira da célula $I_{j+1} = [x_{j+\frac{1}{2}}, x_{j+\frac{3}{2}}]$ ; |
| $c_{ri}$              | constantes do polinômio no estêncil fixo;                                               |
| $V$                   | função primitiva da $v$ ;                                                               |
| $P$                   | polinômio que interpola $V$ ;                                                           |
| $V[.]$                | diferenças divididas;                                                                   |
| $O$                   | ordem do erro do método numérico;                                                       |
| $w_r$                 | coeficientes não-lineares do WENO;                                                      |
| $d_r$                 | coeficientes lineares do WENO;                                                          |
| $\beta_r$             | coeficientes do WENO;                                                                   |
| $\alpha_r$            | coeficientes do WENO;                                                                   |
| $h$                   | aproximação da função de fluxo $f$ ;                                                    |
| $H$                   | primitiva da função $h$ ;                                                               |
| $\alpha$              | coeficiente do método de Lax-Friedrichs;                                                |
| $f^+; f^-$            | componentes do fluxo ‘splitting’;                                                       |
| $L(u)$                | operador do sistema de EDO’s;                                                           |
| $c$                   | número CFL do método Runge-Kutta;                                                       |
| $\alpha_{ik}$         | constantes do método Runge-Kutta;                                                       |
| $\beta_{ik}$          | constantes do método Runge-Kutta;                                                       |
| $\lambda_{ik}$        | constantes do método Runge-Kutta;                                                       |

# Sumário

|                                                                                   |           |
|-----------------------------------------------------------------------------------|-----------|
| Resumo                                                                            | vii       |
| Abstract                                                                          | viii      |
| Lista de Figuras                                                                  | viii      |
| Lista de Tabelas                                                                  | xi        |
| Lista de Símbolos                                                                 | xiii      |
| Introdução                                                                        | 1         |
| <b>1 Leis de Conservação Unidimensionais</b>                                      | <b>4</b>  |
| 1.1 Forma Integral e Forma Diferencial da Lei de Conservação Unidimensional . . . | 4         |
| 1.2 Equações Características . . . . .                                            | 8         |
| 1.3 Formação de Choques . . . . .                                                 | 14        |
| 1.3.1 Solução Fraca . . . . .                                                     | 15        |
| 1.3.2 Condição de Rankine-Hugoniot . . . . .                                      | 17        |
| 1.4 Problema de Riemann . . . . .                                                 | 21        |
| <b>2 Métodos Numéricos</b>                                                        | <b>29</b> |
| 2.1 Métodos Numéricos para Equações Lineares . . . . .                            | 29        |
| 2.1.1 Condição CFL . . . . .                                                      | 31        |
| 2.1.2 Domínio de Dependência . . . . .                                            | 32        |
| 2.1.3 Convergência . . . . .                                                      | 34        |
| 2.1.4 Estabilidade e Consistência . . . . .                                       | 35        |
| 2.1.5 Método de Lax - Friedrichs para Equação de Advecção . . . . .               | 36        |
| 2.1.6 Método Upwind . . . . .                                                     | 40        |
| 2.2 Métodos Numéricos Conservativos . . . . .                                     | 42        |
| 2.3 Métodos Numéricos para Equações Não-Lineares . . . . .                        | 46        |
| 2.3.1 Estabilidade Não-Linear . . . . .                                           | 46        |
| 2.3.2 Convergência para Solução de Entropia . . . . .                             | 54        |

|          |                                                                                                  |            |
|----------|--------------------------------------------------------------------------------------------------|------------|
| <b>3</b> | <b>Discretização Espacial e Temporal</b>                                                         | <b>57</b>  |
| 3.1      | Esquemas ENO via Diferenças Finitas para Leis de Conservação Escalares Unidimensionais . . . . . | 58         |
| 3.2      | Reconstrução Unidimensional . . . . .                                                            | 62         |
| 3.2.1    | Reconstrução via Função Primitiva . . . . .                                                      | 63         |
| 3.3      | Diferenças Divididas . . . . .                                                                   | 72         |
| 3.4      | ENO . . . . .                                                                                    | 75         |
| 3.5      | Esquemas ENO para Leis de Conservação Unidimensionais . . . . .                                  | 79         |
| 3.5.1    | Reconstrução da Função Fluxo Numérica . . . . .                                                  | 79         |
| 3.6      | Métodos Runge-Kutta TVD . . . . .                                                                | 81         |
| <b>4</b> | <b>Resultados Numéricos e Conclusões</b>                                                         | <b>89</b>  |
| 4.1      | Resultados Numéricos . . . . .                                                                   | 89         |
| 4.2      | Conclusões . . . . .                                                                             | 96         |
|          | <b>Apêndices</b>                                                                                 | <b>97</b>  |
| <b>A</b> | <b>WENO</b>                                                                                      | <b>97</b>  |
| <b>B</b> | <b>Esquemas ENO e WENO via Volume Finito para Leis de Conservação Escalares</b>                  | <b>108</b> |
|          | <b>Referências Bibliográficas</b>                                                                | <b>110</b> |
|          | <b>Índice Remissivo</b>                                                                          | <b>113</b> |

# Introdução

Leis de conservação são equações diferenciais parciais (EDP's) de primeira ordem que avaliam a mudança no valor total de uma entidade (substância) física, contidos em qualquer região do espaço, devido ao fluxo da entidade física nas fronteiras da região. As leis de conservação modelam vários fenômenos físicos, tais como, dinâmica dos gases, acústica, ótica, geofísica, biomecânica, aerodinâmica, astrofísica e engenharia de tráfego. Veja Thoe (1986) [32], Cheng e Shu [2], Leveque (2004) [18] e De Azevedo e Eschenazi (2004) [7].

A solução da lei de conservação frequentemente desenvolve descontinuidades conhecidas como choques ou ondas de rarefação. A captura destes choques é uma tarefa difícil de ser realizada pelos métodos numéricos. Desta forma, o estudo das equações mais simples como a equação de advecção e a equação de Burgers, têm um papel importante tanto no entendimento das propriedades das leis de conservação quanto no desenvolvimento dos métodos numéricos que aproximam suas soluções.

Existem vários métodos numéricos convergentes que podem ser aplicados às leis de conservação, porém a convergência destes métodos são de baixa ordem e não apresentam bons resultados na região onde ocorre a formação do choque.

Na busca de se obter métodos numéricos de ordem mais elevada e que não possuam oscilações na região de choque, Harten *et al* (1987) [13] desenvolveu o método numérico Essencialmente Não-Oscilatórios (do inglês Essentially Nonoscillatory - ENO). Esse método é baseado nas técnicas de interpolação polinomial e no método de reconstrução desenvolvido por Colella e Woodwardem (1984) [3]. Harten *et al* uniu esses métodos e adicionou a busca do melhor estêncil, dentro da malha numérica, através do menor valor absoluto das respectivas diferenças divididas, obtendo um método numérico não-oscilatório de alta ordem pertencente à classe dos métodos numéricos de Variação Total Decrescente (do inglês Total Variation diminishing - TVD). A propriedade TVD é importante na construção de métodos numéricos não oscilatórios, pois a variação total da função diminui quando evolui o tempo. O esquema ENO apresentou grandes avanços em relação aos métodos numéricos tradicionais que aproximam as leis de conservação, apresentando resultados de alta ordem.

O esquema numérico Essencialmente Não-Oscilatório Ponderado (do inglês Essentially Non-

oscillatory Weighted - WENO) foi desenvolvido por Liu, Osher e Chan (1994) [22] usando a combinação convexa de todos os possíveis polinômios, obtidos na busca realizada pelo método ENO dentro do estêncil, resultando em um método numérico que apresenta, na região onde a solução é suave, resultados de ordem mais elevada do que a ordem do interpolador ENO e preserva a ordem deste interpolador na região onde a solução é descontínua. O sucesso do WENO está relacionado aos pesos da combinação convexa. Depois disto, Jiang e Shu (1996) [15] introduziram, no método numérico WENO, novos pesos ( $\beta_r$ ) que melhoraram ainda mais a ordem de precisão. Estes coeficientes  $\beta_r$  são chamados de sensores de suavidade, ou seja, são capazes de detectar a regularidade da solução numérica nos pontos de um determinado estêncil  $S_r$ .

Para o cálculo da aproximação numérica da função de fluxo  $f$  será considerado a teoria dos métodos de alta resolução (métodos que possuem no mínimo segunda ordem de precisão em regiões suaves e não possuem oscilação em regiões de descontinuidades) através do fluxo ‘splitting’ de Lax-Friedrichs.

Frequentemente é utilizado os métodos semi-discretos (‘*method of lines*’) para discretização do domínio (Shu (1988) [30]). Estes métodos consideram a discretização em dois estágios, primeiro a discretização é feita somente no espaço considerando o problema contínuo no tempo, este processo transforma a EDP em um sistema de EDO’s no tempo, chamadas de equações semi-discretas. Logo após faz-se a discretização no tempo utilizando um método numérico padrão para sistemas de EDO’s. Esta técnica é utilizada, particularmente, no desenvolvimento de métodos de ordem maior que dois. Pode-se definir alta ordem de aproximação para a função fluxo nas fronteiras das células, no instante de tempo, utilizando ENO ou WENO para interpolação no espaço. Para encontrar alta ordem de precisão no tempo será utilizado o método numérico Runge-Kutta pertencente a classe dos métodos TVD, o que nos garante que o método não será oscilatório. (Leveque (1992) [17]).

Nesta dissertação será considerado o estudo dos métodos numéricos ENO e WENO em malhas estruturadas, igualmente espaçadas, devido a maior facilidade tanto do estudo analítico do método como de sua implementação numérica.

O objetivo maior deste trabalho é estudar a aplicação dos métodos numéricos não-oscilatórios ENO e WENO às leis de conservação hiperbólicas escalares unidimensionais.

Essa dissertação está dividida em quatro capítulos e dois apêndices.

No capítulo 1 será apresentada a teoria para se obter as soluções analíticas das equações lineares e quase-lineares de acordo com Evans (1998) [9], Leveque (1992) [17], Leveque (2004) [18], Smoller (1994) [28], Thoe e Zachmanoglou (1986) [32], Logan (1994) [24] e Marotti (2008) [25].

No capítulo 2, será feito um estudo da teoria dos métodos numéricos e suas propriedades baseado nas técnicas de Diferenças Finitas (DF) e de Volumes Finitos (VF), analisando a convergência desses métodos aplicados a leis de conservação. Será considerado como objeto de estudo:

1. os métodos numéricos conservativos;
2. as funções de fluxo numérico de Lax - Friedrichs;
3. os métodos numéricos monótonos e TVD;
4. os métodos ‘upwind’;
5. função de variação total;
6. os teoremas necessários para convergência dos métodos;
7. o teorema de Lax - Wendroff,

de acordo com Leveque (1992) [17], Leveque (2004) [18], Rezzolla (2004) [26] e Marotti (2008) [25].

No capítulo 3 será apresentada a discussão teórica do esquema numérico ENO, da aproximação (baseada nos esquemas ENO e WENO) para leis de conservação escalares unidimensionais via diferenças finitas, de acordo com Shu (1997) [29] e Krylov (2005) [16]. Também é objeto de estudo deste capítulo os métodos numéricos de Runge-Kutta TVD para aproximação no tempo, de acordo com Butcher (1987) [1], Gottlieb e Shu (1998) [10], Shu e Osher (1998) [30] e Suresh e Huynh (1997) [31].

No capítulo 4 serão apresentados os resultados obtidos pelo método numérico Runge-Kutta de segunda e de terceira ordem (apresentado no capítulo 3), para a equação de advecção e para a equação de Burgers (estudadas no capítulo 1); as conclusões e as perspectivas futuras.

No apêndice A será apresentado o esquema numérico WENO, de acordo com Shu (1997) [29], Liu; Osher e Chan (1994) [22] e Jiang e Shu (1996) [15]. No apêndice B está exposto o esquema via volumes finitos para lei de conservação, através dos esquema ENO e WENO, de acordo com Shu (1997) [29].

# Capítulo 1

## Leis de Conservação Unidimensionais

Nesse capítulo será apresentado:

- A dedução das leis de conservação, de acordo com seus princípios físicos;
- O desenvolvimento das soluções analíticas;
- As propriedades físicas e matemáticas referente à solução analítica.

Esse estudo será iniciado com uma equação mais simples do tipo linear. Logo após será analisada uma equação quase-linear, explorando uma solução analítica a qual deve obedecer à condição de entropia. Para equações não lineares existem dificuldades adicionais no cálculo das soluções analíticas em todo o domínio. Por exemplo, a introdução de efeitos não lineares na lei de conservação pode ocasionar a formação de choque, ilustrado na Figura 1.11, ou a formação de leques de rarefação, ilustrado na Figura 1.14, que são fenômenos a serem considerados na solução da equação. Tais fenômenos também aparecem quando a condição inicial é descontínua, ver Eq. (1.55). Assim surge a necessidade de se estudar novas técnicas matemáticas para se obter soluções analíticas, em todo o domínio, quando ocorrer tais fenômenos.

Será desenvolvida a teoria das equações características para obter a solução analítica da lei de conservação. Para os problemas que apresentam a formação de choque ou a formação de leque de rarefação se faz necessário utilizar a teoria das soluções fracas, cuja formulação também será exposta.

### 1.1 Forma Integral e Forma Diferencial da Lei de Conservação Unidimensional

Considerando um problema de dinâmica dos fluidos será feita a dedução da equação da lei de conservação na forma diferencial e na forma integral. Considera-se um tubo unidimensional, de fronteira impermeável, através do qual flui um gás com velocidade  $\phi(x, t)$  conhecida no ponto  $x$  e no instante  $t$ , Figura 1.1. Em problemas de dinâmica dos fluidos deve-se considerar a função

velocidade  $\phi(x, t)$  ( $\phi$  representa o movimento do fluido) variando apenas em função do tempo  $t$  e da distância percorrida ao longo do tubo.

Sendo  $\phi(x, t)$  uma função conhecida deseja-se, então, apenas modelar a concentração ou densidade de alguma espécie química presente no gás. Será denotada por  $u(x, t)$  a densidade (ou concentração) do traçador químico (veja observação 1.1) na posição  $x$  e no tempo  $t$ . Para intervalos de tempo  $t \geq 0$  existe a densidade de massa  $u(x, t)$ , bem determinada tal que seja possível calcular a massa total  $m(I, t)$  no intervalo  $I = [x_1, x_2]$  no instante  $t$ .

**Observação 1.1** • *densidade* =  $\frac{\text{massa}}{\text{comprimento do intervalo}}$  logo  $m(x, t) = u(x, t)\Delta x$ .

- *Traçador é uma substância presente, no fluido em estudo, em pequenas concentrações de modo que sua magnitude não afeta a concentração da dinâmica do fluido. (Ver Leveque (2004) [18]).*

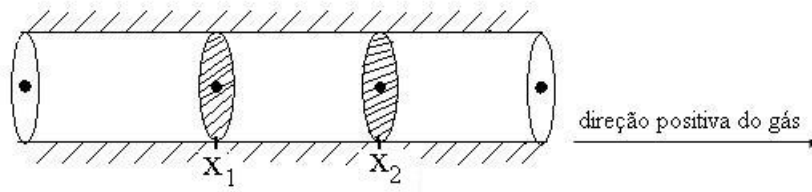


Figura 1.1: Tubo de gás

Considere uma partição do intervalo  $[x_1, x_2]$ , em  $n$  subintervalos de comprimento  $\Delta x = \frac{x_2 - x_1}{n}$ . Para o caso discreto obtém-se

$$m(I, t) = \sum_{i=0}^n u(x_i, t) \cdot \Delta x \quad (1.1)$$

fazendo  $\Delta x \rightarrow 0$  tem-se o caso contínuo

$$m(I, t) = \int_{x_1}^{x_2} u(x, t) dx, \quad (1.2)$$

pois o somatório na Eq. (1.1) aproxima a integral de Riemann.

Caso se estude uma substância que não é criada ou extinta na seção  $x_1 < x < x_2$  e considerando que a integral na Eq. (1.2) varia com o tempo, então a massa total dentro da seção pode mudar apenas devido ao fluxo de partículas através dos extremos da seção  $x_1$  e  $x_2$ .

**Definição 1.1** O *fluxo de massa no ponto*  $(x, t)$  é definido por

$$u(x, t) \cdot \phi(x, t) = \mathbb{F}(x, t). \quad (1.3)$$

Por convenção  $\mathbb{F}(x, t) > 0$  corresponde ao fluxo da esquerda para a direita e  $\mathbb{F}(x, t) < 0$  corresponde ao fluxo da direita para a esquerda. Para esta construção será considerado  $\mathbb{F}(x, t) > 0$ . A taxa de variação de massa no intervalo  $I = [x_1, x_2]$  se mede pela diferença dos fluxos em  $x_1$  e  $x_2$  no instante  $t$ . Pela Eq. (1.2) tem-se

$$\frac{d}{dt}m(I, t) = \frac{d}{dt} \int_{x_1}^{x_2} u(x, t) dx = \mathbb{F}(x_1, t) - \mathbb{F}(x_2, t). \quad (1.4)$$

Pela definição 1.1 tem-se

$$\frac{d}{dt} \int_{x_1}^{x_2} u(x, t) dx = u(x_1, t) \cdot \phi(x_1, t) - u(x_2, t) \cdot \phi(x_2, t). \quad (1.5)$$

Portanto, integrando em  $t$  a Eq. (1.5) obtém-se

$$\int_{t_1}^{t_2} \frac{d}{dt} \int_{x_1}^{x_2} u(x, t) dx dt = \int_{t_1}^{t_2} (u(x_1, t) \cdot \phi(x_1, t) - u(x_2, t) \cdot \phi(x_2, t)) dt,$$

trocando os limites de integração obtemos

$$\int_{x_1}^{x_2} (u(x, t_2) - u(x, t_1)) dx = \int_{t_1}^{t_2} (u(x_1, t) \cdot \phi(x_1, t) - u(x_2, t) \cdot \phi(x_2, t)) dt. \quad (1.6)$$

Se  $u, \phi$  são suficientemente diferenciáveis, a Eq. (1.6) pode ser reescrita como

$$\int_{x_1}^{x_2} \int_{t_1}^{t_2} \frac{\partial u(x, t)}{\partial t} dt dx = - \int_{t_1}^{t_2} \int_{x_1}^{x_2} \frac{\partial (u(x, t) \phi(x, t))}{\partial x} dx dt. \quad (1.7)$$

Assim, invertendo a ordem das integrais em relação ao tempo e ao espaço a Eq. (1.7) se reescreve como

$$\int_{t_1}^{t_2} \int_{x_1}^{x_2} \left[ \frac{\partial u(x, t)}{\partial t} + \frac{\partial (u(x, t) \phi(x, t))}{\partial x} \right] dx dt = 0. \quad (1.8)$$

Como  $[x_1, x_2]$  e  $[t_1, t_2]$  são intervalos quaisquer, pode-se concluir que o integrando, na Eq. (1.8), deve ser identicamente nulo, ou seja

$$\frac{\partial u(x, t)}{\partial t} + \frac{\partial (u(x, t) \phi(x, t))}{\partial x} = 0. \quad (1.9)$$

Sabendo que  $\phi(x, t)$  é uma função conhecida, pode-se reescrever a função fluxo  $\mathbb{F}$  na definição 1.1 como  $\mathbb{F}(x, t) = f(u, x, t)$ . Substituindo nas Eqs. (1.9) e (1.4) obtém-se, respectivamente,

$$\frac{\partial u(x, t)}{\partial t} + \frac{\partial f(u(x, t))}{\partial x} = 0 \quad (1.10)$$

e

$$\frac{d}{dt} \int_{x_1}^{x_2} u(x, t) dx = f(u(x_1, t)) - f(u(x_2, t)). \quad (1.11)$$

As Eqs. (1.10) e (1.11) são chamadas de **forma diferencial conservativa** e **forma integral**, respectivamente, da lei de conservação.

Quando as derivadas parciais são denotadas por subscritos, a Eq. (1.10) se reescreve como

$$u_t + (f(u))_x = 0. \quad (1.12)$$

Sendo  $f$  uma função suave, pode-se reescrever a Eq. (1.12) como

$$u_t + f'(u)(u)_x = 0. \quad (1.13)$$

Essa equação é conhecida como **forma diferencial não-conservativa** da lei de conservação.

Nesse capítulo serão estudadas duas leis de conservação escalares da forma Eq. (1.12), sendo  $x \in \mathbb{R}$  e  $f : \mathbb{R} \rightarrow \mathbb{R}$  uma função linear no primeiro caso e não-linear no segundo caso. Tais equações são EDP's hiperbólicas de uma incógnita  $u = u(x, t)$ , ou seja, leis de conservação unidimensionais.

O estudo das leis de conservação se inicia com o caso mais simples, quando a EDP é linear com coeficiente constante. Quando a velocidade é constante, não depende de  $x$  ou de  $t$ , a função fluxo é dada por  $\mathbb{F}(x, t) = f(u) = a.u$ , sendo  $a$  constante. Substituindo essa função fluxo na Eq. (1.12), obtém-se a **equação de advecção** dada por

$$u_t + a u_x = 0. \quad (1.14)$$

A equação Eq. (1.14) possui esse nome porque modela a advecção (veja observação 1.2) de um traçador ao longo da corrente fluída.

**Observação 1.2** *Na engenharia mecânica ou na engenharia química, advecção é um mecanismo de transporte de uma substância, ou uma propriedade conservada, com um fluído em movimento. (Ver Leveque (2004) [18]).*

Logo após será considerado como objeto de estudo a equação de Burgers na sua forma conservativa

$$u_t + \left( \frac{u^2}{2} \right)_x = 0, \quad (1.15)$$

ou na sua forma não-conservativa

$$u_t + uu_x = 0, \quad (1.16)$$

tal que  $f(u)$  é uma função escalar não-linear dada por  $f(u) = \frac{u^2}{2}$ .

**Observação 1.3** *A equação de Burgers pode ser vista como um protótipo unidimensional da equação de Euler e, neste caso, a incógnita  $u = u(x, t)$  é interpretada como sendo a velocidade unidimensional de uma partícula de fluido que passa pela posição  $x$  no instante  $t$ . Do ponto de vista da mecânica Lagrangeana, a equação de Burgers descreve o fato de que num escoamento incompressível e sem viscosidade a velocidade de uma partícula fixada se mantém constante ao longo do caminho percorrido por ela no campo de escoamento. Veja Marotti (2008) [25].*

## 1.2 Equações Características

Uma lei de conservação associada a uma condição inicial compõem o que chamamos de **problema de valor inicial** ou **problema de Cauchy**. Para o estudo das equações características será considerado o problema de Cauchy unidimensional composto pela lei de conservação unidimensional na forma não conservativa

$$u_t + f'(u)(u)_x = 0, \quad (1.17)$$

no domínio  $-\infty < x < +\infty$ ,  $t \geq 0$ , acrescido de uma condição inicial

$$u(x, 0) = u(x_0) = u_0(x). \quad (1.18)$$

As **características** são curvas que transportam informações da equação diferencial parcial no plano  $(x, t)$ . E essas curvas são importantes no estudo de problemas de valor inicial pois permitem um tratamento mais simples das equações, mesmo para equações não-lineares, que em geral são de tratamento difícil. De acordo com Thoe e Zachmanoglou (1986) [32] e Evans (1998) [9], o método das equações características consiste em transformar a EDP em um sistema de equações diferenciais ordinárias (EDO's). Dada a solução  $u$  no ponto inicial  $x_0$ , utilizando o método das equações características, deseja-se calcular a solução analítica em todo o domínio.

Nesta seção será apresentada a teoria das equações características para equações diferenciais parciais quase-lineares de primeira ordem. De maneira geral estas equações se escrevem como

$$P(x, y, z)z_x + Q(x, y, z)z_y = R(x, y, z), \quad (1.19)$$

sendo  $P, Q, R$  de classe  $C^1$  em algum domínio  $\tilde{\Omega} \subset \mathbb{R}^3$  e não se anulam simultaneamente nos pontos de  $\tilde{\Omega}$ . A solução da equação quase-linear Eq. (1.19) é uma função definida implicita-

mente por  $z = g(x, y)$ , de classe  $C^1$  em  $(x, y) \in \Omega \subset \mathbb{R}^2$ , que satisfaz as seguintes condições:

1. Para todo ponto  $(x, y) \in \Omega$ , o ponto  $(x, y, g(x, y))$  pertence ao domínio  $\tilde{\Omega}$  das funções  $P, Q, R$ .
2. Quando  $z$  é substituída em Eq. (1.19) a equação resultante é uma identidade em  $x, y$  para todo  $(x, y) \in \Omega$ .

A solução  $z = g(x, y)$  pode ser vista como uma superfície  $\mathbb{S}$  em  $\mathbb{R}^3$ , sendo  $V = (P, Q, R)$  um campo de vetores.

Para que  $z$  seja solução da Eq. (1.19), o vetor normal a superfície  $\mathbb{S}$  deve ser ortogonal ao campo de vetores  $V$ . Por definição, o vetor normal à superfície dada por  $z = g(x, y)$  é o gradiente da função  $G = g(x, y) - z$ , logo  $\vec{n} = (z_x, z_y, -1)$ .

**Observação 1.4** *Temos a seguinte condição de ortogonalidade*

$$\begin{aligned} P(x, y, z)(z_x) + Q(x, y, z)(z_y) - R(x, y, z)(-1) &= 0 \\ \Leftrightarrow \langle V, \vec{n} \rangle &= 0 \\ \Leftrightarrow \langle (P, Q, R), (z_x, z_y, -1) \rangle &= 0. \end{aligned}$$

A solução da EDP quase-linear Eq. (1.19) está associada ao cálculo de soluções, ou superfícies integrais, do campo  $V$  da seguinte EDP

$$Pu_x + Qu_y + Ru_z = 0, \quad (1.20)$$

sendo

$$u = u(x, y, z), \quad \text{para todo } (x, y, z) \in \tilde{\Omega} \quad (1.21)$$

solução de Eq. (1.20).

De acordo com Thoe e Zachnomanoglou (1986) [32], a Eq. (1.20) pode ser resolvida para  $z$  em termos de  $x$  e  $y$  e assim a função  $z$  é solução da Eq. (1.19). Esse fato será mostrado no Lema 1.1.

**Lema 1.1** *Seja  $u$  uma função de classe  $C^1$  em  $\tilde{\Omega} \subset \mathbb{R}^3$  e suponha que, para todo ponto, a superfície integral Eq. (1.21) satisfaz as seguintes condições:*

1.  $Pu_x + Qu_y + Ru_z = 0, \quad \forall (x, y, z) \in \tilde{\Omega};$
2.  $u_z \neq 0.$

A Eq. (1.21) implica que  $g(x, y) = z$  e satisfaz a EDP (1.19). Ou seja,  $Pz_x + Qz_y = R$ .

**Demonstração:**

Por hipótese  $u_z \neq 0$ , então segue pelo teorema da função implícita (veja Lima (2006) [21]) que  $g(x, y) = z$ . Seja  $U(x, y) = u(x, y, g(x, y))$ ,  $\forall (x, y) \in \Omega$ . Calculando as derivadas parciais de  $U$  tem-se:

$$U_x = \frac{\partial u}{\partial x} + \frac{\partial u}{\partial z} z_x, \quad (1.22)$$

e

$$U_y = \frac{\partial u}{\partial y} + \frac{\partial u}{\partial z} z_y. \quad (1.23)$$

Substituindo as derivadas parciais no primeiro membro da Eq. (1.19), obtém-se

$$P(U_x - u_x) \frac{1}{u_z} + Q(U_y - u_y) \frac{1}{u_z} = \frac{1}{u_z} (PU_x + QU_y) - \frac{1}{u_z} (Pu_x + Qu_y) = -\frac{1}{u_z} (Ru_z) = R$$

sendo que na penúltima passagem foi usado o item (1) do lema. ■

O cálculo das soluções analíticas obtidas nesse trabalho usa a teoria das equações características, de acordo com a definição 1.2.

**Definição 1.2** Uma curva  $C$  em  $\tilde{\Omega} \subset \mathbb{R}^3$  é uma curva integral do campo  $V = (P, Q, R)$  se  $V$  é tangente a curva  $C$  em cada um de seus pontos. Define-se como equações características o sistema de EDO's associadas a  $V$ :

$$\begin{cases} \frac{dx}{dt} = P(x, y, z), \\ \frac{dy}{dt} = Q(x, y, z), \\ \frac{dz}{dt} = R(x, y, z), \end{cases} \quad (1.24)$$

com  $t \in [-\epsilon, +\epsilon]$ ,  $\epsilon > 0$ .

**Observação 1.5** Se  $P \neq 0$  (ou seja  $(\frac{dx}{dt} \neq 0)$ ) então o sistema dado em Eq. (1.24) pode ser reescrito eliminando a variável  $t$ . De fato,

se

$$\begin{cases} y = y(x(t)), \\ z = z(x(t)), \end{cases}$$

então

$$\begin{cases} y' = y_x \dot{x} = y_x P, \\ z' = z_x \dot{x} = z_x P, \end{cases}$$

de onde segue que

$$\begin{cases} y_x = \frac{Q}{P}, \\ z_x = \frac{R}{P}. \end{cases} \quad (1.25)$$

Uma construção análoga pode ser feita se  $Q \neq 0$  ou se  $R \neq 0$ . Note que o sistema dado na Eq. (1.24) é equivalente ao sistema da Eq. (1.25). Sendo  $Q = Q(x, y(x), z(x))$  e  $R = R(x, y(x), z(x))$ .

**Definição 1.3** Define-se como curva característica da EDP hiperbólica  $u_t + f'(u)(u)_x = 0$  a solução da EDO característica  $\frac{d}{dt}x(t) = f'(u(x(t)))$ .

Uma curva característica é uma curva ao longo da qual a equação diferencial parcial (1.17) se transforma em uma equação diferencial ordinária.

Se  $u = u(x, t)$  é solução da lei de conservação Eq. (1.17), então ao longo das curvas características  $x(t)$  tem-se que  $u$  é constante. De fato, seja  $u(x, t)$  uma solução da lei de conservação Eq. (1.17). Calculando as derivadas parciais de  $u$  obtém-se

$$\frac{d}{dt}u(x(t), t) = u_t(x(t), t) + u_x(x(t), t) \frac{dx}{dt},$$

ou seja

$$\frac{d}{dt}u(x(t), t) = u_t(x(t), t) + f'(u(x(t), t)) u_x(x(t), t).$$

Então

$$\frac{d}{dt}u(x(t), t) = 0, \tag{1.26}$$

o que conclui a afirmação.

O fato da solução  $u$  ser constante sobre as curvas características diz que estas curvas devem ser retas, independente da linearidade da EDP. A diferença é apenas a inclinação das retas: no caso linear a inclinação das retas é a mesma ao passo que para as equações não-lineares, o ângulo de inclinação varia. Para o caso linear veja as Figuras 1.2 e 1.3 (página 13) e para o caso não-linear veja as Figuras 1.11 e 1.13 (páginas 24 e 26, respectivamente).

Vamos agora calcular a solução analítica para o problema de valor inicial, considerando a equação de advecção

$$u_t + au_x = 0, \quad -\infty < x < +\infty, \quad t \geq 0, \tag{1.27}$$

$$u(x, 0) = u_0(x). \tag{1.28}$$

Fazendo uma analogia entre as Eqs. (1.20) e (1.27) tem-se

$$P = a, \quad Q = 1 \quad \text{e} \quad R = 0.$$

Então, resolver o problema de Cauchy dado pelas Eqs. (1.27) e (1.28) é equivalente a resolver os sistemas de equações características

$$\begin{cases} \frac{dx}{dt} = a, \\ x(0) = x_0, \end{cases} \quad (1.29)$$

sendo  $x_0$  uma constante, e

$$\begin{cases} \frac{du}{dt} = 0, \\ u(0) = u(x(0), 0) = u(x_0, 0) = u_0(x_0). \end{cases} \quad (1.30)$$

Integrando as EDO's em relação a  $t$  nas Eqs. (1.29) e (1.30) obtém-se

$$\begin{cases} x(t) = a.t + c_1, \\ u(x, t) = c_2, \end{cases} \quad (1.31)$$

sendo  $c_1$  e  $c_2$  constantes.

Das condições iniciais,  $u(0) = u_0(x_0)$ , e da Eq. (1.31) tem-se que

$$u(0) = u_0(x_0) = c_2. \quad (1.32)$$

Das Eqs. (1.29) e (1.31) tem-se que  $x_0 = c_1$ . Logo

$$x_0 = c_1 = x(t) - at. \quad (1.33)$$

Das Eqs. (1.31) e (1.32) tem-se que

$$u(x, t) = u_0(x - at), \quad (1.34)$$

é a solução analítica do problema de valor inicial composto pelas Eqs. (1.27) e (1.28), em todo o domínio. De fato,

$$\frac{\partial u}{\partial t} + a \frac{\partial u}{\partial x} = \frac{\partial u_0}{\partial t} + a \frac{\partial u_0}{\partial x} = \frac{du_0}{ds} \cdot \frac{ds}{dt} + a \frac{du_0}{ds} \cdot \frac{ds}{dx} = u'_0 \cdot (-a) + a \cdot u'_0 \cdot 1 = 0,$$

sendo  $s = x - at$ .

A equação de advecção definida na Eq. (1.14) é uma EDP linear, escalar, do tipo hiperbólica. Suas curvas características são retas cujas equações são dadas pela solução da equação característica  $\frac{dx}{dt} = a$ , isto é, as curvas características da equação de advecção são dadas por  $x(t) = at + x_0$ , no plano  $(x, t)$ , sendo que  $x_0$  faz parte da condição inicial. A Eq (1.26) diz que ao longo das curvas características a solução  $u$  é constante.

Da Eq. (1.33) tem-se que, no plano  $(x, t)$ , o coeficiente angular das curvas características é dado por

$$\tan \theta = \frac{1}{a},$$

e  $x_0$  é o ponto de interseção da curva característica com o eixo  $x$ . Como o coeficiente angular das curvas características não varia, as retas são paralelas. As curvas características da equação de advecção são ilustradas nas Figuras 1.2 e 1.3 considerando o sinal do coeficiente  $a$ .

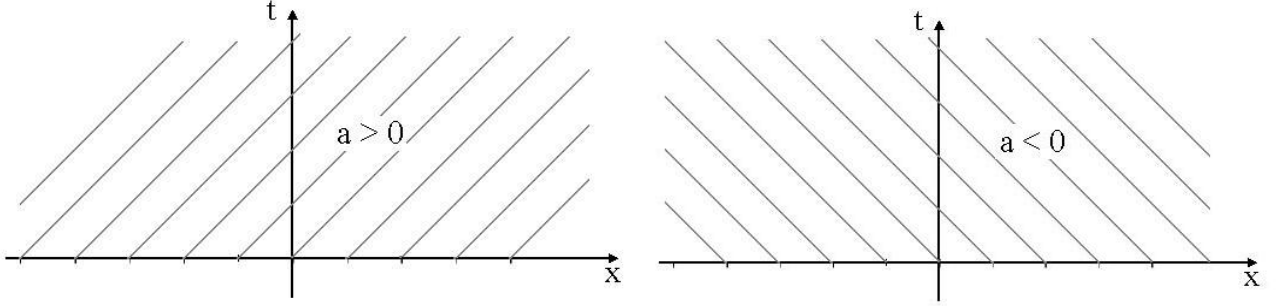


Figura 1.2: Curvas características da equação de advecção,  $a > 0$       Figura 1.3: Curvas características da equação de advecção,  $a < 0$

A interpretação dessa solução é que, dado uma informação inicial,  $u_0(x)$ , a equação de advecção simplesmente desloca essa informação para direita ( $a > 0$ ) ou para esquerda ( $a < 0$ ) com velocidade  $a$ .

Através de um procedimento inteiramente análogo ao descrito para equação de advecção, obtém-se a solução analítica para um problema de Cauchy qualquer definido em Eq. (1.17) e Eq. (1.18). Logo os sistemas de equações características para este problema é dado por

$$\begin{cases} \frac{dx}{dt} = f'(u), \\ \frac{du}{dt} = 0, \\ x(0) = x_0, \\ u(0) = u(x(0), 0) = u(x_0, 0) = u_0(x_0). \end{cases}$$

Integrando as EDO's obtém-se

$$\begin{cases} x = f'(u)t + c_1, \\ u = c_2, \end{cases} \quad (1.35)$$

sendo  $c_1, c_2$  constantes. Das condições iniciais tem-se

$$\begin{cases} u(0) = u_0(x_0) = c_2, \\ x(0) = c_1 = x_0, \end{cases}$$

logo  $x_0 = c_1 = x - f'(u)t$ , e substituindo  $x_0$  na equação anterior, obtém-se  $u_0(x_0) = u_0(x - f'(u)t)$ . Portanto a solução analítica do problema de Cauchy, para leis de conservação escalares, é dada por

$$u(x, t) = u_0(x - f'(u)t). \quad (1.36)$$

Note que, para o problema de Cauchy, as curvas características são dada por

$$x(t) = f'(u)t + x_0. \quad (1.37)$$

Assim, pode-se concluir que a solução da EDP é constante e igual a  $u(x_0, 0)$ , ao longo da curva característica, de acordo com Eq. (1.26).

Serão apresentadas a teoria da formação de choque e a formulação matemática das soluções fracas para que seja possível fazer um estudo da equação não-linear, utilizando o procedimento das equações características.

### 1.3 Formação de Choques

A solução de uma lei de conservação pode ser suave em todo o domínio ou apenas em parte dele. O surgimento de descontinuidades pode estar relacionado a dados iniciais descontínuos ou à inclusão de efeitos não lineares na equação ao longo do tempo (para soluções suaves, veja Figuras 1.4 e 1.6). Esse efeito é conhecido como a formação de **choque**. A curva onde a solução possui descontinuidade é formada pelos pontos nos quais as curvas características se intersectam, como será visto na equação de Burgers. Quando ocorre a formação do choque, não se tem solução clássica ao longo da curva de descontinuidade da EDP. Quando não houver o cruzamento das curvas características ocorre o fenômeno conhecido com **ondas de rarefação**.

A solução é procurada em todo o domínio e não apenas em parte dele, logo a análise matemática do choque formado envolve a generalização da solução da lei de conservação ao longo da descontinuidade. Surge então a teoria das soluções fracas. Quando ocorre a formação do leque de rarefação o problema em questão possui infinitas soluções. Como as leis de conservação são equações que modelam fenômenos físicos, deseja-se que essas soluções também tenham significado físico. Com a teoria das soluções fracas ganha-se no encontro de uma solução, mas perde-se na unicidade da solução. Para determinar qual será a melhor solução fraca utiliza-se uma condição chamada de ‘entropia’. Uma solução entrópica é uma solução da EDP que possui significado físico (de acordo com o problema estudado).

**Exemplo 1.1** *Como ilustração da formação de choque para condições iniciais suaves, serão apresentados neste exemplo apenas os resultados numéricos, pois o cálculo da solução analítica deste tipo de problemas não é objeto de estudo desta dissertação. Para mais detalhes veja Harten et al (1987) [14]. Considere, então, o problema de Cauchy, composto pela equação de Burgers  $u_t + uu_x = 0$  e pela condição inicial suave*

$$u_0(x) = \sin(\pi(x+1)) \quad -1 \leq x \leq 1. \quad (1.38)$$

*A Figura 1.4 exibe a condição inicial suave dada na Eq. 1.38; a Figura 1.5 exibe a solução antes da formação do choque, em  $t = 0.2$ ; a Figura 1.6 exibe a formação do choque que ocorre*

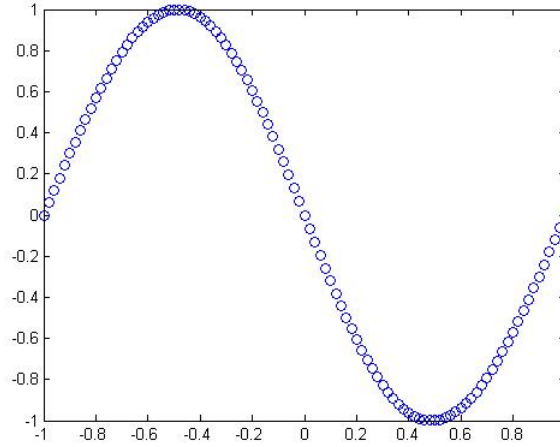


Figura 1.4: Condição inicial suave  $N = 100$ ,  $CFL = 0.6$ ,  $t = 0$

em  $t = \frac{1}{\pi}$  e em  $x = 0$ . Na fronteira esquerda do domínio utilizou-se a seguinte condição:  $u(-1, t) = 0$ . Em todas as simulações foram utilizados  $N = 100$  subintervalos de  $[-1, 1]$  e condição  $CFL = 0.6$ .

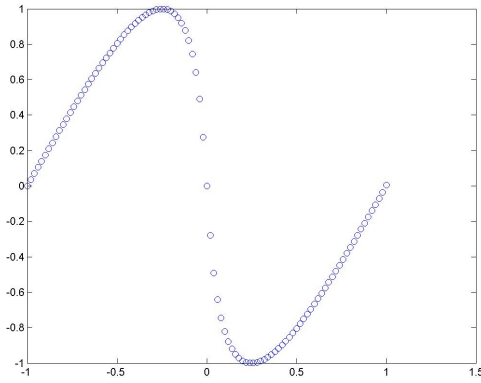


Figura 1.5: Solução antes da formação do choque:  $t = 0.2$ ;  $N = 100$  e  $CFL = 0.6$ .

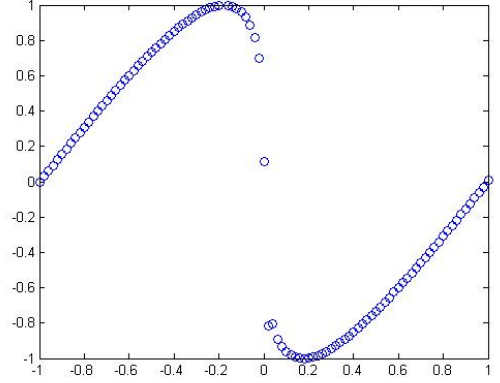


Figura 1.6: Solução na formação de choque:  $t = \frac{1}{\pi}$ ,  $N = 100$ ,  $CFL = 0.6$ .

### 1.3.1 Solução Fraca

Quando ocorre a formação de choque (ver Figura 1.11) ou a formação de um leque de rarefação (ver Figuras 1.13 e 1.16), a lei de conservação não possui solução clássica em todo seu domínio. A perda da solução clássica é real, ou seja, não é devido à técnica utilizada para encontrar as soluções, mas sim devido à não linearidade da função fluxo  $f$  (dependência de  $f$  sobre  $u$ ). A técnica matemática utilizada para se obter uma solução fraca consiste em multiplicar a lei de conservação, na forma diferencial dada na Eq. (1.12), por uma função teste, de forma que a equação modificada requeira menos suavidade da solução  $u$ . Solução fraca é uma solução a qual obedece a lei de conservação modificada e não necessariamente a original.

Seja  $\Phi \in C_0^1(\mathbb{R} \times \mathbb{R})$  uma função teste, sendo  $C_0^1$  o espaço das funções continuamente diferenciáveis com suporte compacto. Neste caso o suporte da função  $\Phi$ ,  $\text{supp } \Phi$ , é um conjunto compacto  $S = [x_1, x_2] \times [t_1, t_2] \subset \mathbb{R} \times (0, +\infty)$  tal que  $\Phi \equiv 0$ , para todo  $x \in S^C$  (complementar de  $S$ ).

Multiplicando a Eq. (1.12) por  $\Phi$  e integrando em relação a  $x$  e em relação a  $t$ , obtém-se

$$\int_0^{+\infty} \int_{-\infty}^{+\infty} [u_t + (f(u))_x] \Phi \, dx dt = 0, \quad (1.39)$$

$$\int_{-\infty}^{+\infty} \int_0^{+\infty} u_t \Phi \, dt dx + \int_0^{+\infty} \int_{-\infty}^{+\infty} (f(u))_x \Phi \, dx dt = 0. \quad (1.40)$$

Como  $\Phi$  é suficientemente diferenciável, usando integração por partes, pode-se escrever

$$\begin{aligned} \Phi(x, t) u \Big|_0^{+\infty} - \int_0^{+\infty} \Phi_t u dt &= \int_0^{+\infty} \Phi(x, t) u_t dt, \\ \Phi(x, t) f(u) \Big|_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} \Phi_x f(u) dx &= \int_{-\infty}^{+\infty} \Phi(x, t) f(u)_x dx. \end{aligned}$$

Como  $\Phi$  tem suporte compacto, tem-se

$$-\Phi(x, 0) u(x, 0) - \int_0^{+\infty} \Phi_t u dt = \int_0^{+\infty} \Phi u_t dt, \quad (1.41)$$

$$0 - \int_{-\infty}^{+\infty} \Phi_x f(u) dx = \int_{-\infty}^{+\infty} \Phi f(u)_x dx. \quad (1.42)$$

Substituindo as Eqs. (1.41) e (1.42) na Eq. (1.40), obtém-se

$$\int_0^{+\infty} \int_{-\infty}^{+\infty} [\Phi_t u + \Phi_x f(u)] dt dx = - \int_{-\infty}^{+\infty} \Phi(x, 0) u(x, 0) dx. \quad (1.43)$$

**Definição 1.4** A função  $u$  é chamada solução fraca do problema de valor inicial Eq. (1.17), Eq. (1.18) se satisfaz a Eq. (1.43), para todo  $\Phi \in C_0^1$ .

É possível mostrar que toda solução clássica é uma solução fraca, no entanto uma solução fraca pode não ser uma solução clássica, pois essas últimas podem não ser diferenciáveis. A solução fraca não é única. Para resolver a questão da falta de unicidade, surge a definição de condição de entropia. Existem algumas definições de condição de entropia, porém é possível provar que sob certas condições todas são equivalentes, veja Smoller (1994) [28]. Será apresentado, na definição 1.5, a condição de entropia utilizada no estudo das equações apresentadas neste trabalho. Para um estudo mais completo das soluções fracas é necessário que a velocidade de propagação do choque seja calculada pela condição de Rankine - Hugoniot.

### 1.3.2 Condição de Rankine-Hugoniot

Quando ocorre a formação de choque as curvas características se cruzam, assim é preciso calcular uma curva ao longo da qual se conheça a solução da equação. Essa curva é calculada pela condição de Rankine-Hugoniot.

Para o cálculo da condição de Rankine - Hugoniot, considera-se a situação na qual  $u$  não é contínua, porém possui uma estrutura simples em uma região aberta  $V \subset \mathbb{R} \times (0, +\infty)$ . Seja  $u$  uma solução fraca, satisfazendo Eq. (1.43) em  $V$ . Suponha que  $u$  seja suave em cada uma das duas regiões  $V_L, V_R$  de  $V$  separadas por uma curva suave  $C$ , sobre a qual  $u$  é descontínua.

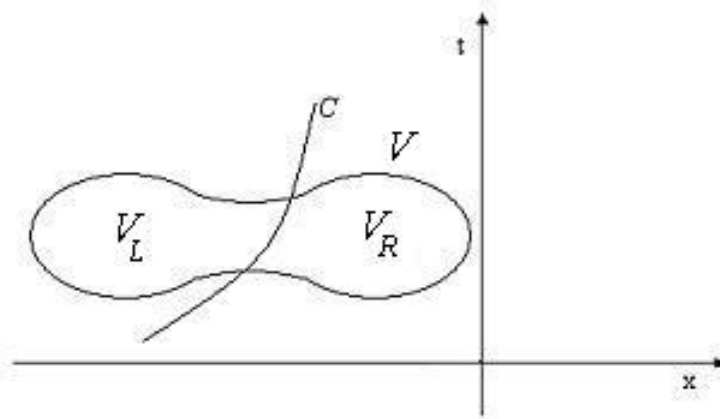


Figura 1.7: Região aberta  $V \subset \mathbb{R} \times (0, +\infty)$  para condição de Rankine-Hugoniot.

Suponha, ainda, que  $u$  e suas primeiras derivadas sejam uniformemente contínuas em  $V_L$  e  $V_R$ . Primeiramente escolha  $\Phi$  com suporte compacto em  $V_L$ , a Eq. (1.43) pode assumir sua forma original na Eq. (1.39), como segue

$$\int_0^{+\infty} \int_{-\infty}^{+\infty} \Phi [u_t + (f(u))_x] dx dt = 0.$$

Para o cálculo da condição de Rankine-Hugoniot será apresentado um lembrete da teoria do divergente. Seja o campo de vetores  $G = \Phi(f(u), u) = (\Phi f(u), \Phi u)$ , então

$$\nabla_{x,t} G = \Phi_x f(u) + \Phi f(u)_x + \Phi_t u + \Phi u_t,$$

e o teorema da divergência diz que

$$\int_{V_L} \nabla_{x,t} G dx dt = \int_{\partial V_L} G \vec{n} ds,$$

sendo  $\vec{n} = (n_1, n_2)$  é o vetor normal exterior à fronteira de  $V_L$ . Então

$$\int_{V_L} [\Phi_t u + \Phi_x f(u)] dx dt + \int_{V_L} \Phi [u_t + f(u)_x] dx dt = \int_{\partial V_L} (\Phi u n_2 + \Phi f(u) n_1) ds, \quad (1.44)$$

$$\int_{V_L} [\Phi_t u + \Phi_x f(u)] dx dt = - \int_{V_L} \Phi [u_t + f(u)_x] dx dt + \int_{\partial V_L} (\Phi u n_2 + \Phi f(u) n_1) ds.$$

**Observação 1.6** Quando fronteira de  $V_L$  for composta pela curva  $C$  e pelo intervalo  $(-\infty, x_0)$ , sendo  $x_0$  o ponto sobre a curva  $C = C(t)$  no ponto  $t = 0$ . Obtém a situação está ilustrada na Figura 1.8. Note que  $\vec{n} = (0, -1)$ , na fronteira inferior de  $V_L$ .

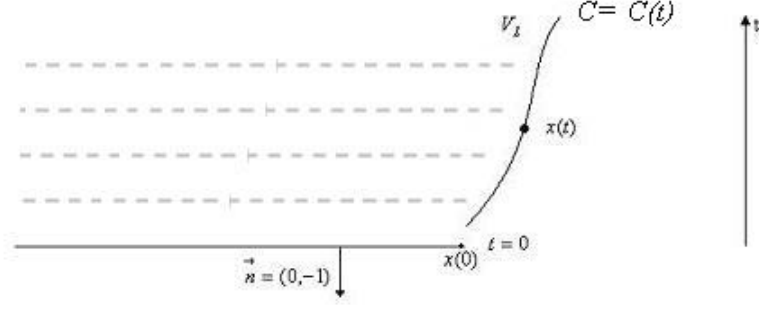


Figura 1.8: Caso particular para fronteira da condição de Rankine-Hugoniot

De acordo com o teorema da divergência e sabendo que  $\Phi$  possui suporte compacto em  $V_L$  concluímos  $\Phi(x(t), t) = 0$ , para todo  $(x(t)) \in C(t)$  ( $x(t)$  está sobre a curva  $C$ ). Assim

$$\int_{\partial V_L} G \vec{n} ds = \int_C (\Phi(x(t), t) n_1(x(t), t) f(u)(x(t), t) + \Phi(x(t), t) n_2(x(t), t) U) dt + \int_{-\infty}^{x_0} -\Phi(x, 0) u(x, 0) dx. \quad (1.45)$$

Como  $\Phi(x(t), t) = 0$ ,  $\forall (x(t)) \in C(t)$ , temos que

$$\int_{\partial V_L} G \vec{n} ds = - \int_{-\infty}^{+\infty} \Phi(x, 0) u(x, 0) dx.$$

Portanto pela Eq. (1.44), tem-se

$$\int_{V_L} [\Phi_t u + \Phi_x f(u)] dx dt = - \int_{V_L} \Phi [u_t + f(u)_x] dx dt - \int_{-\infty}^{+\infty} \Phi(x, 0) u(x, 0) dx. \quad (1.46)$$

Para que  $u$  seja solução da lei de conservação em  $V_L$  tem-se, pela Eq. (1.44), que a primeira integral, no segundo membro da Eq. (1.46) deve se anular, ou seja

$$\int_{V_L} [\Phi_t u + \Phi_x f(u)] dx dt + \int_{-\infty}^{+\infty} \Phi(x, 0) u(x, 0) dx = 0. \quad (1.47)$$

Para o caso geral, se o suporte de  $\Phi$  não possui pontos da forma  $(x, 0)$ , então  $\Phi(x, t) = 0$ . Em geral pode-se considerar o suporte de  $\Phi$  longe da fronteira contendo os pontos  $(x, 0)$ , e portanto vale a Eq. (1.48)

$$\int_{V_L} [\Phi_t u + \Phi_x f(u)] dx dt = 0. \quad (1.48)$$

Analogamente, se  $\Phi$  tem suporte compacto em  $V_R$  tem-se que

$$\int_{V_R} [\Phi_t u + \Phi_x f(u)] dx dt = 0. \quad (1.49)$$

Agora, seja uma função teste com suporte compacto em  $V$ , mas que não necessariamente se anula em  $C$ . Seja  $G = (\Phi f(u), \Phi u)$ ; o divergente do campo  $G$  é dado por  $\nabla_{x,t} \cdot G = \Phi_x f(u) + \Phi f(u)_x + \Phi_t u + \Phi u_t$ , então pelo teorema da divergência tem-se

$$\int_0^{+\infty} \int_{-\infty}^{+\infty} \nabla_{x,t} \cdot G dx dt = \int_{V_L} \nabla_{x,t} \cdot G dx dt + \int_{V_R} \nabla_{x,t} \cdot G dx dt + \int_{\Omega} \nabla_{x,t} \cdot G dx dt, \quad (1.50)$$

sendo  $\Omega = \mathbb{R} \times [0, +\infty) - V_L - V_R$ .

Tem-se que  $\Phi$  se anula fora de  $V$  então  $\Phi_x$  e  $\Phi_t$  também se anulam fora de  $V$ , portanto a Eq. (1.50) se reduz à

$$\begin{aligned} \int_0^{+\infty} \int_{-\infty}^{+\infty} \nabla_{x,t} \cdot G dx dt &= \int_{\partial V_L} G n_{V_L} ds + \int_{\partial V_R} G n_{V_R} ds \\ &= \int_{C_-} [\Phi(x(t)) f(u(x(t), t)) n_{1,L} + \Phi(x(t)) u(x(t), t) n_{2,L}] dt + \\ &\quad \int_{C_+} [-\Phi(\tilde{x}(t)) f(u(\tilde{x}(t), t)) n_{1,L} + \Phi(\tilde{x}(t)) u(\tilde{x}(t), t) n_{2,L}] dt. \end{aligned}$$

**Observação 1.7** *Tem-se que:*

1.  $u_L(x(t)) = \lim_{n \rightarrow +\infty} u(x_n(t))$  tal que  $\lim_{n \rightarrow +\infty} x_n(t) = x(t) \in C(t)$  sendo  $x_n \in V_L$  para todo  $n \in [0, +\infty)$ .
2.  $u_R(\tilde{x}(t)) = \lim_{n \rightarrow +\infty} u(\tilde{x}_n(t))$  tal que  $\lim_{n \rightarrow +\infty} \tilde{x}_n(t) = \tilde{x}(t) \in C(t)$  sendo  $\tilde{x}_n \in V_R$  para todo  $n \in [0, +\infty)$ .

De acordo com o provado até o momento, tem-se que

$$\int_{V_L} \Phi_x f(u_L) + \Phi_t u_L + \int_{-\infty}^{+\infty} \Phi(x, 0) u(x, 0) dx = \int_0^{+\infty} \int_{-\infty}^{+\infty} \Phi [u_t^L + f(u^L)_x] dx dt$$

e

$$\int_{V_R} \Phi_x f(u_R) + \Phi_t u_R + \int_{-\infty}^{+\infty} \Phi(x, 0) u(x, 0) dx = \int_0^{+\infty} \int_{-\infty}^{+\infty} \Phi [u_t^R + f(u^R)_x] dx dt.$$

Usando o fato da função teste  $\Phi$  ter suporte compacto, obtém-se

$$\int_V \Phi f(u) + \Phi_t u dx dt = \int_V \Phi [u_t + f(u)_x] dx dt.$$

Portanto,

$$\int_V \Phi f(u) + \Phi_t u dx dt = 0$$

se, e somente se,

$$\int_V \Phi [u_t + f(u)_x] dx dt = 0.$$

Assim

$$\int_C [\Phi f(u_L) n_1 + \Phi u_L n_2] dt = \int_C [\Phi f(u_R) n_1 + \Phi u_R n_2] dt$$

portanto

$$\int_{C(t)} \{[f(u_L) - f(u_R)] n_1 + [u_L - u_R] n_2\} \Phi(x, t) dt = 0.$$

Como  $\Phi$  não é a função nula, tem-se que

$$[f(u_L) - f(u_R)] n_1 + [u_L - u_R] n_2 = 0, \quad (1.51)$$

para todo  $t$  tal que  $x(t) \in C(t)$ . Como a curva  $C(t) = C(x(t), t)$ , tem-se que  $C'(t) = C(x'(t), 1)$  e o vetor normal apontando para fora da curva  $C$  (na direção de  $V_L$  para  $V_R$ ) é dado por

$$\vec{n}(t) = \frac{(1, -x'(t))}{\sqrt{1 + (-x'(t))^2}}. \quad (1.52)$$

Substituindo a Eq. (1.52) na Eq. (1.51), obtém-se

$$[f(u_L) - f(u_R)] = [u_L - u_R] x'(t).$$

Seja  $s = x'(t)$ , então

$$s = \frac{f(u_L) - f(u_R)}{u_L - u_R}. \quad (1.53)$$

A Eq. (1.53) é conhecida como condição de Rankine - Hugoniot e pode ser escrita como

$$s = \frac{[f]}{[u]},$$

sendo  $s$  a velocidade na qual viaja a descontinuidade da solução da lei de conservação.

**Definição 1.5 :** (*Condição de Entropia*)

*Para a lei de conservação escalar, a propagação da descontinuidade com velocidade  $s$  satisfaz*

a condição de entropia se

$$f'(u_l) > s > f'(u_r), \quad (1.54)$$

sendo  $f'(u_l)$  e  $f'(u_r)$  a velocidade característica imediatamente à esquerda e a direita de  $s$  respectivamente.

Pode-se verificar que para funções convexas (ou côncavas), a condição de entropia Eq. (1.54) é sempre satisfeita, pois a derivada de uma função convexa (ou côncava) é sempre crescente (decrecente). A solução  $u$  que satisfaz a condição de entropia é chamada solução entrópica e é a melhor solução, pois é a mais próxima da solução com significado físico.

**Teorema 1.1** *Se a função de fluxo  $f$  é convexa e suave, a menos de um conjunto de medida nula, a solução de entropia é única.*

**Demonstração:** *Veja Evans (1998) [9].*

## 1.4 Problema de Riemann

Uma lei de conservação associada a uma condição inicial constante por partes, com uma descontinuidade simples, é conhecido como **Problema de Riemann**, Figura 1.9.

O objetivo é resolver os problemas de Riemann, descritos nas Eqs. (1.55) e (1.57) para uma função fluxo convexa, isto é,  $f'' > 0$  ( $f'$  é crescente) para que se satisfaça a Teorema 1.1.

Para complementar o estudo feito na seção 1.2 será realizado o estudo do problema de Riemann composto pela equação de advecção

$$u_t + au_x = 0,$$

com condição inicial

$$u(x, 0) = u_0(x) = \begin{cases} u_l, & \text{se } x < 0, \\ u_r, & \text{se } x > 0, \end{cases} \quad (1.55)$$

no domínio  $-\infty < x < +\infty$  e  $t > 0$ , sendo  $u_l$  e  $u_r$  constantes.

A solução analítica da equação de advecção é dada por  $u(x, t) = u_0(x - at)$ , veja Eq. (1.34). Assumindo que  $a > 0$  (procedimento análogo pode ser feito para  $a < 0$ ), a descontinuidade da condição inicial em  $x = 0$ , ilustrada na Figura 1.9, se propaga para o interior do domínio, da esquerda para a direita, com velocidade  $a$  e distância  $at$ , no tempo  $t$ .

Para o problema de Riemann, Eq. (1.55), a curva característica que passa por  $x = 0$  é importante, pois é a única através da qual a solução muda. Essa curva característica é dada por  $x = at$  e separa as outras curvas características em duas partes:

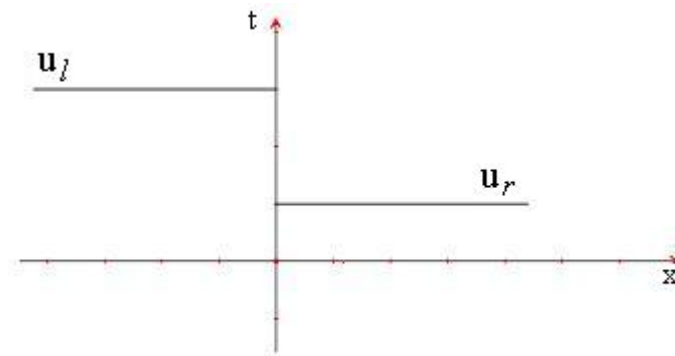


Figura 1.9: Condição inicial do Problema de Riemann

- à esquerda, cuja solução é  $u_l$
- e
- à direita, cuja solução é  $u_r$ ,
- como ilustra a Figura 1.10.

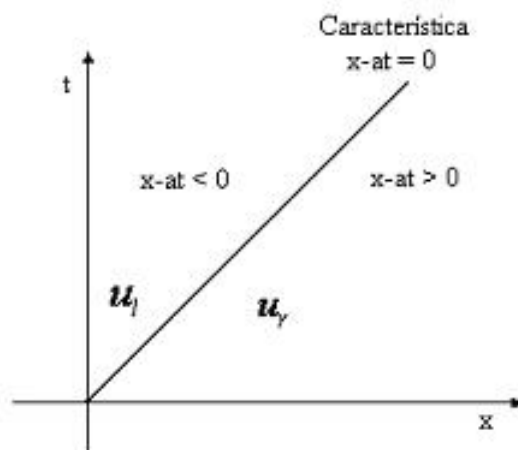


Figura 1.10: Curva característica e solução da equação de advecção.

De acordo com o exposto acima, a solução do problema de Riemann, Eq. (1.55), é dada por

$$u(x, t) = u_0(x - at) = \begin{cases} u_l, & \text{se } x - at < 0, \\ u_r, & \text{se } x - at > 0. \end{cases} \quad (1.56)$$

Por qualquer ponto  $x_0$ , no eixo  $x$ , passa uma reta característica e como  $a$  é constante todas as retas são paralelas umas as outras, não ocorrendo a formação de choque. Como ilustra a Figura 1.2.

A equação de Burgers é uma equação escalar, quase-linear de classe  $C^1$  em  $t > 0$ , com função de fluxo  $f$  convexa, pois  $f''(u) = 1 > 0$  para todo  $t > 0$  ( $f'$  é crescente). Ao contrário da equação de advecção, para a equação de Burgers não existe uma solução analítica global. Tem-se o cruzamento das curvas características o que diz que ocorre, na solução da equação, a

formação de choque ou a formação de leque de rarefação.

Considere o problema de Riemann composto pela equação de Burgers na forma não conservativa

$$u_t + uu_x = 0$$

e a sua condição inicial

$$u(x, 0) = u_0(x) = \begin{cases} u_l, & \text{se } x < 0, \\ u_r, & \text{se } x > 0. \end{cases} \quad (1.57)$$

no domínio  $-\infty < x < +\infty$  e  $t > 0$ , sendo  $u_l$  e  $u_r$  constantes.

Se  $u$  é solução desse problema então, para todo  $\lambda > 0$ , a função  $u_\lambda$  dada por  $u_\lambda = u(\lambda x, \lambda t)$  também é solução da mesma equação. Como o interesse é obter a solução de entropia do problema, devido a sua unicidade, é natural considerar somente soluções as que dependam do raio  $x/t$ . Veja Smoller (1994) [28].

Da Eq. (1.36), tem-se que a solução da equação de Burgers, para uma condição inicial  $u_0$  qualquer, é dada por

$$u(x, t) = u_0(x - u.t), \quad (1.58)$$

pois  $f(u) = \frac{u^2}{2}$ . De acordo com a Eq. (1.37), tem-se que as curvas características da equação de Burgers são dadas por

$$x(t) = u.t + x_0, \quad (1.59)$$

sendo  $u$  constante ao longo das curvas características, logo pode-se concluir que essas curvas são retas. O valor constante de  $u$  ao longo das retas características, as quais se originam no ponto  $(x_0, 0)$  do plano  $(x, t)$ , é dado por  $u_0(x_0)$  e o coeficiente angular dessas retas é dado por

$$\tan \theta = \frac{1}{u}. \quad (1.60)$$

Como o coeficiente angular depende de  $u$ , é de se esperar que as curvas características se cruzem, pois  $u$  varia, formando ondas de choques ou ondas de rarefação.

Para resolver o problema de Riemann, Eq. (1.57), é preciso considerar dois casos:

1.  $u_l > u_r$

Como  $f'$  é crescente tem-se que

$$f'(u_l) > f'(u_r). \quad (1.61)$$

Isso diz que a informação à esquerda se propaga mais rapidamente do que a informação à direita, resultando no cruzamento das curvas características e na formação de choque.

Sejam  $\theta_l$  o ângulo correspondente à  $u_l$  e  $\theta_r$  o ângulo correspondente à  $u_r$ . Para a equação de Burgers, tem-se que  $f'(u) = u$ . Das Eqs. (1.60) e (1.61) deduzimos que

$$\tan \theta_l \leq \tan \theta_r. \quad (1.62)$$

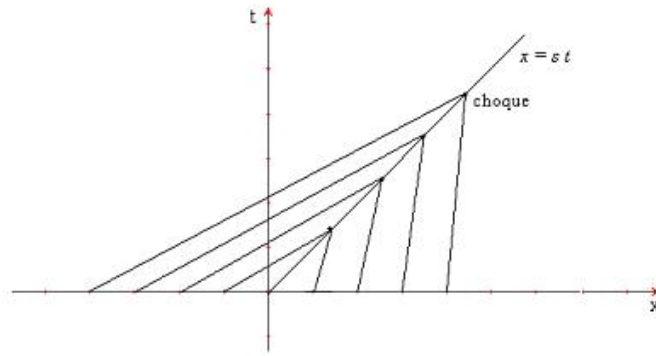


Figura 1.11: Formação de choque através da colisão das curvas características.

Quando ocorre o cruzamento das curvas características, como ilustra a Figura 1.11, tem-se a formação de **choque**. Nesse ponto  $f'(u_l) = f'(u_r)$ , o que contradiz a desigualdade da condição de entropia. Isso sugere que a solução analítica da equação de Burgers não é encontrada ao longo de todo domínio. Essa argumentação indica que, na presença de choques, a solução procurada deve ser descontínua. Essa descontinuidade é calculada pela condição de Rankine-Hugoniot na Eq. (1.53), sendo  $s$  a velocidade com a qual viaja a descontinuidade da solução dentro do domínio. Para a equação de Burgers temos

$$s = \frac{u_l + u_r}{2}. \quad (1.63)$$

De acordo com a Eq. (1.58) e com a condição de Rankine-Hugoniot, tem-se que a solução da equação de Burgers para  $u_l > u_r$  é dada por

$$u(x, t) = u_0(x - ut) = \begin{cases} u_l, & \text{se } \frac{x}{t} \leq s, \\ u_r, & \text{se } \frac{x}{t} > s. \end{cases} \quad (1.64)$$

Pode-se verificar que essa solução satisfaz a condição de entropia Eq. (1.54). Como  $u_r < u_l$  então

$$2u_r < u_r + u_r < 2u_l. \quad (1.65)$$

Dividindo por 2

$$u_r < \frac{u_r + u_l}{2} < u_l. \quad (1.66)$$

Da Eq. (1.63), obtém-se

$$u_r < s < u_l. \quad (1.67)$$

A Eq. (1.67) mostra que esta solução satisfaz a condição de entropia Eq. (1.54).

Para exemplificar o processo da formação de choque considere como condição inicial do problema de Riemann

$$u(x, 0) = u_0(x) = \begin{cases} 1, & \text{se } x < 0, \\ 0, & \text{se } x > 0, \end{cases} \quad (1.68)$$

no domínio  $-\infty < x < +\infty$  e  $t > 0$ .

Tem-se que:

- se  $x < 0$  as curvas características emanam do eixo  $x$  com coeficiente angular  $\tan \theta = \frac{1}{u_l} = 1$  logo  $\theta = 45^\circ$ ,

e

- se  $x > 0$  as curvas características emanam do eixo  $x$  com coeficiente angular  $\tan \theta = \frac{1}{u_r} = \frac{1}{0}$  logo  $\theta = 90^\circ$ .

Assim, pela Eq. (1.63), a condição de Rankine-Hugoniot é  $s = 1/2$ .

Este exemplo é ilustrado na Figura 1.12.

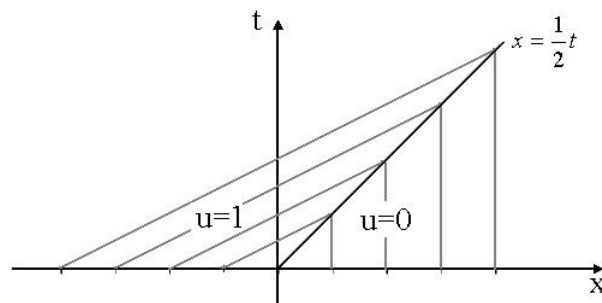


Figura 1.12: Formação de choque - equação 1.68

2.  $u_l < u_r$

Como  $f'$  é crescente, tem-se que

$$f'(u_l) < f'(u_r), \quad (1.69)$$

o que diz que a informação à esquerda se propaga mais lentamente do que a informação à direita. Neste caso as curvas características nunca se cruzam e o que ocorre é a formação do que se define como **leque de rarefação**. Com mesma análise realizada no caso anterior tem-se das Eqs. (1.60) e (1.69) que

$$\tan \theta_l > \tan \theta_r, \quad (1.70)$$

o que garante que realmente as curvas características não se cruzarão. Como ilustra a Figura 1.13.

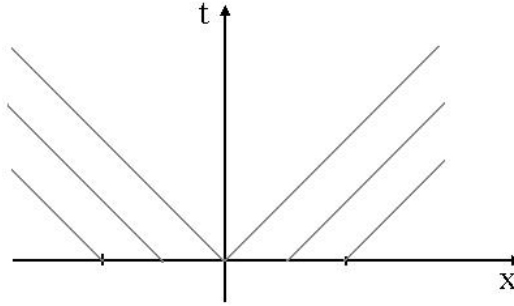


Figura 1.13: Leque de rarefação

A função  $u$  dada em Eq. (1.64) também é solução do problema de Riemann Eq. (1.57), porém não satisfaz a condição de entropia.

É possível construir uma solução que satisfaça a condição de entropia. Essa solução será construída para um caso particular, para melhor compreensão do problema. Considere como condição inicial para o problema de Riemann

$$u(x, 0) = u_0(x) = \begin{cases} 0, & \text{se } x \leq 0, \\ 1, & \text{se } x > 0. \end{cases} \quad (1.71)$$

Analogamente ao caso anterior, tem-se:

- se  $x < 0$  as curvas características emanam do eixo  $x$  com coeficiente angular  $\tan \theta = \frac{1}{u_r}$  logo  $\theta = 90^\circ$ ;
- se  $x > 0$  as curvas características emanam do eixo  $x$  com coeficiente angular  $\tan \theta = \frac{1}{u_l} = 1$  logo  $\theta = 45^\circ$ .

Pela Eq. (1.63) a condição de Rankine-Hugoniot é  $s = 1/2$ . Na região do plano  $(x, t)$  que possui curvas características elas são dadas, de acordo com Eq. (1.59), por

$$x(t) = \begin{cases} x_0, & \text{se } x \leq 0, \\ t + x_0, & \text{se } x > 0. \end{cases}$$

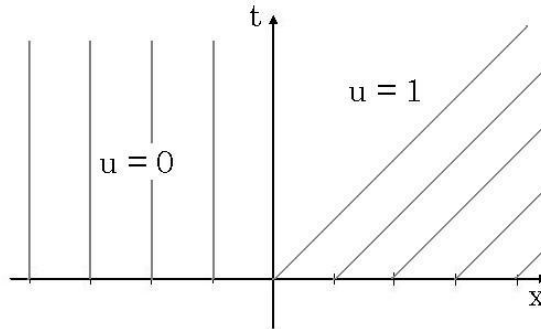


Figura 1.14: Curvas características - Equação 1.71

Uma solução do problema é dada por

$$u(x, t) = \begin{cases} 0, & \text{se } x \leq 0 < t, \\ 1, & \text{se } 0 < t < x, \end{cases}$$

porém não satisfaz a condição de entropia.

É preciso construir uma solução que satisfaça a condição de entropia. Para  $0 < x \leq t$  (região na qual o problema não possui curvas características) pode-se verificar que a solução  $u(x, t) = x/t$  satisfaz a equação de Burgers. A solução  $u(x, t) = x/t$  é encontrada inserindo curvas características partindo da origem (descontinuidade) de forma que  $u$  seja constante ao longo delas e  $u$  varie continuamente de 0 para 1, como ilustra a Figura 1.15.

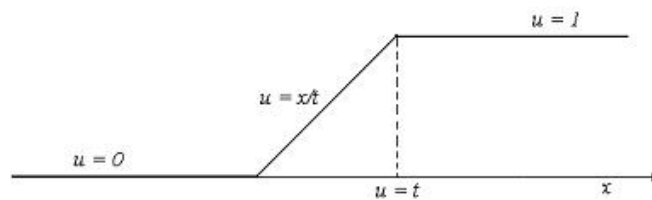


Figura 1.15: Construção da solução que satisfaz a condição de entropia

Então a solução entrópica do problema de Riemann Eq. (1.57), com condição inicial dada em Eq. (1.71), em todo plano  $(x, t)$  é dada pela Eq. (1.72) e ilustrada na Figura 1.16.

$$u(x, t) = \begin{cases} 0, & \text{se } x \leq 0, \\ x/t, & \text{se } 0 \leq x < t, \\ 1, & \text{se } t < x, \end{cases} \quad (1.72)$$

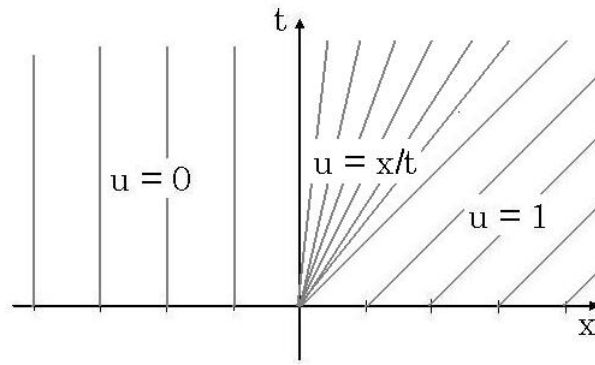


Figura 1.16: Solução entrópica para equação de Burgers

Para verificação analítica que essa solução satisfaz a condição de entropia ver os livros Evans (1998) [9] e Smoller (1998) [28].

Pode-se construir outra solução, descontínua, preenchendo as lacunas sem curvas características ao longo de uma curva diferenciável  $x/t$  iniciando no ponto  $(0,0)$  com inclinação igual a  $1/2$ , da condição de Rankine - Hugoniot. A solução procurada é dada pela Eq. (1.73) e ilustrada na Figura 1.17.

$$u(x, t) = \begin{cases} 0, & \text{se } x \leq \frac{t}{2}, \\ 1, & \text{se } x > \frac{t}{2}. \end{cases} \quad (1.73)$$

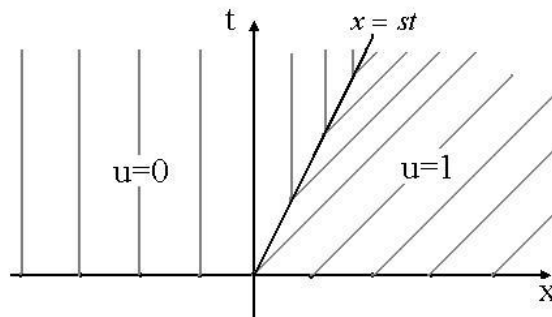


Figura 1.17: Solução não entrópica para equação de Burgers

Pode-se verificar que essa solução não satisfaz a condição de entropia. (Veja Smoller (1994) [28]).

É possível calcular o ponto em que as curvas características se cruzam, porém não é objetivo desse trabalho, para mais informações ver Marotti (2008) [25].

# Capítulo 2

## Métodos Numéricos

Mesmo para modelos mais simples de equações diferenciais parciais pode ser difícil obter soluções analíticas. Neste caso, é fundamental o desenvolvimento de métodos numéricos para se obter soluções aproximadas destas equações. Contudo, outra dificuldade surge ao se analisar os resultados numéricos. Devido à inexistência de uma teoria matemática que abranja todas as complexidades envolvidas nas equações diferenciais parciais hiperbólicas, principalmente as equações não-lineares. Muitos resultados não podem ser rigorosamente comprovados através de teoremas de análise numérica. Alguns resultados são demonstrados para as equações lineares e transferidos para as equações não-lineares. Mesmo sem a prova desses resultados, na prática os métodos numéricos têm apresentado bons resultados para as equações não-lineares. Várias são as dificuldades que podem surgir na busca das soluções exatas: complexidade da região de busca, os coeficientes da equação diferencial podem variar ponto a ponto e até mesmo depender da própria solução (problemas não-lineares).

Nesse capítulo se dará o início do estudo dos métodos numéricos que permitem encontrar aproximações para as soluções analíticas das leis de conservação (estudadas no capítulo 1), baseadas no método das diferenças finitas tradicionais.

Será apresentada a teoria necessária para análise da consistência, da estabilidade e da convergência dos métodos numéricos para equações lineares e equações não-lineares. Como principais resultados deste capítulo tem-se o teorema de Lax-Wendroff (teorema 2.2), o teorema que garante a convergência dos métodos numéricos (teorema 2.3) e o teorema que garante a convergência para a solução de entropia (teorema 2.5).

### 2.1 Métodos Numéricos para Equações Lineares

Seja o problema de Cauchy unidimensional

$$\begin{cases} u_t + au_x = 0, \\ u(x, 0) = u_0(x), \end{cases} \quad (2.1)$$

no domínio  $-\infty < x < +\infty, \quad t \geq 0$ .

A discretização do problema de Cauchy unidimensional Eq. (2.1), é feita em uma malha

uniforme, no espaço e no tempo, escolhendo como respectivos passos  $\Delta x$  e  $\Delta t$ , ou seja, nos pontos discretos da malha tem-se

$$\begin{aligned} x_j &= j \cdot \Delta x \quad j = \dots, -1, 0, 1, \dots, \\ t_n &= n \cdot \Delta t \quad n = 0, 1, 2, \dots, \end{aligned}$$

é preciso definir também o ponto médio do intervalo  $I = [x_j, x_{j+1}]$  como sendo

$$x_{j+\frac{1}{2}} = x_j + \frac{\Delta x}{2} = \left(j + \frac{1}{2}\right) \Delta x.$$

Os métodos numéricos discutidos podem ser aplicados em malhas variáveis, mas por simplicidade teórica e computacional serão usadas malhas uniformes sendo  $\Delta x$  e  $\Delta t$  constantes. Seja  $U_j^n \in \mathbb{R}$  a aproximação da solução  $u(x_j, t_n)$  nos pontos da malha numérica. O valor da solução verdadeira, nos pontos da malha, será denotado por  $u_j^n = u(x_j, t_n)$ .

No desenvolvimento dos métodos numéricos para leis de conservação é preferível considerar  $U_j^n$  como uma aproximação média de  $u(x, t)$  na célula  $I_j = [x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}]$  definida por

$$\bar{u}_j^n \equiv \frac{1}{\Delta x} \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} u(x, t_n) dx, \quad (2.2)$$

a qual aproxima melhor o valor da solução verdadeira  $u_j^n$  do que o método de diferenças finitas, porém para simplificar os cálculos será considerado o método das diferenças finitas. A interpretação por aproximação dos valores médios das células é natural uma vez que a forma integral da lei de conservação Eq. (1.11) descreve a evolução no tempo da integral Eq. (2.2).

Como aproximação para o dado inicial  $u_0(x)$  será utilizado  $U_j^0$ , ou seja,  $u_j^0 = U_j^0$  ou preferencialmente  $U_j^0 = \bar{u}_j^0$ . Será definido, ainda uma função constante por partes  $U_{\Delta t}(x, t)$  da seguinte forma

$$U_{\Delta t}(x, t) = U_j^n, \quad \forall (x, t) \in \left[x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}\right) \times [t_n, t_{n+1}). \quad (2.3)$$

De acordo com Leveque (1992) [17], para equações hiperbólicas dependentes do tempo, é comum assumir que a razão  $\frac{\Delta t}{\Delta x} = cte$  fixa quando  $\Delta t \rightarrow 0$ , que é o caso das leis de conservação. Para o estudo numérico da Eq. (2.1), **na prática, é computado um domínio numérico finito**  $a < x < b$ . O procedimento de obter as aproximações é recursivo; define-se  $U^0$  como uma aproximação do dado inicial  $u_0$  e a partir de  $U^0$  encontra-se  $U^1$ , a partir de  $U^1$  encontra-se  $U^2$  e assim sucessivamente, até obter  $U^{n+1}$ .

Sempre que métodos numéricos são usados para resolver uma EDP, deve-se ter uma preocupação com as propriedades de precisão e convergência das aproximações.

O método numérico, e principalmente o código implementado, deve ser testado em problemas simples, para os quais a solução exata é conhecida, ou em problemas onde comparações com soluções de alta precisão podem ser realizadas. Em alguns casos soluções experimentais podem estar disponíveis.

Em problemas reais, geralmente não se possui a solução verdadeira para fins de comparação. Então, na prática o que se faz é estudar a convergência de um método numérico para equações as quais se conhece a solução (geralmente problemas lineares) e aplicar esses métodos numéricos aos problemas para os quais as soluções não são conhecidas, mesmo sem a garantia de convergência para tais equações.

O ideal é provar que o método escolhido converge para a solução correta quando se refina a malha.

O teorema fundamental da convergência dos métodos numéricos para equações diferenciais lineares, pode ser resumido em:

$$\text{consistência} + \text{estabilidade} \iff \text{convergência}.$$

Logo, para se estudar convergência de um método numérico é preciso estudar sua estabilidade e sua consistência.

### 2.1.1 Condição CFL

De acordo com Leveque (1992) [17], o primeiro artigo sobre diferenças finitas para EDP's foi escrito por Courant, Friedrichs e Lewy (1928) [5], do qual existe uma tradução em inglês (1967) [6]. Eles usaram o método das diferenças finitas na prova da existência de soluções analíticas para certas EDP's. A idéia é definir uma sequência de aproximações, via diferenças finitas da equação e provar sua convergência para solução exata quando a malha é refinada. Eles mostraram que os termos da sequência satisfazem a EDP, no limite, obtendo a existência da solução.

No decorrer da demonstração da convergência desta sequência, eles identificaram que uma condição necessária para estabilidade do método numérico é que o domínio de dependência do método de diferenças finitas deve estar contido no domínio de dependência da EDP (seção 2.1.2) ao menos em limite, ou seja, quando  $\Delta t, \Delta x \rightarrow 0$ . Esta condição é conhecida como condição CFL devido à Courant, Friedrichs e Lewy.

**Definição 2.1 *Condição CFL:*** *Para que um método numérico seja convergente é necessário que o domínio de dependência numérico esteja contido no domínio de dependência da EDP, ao menos quando  $\Delta t$  e  $\Delta x$  tendem a zero. Leveque (2004) [18].*

É preciso observar que a condição CFL não é uma condição suficiente para estabilidade, pois pode existir um ponto  $\xi$  dentro do domínio de dependência da EDP que não pertença ao domínio de dependência numérico. Se alterar o valor do dado inicial em  $\xi$  pode-se afetar a solução verdadeira, mas não a solução numérica (ver Leveque (1992) [17] e Leveque (2004) [18]). A condição CFL estabelece que um método deve ser aplicado de forma que a informação tenha a chance de se propagar com velocidade fisicamente correta, como determinado pelos valores de  $f'(U)$ .

Supondo que se aplique um método numérico para a lei de conservação de forma que a solução exata se mova, dentro do domínio, a uma velocidade  $f'(u)$  e a uma distância  $f'(u) \cdot \Delta t$  durante um intervalo de tempo. A condição CFL exige que a distância percorrida, durante um intervalo de tempo, deve ser menor do que uma célula (ou seja,  $|f'(u) \cdot \Delta t| \leq \Delta x$ , de acordo com Eq. (2.8)).

**Observação 2.1** *Pode-se provar que para o método de Lax - Friedrichs, Eq. (2.26) a condição CFL é também uma condição suficiente para estabilidade do método (ver Leveque (1992) [17]).*

**Definição 2.2** *Define-se como **estêncil** o conjunto de pontos, da malha numérica, envolvidos na cálculo da aproximação numérica de  $U_j^{n+1}$ .*

A Figura 2.1 ilustra um estêncil contendo três pontos  $\{U_{j-1}^n, U_j^n, U_{j+1}^n\}$  com os quais se constrói a aproximação para a solução no instante  $n + 1$ ; ( $U_j^{n+1}$ ).

## 2.1.2 Domínio de Dependência

Domínio de dependência da EDP é o conjunto de pontos, dentro do domínio, os quais dependem apenas da condição inicial. Se  $u$  é a solução do problema de Cauchy definido nas Eqs. (1.17) e (1.18), então de acordo com Eq. (1.36)  $u$  deve satisfazer

$$u(x, t) = u_0(x - f'(u)t), \quad x \in \mathbb{R} \quad t > 0, \quad (2.4)$$

Seja  $(X, T)$  um ponto fixo no plano  $(x, t)$ . Então a solução  $u(X, T)$  depende apenas da condição inicial  $u_0$  e do ponto  $(X - f'(u)T)$  e assim define domínio de dependência da EDP como sendo o conjunto

$$D(X, T) = \{x \in \mathbb{R}; x = X - f'(u)T\}. \quad (2.5)$$

O domínio de dependência numérico é definido de forma similar, ou seja, é o conjunto de pontos da malha computacional que dependem apenas da condição inicial:

$$D_{\Delta t}(X, T) = \{x_j; u_0(x_j) \text{ é usado para o cálculo da aproximação numérica de } u(X, T)\}, \quad (2.6)$$

sendo  $u$  a solução numérica em  $(X, T)$ .

Segue a apresentação da condição necessária, consequência da condição CFL, para estabilidade de um método numérico, em uma malha numérica igualmente espaçada, baseado no método de diferenças finitas, conforme o estêncil dado na Figura 2.1.

A aproximação  $U_{\Delta t}(x_j, t_n)$  depende dos valores  $x_{j-1}, x_j, x_{j+1}$  no tempo  $t_{n-1}$ , ou seja, depende de  $x_{j+q}$ ,  $q = -1, 0, 1$ . Por sua vez esses valores dependem de  $q = -2, -1, 0, 1, 2$  no

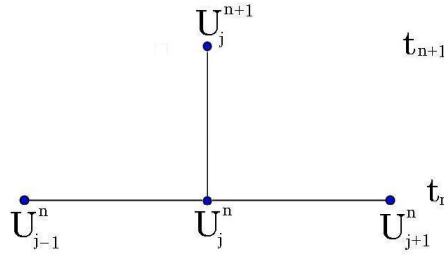


Figura 2.1: Diagrama para condição CFL da aproximação  $U_j^{n+1}$ .

tempo  $t_{n-2}$ . Continuando esse processo até o tempo  $t_0 = t_{n-n}$  vemos que a solução no tempo  $t_n$  depende apenas do dado inicial nos pontos  $x_{j+q}$ ,  $q = -n, \dots, -2-1, 0, 1, 2, \dots, n$  como ilustra a Figura 2.2.

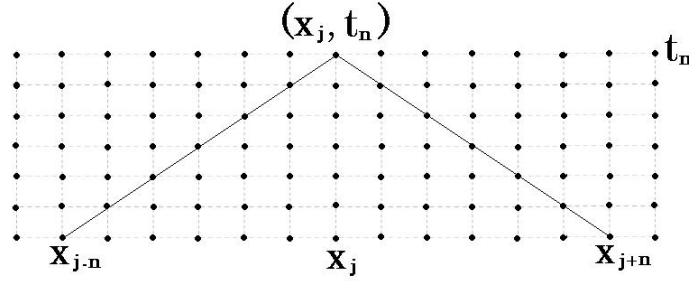


Figura 2.2: Triângulo computacional da condição CFL para um esquema de diferenças finitas de três pontos, ilustrado no diagrama 2.1.

Para o caso  $n = 2$ , de acordo com a Figura 2.2, o domínio de dependência numérico satisfaz

$$D_{\Delta t}(x_j, t_2) \subset \{x; |x - x_j| \leq 2\Delta x\}.$$

Para o caso geral tem-se que o domínio de dependência numérico é dado por

$$D_{\Delta t}(x_j, t_n) \subset \{x; |x - x_j| \leq n\Delta x\}. \quad (2.7)$$

Agora considere um ponto fixo  $(X, T)$ , então  $D_{\Delta t}(X, T) \subset \{x; |x - X| \leq \frac{T}{\Delta t} \Delta x\}$ , pois  $T = n\Delta t$ .

Se refinar a malha com  $\Delta t/\Delta x$  fixo, ou seja,  $\Delta t/\Delta x = r$  sendo  $r$  constante, então

$$D_{\Delta t}(X, T) \subset \left\{x; |x - X| \leq \frac{T}{r}\right\},$$

para todo passo de tempo  $\Delta t$ .

A **condição CFL** requer que

$$D(X, T) \subset D_{\Delta t}(X, T),$$

logo, usando a definição do domínio de dependência expressa na Eq. (2.5), tem-se

$$|(X - f'(u)T) - X| \leq \frac{T}{r},$$

ou seja

$$|f'(u)T| \leq \frac{T}{r},$$

que é equivalente a

$$\left| f'(u) \frac{\Delta t}{\Delta x} \right| \leq 1. \quad (2.8)$$

**Definição 2.3** A quantidade  $\alpha_{CFL} = \max |f'(u) \frac{\Delta t}{\Delta x}|$  é chamada de número de Courant ou número CFL.

Segundo Leveque (1992) [17], uma condição necessária para que um método numérico, com estêncil definido como acima, seja estável é que o número de Courant seja no máximo 1, ou seja,

$$\alpha_{CFL} = \max \left| f'(u) \frac{\Delta t}{\Delta x} \right| \leq 1. \quad (2.9)$$

### 2.1.3 Convergência

Para saber o quanto  $U_j^n$  aproxima a solução verdadeira  $u(x_j, t_n) = u_j^n$ , primeiro é preciso encontrar uma forma de quantificar o erro numérico. Para soluções suaves é interessante escolher o valor pontual  $u_j^n$  obtido pelo método de diferenças finitas, assim o erro pontual é dado por

$$E_j^n = U_j^n - u_j^n. \quad (2.10)$$

Para leis de conservação, em alguns momentos, é mais conveniente considerar o erro relativo dado por

$$\overline{E}_j^n = U_j^n - \overline{u}_j^n, \quad (2.11)$$

onde  $\overline{u}_j^n$ , dado na Eq. (2.2), obtido pelo método dos volumes finitos .

Se a função  $u$  é suficientemente suave o valor pontual  $U_j^n$  concorda com o valor médio da célula  $I_j$  em Eq. (2.2) na ordem  $\Delta x^2$ . Esta propriedade está demonstrada na Eq. (3.24).

Para unificar os casos acima, define-se a função erro dada por

$$E_{\Delta t}(x, t) = U_{\Delta t}(x, t) - u_{\Delta t}(x, t), \quad (2.12)$$

sendo  $E_j^n$  o valor nos pontos da malha de  $E_{\Delta x}(x_j, t_n)$ ,  $\overline{E}_j^n$  é o valor médio de  $E_{\Delta x}$  no tempo  $t_n$  e  $U_{\Delta t}(x, t) = U_j^n$  para  $(x, t) \in \left[x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}\right) \times [t_n, t_{n+1})$ .

Para quantificar o erro deve-se escolher alguma norma, para que seja medido o erro em um tempo fixo e também para que se verifique a convergência do método numérico.

**Definição 2.4** *Um método numérico é dito convergente em uma norma  $\|\cdot\|$  se*

$$\|E_{\Delta t}(\cdot, t)\| \rightarrow 0 \quad \text{quando} \quad \Delta t \rightarrow 0, \quad (2.13)$$

para todo  $t > 0$  fixo e para qualquer dado inicial  $u_0$ .

A norma mais natural para leis de conservação, em geral, é a norma  $L_1$ ; ao passo que para equações lineares a norma mais apropriada é a norma  $L_2$ . Para o caso discreto a norma  $L_1$  é definida como

$$\|U^n\|_1 = \Delta x \sum_{j=-\infty}^{+\infty} |U_j^n|. \quad (2.14)$$

Segundo Leveque (2004) [18], o fator  $\Delta x$  na Eq. (2.14) é muito importante para corrigir a escala e a ordem de precisão, à medida que a malha é refinada. Para o caso contínuo as normas  $L_1$  e  $L_2$  são definidas nas Eqs. (2.15) e (2.16), respectivamente.

$$\|v\|_1 = \int_{-\infty}^{+\infty} |v(x)| dx \quad \text{e} \quad (2.15)$$

$$\|v\|_2 = \left[ \int_{-\infty}^{+\infty} |v(x)|^2 dx \right]^{1/2}. \quad (2.16)$$

Neste trabalho será considerada apenas a norma  $L_1$ , dispensando assim o subscrito.

Para análise da estabilidade e da consistência do método numérico será utilizado o erro de truncamento local definido na subseção 2.1.4.

### 2.1.4 Estabilidade e Consistência

Geralmente é impossível obter uma forma para o erro global (Eqs. (2.10) e (2.11)) após centenas ou milhares de passos de tempos. Ao invés de se obter o erro global diretamente, o estudo do erro numérico pode ser obtido em um passo de tempo simples resultando em um erro local (Leveque (2004) [18]). O **erro de truncamento local** fornece uma indicação da magnitude do erro global, e particularmente da ordem de precisão, em casos onde o método é estável.

O erro de truncamento local, que será indicado por  $L_{\Delta t}$ , é um operador que avalia se  $U_j^n$  é uma boa aproximação da solução verdadeira  $u(x_j, t_n)$  nos pontos da malha computacional. Assim erro de truncamento local computa, localmente, a discrepância entre esses valores. Para

o cálculo do erro de truncamento local, considera-se que as soluções  $u$  são suaves e substitui-se o valor da aproximação  $U_j^n$  pelo valor da solução da equação  $u(x, t)$ . Será visto um exemplo na seção 2.1.5. (Veja Leveque (1992) [17] e Leveque (2004) [18]).

**Definição 2.5** Um método numérico é dito consistente em uma norma  $\|\cdot\|$  se

$$\|L_{\Delta t}(\cdot, t)\| \rightarrow 0 \quad \text{quando} \quad \Delta t \rightarrow 0, \quad (2.17)$$

para todo  $t > 0$  fixo.

**Definição 2.6** Um método numérico é dito de ordem  $p$  se

$$\lim_{\Delta t \rightarrow 0} \frac{|L_{\Delta t}(x, t)|}{\Delta t^p} = \text{constante}, \quad (2.18)$$

para todo  $t > 0$  fixo.

**Observação 2.2** Seja  $f$  uma função real, dizemos que  $f(h) = O(h^k)$  se  $\lim_{h \rightarrow 0} \left| \frac{f(h)}{h^k} \right| = \text{constante}$ .

A definição formal de estabilidade pode ser encontrada nos livros Leveque (1992) [17] e Leveque (2004) [18]. Neste trabalho considera-se que um método numérico é estável se

$$\|U^{n+1}\| \leq \|U^n\|. \quad (2.19)$$

### 2.1.5 Método de Lax - Friedrichs para Equação de Advecção

O método de Lax-Friedrichs é um método explícito de nível dois para funções  $u$  suaves, baseado na expansão em série de Taylor do estêncil  $\{U_{j-1}^n, U_{j+1}^n\}$ , ilustrado na Figura 2.3, esse estêncil corresponde a  $\{u(x - \Delta x, t), u(x + \Delta x, t)\}$  no caso contínuo.

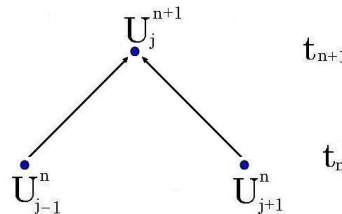


Figura 2.3: Diagrama para o Método de Lax-Friedrichs

Fazendo a expansão em série de Taylor de  $u(x, t)$  na variável  $x$ , para aproximar  $u_x$ , obtém-se

$$u(x + \Delta x, t) = u(x, t) + u_x(x, t) \cdot \Delta x + u_{xx}(x, t) \cdot \frac{\Delta x^2}{2!} + u_{xxx}(\tilde{x}, t) \cdot \frac{\Delta x^3}{3!},$$

$\tilde{x}$  entre  $x$  e  $x + \Delta x$ .

$$u(x - \Delta x, t) = u(x, t) - u_x(x, t) \cdot \Delta x + u_{xx}(x, t) \cdot \frac{\Delta x^2}{2!} - u_{xxx}(\hat{x}, t) \cdot \frac{\Delta x^3}{3!},$$

$\hat{x}$  entre  $x$  e  $x + \Delta x$ . Portanto temos

$$u(x + \Delta x, t) - u(x - \Delta x, t) = 2u_x(x, t) \cdot \Delta x + O(\Delta x^3),$$

de onde concluimos

$$u_x(x, t) = \frac{u(x + \Delta x, t) - u(x - \Delta x, t)}{2\Delta x} - O(\Delta x^2). \quad (2.20)$$

Fazendo a expansão em série de Taylor de  $u(x, t)$  na variável  $t$ , para aproximar  $u_t$ , obtém-se

$$u(x, t + \Delta t) = u(x, t) + u_t(x, t) \cdot \Delta t + u_{tt}(x, \tilde{t}) \cdot \frac{\Delta t^2}{2!},$$

para  $\tilde{t}$  entre  $t$  e  $t + \Delta t$ . Então, devemos ter

$$u_t(x, t) = \frac{u(x, t + \Delta t) - u(x, t)}{\Delta t} - O(\Delta t). \quad (2.21)$$

Substituindo as Eqs. (2.21) e (2.20) na equação de advecção  $u_t + au_x = 0$ , obtém-se

$$\frac{u(x, t + \Delta t) - u(x, t)}{\Delta t} - O(\Delta t) + a \cdot \left[ \frac{u(x + \Delta x, t) - u(x - \Delta x, t)}{2\Delta x} - O(\Delta x^2) \right] = 0,$$

ou seja

$$\begin{aligned} u(x, t + \Delta t) - u(x, t) &= -a \cdot \frac{\Delta t}{2\Delta x} [u(x + \Delta x, t) - u(x - \Delta x, t)] + \\ &O(\Delta x^2) \cdot a \cdot \Delta t + \Delta t \cdot O(\Delta t). \end{aligned} \quad (2.22)$$

Fazendo, novamente, a expansão em série de Taylor de  $u(x, t)$  na variável  $x$ , escrevemos

$$u(x + \Delta x, t) = u(x, t) + u_x(\tilde{x}, t) \cdot \Delta x,$$

com  $\tilde{x}$  entre  $x$  e  $x + \Delta x$  e

$$u(x - \Delta x, t) = u(x, t) - u_x(\hat{x}, t) \cdot \Delta x,$$

com  $\hat{x}$  entre  $x$  e  $x + \Delta x$ . Portanto, reunindo estas conclusões

$$u(x, t) = \frac{u(x + \Delta x, t) + u(x - \Delta x, t)}{2} + O(\Delta x). \quad (2.23)$$

Substituindo a Eq. (2.23) na Eq. (2.22), obtém-se

$$u(x, t + \Delta t) = \frac{u(x + \Delta x, t) + u(x - \Delta x, t)}{2} - a \frac{\Delta t}{2\Delta x} [u(x + \Delta x, t) - u(x - \Delta x, t)] + O(\Delta x^2) \cdot a \cdot \Delta t + O(\Delta t^2) + O(\Delta x) \quad (2.24)$$

ou seja, o método de Lax-Friedrichs na forma contínua é dado por

$$u(x, t + \Delta t) = \frac{1}{2} (u(x + \Delta x, t) + u(x - \Delta x, t)) - a \frac{\Delta t}{2\Delta x} [u(x + \Delta x, t) - u(x - \Delta x, t)] + O(\Delta x^2) + O(\Delta t^2). \quad (2.25)$$

No caso discreto é dado por

$$U_j^{n+1} = \frac{1}{2} (U_{j-1}^n + U_{j+1}^n) - \frac{\Delta t}{2\Delta x} a (U_{j+1}^n - U_{j-1}^n). \quad (2.26)$$

O operador erro de truncamento local,  $L_{\Delta t}(x, t)$ , para o método de Lax - Friedrichs, obtido da Eq. (2.25), desconsiderando os termos  $O(\Delta t^2)$  e  $O(\Delta x^2)$ , ou seja

$$L_{\Delta t}(x, t) = \left[ u(x, t + \Delta t) - \frac{1}{2} (u(x + \Delta x, t) + u(x - \Delta x, t)) \right] \frac{1}{\Delta t} + \frac{a}{2\Delta x} (u(x + \Delta x, t) + u(x - \Delta x, t)). \quad (2.27)$$

Como  $u$  é suave pode-se expandir  $L_{\Delta t}(x, t)$  em série de Taylor (para facilidade de notação será omitido  $(x, t)$ ), e eliminando os termos opostos, obtém-se

$$L_{\Delta t}(x, t) = (u_t + au_x) + \frac{1}{2} \left( \Delta t u_{tt} - \frac{\Delta x^2}{\Delta t} u_{xx} \right) + O(\Delta x^2). \quad (2.28)$$

Se  $u$  é solução de  $(u_t + au_x) = 0$ , então Eq. (2.28) se reduz a

$$L_{\Delta t}(x, t) = \frac{1}{2} \Delta t \left( u_{tt} - \frac{\Delta x^2}{\Delta t^2} u_{xx} \right) + O(\Delta x^2). \quad (2.29)$$

Por outro lado, sabendo que  $u_t = -au_x$ , tem-se que

$$u_{tt} = -au_{xt} = -au_{tx} = -a(-au_{xx}) = a^2 u_{xx}. \quad (2.30)$$

Substituindo este resultado na Eq. (2.29), obtém-se

$$L_{\Delta t}(x, t) = \frac{1}{2} \Delta t \left( a^2 - \frac{\Delta x^2}{\Delta t^2} \right) u_{xx} + O(\Delta x^2). \quad (2.31)$$

Como a razão  $\frac{\Delta t}{\Delta x}$  é constante podemos escrever, por exemplo,  $\Delta x = c\Delta t$  onde  $c$  é uma constante. Então  $\frac{\Delta x^2}{\Delta t^2} = c^2$  também é constante. Portanto a Eq. (2.31) pode ser escrita em função de  $\Delta t$  ao invés de  $\Delta x$ , isto é,

$$L_{\Delta t}(x, t) = \frac{1}{2} \Delta t (a^2 - c^2) u_{xx} + O(\Delta t^2). \quad (2.32)$$

Agora, usando a Eq. (2.18), tem-se que

$$\lim_{\Delta t \rightarrow 0} \frac{|L_{\Delta t}(x, t)|}{\Delta t} = \lim_{\Delta t \rightarrow 0} \left[ \frac{1}{2} (a^2 - c^2) u_{xx}(x, t) + O(\Delta t) \right] = cte = \frac{1}{2} (a^2 - c^2) u_{xx}(x, t). \quad (2.33)$$

Observe que na Eq. (2.33) o ponto  $(x, t)$  está fixado. Assim, de acordo com a definição 2.6, o método de Lax - Friedrichs para a equação de advecção é de ordem 1 no tempo, ou seja,  $L_{\Delta t}(x, t) = O(\Delta t)$  quando  $\Delta t \rightarrow 0$ . Com análise semelhante, verifica-se que esse método é de ordem  $O(\Delta x^2)$  no espaço, logo a ordem do método é  $O(\Delta t, \Delta x^2)$  quando  $\Delta t \rightarrow 0$ .

De acordo com Leveque (1992) [17] é possível mostrar que o método de Lax-Friedrichs, para a equação de advecção, é consistente.

Será analisada, agora, a estabilidade do método de Lax-Friedrichs. De acordo com a condição CFL, Eq. (2.8), para a equação de advecção, o método de Lax - Friedrichs é estável se

$$\left| a \frac{\Delta t}{\Delta x} \right| \leq 1. \quad (2.34)$$

Será usado a condição Eq. (2.19) para mostrar a estabilidade do método, usando a forma discreta do método de Lax-Friedrichs Eq. (2.26), logo tem-se

$$\begin{aligned} \|U^{n+1}\|_1 &= \Delta x \sum_{j=-\infty}^{+\infty} |U_j^{n+1}| = \Delta x \sum_{j=-\infty}^{+\infty} \left| \frac{1}{2} (U_{j-1}^n + U_{j+1}^n) - \frac{\Delta t}{2\Delta x} a (U_{j+1}^n - U_{j-1}^n) \right| \\ &= \Delta x \sum_{j=-\infty}^{+\infty} \left| U_{j+1}^n \left( \frac{1}{2} - a \frac{\Delta t}{2\Delta x} \right) + U_{j-1}^n \left( \frac{1}{2} + a \frac{\Delta t}{2\Delta x} \right) \right| \\ &\leq \frac{\Delta x}{2} \left[ \sum_{j=-\infty}^{+\infty} \left| U_{j+1}^n \left( 1 - a \frac{\Delta t}{\Delta x} \right) \right| + \sum_{j=-\infty}^{+\infty} \left| U_{j-1}^n \left( 1 + a \frac{\Delta t}{\Delta x} \right) \right| \right]. \end{aligned}$$

De Eq. (2.34) tem-se que

$$1 - a \frac{\Delta t}{2\Delta x} \geq 0$$

e

$$1 + a \frac{\Delta t}{2\Delta x} \geq 0,$$

logo podemos concluir

$$\begin{aligned} \|U^{n+1}\|_1 &\leq \frac{\Delta x}{2} \left[ \left( 1 - a \frac{\Delta t}{\Delta x} \right) \sum_{j=-\infty}^{+\infty} |U_{j+1}^n| + \left( 1 + a \frac{\Delta t}{\Delta x} \right) \sum_{j=-\infty}^{+\infty} |U_{j-1}^n| \right] \\ &\leq \frac{1}{2} \left[ \left( 1 - a \frac{\Delta t}{\Delta x} \right) \|U^n\|_1 + \left( 1 + a \frac{\Delta t}{\Delta x} \right) \|U^n\|_1 \right] = \|U^n\|_1 \end{aligned}$$

e portanto

$$\|U^{n+1}\|_1 \leq \|U^n\|_1. \quad (2.35)$$

Assim, conclui-se que o método de Lax - Friedrichs para a equação de advecção é estável. De acordo com o teorema fundamental da convergência dos métodos numéricos lineares, tem-se que o método numérico de Lax - Friedrichs para a equação de advecção é convergente.

### 2.1.6 Método Upwind

Para problemas hiperbólicos é esperado que a informação se propague ao longo das curvas características. Tanto a direção quanto a velocidade de propagação da informação, ao longo das curvas características, varia de equação para equação, então é preciso considerar cada caso separadamente.

Para a lei de conservação escalar de advecção, se  $a > 0$  a propagação da informação ocorre da esquerda para a direita com velocidade  $a$  ao passo que se  $a < 0$  a propagação ocorre da direita para a esquerda, também com velocidade  $a$ . Logo, um método numérico utilizando um estêncil centrado normalmente não é a melhor escolha, veja Figuras 2.5 e 2.6.

**Métodos ‘upwind’** são métodos numéricos os quais buscam os resultados numéricos na melhor direção, ou seja, a direção com a qual se propaga a informação característica.

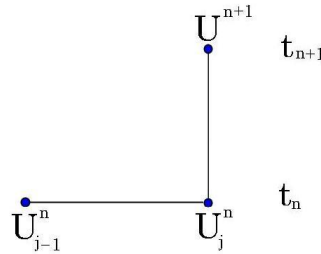


Figura 2.4: Diagrama do método ‘upwind’ quando  $a > 0$

O método ‘upwind’ explícito de nível dois, para a equação de advecção com  $a > 0$ , definido abaixo, na Eq. (2.36), para funções  $u$  suaves, foi desenvolvido utilizando-se o estêncil  $\{U_{j-1}^n, U_j^n\}$  e a expansão em série de Taylor.

$$U_j^{n+1} = U_j^n - \frac{\Delta t}{\Delta x} a (U_j^n - U_{j-1}^n) \quad (2.36)$$

Com análise semelhante à descrita para o método de Lax-Friedrichs, tem-se que o método ‘upwind’ é estável (se  $0 \leq a \frac{\Delta t}{\Delta x} \leq 1$ ), é consistente e portanto convergente. A ordem do método ‘upwind’ Eq. (2.36) é de  $O(\Delta t^2, \Delta x)$ .

Para leis de conservação o interesse é calcular a convergência no intervalo no qual ocorre a formação do choque, ou seja, no intervalo no qual ocorre a descontinuidade da solução. Será

ilustrado nas Figuras 2.5 e 2.6 o método de Lax-Friedrichs (Eq. (2.26)), e o método ‘upwind’ (Eq. (2.36)), respectivamente, para o problema de Riemann definido pela equação de advecção Eq. (2.1) para  $a = 1$ :

$$u_t + u_x = 0,$$

com o dado inicial  $u_0$  dado por

$$u(x, 0) = u_0(x) = \begin{cases} 1, & \text{se } x \leq 0; \\ 0, & \text{se } x > 0, \end{cases} \quad (2.37)$$

no domínio  $0 \leq x \leq 1$ ,  $t \geq 0$ .

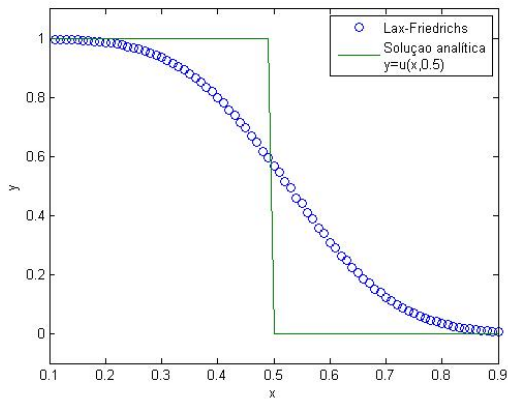


Figura 2.5: Método de Lax-Friedrichs: no domínio  $[0, 1]$ ,  $\Delta x = 0.01$ ,  $t = 0.5$ ,  $\Delta t = 0.02$  e  $CFL = 0.2$ .

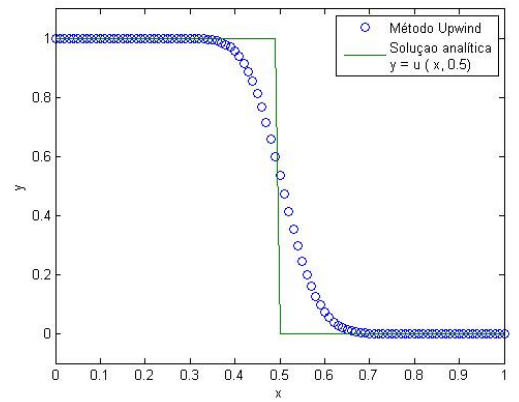


Figura 2.6: Método ‘Upwind’: no domínio  $[0, 1]$ ,  $\Delta x = 0.01$  e  $t = 0.5$ ,  $\Delta t = 0.02$ ,  $CFL = 0.2$ .

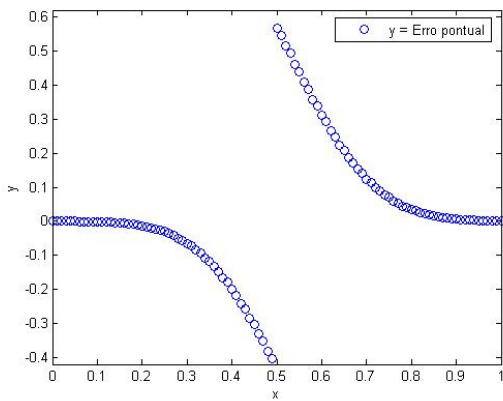


Figura 2.7: Ilustração do erro pontual do método de Lax-Friedrichs para  $N = 100$  no intervalo  $[0, 1]$ .

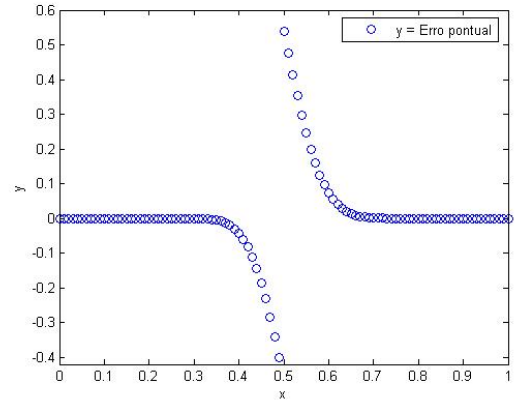


Figura 2.8: Ilustração do erro pontual do método ‘upwind’ para  $N = 100$  no intervalo  $[0, 1]$ .

As Figuras 2.5 e 2.6 ilustram a solução numérica do problema de Riemann, Eq. (4.2), calculada com os métodos numéricos descritos acima, com os seguintes dados:  $t = 0.5$ ,  $\Delta t = 0.02$ ,

$CFL = \Delta t / \Delta x = 0.2$ ; no domínio  $[0, 1]$  e  $\Delta x = 0.01$ .

Para comparação dos métodos numéricos usou-se a solução analítica do problema, de acordo com Eq. (1.56) dada por

$$u(x, t) = \begin{cases} 1, & \text{se } x < t, \\ 0, & \text{se } x > t. \end{cases} \quad (2.38)$$

| Método         | Erro: $L_1$  |
|----------------|--------------|
| Lax-Friedrichs | $1.219e - 1$ |
| ‘Upwind’       | $5.123e - 2$ |

Tabela 2.1: Tabela de erros para equação de advecção, na norma  $L_1$ , com  $N = 100$ .

Pode-se observar, nas Figuras 2.5 e 2.6, a suavização da solução numérica na região de descontinuidade, especialmente no método de Lax-Friedrichs. A tabela 2.1 mostra o erro, na norma  $L_1$  (Eq. (2.14)), entre as soluções numérica e analítica da equação de advecção, relativo ao método de Lax-Friedrichs e ao ‘upwind’. Os valores apresentados nesta tabela não são muito pequenos porque, na região onde ocorre o choque, qualquer método numérico possui baixa ordem de precisão e isto é evidenciado pela norma  $L_1$  do erro. As Figuras 2.7 e 2.8 ilustram os comportamentos dos erros pontuais para o método de Lax-Friedrichs e para o método ‘upwind’, respectivamente. Neste estudo, foi considerada uma malha computacional de  $N = 100$  subintervalos. Observe que na região onde a solução é suave o erro entre a solução numérica, de ambos os métodos, e a analítica está próximo de zero. Na região de descontinuidade, o método ‘upwind’ apresenta resultados mais precisos do que o método de Lax-Friedrichs, conforme mostra a tabela 2.1.

## 2.2 Métodos Numéricos Conservativos

Quando se resolvem leis de conservação não-lineares encontram-se dificuldades adicionais, tanto teóricas quanto numéricas, comparado com as equações lineares. Para soluções suaves  $u$ , os métodos numéricos de problemas não-lineares podem ser linearizados e os resultados da teoria dos métodos de diferenças finitas podem ser aplicados para se obter resultados de convergência. Porém este trabalho trata de problemas com soluções descontínuas e o uso destes métodos pode apresentar resultados desastrosos, de acordo com a literatura de leis de conservação. (Veja Leveque (1992) [17]).

As dificuldades adicionais referentes às equações não-lineares podem ser resumidas em:

- o método numérico pode ser instável;
- o método numérico pode convergir para uma solução errada, ou seja, pode convergir para uma solução que não é solução de entropia.

Um método na forma conservativa garante que a solução discreta seja conservativa. Para soluções fracas envolvendo choques, a forma integral da lei de conservação possui mais fundamentos do que a equação na forma diferencial e também forma a base da teoria matemática das soluções fracas, juntamente com a condição de Rankine-Hugoniot, a qual modela a forma e a velocidade das ondas de choques. Isto justifica o fato de que um método conservativo baseado na forma integral da lei de conservação pode ter mais sucesso do que métodos baseados na forma diferencial da equação. Para leis de conservação não-lineares é muito importante que o método numérico esteja na *forma conservativa* para possibilitar que a solução fraca, da lei de conservação, seja aproximada corretamente.

Para dedução formal da forma conservativa de um método numérico, considere a forma integral da lei de conservação no intervalo  $I_j = [x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}]$  dada por

$$\frac{d}{dt} \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} u(x, t) dx = f(u(x_{j-\frac{1}{2}}, t)) - f(u(x_{j+\frac{1}{2}}, t)). \quad (2.39)$$

Integrando a Eq. (2.39) em relação a  $t$  de  $t_n$  a  $t_{n+1}$ , obtém-se

$$\begin{aligned} \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} u(x, t_{n+1}) dx - \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} u(x, t_n) dx &= \int_{t_n}^{t_{n+1}} f(u(x_{j-\frac{1}{2}}, t)) dt - \\ &\int_{t_n}^{t_{n+1}} f(u(x_{j+\frac{1}{2}}, t)) dt \end{aligned} \quad (2.40)$$

e dividindo por  $\Delta x$  ambos os membros da Eq. (2.40), temos

$$\begin{aligned} \frac{1}{\Delta x} \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} u(x, t_{n+1}) dx &= \frac{1}{\Delta x} \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} u(x, t_n) dx + \frac{1}{\Delta x} \left[ \int_{t_n}^{t_{n+1}} f(u(x_{j-\frac{1}{2}}, t)) dt - \right. \\ &\left. \int_{t_n}^{t_{n+1}} f(u(x_{j+\frac{1}{2}}, t)) dt \right]. \end{aligned} \quad (2.41)$$

Usando a aproximação média, conforme a Eq. (2.2), pode-se escrever

$$U_j^n \equiv \frac{1}{\Delta x} \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} u(x, t_n) dx \quad (2.42)$$

e fazendo a função de fluxo numérico como sendo

$$F_{j+\frac{1}{2}} = \frac{1}{\Delta t} \int_{t_n}^{t_{n+1}} f(u(x_{j+\frac{1}{2}}, t)) dt, \quad (2.43)$$

a Eq. (2.41) se reduz a

$$U_j^{n+1} = U_j^n - \frac{\Delta t}{\Delta x} (F_{j+\frac{1}{2}} - F_{j-\frac{1}{2}}). \quad (2.44)$$

Para que um método numérico esteja na forma conservativa ele deve ser escrito de acordo com

a Eq. (2.44).

Segundo Leveque (2004) [18], a função de fluxo numérico  $F$  pode depender de vários argumentos, mas como para problemas hiperbólicos a informação se propaga com velocidade finita é razoável que a função de fluxo numérico dependa apenas de dois argumentos e a Eq. (2.44) pode ser escrita como

$$U_j^{n+1} = U_j^n - \frac{\Delta t}{\Delta x} (F(U_j^n, U_{j+1}^n) - F(U_{j-1}^n, U_j^n)). \quad (2.45)$$

Para ilustrar a importância do método numérico conservativo será exposto, para a equação de Burgers, um método ‘upwind’ na forma não conservativa e um método ‘upwind’ na forma conservativa.

### Forma não conservativa

Dada a equação de Burgers na forma quase-linear  $u_t + uu_x = 0$  (assumindo que  $U_j^n > 0$  para todo  $n, j$ ), fazendo-se uma pequena modificação no método ‘upwind’ (Eq. (2.36)) para a equação de advecção, obtém-se o método **‘upwind’ não conservativo** para a equação de Burgers

$$U_j^{n+1} = U_j^n - \frac{\Delta t}{\Delta x} U_j^n (U_j^n - U_{j-1}^n). \quad (2.46)$$

### Forma conservativa

Para obter um método numérico conservativo é preciso considerar a equação de Burgers na sua forma conservativa

$$u_t + \left( \frac{u^2}{2} \right)_x = 0,$$

onde a função fluxo é dada por  $f(u) = \frac{u^2}{2}$ . Neste caso, substituindo a função de fluxo numérico, dada por  $F(v, w) = \frac{v^2}{2}$ , na Eq. (2.45), obtém-se o método **‘upwind’ na sua forma conservativa**

$$U_j^{n+1} = U_j^n - \frac{\Delta t}{2\Delta x} \left( (U_j^n)^2 - (U_{j-1}^n)^2 \right). \quad (2.47)$$

Como condição inicial será considerado

$$u(x, 0) = u_0(x) = \begin{cases} 1.2, & \text{se } x \leq 0, \\ 0.4, & \text{se } x > 0, \end{cases} \quad (2.48)$$

no domínio  $0 \leq x \leq 1, \quad t \geq 0$ .

O estudo numérico do problema de Riemann, composto pela equação de Burgers e pela

Eq. (2.48), através dos métodos numéricos ‘upwind’ conservativo e não-conservativo, foi realizado no domínio  $[0, 1]$  com  $\Delta t/\Delta x = 0.5$ ,  $\Delta x = 0.01$ ;  $t = 1$ ,  $\Delta t = 0.005$ . Os resultados numéricos são ilustrados nas Figuras 2.9 e 2.10.

Para comparação dos métodos numéricos usou-se a solução analítica do problema, de acordo com Eq. (1.64), dada por:

$$u(x, t) = \begin{cases} 1.2, & \text{se } x < 0.8t, \\ 0.4, & \text{se } x > 0.8t. \end{cases} \quad (2.49)$$

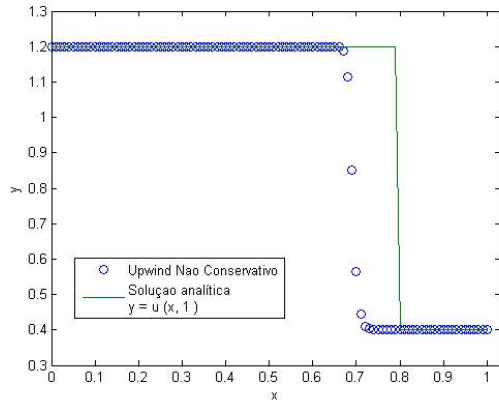


Figura 2.9: ‘Upwind’ não conservativo: no domínio  $[0, 1]$  com  $\Delta t/\Delta x = 0.5$ ,  $\Delta x = 0.01$ ;  $t = 1$ ,  $\Delta t = 0.005$ .

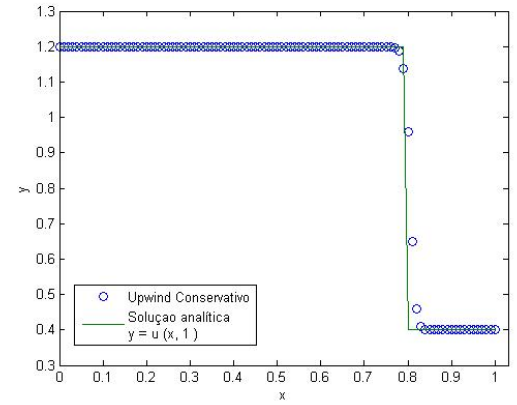


Figura 2.10: ‘Upwind’ conservativo: no domínio  $[0, 1]$  com  $\Delta t/\Delta x = 0.5$ ,  $\Delta x = 0.01$ ;  $t = 1$ ,  $\Delta t = 0.005$ .

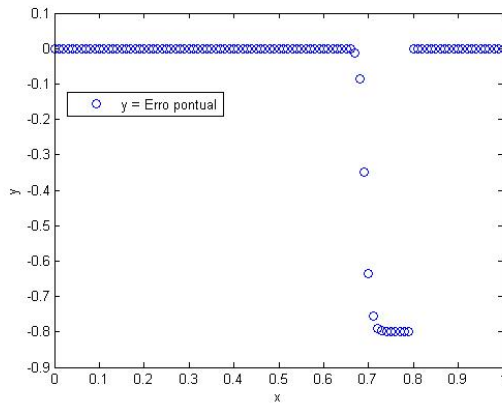


Figura 2.11: Ilustração do erro pontual do método ‘upwind’ não conservativo no intervalo  $[0, 1]$ .

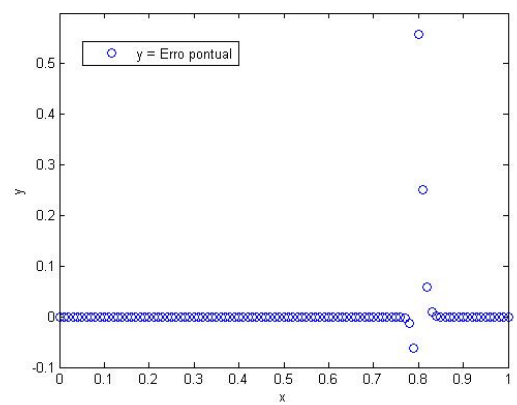


Figura 2.12: Ilustração do erro pontual para método ‘upwind’ conservativo no intervalo  $[0, 1]$ .

Pode-se observar que o método numérico não conservativo, ilustrado na Figura 2.9, apresentou um retardo da solução na região da formação de choque, ao passo que o método numérico conservativo, ilustrado na Figura 2.10, reproduziu com maior precisão a descontinuidade da solução, porém ainda apresenta suavização nesta região. A tabela 2.2 mostra o erro, na norma  $L_1$  (Eq. (2.14)), entre as soluções numérica e analítica da equação de Burgers, relativo ao

| Método                    | Erro: $L_1$  |
|---------------------------|--------------|
| ‘Upwind’ Conservativo     | $9.574e - 3$ |
| ‘Upwind’ Não Conservativo | $8.223e - 2$ |

Tabela 2.2: Tabela de erro para equação de Burgers, na norma  $L_1$ , com  $N = 100$ .

método ‘upwind’ conservativo e ao ‘upwind’ não conservativo. Os valores apresentados nesta tabela não são muito pequenos porque, na região onde ocorre o choque, qualquer método numérico possui baixa ordem de precisão e isto é evidenciado pela norma  $L_1$  do erro, especialmente para o método ‘upwind’ não conservativo devido ao atraso na solução numérica. As Figuras 2.11 e 2.12 ilustram os comportamentos dos erros pontuais para o método ‘upwind’ não conservativo e para o método ‘upwind’ conservativo, respectivamente. Neste estudo, foi considerada uma malha computacional de  $N = 100$  subintervalos.

## 2.3 Métodos Numéricos para Equações Não-Lineares

**Definição 2.7** *Um método numérico conservativo é consistente com a lei de conservação se a função numérica  $F$  se reduz à função fluxo verdadeira  $f$  para o fluxo constante, ou seja, se  $u(x, t) = \bar{u}$  então  $F(\bar{u}, \bar{u}) = f(\bar{u})$ .*

**Definição 2.8** *Uma função  $F(u, v)$  satisfaz à condição de Lipschitz em relação a  $\bar{u}$  se existe uma constante  $L \geq 0$  tal que*

$$|F(w, v) - f(\bar{u})| \leq L \max(|w - \bar{u}|, |v - \bar{u}|), \quad (2.50)$$

para todo  $|w - \bar{u}|$  e  $|v - \bar{u}|$  suficientemente pequenos. A constante  $L$  é chamada de constante de Lipschitz.

A constante  $L$  pode depender de  $\bar{u}$ .

Dizemos que  $F$  é Lipschitz contínua se é Lipschitz em todo ponto do domínio.

Segundo Leveque (1992) [17], para que a função de fluxo numérico  $F$  seja consistente é suficiente que  $F$  seja Lipschitz contínua em cada variável.

### 2.3.1 Estabilidade Não-Linear

O objetivo dessa seção é provar a convergência dos métodos numéricos não lineares, para isso é preciso que os métodos numéricos sejam estáveis. Será apresentada uma forma de estabilidade para uma ampla classe de métodos, no entanto essa técnica têm obtido sucesso completo apenas para o caso escalar, segundo Leveque (1992) [17]. Já existem resultados sobre convergência em alguns casos especiais, veja por exemplo DiPerna (1982) [8], Leveque e Temple (1985) [19] e Liu (1977) [23].

O fato de provar a convergência para equações escalares não garante que os métodos numéricos para equações não-lineares também sejam convergentes, mas na prática os métodos

numéricos que obedecem aos critérios impostos abaixo tem apresentado bons resultados. (Veja Leveque (1992) [17]).

A prova da convergência dos métodos numéricos para leis de conservação apresenta algumas dificuldades. A primeira delas é que o erro global dado por

$$E_{\Delta t}(x, t) = U_{\Delta t}(x, t) - u(x, t),$$

não é bem definido quando a solução  $u$  não é única. Para resolver esse problema, o erro global é medido como sendo a distância da aproximação numérica  $U_{\Delta t}(x, t)$  ao conjunto de todas as soluções fracas (Definição 1.4), definido como

$$W = \{v(x, t); v \text{ é uma solução fraca da lei de conservação}\}. \quad (2.51)$$

Para se medir essa distância a norma utilizada é a norma  $L_1$  sobre um intervalo de tempo finito  $[0, T]$  definida por:

$$\|v\|_{1,T} = \int_0^T \|v(\cdot, t)\|_1 dt = \int_0^T \int_{-\infty}^{+\infty} |v(x, t)| dx dt. \quad (2.52)$$

Essa norma é chamada de **norma-T** e o erro global é definido por

$$E_{\Delta t}(x, t) = \text{dist}(U_{\Delta t}, W) = \inf_{v \in W} \|U_{\Delta t} - v\|_{1,T}. \quad (2.53)$$

Provar a convergência dos métodos numéricos é o mesmo que provar que  $\text{dist}(U_{\Delta t}, W) \rightarrow 0$  quando  $\Delta t \rightarrow 0$ , sendo  $W$  definido na Eq. (2.51).

Para provar a convergência dos métodos numéricos para equações não-lineares se faz necessário o uso da definição apropriada de estabilidade para tais equações. A convergência desses métodos só é possível em conjuntos compactos. Para todo espaço linear normado de dimensão finita, conjunto compacto é um conjunto fechado e limitado; porém em espaços de dimensão infinita não se têm essa caracterização, logo é preciso uma definição geral, que seja equivalente à citada anteriormente.

**Definição 2.9** *Um conjunto  $K$  é um conjunto compacto se toda cobertura aberta de  $K$  admite subcobertura finita.*

Veja Lima (2006) [20] e Conway (1990) [4] para maiores detalhes.

**Proposição 2.1** *Seja  $K$  um conjunto compacto definido em um espaço normado. Toda sequência infinita definida em  $K$  possui subsequência convergente para algum elemento de  $K$ .*

**Demonstração:** *Veja Conway (1990) [4].*

Toda construção definida abaixo segue na direção de encontrar um conjunto compacto para que se possa provar a convergência dos métodos numéricos, usando a proposição 2.1. Para isso,

seguem algumas definições.

Para evitar que apareçam oscilações que não sejam físicas, o método numérico precisa medir as oscilações da solução, isso se faz sob a teoria de variação total de uma função.

**Definição 2.10** Para uma função discreta  $U$  a variação total é definida como

$$TV(U) = \sum_{j=-\infty}^{+\infty} |U_j - U_{j-1}|. \quad (2.54)$$

**Definição 2.11** Para uma função arbitrária  $u(x)$  a variação total é definida como

$$TV(u) = \sup \sum_{j=1}^N |u(x_j) - u(x_{j-1})|, \quad (2.55)$$

sendo que o supremo é tomado sobre toda subdivisão da reta real  $-\infty = x_0 < x_1 < \dots < x_N = +\infty$ .

**Definição 2.12** O espaço de funções  $L_1$  consiste de todas as funções de  $x$  com norma  $\|v\|_1$  finita, ou seja,

$$L_1 = \{v : \mathbb{R} \longrightarrow \mathbb{R}; \|v\|_1 < \infty\}.$$

**Definição 2.13** O espaço de funções  $L_{1,T}$  consiste de todas as funções de  $x$  e de  $t$  com a norma  $\|v\|_{1,T}$  finita, ou seja,

$$L_{1,T} = \left\{ v : \mathbb{R} \longrightarrow \mathbb{R}; \|v\|_{1,T} < \infty \right\}.$$

O objetivo é definir um conjunto compacto em  $L_{1,T}$ ; porém primeiramente será construído um conjunto compacto em  $L_1$  para maior esclarecimento. O primeiro passo é considerar o conjunto

$$A = \{v \in L_1 : TV(v) \leq R \text{ e } \text{supp}(v) \subset [-M, M]\}, \quad (2.56)$$

que é um conjunto compacto.

**Proposição 2.2** Seja o conjunto  $A$  definido na Eq. (2.56). Então  $\|v\|_{L_1(\mathbb{R})} \leq MR$ , para toda  $v \in A$ .

**Demonstração:**

Considere  $x \in [-M, M]$  e uma partição do intervalo  $[-M, M]$  dada por  $P = \{-M, x, M\}$ . Como  $v$  têm suporte compacto em  $[-M, M]$  existe  $\delta > 0$  tal que  $v(-M - \delta) = 0$  e  $v(M + \delta) = 0$  logo

$$2|v(x)| = |v(-M - \delta) - v(x)| + |v(x) - v(M + \delta)| \leq TV(v) \leq R,$$

para  $\delta > 0$ . Assim pode-se concluir que

$$|v(x)| \leq \frac{R}{2}, \quad \text{para cada } x \in [-M, M].$$

Portanto

$$\|v\|_{L_1(\mathbb{R})} = \int_{-M}^M |v(x)| dx \leq \int_{-M}^M \frac{R}{2} dx = MR. \quad (2.57)$$

■

**Definição 2.14** define-se a variação total da função  $u \in L_{1,T}$ , no intervalo  $[0, T]$ , no espaço e no tempo como

$$\begin{aligned} TV_T(u) = & \limsup_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \int_0^T \int_{-\infty}^{+\infty} |u(x + \epsilon, t) - u(x, t)| dx dt + \\ & \limsup_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \int_0^T \int_{-\infty}^{+\infty} |u(x, t + \epsilon) - u(x, t)| dx dt. \end{aligned} \quad (2.58)$$

Desde que as funções  $U_{\Delta t}(x, t)$  sejam sempre constantes por partes, a definição 2.14 pode ser escrita como:

**Definição 2.15** Para dados discretos a definição 2.14 se reduz à

$$TV_T(U)^n = \sum_{n=0}^{T/\Delta t} \sum_{j=-\infty}^{+\infty} [\Delta t |U_{j+1}^n - U_j^n| + \Delta x |U_j^{n+1} - U_j^n|].$$

A definição acima pode ser escrita em termos da variação total, ou seja:

$$\textbf{Definição 2.16} \quad TV_T(U)^n = \sum_{n=0}^{T/\Delta t} [\Delta t TV(U^n) + \|U^{n+1} - U^n\|].$$

Para provar a convergência do método numérico no conjunto  $L_{1,T}$  trabalha-se com o seguinte conjunto

$$K = \{u \in L_{1,T}; TV_T(u) \leq R \quad e \quad \text{supp}(u(., t)) \subset [-M, M], \quad \forall \quad t \in [0, T]\}, \quad (2.59)$$

o qual é um conjunto compacto em  $L_{1,T}$ , de acordo com a definição 2.9. (Leveque (1992) [17]).

**Definição 2.17** Dizemos que um método numérico é de Variação Total Estável, ou simplesmente TV-estável, se toda aproximação  $U_{\Delta t}$  com  $\Delta t < \Delta t_0$ , está contido em algum conjunto fixo  $K$  da forma Eq. (2.59). Os números  $R$  e  $M$  podem depender do dado inicial  $u_0$ , da função fluxo  $f(u)$ , mas não do valor de  $\Delta t$ .

O teorema 2.1 exhibe uma forma mais simples de verificar se um método numérico é TV-estável.

**Teorema 2.1** *Considere um método numérico conservativo com uma função fluxo numérico  $F(U; j)$  Lipschitz contínua e suponha que, para cada dado inicial  $u_0$ , existem constantes  $\Delta t_0, R > 0$  tais que*

$$TV(U^n) \leq R, \text{ para todo } n, \Delta t \text{ com } \Delta t < \Delta t_0 \text{ e } n\Delta t \leq T.$$

*Então o método é TV-estável.*

Para a demonstração do teorema 2.1 é preciso o lema 2.1.

**Lema 2.1** *Considere um método numérico com as hipóteses do teorema 2.1. Então existe  $\alpha > 0$  tal que então  $\|U^{n+1} - U^n\|_1 \leq \alpha \Delta t$ , para todo  $n, \Delta t$  com  $\Delta t < \Delta t_0$  e  $n\Delta t \leq T$ .*

**Demonstração:**

*Um método numérico conservativo é da forma*

$$U_j^{n+1} = U_j^n - \frac{\Delta t}{\Delta x} [F(U_{j+1}^n, U_j^n) - F(U_j^n, U_{j-1}^n)], \quad (2.60)$$

*e trabalhando com essa equação, para todo  $j$ , obtém-se*

$$\Delta x \sum_{j=-\infty}^{+\infty} |U_j^{n+1} - U_j^n| = \Delta t \sum_{j=-\infty}^{+\infty} |F(U_{j+1}^n, U_j^n) - F(U_j^n, U_{j-1}^n)|.$$

*Pela definição de norma  $L_1$ , temos que*

$$\|U_j^{n+1} - U_j^n\|_1 = \Delta t \sum_{j=-\infty}^{+\infty} |F(U_{j+1}^n, U_j^n) - F(U_j^n, U_{j-1}^n)|. \quad (2.61)$$

*Da demonstração da proposição 2.2, considerando o caso discreto, tem-se que*

$$|U_j^n| \leq \frac{R}{2}, \text{ para todo } j, n \text{ com } nk < T,$$

*ou seja,  $U_j^n$  é uniformemente limitada. Como  $F$  é contínua e  $F(U^n; j)$  é Lipschitz contínua, tem-se que*

$$|F(U_{j+1}^n, U_j^n) - F(U_j^n, U_{j-1}^n)| \leq L \max \{|U_{j+1}^n - U_j^n|, |U_j^n - U_{j-1}^n|\},$$

*ou seja*

$$|F(U_{j+1}^n, U_j^n) - F(U_j^n, U_{j-1}^n)| \leq L \sum_{i=-1}^1 |U_{j+i}^n - U_{j+i-1}^n|.$$

*Como a equação acima é válida para todo  $j$ , concluímos que*

$$\sum_{j=-\infty}^{+\infty} |F(U_{j+1}^n, U_j^n) - F(U_j^n, U_{j-1}^n)| \leq L \sum_{i=-1}^1 \sum_{j=-\infty}^{+\infty} |U_{j+i}^n - U_{j+i-1}^n|.$$

Agora, multiplicando essa equação por  $\Delta t$ , usando a Eq. (2.61) e a definição de variação total em Eq. (2.54), obtém-se

$$\|U_j^{n+1} - U_j^n\|_1 \leq L \Delta t \sum_{i=-1}^1 TV(U^n), \quad (2.62)$$

logo devemos ter

$$\|U_j^{n+1} - U_j^n\|_1 \leq \Delta t (3 L R). \quad (2.63)$$

Portanto, existe  $\alpha = 3 L R$  tal que  $\|U_{n+1} - U_n\|_1 \leq \alpha \Delta t$ , para todo  $n, \Delta t$  com  $\Delta t < \Delta t_0$  e  $n\Delta t \leq T$ . ■

Agora procederemos à demonstração do teorema 2.1.

### Demonstração:

Devemos provar que a aproximação numérica  $U_{\Delta t}$  está contida em algum conjunto compacto  $K$  definido em Eq. (2.59).

Por hipótese temos  $TV(U^n) \leq R$ , logo pelo lema 2.1,  $\|U^{n+1} - U^n\|_1 \leq \alpha \Delta t$ . Considerando a variação total discreta da definição 2.15 tem-se que

$$\begin{aligned} TV_T(U^n) &= \sum_{n=0}^{T/\Delta t} [\Delta t TV(U^n) + \|U^{n+1} - U^n\|_1] \\ &\leq \sum_{n=0}^{T/\Delta t} (\Delta t R + \alpha \Delta t) \\ &\leq \Delta t (R + \alpha) \frac{T}{\Delta t} \\ &\leq (R + \alpha) T, \end{aligned}$$

para todo  $\Delta t < \Delta t_0$ . Pode-se, então se dizer que  $TV(U_n)$  é uniformemente limitada quando  $\Delta t \rightarrow 0$ . Por outro lado, como a lei de conservação tem velocidade de propagação finita, tem-se que o suporte de  $U^n$  é compacto e como consequência  $U_{\Delta t}$  está contido em um conjunto  $K$  compacto. Além disso, por definição o método é TV-estável. ■

Como parte preliminar do objetivo do capítulo, segue o teorema de Lax-Wendroff, o qual garante a convergência de um método numérico para uma solução fraca da lei de conservação. Porém a convergência é hipótese do teorema e não sua conclusão. Para demonstração desse teorema usa-se **soma por partes**, o que nada mais é do que uma reorganização dos termos de

um somatório. Dados quaisquer  $a_j$  e  $b_j$

$$\begin{aligned}
 \sum_{j=1}^m a_j (b_j - b_{j-1}) &= (a_1 b_1 - a_1 b_0) + (a_2 b_2 - a_2 b_1) + \dots + (a_m b_m - a_m b_{m-1}) \\
 &= -a_1 b_0 + (a_1 b_1 - a_2 b_1) + \dots + (a_{m-1} b_{m-1} - a_m b_{m-1}) + a_m b_m \\
 &= a_m b_m - a_1 b_0 - \sum_{j=1}^{m-1} (a_{j+1} - a_j) b_j.
 \end{aligned} \tag{2.64}$$

**Teorema 2.2** ( Teorema Lax-Wendroff) *Considere uma sequência de grelhas indexadas por  $l = 1, 2, \dots$ , com os respectivos parâmetros  $\Delta x_l, \Delta t_l \rightarrow 0$  quando  $l \rightarrow \infty$ . Se  $U_l(x, t)$  denota a aproximação numérica computada com um método numérico conservativo e consistente na  $l$ -ésima grelha. E ainda, se  $U_l(x, t)$  converge para uma função  $u(x, t)$  quando  $l \rightarrow \infty$  então  $u$  é uma solução fraca da lei de conservação.*

**Demonstração:**

Para demonstrar esse teorema assume-se que  $U_l$  converge quase-sempre para  $u$ , ou seja,  $U_l \rightarrow u$ , a menos de um conjunto de medida nula, e que  $TV(U_l) \leq R$ ,  $\forall 0 \leq t \leq T$ ,  $l = 1, 2, \dots$

Será mostrado que  $u$  é uma solução fraca da lei de conservação, ou seja, para toda  $\Phi \in C_0^1$  vale a Eq. (2.65)

$$\int_0^{+\infty} \int_{-\infty}^{+\infty} (\Phi_t u + \Phi_x f(u)) dx dt = - \int_{-\infty}^{+\infty} \Phi(x, 0) u. \tag{2.65}$$

Por hipótese o método numérico é conservativo, ver a Eq. (2.60). Multiplicando essa equação por  $\Phi(x_j, t_n)$ , sendo  $\Phi \in C_0^1$ , obtém-se

$$\Phi(x_j, t_n) U_j^{n+1} = \Phi(x_j, t_n) U_j^n - \Phi(x_j, t_n) \frac{\Delta t}{\Delta x} [F(U_{j+1}^n, U_j^n) - F(U_j^n, U_{j-1}^n)], \tag{2.66}$$

para cada  $n, j$ .

Para simplificar a notação, a indexação  $l$  da grelha em  $\Delta x^l$ ,  $\Delta t^l$  e  $U_l(x, t)$  será omitida e consequentemente  $U_{l,j}^n(x, t)$  será representado por  $U_j^n(x, t)$ .

Se a Eq. (2.66) for avaliada  $\forall j$  e  $\forall n \geq 0$  e realizando algumas modificações, na respectiva equação, será obtido

$$\sum_{n=0}^{+\infty} \sum_{j=-\infty}^{+\infty} \Phi(x_j, t_n) (U_j^{n+1} - U_j^n) = - \frac{\Delta t}{\Delta x} \sum_{n=0}^{+\infty} \sum_{j=-\infty}^{+\infty} \Phi(x_j, t_n) [F(U_{j+1}^n, U_j^n) - F(U_j^n, U_{j-1}^n)]. \tag{2.67}$$

Usando soma por partes na Eq. (2.67), em ambos os lados, sendo que no primeiro termo a soma será feita em  $n$  e no segundo termo em  $j$  e usando o fato de que  $\Phi$  tem suporte compacto em  $[a, b] \times [0, T]$  obtém-se

$$- \sum_{j=-\infty}^{+\infty} \Phi(x_j, t_0) U_j^0 - \sum_{j=-\infty}^{+\infty} \sum_{n=1}^{+\infty} (\Phi(x_j, t_{n+1}) - \Phi(x_j, t_n)) U_j^n = \frac{\Delta t}{\Delta x} \sum_{n=0}^{+\infty} \sum_{j=-\infty}^{+\infty} (\Phi(x_{j+1}, t_n) - \Phi(x_j, t_n)) F_{j-\frac{1}{2}}^n. \tag{2.68}$$

Agora multiplicando essa equação por  $\Delta x$  e rearranjando os termos, obtém-se:

$$\Delta t \Delta x \left[ \sum_{j=-\infty}^{+\infty} \sum_{n=1}^{+\infty} \frac{(\Phi(x_j, t_{n+1}) - \Phi(x_j, t_n))}{\Delta t} U_j^n + \sum_{n=0}^{+\infty} \sum_{j=-\infty}^{+\infty} \frac{(\Phi(x_{j+1}, t_n) - \Phi(x_j, t_n))}{\Delta x} F_{j-\frac{1}{2}}^n \right] = -\Delta x \sum_{j=-\infty}^{+\infty} \Phi(x_j, t_0) U_j^0. \quad (2.69)$$

essa equação é análoga a derivação da Eq. (2.65), ou seja, fazendo  $l \rightarrow \infty$  com  $\Delta t_l, \Delta x_l \rightarrow 0$  na Eq. (2.69), obtém-se a Eq. (2.65). Lembrando que  $U_{l_j}^n(x, t) = U_j^n(x, t)$ , que por hipótese  $\|U_l - u\|_{1,T} \rightarrow 0$  e que  $\Phi$  é suave, concluímos que

$$\lim_{\Delta t_l, \Delta x_l \rightarrow 0} \Delta t_l \Delta x_l \sum_{n=1}^{+\infty} \sum_{j=-\infty}^{+\infty} \frac{\Phi_j^{n+1} - \Phi_j^n}{\Delta t_l} U_j^n = \int_0^T \int_{-\infty}^{+\infty} \Phi_t(x, t) u(x, t) dx dt. \quad (2.70)$$

Se  $U_j^n$  é a aproximação do dado inicial  $u_0$ , o lado direito da Eq. (2.69) se reduz a

$$\lim_{\Delta x_l \rightarrow 0} -\Delta x_l \sum_{j=-\infty}^{\infty} \Phi_j^0 U_j^0 = - \int_{-\infty}^{+\infty} \Phi(x, 0) u(x, 0) dx. \quad (2.71)$$

Para mostrar a convergência do termo restante na Eq. (2.69), o qual envolve  $F_{j-\frac{1}{2}}$ , precisa-se de mais exigências sobre  $F$  e  $U$ . O valor de  $F_{j-\frac{1}{2}}$  depende de dois argumentos:  $U_{j-1}$  e  $U_j$ . Usando a consistência da  $F$ , temos que

$$|F(U_{j-1}, U_j) - f(U_l(x_j, t_n))| \leq L |U_j^n - U_{j-1}^n|, \quad (2.72)$$

sendo  $L$  é a constante de Lipschitz. Além disso, para todo  $T$  existe  $R > 0$  tal que  $TV(U_l) \leq R$ , ou seja,  $|U_l(x_j, t_n) - U_l(x_{j-1}, t_n)| \rightarrow 0$  quando  $l \rightarrow \infty$  quase-sempre. Como  $U_l \rightarrow u$  e  $f$  é suave então  $f(U_l) \rightarrow f(u)$  e conclui-se que

$$F_{j-\frac{1}{2}} = F(U_{j-1}, U_j) \rightarrow f(U_l(x_j, t_n)) \rightarrow f(u) \quad (2.73)$$

e assim

$$\lim_{\Delta t_l, \Delta x_l \rightarrow 0} \Delta t_l \Delta x_l \sum_{n=0}^{\infty} \sum_{j=-\infty}^{\infty} \frac{\Phi_{j+1}^n - \Phi_j^n}{\Delta x_l} F_{j-\frac{1}{2}} = \int_0^T \int_{-\infty}^{+\infty} \Phi_x(x, t) f(u) dx dt. \quad (2.74)$$

Finalmente, substituindo as Eqs. (2.74), (2.71) e (2.70) na Eq. (2.69), obtém-se a Eq. (2.65). Portanto  $u$  é uma solução fraca. ■

**Teorema 2.3** Considere uma aproximação  $U_{\Delta t}$  computada por um método numérico conservativo com uma função fluxo numérico  $F(U; j)$  Lipschitz, consistente com uma lei de conservação escalar. Se o método numérico é TV-estável então ele converge.

**Demonstração:**

Será provado que  $\text{dist}(U_{\Delta t}, W) \rightarrow 0$  quando  $\Delta t \rightarrow 0$ .

Suponhamos, por absurdo, que  $\text{dist}(U_{\Delta t}, W)$  não converge para 0 quando  $\Delta t \rightarrow 0$ , ou seja,  $\forall \epsilon > 0$  existe uma sequência de aproximações  $\{U_{\Delta t_1}, U_{\Delta t_2}, \dots\}$  tal que  $\Delta t_j \rightarrow 0$  quando  $j \rightarrow \infty$  e

$$\text{dist}(U_{\Delta t_j}, W) > \epsilon \quad \text{quando } j \rightarrow \infty.$$

Como  $U_{\Delta t}$  é TV-estável,  $U_{\Delta t_j} \in K \quad \forall j$  e  $\Delta t$  é compacto, então  $U_{\Delta t_j}$  possui subsequência convergente em  $K$ . Pode-se supor, sem perda de generalidade, que  $U_{\Delta t_j} \rightarrow v$  em  $K$ , logo

$$\text{dist}(U_{\Delta t_j}, W) < \epsilon$$

quando  $j \rightarrow \infty$ . Pelo teorema de Lax-Wendroff temos que  $v \in W$ , pois é uma solução fraca. Absurdo, portanto devemos ter

$$\lim_{\Delta t \rightarrow 0} \text{dist}(U_{\Delta t}, W) = 0.$$

■

### 2.3.2 Convergência para Solução de Entropia

O restante deste capítulo se destina a obter mais propriedades para uma importante classe de métodos, bem como a construção da teoria necessária para que se tenha a convergência para a solução de entropia.

Para que um método numérico seja convergente é necessário que ele seja TV-estável. Um caminho para assegurar este fato é exigir que tais métodos tenham variação total não-crescente quando o tempo evolui. Assim a variação total será uniformemente limitada pela variação total do dado inicial, quando evoluir o tempo. Os métodos numéricos que têm variação total não-crescente são conhecidos como métodos **TVD** (do inglês Total Variation Diminishing) e garante que o método numérico não será oscilatório.

**Definição 2.18** Um método numérico é chamado de Variação Total Decrescente (TVD) se

$$TV(U^{n+1}) \leq TV(U^n),$$

para toda função numérica discreta  $U^n$ .

Como mencionado anteriormente, se um método numérico é TVD então

$$TV(U^n) \leq TV(U^0) \leq TV(u_0), \quad \forall \quad n \geq 0.$$

Pode-se mostrar que a solução verdadeira da lei de conservação escalar tem a propriedade TVD, isto é,

$$TV(u(., t_2)) \leq TV(u(., t_1)), \quad \forall \quad t_2 \geq t_1.$$

Uma dificuldade associada às aproximações numéricas é o surgimento de oscilações nas descontinuidades. Uma exigência natural para eliminar essa possibilidade é que o método numérico preserve a monotonicidade.

Se a aproximação numérica do dado inicial  $U_j^0$  é uma função monótona de  $j$  (não-crescente ou não-decrescente), então, se a solução  $U_j^n$  possui a mesma propriedade de monotonicidade para todo  $n$  dizemos que o método numérico **preserva monotonicidade**. Ou seja, se  $U_j^0 \geq U_{j+1}^0$ , para todo  $j$  então  $U_j^n \geq U_{j+1}^n$ , para todo  $n, j$ . Um exemplo de método numérico que preserva monotonicidade é o método ‘upwind’ conservativo, para equação de Burgers, dado pela Eq. (2.47) e representado na Figura 2.10. Dessa forma as oscilações não surgem na propagação das descontinuidades.

**Teorema 2.4** *Todo método numérico TVD preserva monotonicidade.*

**Demonstração:** Ver Leveque (1992) [17].

**Definição 2.19** *Dizemos que um método numérico  $U$  é monótono se toda aproximação numérica obedecer*

$$V_j^n \geq U_j^n \quad \Rightarrow \quad V_j^{n+1} \geq U_j^{n+1}, \quad (2.75)$$

para todo  $j$ .

Para verificar se um método numérico é monótono é suficiente provar que

$$\frac{\partial U_j^{n+1}}{\partial U_i^n} \geq 0, \quad \text{para todo } i, j, U^n, U^{n+1}, \quad (2.76)$$

ou seja, se o valor de  $U_i^n$  cresce então o valor de  $U_j^{n+1}$  não pode decrescer.

A generalização do método de Lax-Friedrichs para equações não-lineares é feita considerando a função fluxo numérica

$$F(U_j, U_{j+1}) = \frac{\Delta x}{2\Delta t} (U_j - U_{j+1}) + \frac{1}{2} (f(U_j) + f(U_{j+1})), \quad (2.77)$$

no método conservativo Eq. (2.45), obtendo

$$U_j^{n+1} = \frac{1}{2} (U_{j+1}^n + U_{j-1}^n) - \frac{\Delta t}{2\Delta x} (f(U_{j+1}^n) - f(U_{j-1}^n)). \quad (2.78)$$

**Exemplo 2.1** O método numérico de Lax-Friedrichs Eq. (2.78) é monótono sob a condição CFL.

O método de Lax-Friedrichs é dado na Eq. (2.78), logo

$$\frac{\partial}{\partial U_i^n} U_j^{n+1} = \begin{cases} \frac{1}{2} \left[ 1 + \frac{\Delta t}{\Delta x} f' (U_{j-1}^n) \right], & \text{se } i = j - 1 \\ \frac{1}{2} \left[ 1 - \frac{\Delta t}{\Delta x} f' (U_{j+1}^n) \right], & \text{se } i = j + 1 \\ 0 & \text{nos outros casos.} \end{cases}$$

e a condição CFL garante que  $\left[ 1 \pm \frac{\Delta t}{\Delta x} f' (U_i^n) \right] \geq 0$  para todo  $i$ . Então

$$\frac{\partial}{\partial U_i^n} U_j^{n+1} \geq 0, \quad \forall i, j. \quad (2.79)$$

De acordo com Eq. (2.76) o método de Lax-Friedrichs é monótono.

Termina-se o capítulo com o teorema que garante a convergência do método numérico para a solução fracamente correta, ou seja, para a solução de entropia.

**Teorema 2.5** A solução numérica computada com um método monótono consistente com  $\Delta t / \Delta x$  fixo, converge para a solução de entropia quando  $\Delta t \rightarrow 0$ .

**Demonstração:** Ver Grandall e Majda (1980) [11] e Harten; Hyman e Lax (1976) [12].

## Capítulo 3

# Discretização Espacial e Temporal

Tanto os métodos numéricos de diferenças finitas (ou volumes finitos) quanto a interpolação polinomial tradicionais são baseados em estênceis fixos. Tais métodos funcionam bem para problemas cujas funções são suaves, porém para funções descontínuas esses métodos não conseguem boas aproximações nas regiões das discontinuidades. Podem suavizar a solução na região da discontinuidade, como visto nas Figuras 2.5 e 2.6. O método numérico Essencialmente Não-Oscilatório (ENO), proposto por Harten *et. al.* (1987) [13],[14], apresenta resultados significativamente melhores com respeito a diminuição da dissipação numérica além de produzirem soluções essencialmente não-oscilatórias (ou seja, além de evitarem ao máximo oscilações) nas regiões de discontinuidades (choques) da solução. O esquema ENO nada mais é do que uma interpolação polinomial, baseada no método clássico de diferenças finitas sobre um estêncil adaptativo, sendo que o número de pontos do estêncil é fixado a priori. A ordem de precisão do esquema ENO na região da discontinuidade pode ser escolhida tão alta quanto se queira. O melhor estêncil, dentro das possibilidades existentes com a mesma quantidade de pontos na malha numérica, é escolhido de acordo com o valor das diferenças divididas. Supondo que  $k$  é a quantidade de pontos escolhidos para o estêncil veremos que a ordem do esquema ENO será  $O(\Delta x^k)$ .

Na literatura existem diferentes abordagens para a construção desses métodos numéricos. Neste trabalho será apresentado o procedimento baseado na reconstrução por função primitiva, segundo Chi-Wang Shu (1997) [29].

O esquema numérico Essencialmente Não-Oscilatório Ponderado (do inglês Essentially Nonoscillatory Weighted - WENO) foi desenvolvido por Liu, Osher e Chan (1994) [22] usando a combinação convexa de todos os possíveis polinômios, obtidos na busca realizada pelo método ENO dentro do estêncil, resultando em um método numérico que apresenta, na região onde a solução é suave, resultados de ordem mais elevada do que a ordem do interpolador ENO e preserva a ordem deste interpolador na região onde a solução é descontínua. O sucesso do WENO está relacionado aos pesos da combinação convexa. Depois disto, Jiang e Shu (1996) [15] introduziram, no método numérico WENO, novos pesos ( $\beta_r$ ) que melhoraram ainda mais

a ordem de precisão. Estes coeficientes  $\beta_r$  são chamados de sensores de suavidade, ou seja, são capazes de detectar a regularidade de uma função em relação aos pontos de um determinado estêncil  $S_r$ .

Os esquemas numéricos ENO e WENO (apêndice A) foram construídos de modo que a propriedade TVD fosse satisfeita. Na prática é o que realmente acontece nos resultados numéricos, porém não se conseguiu obter esse resultado analiticamente. O fato é que acerca deste assunto pouco se têm demonstrado e o que se conseguiu provar, até o momento, é apenas para o caso unidimensional. (Veja Chi-Wang Shu (1997) [29]).

O processo de discretização do domínio será realizado em dois estágios, primeiro a discretização é feita somente no espaço considerando o problema contínuo no tempo (métodos semi-discretos). Isto leva a EDP à um sistema de EDO's no tempo, chamadas de equações semi-discretas. A discretização no tempo é feita utilizando um método numérico padrão para sistemas de EDO's. Esta técnica é utilizada, particularmente, no desenvolvimento de métodos de ordem maior que dois. Pode-se definir alta ordem de aproximação para a função fluxo nas fronteiras das células, no instante de tempo, utilizando ENO ou WENO para interpolação no espaço. Para encontrar alta ordem de precisão no tempo será utilizado o método numérico Runge-Kutta pertencente a classe dos métodos TVD, o que nos garante que o método não será oscilatório. (Leveque (1992) [17]). Os métodos Runge-Kutta TVD apresentados estão escritos de maneira que se requeira menor armazenamento computacional. (Shu (1988) [30]).

### 3.1 Esquemas ENO via Diferenças Finitas para Leis de Conservação Escalares Unidimensionais

Para o método das diferenças finitas resolve-se a lei de conservação na forma diferencial conservativa Eq. (1.12), diretamente e não na sua forma integral como no método dos volumes finitos (ver apêndice B). Considerando uma grelha uniforme usa-se a aproximação conservativa, de acordo com Eq. (2.44), para aproximar a derivada da função fluxo  $f(u)_x$  no espaço, como mostra a Eq. (3.1)

$$\frac{du_j(t)}{dt} = -\frac{1}{\Delta x} \left( F_{j+\frac{1}{2}} - F_{j-\frac{1}{2}} \right), \quad (3.1)$$

sendo  $u_j$  é a aproximação numérica do ponto  $u(x_j, t)$  e  $F$  a aproximação da função fluxo  $f$ . A função de fluxo numérica  $F_{j+\frac{1}{2}}$  depende dos argumentos  $u_{j-r}, \dots, u_{j+s}$ , de acordo com Eq. (3.66).

Precisa-se obter a função de fluxo numérica de maneira que as seguintes condições sejam satisfeitas:

- $F$  é Lipschitz contínua em todos seus argumentos;

- $F$  é consistente com a função de fluxo física  $f(u)$ .

O teorema de Lax-Wendroff garante que: se a solução converge então ela converge para uma solução fraca do problema.

A função de fluxo numérica  $F_{j+\frac{1}{2}}$  é obtida pelo esquema ENO (ou pelo esquema WENO o qual se encontra no apêndice A), fazendo

$$\bar{v}(x) = f(u(x, t)) \quad (3.2)$$

em Eq. (3.22). Para a reconstrução da função de fluxo através do método das diferenças finitas é usado

$$\bar{v}_j = f(u(x_j, t)). \quad (3.3)$$

Existem algumas maneiras de se calcular a função de fluxo numérica  $h$ , Eq. (3.70). De acordo com Shu (1997)[29] é mais robusto usar ‘flux splitting’:

$$f(u) = f^+(u) + f^-(u), \quad (3.4)$$

tal que

$$\frac{df^+(u)}{du} \geq 0 \quad \text{e} \quad \frac{df^-(u)}{du} \leq 0, \quad (3.5)$$

sendo que tanto a parte positiva  $f^+$  quanto a parte negativa  $f^-$  da função, devem possuir derivadas da ordem do esquema. O ‘splitting’ separa a função  $f$  nas funções  $f^+$  e  $f^-$  obedecendo o movimento dos fluxos, de acordo com a direção ‘upwind’. Na ocorrência do choque o movimento do fluxo ocorre da esquerda para direita antes do choque e da direita para esquerda após o choque, justificando a Eq. (3.5).

Um exemplo de fluxo suave é o fluxo ‘splitting’ de Lax-Friedrichs:

$$f^+(u) = \frac{1}{2}(f(u) + \alpha u) \quad \text{e} \quad f^-(u) = \frac{1}{2}(f(u) - \alpha u), \quad (3.6)$$

tal que  $\alpha$  é dado por

$$\alpha = \max |f'(u)|. \quad (3.7)$$

Pois do número CFL, definição 2.3,  $\alpha = \max |f'(u) \frac{\Delta t}{\Delta x}| \leq 1$ . De acordo com Leveque (2004) [18], pode-se tomar  $\alpha$  como na Eq. (3.7) sem prejuízo para o método numérico.

Observe que todo fluxo ‘splitting’ da forma Eq. (3.4), satisfazendo Eq. (3.5), é um fluxo monótono (veja apêndice B), tal que a função de fluxo numérica  $h$  (Eq. (3.70)) é dada por

$$h(y, z) = f^+(y) + f^-(z). \quad (3.8)$$

É importante considerar que nem todo fluxo monótono que pode ser escrito na forma Eq. (3.4). Veja Shu (1997) [29].

Com o ‘splitting’ aplica-se o esquema ENO (seção 3.4) ou o esquema WENO (apêndice A) para aproximar

$$f^+(u(x, t)) \quad (3.9)$$

e

$$f^-(u(x, t)) \quad (3.10)$$

O fluxo ‘splitting’ de Lax-Friedrichs é utilizado para obter as aproximações numéricas  $F_{j+\frac{1}{2}}^+$  e  $F_{j+\frac{1}{2}}^-$  cuja soma aproxima  $F_{j+\frac{1}{2}}$  em Eq. (3.1), ou seja,

$$F_{j+\frac{1}{2}} = F_{j+\frac{1}{2}}^+ + F_{j+\frac{1}{2}}^-. \quad (3.11)$$

Analogamente obtém-se a aproximação numérica  $F_{j-\frac{1}{2}}$ , ou seja

$$F_{j-\frac{1}{2}} = F_{j-\frac{1}{2}}^+ + F_{j-\frac{1}{2}}^-. \quad (3.12)$$

### Procedimento para o esquema de diferenças finitas no caso escalar unidimensional

Será descrito o processo para encontrar um fluxo ‘splitting’ Eq. (3.4), satisfazendo Eq. (3.5)

1. Identifique

$$\overline{v}_j = f^+(u_j) \quad (3.13)$$

e use o processo de reconstrução através dos esquemas ENO (seção 3.4) ou WENO (apêndice A) para obter as aproximações em cada fronteira das células  $v_{j+\frac{1}{2}}^e = p\left(x_{j+\frac{1}{2}}\right)$ , para todo  $j = 1, \dots, N$ . Veja Figura 3.1.

2. Faça a função de fluxo numérica positiva  $F_{j+\frac{1}{2}}^+$  como

$$F_{j+\frac{1}{2}}^+ = v_{j+\frac{1}{2}}^e. \quad (3.14)$$

3. Identifique

$$\overline{v}_j = f^-(u_j) \quad (3.15)$$

e use o processo de reconstrução através dos esquemas ENO ou WENO para obter as aproximações em cada fronteira das células  $v_{j+\frac{1}{2}}^d = \bar{p}\left(x_{j+\frac{1}{2}}\right)$ , para todo  $j = 1, \dots, N$ .

4. Faça a função de fluxo numérica negativa  $F_{j+\frac{1}{2}}^-$  como

$$F_{j+\frac{1}{2}}^- = v_{j+\frac{1}{2}}^d. \quad (3.16)$$

5. Forme a função de fluxo numérica, de acordo com Eq. (3.4), como

$$F_{j+\frac{1}{2}} = F_{j+\frac{1}{2}}^+ + F_{j+\frac{1}{2}}^-. \quad (3.17)$$

6. Forme a aproximação (3.1), através da Eq. (3.17).

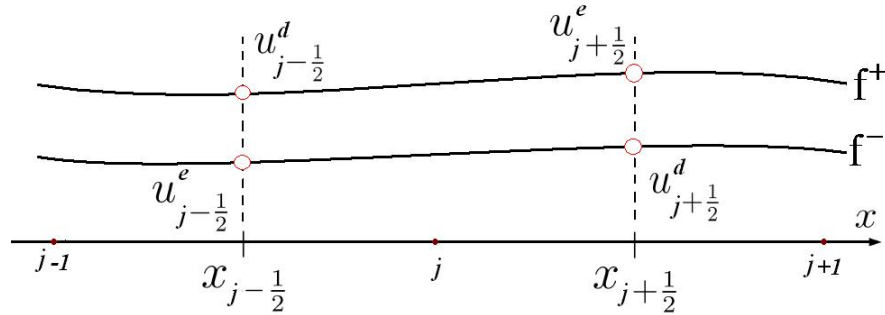


Figura 3.1: Aproximação ENO: fluxo 'splitting' Lax-Friedrichs

**Observação 3.1** Resumindo o procedimento acima tem-se:

$$\left. \begin{aligned} \bar{v}_j = f^+(u_j) &= \frac{1}{2} [f(u) + \alpha u] \xrightarrow{\text{ENO} / \text{WENO}} v_{j+\frac{1}{2}}^e = F_{j+\frac{1}{2}}^+ \\ \bar{v}_j = f^-(u_j) &= \frac{1}{2} [f(u) - \alpha u] \xrightarrow{\text{ENO} / \text{WENO}} v_{j+\frac{1}{2}}^d = F_{j+\frac{1}{2}}^- \end{aligned} \right\} \Rightarrow F_{j+\frac{1}{2}} = F_{j+\frac{1}{2}}^+ + F_{j+\frac{1}{2}}^-.$$

**Observação 3.2** Para se obter a aproximação da derivada da função fluxo numérico ( $F'_j$ ) (de acordo com a Eq. 3.1) nas fronteiras esquerda ( $x_j = x_{\text{inicial}} = x_1$ ) e direita ( $x_j = x_{\text{final}} = x_N$ ) do domínio, foram consideradas as expressões Eqs. (3.18) e (3.19), respectivamente. Veja Figura 3.2.

Considere  $F(u(x)) = F \circ u(x) = G(x)$ . Fazendo a expansão em série de Taylor de  $G\left(x + \frac{\Delta x}{2}\right)$  e considerando  $x = x_1$ , obtém-se uma aproximação da derivada da função fluxo na fronteira esquerda do domínio. Note que

$$G\left(x + \frac{\Delta x}{2}\right) = G(x) + \frac{\Delta x}{2} G'(x) + O(\Delta x^2)$$

então

$$G'(x_1) = \frac{2}{\Delta x} (F(u_{m_1}) - F(u_1)) + O(\Delta x).$$

Como os valores da função  $F$  nos pontos inteiros da malha computacional são conhecidos e coincidem com os valores da função  $f$ , obtém-se

$$F'(u_1) = G'(x_1) = \frac{2}{\Delta x} (F(u_{m_1}) - f(u_1)), \quad (3.18)$$

onde  $u_{m_1}$  é o valor da função  $u$  no primeiro ponto da malha metade (malha numérica composta pelos pontos médios  $j + \frac{1}{2}$  para todo  $j$ ) e  $u_1$  é o valor da função  $u$  no primeiro ponto da malha inteira (malha numérica composta pelos pontos inteiros  $j$  para todo  $j$ ).

De forma análoga se obtém uma expressão para a aproximação na fronteira direita. Basta observar que

$$G\left(x - \frac{\Delta x}{2}\right) = G(x) - \frac{\Delta x}{2} G'(x) + O(\Delta x^2),$$

então, para  $x = x_N$ , tem-se que:

$$F'(u_N) = -\frac{2}{\Delta x} (F(u_{m_N}) - f(u_N)), \quad (3.19)$$

onde  $u_{m_N}$  é o valor da função  $u$  no último ponto da malha metade e  $u_N$  é o valor da função  $u$  no último ponto da malha inteira.

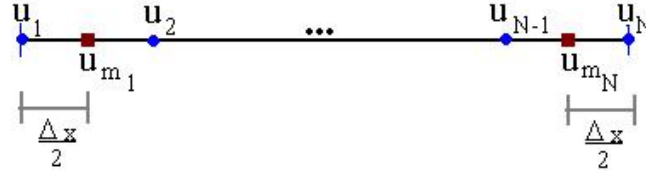


Figura 3.2: Malha numérica apresentando os primeiros e os últimos pontos, da malhas numéricas inteira e metade, para ilustrar a condição de fronteira do método de diferenças finitas.

Para EDP's lineares unidimensionais com coeficientes constantes, o esquema de diferenças finitas e o esquema de volumes finitos, apresentado no apêndice A, são equivalentes. Isto não ocorre para equações não-lineares. Veja Shu (1997) [29].

## 3.2 Reconstrução Unidimensional

Para obter uma aproximação espacial de alta ordem para a lei de conservação Eq. (1.12) é preciso definir a função de fluxo numérica  $F$  de forma que  $F_{j+\frac{1}{2}}$  seja uma boa aproximação para a função fluxo  $f\left(u\left(x_{j+\frac{1}{2}}, t\right)\right) = f_{j+\frac{1}{2}}$ . Seguindo a linha de construir um método numérico conservativo Eq. (2.44) para a lei de conservação Eq. (1.12), recaímos na construção pelos valores

médios Eq. (2.2) em cada célula  $I_j = [x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}]$ .

A questão é: dado somente a aproximação média da célula  $I_j$ , Eq. (2.2), como construir uma aproximação polinomial de alta ordem, para  $F_{j+\frac{1}{2}}$  e  $F_{j-\frac{1}{2}}$ , usando esses valores?

Uma elegante resposta a esta questão é dada por Colella e Woodward (1997) com o método PPM (do inglês Piecewise Parabolic Method) em [3], conhecido como reconstrução via função primitiva. Essa reconstrução utiliza o método clássico dos volumes finitos e a forma integral da lei de conservação dada pela Eq. (2.39).

Dada a malha numérica

$$a = x_{\frac{1}{2}} < x_{\frac{3}{2}} < \dots < x_{N+\frac{1}{2}} = b,$$

define-se como centro das células e o tamanho das células, respectivamente

$$x_j = \frac{x_{j-\frac{1}{2}} + x_{j+\frac{1}{2}}}{2} \quad \text{e} \quad (3.20)$$

$$\Delta x_j = x_{j+\frac{1}{2}} - x_{j-\frac{1}{2}} \quad (3.21)$$

dentro do intervalo  $[a, b]$ , sendo  $N$  é o número de células da malha computacional.

O processo da reconstrução via função primitiva é baseada no método dos volumes finitos, porém os esquemas não-oscilatórios, ENO e WENO (apêndice A), podem trabalhar tanto com volumes finitos quanto com diferenças finitas, mudando o enfoque da função primitiva. Ou seja, para o método de volume finito é considerado a aproximação média da célula dada na Eq. (2.2) ao passo que, para o método das diferenças finitas, considera-se a média  $\bar{u}$  como uma aproximação para a função  $f(u)$ , dada na Eq. (3.3). Esse processo será visto detalhadamente para o caso de diferenças finitas na seção 3.1 e para volumes finitos no apêndice A.

### 3.2.1 Reconstrução via Função Primitiva

Os esquemas numéricos ENO e WENO (apêndice A) podem ser formulados tanto no método das diferenças finitas quanto dos volumes finitos. Tendo em vista o diferente enfoque dos esquemas ENO e WENO e para melhor compreensão, o processo de reconstrução será feito para uma função qualquer  $v$ .

Dada uma função suave  $v(x, t)$ , define-se a aproximação média de  $v(x, t)$  na célula  $I_j$ , de acordo com Eq. (2.2), como

$$\bar{v}_j = \frac{1}{\Delta x_j} \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} v(\xi, t) d\xi, \quad \text{com} \quad \xi \in \left(x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}\right), \quad (3.22)$$

para todo  $j = 1, \dots, N$ .

Sendo  $v(x, t)$  uma função integrável para  $x \in (a, b)$ , é possível mostrar que  $\bar{v}_j$  está bem definido. Seja  $v$  uma função integrável e  $V(x)$  sua primitiva  $v(x)$ , isto é,  $V'(x) = v(x)$ . Do teorema do valor médio existe  $c \in [a, b]$  tal que

$$V'(c) = \frac{1}{b-a} \int_a^b v(x) dx \equiv \bar{v}. \quad (3.23)$$

Portanto  $V'(c) = \bar{v}$  para algum  $c \in [a, b]$ . Por outro lado  $c \in [a, b]$  se e somente se existe  $z \in \left[-\frac{1}{2}, \frac{1}{2}\right]$  tal que  $c = \frac{a+b}{2} + z(b-a)$ .

De fato,

$(\Rightarrow)$  se  $c \in [a, b]$  então existe  $z \in \left[-\frac{1}{2}, \frac{1}{2}\right]$  tal que  $c = \frac{a+b}{2} + z(b-a)$ . Seja

$$z = \frac{c - \frac{a+b}{2}}{b-a},$$

com

$$|z| = \left| \frac{c - \frac{a+b}{2}}{b-a} \right| \leq \frac{1}{b-a} \left| c - \frac{a+b}{2} \right| \leq \frac{b-a}{2} \frac{1}{b-a} = \frac{1}{2},$$

então temos

$$c = \frac{a+b}{2} + \frac{c - \frac{a+b}{2}}{b-a} \cdot (b-a)$$

$(\Leftarrow)$  Temos que mostrar que  $c \geq a$  e  $c \leq b$ . De fato,

$$c - a = \frac{a+b}{2} + z(b-a) - a = (b-a) \left( z + \frac{1}{2} \right) \geq 0 \quad \Rightarrow \quad c \geq a \quad \text{e}$$

$$c - b = \frac{a+b}{2} + z(b-a) - b = (b-a) \left( z - \frac{1}{2} \right) \leq 0 \quad \Rightarrow \quad c \leq b,$$

logo  $c \in (a, b)$ . Assim,

$$\bar{v} = V'(c) = V' \left( \frac{a+b}{2} + z(b-a) \right), \quad -\frac{1}{2} \leq z \leq \frac{1}{2}.$$

Portanto, se  $v$  for contínua temos que

$$\bar{v} = \lim_{a \rightarrow b} V'(c) = V' \left( \frac{a+b}{2} \right) = v \left( \frac{a+b}{2} \right).$$

Por outro lado, a aproximação  $\bar{v}_j$  concorda com o valor pontual de  $v_j$  na ordem  $O(\Delta x^2)$ , ou

seja,  $\bar{v}_j = v_j + O(\Delta x^2)$ . De fato, dada uma malha uniforme, se  $V$  é uma primitiva de  $v$  então

$$\frac{V\left(x_{j+\frac{1}{2}}\right) - V\left(x_{j-\frac{1}{2}}\right)}{\Delta x_j} = \bar{v}_j.$$

Por outro lado, fazendo a expansão em série de Taylor de  $V(x_j)$  para aproximar  $V_{j+\frac{1}{2}}$  e  $V_{j-\frac{1}{2}}$ , tem-se

$$\frac{V\left(x_{j+\frac{1}{2}}\right) - V\left(x_{j-\frac{1}{2}}\right)}{\Delta x_j} = V'_j + O(\Delta x^2),$$

ou seja

$$\bar{v}_j = v_j + O(\Delta x^2). \quad (3.24)$$

O objetivo é encontrar um polinômio  $p_j(x)$  de grau no máximo  $k-1$  para cada célula  $I_j$ , ou seja, deve-se obter o polinômio  $p_j(x)$  que aproxime a função suave  $v(x)$  com ordem de precisão  $O(\Delta x^k)$  no interior de cada intervalo  $I_j$ , para todo  $j = 1, \dots, N$ ,

$$p_j(x) = v(x) + O(\Delta x^k), \quad x \in I_j \quad j = 1, \dots, N. \quad (3.25)$$

A definição de ordem de erro é a definição usual de interpolação polinomial.

Em particular, as aproximações nas fronteiras da célula  $I_j = (x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}})$  são dadas por:

$$v_{j+\frac{1}{2}}^e = p_j\left(x_{j+\frac{1}{2}}\right) \quad \text{e} \quad v_{j-\frac{1}{2}}^d = p_j\left(x_{j-\frac{1}{2}}\right), \quad j = 1, \dots, N. \quad (3.26)$$

as quais possuem ordem de precisão igual a  $k$ , pois

$$v_{j+\frac{1}{2}}^e = v\left(x_{j+\frac{1}{2}}\right) + O(\Delta x^k) \quad \text{e} \quad v_{j-\frac{1}{2}}^d = v\left(x_{j-\frac{1}{2}}\right) + O(\Delta x^k), \quad j = 1, \dots, N. \quad (3.27)$$

**Observação 3.3** 1. O polinômio  $v_{j+\frac{1}{2}}^e$  é a aproximação realizada através da **célula à esquerda**,  $I_j = (x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}})$ , do ponto  $x_{j+\frac{1}{2}}$  e  $v_{j-\frac{1}{2}}^d$  é a aproximação realizada através da **célula à direita**,  $I_j = (x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}})$ , do ponto  $x_{j-\frac{1}{2}}$ .

2. O polinômio  $v_{j+\frac{1}{2}}^d$  tem como referência a célula  $I_{j+1} = (x_{j+\frac{1}{2}}, x_{j+\frac{3}{2}})$  e o polinômio  $v_{j-\frac{1}{2}}^e$  tem como referência a célula  $I_{j-1} = (x_{j-\frac{3}{2}}, x_{j-\frac{1}{2}})$ , como ilustra a Figura 3.3.

As condições de fronteira serão discutidas na observação 3.6 da seção 3.4.

Para determinar o polinômio  $p(x)$ , o primeiro passo é escolher o estêncil para se trabalhar. Dada a ordem de precisão  $k$  e a célula  $I_j$  escolhemos um estêncil, dentro da malha computacional, baseado em  $r$  células à esquerda e  $s$  células à direita da célula  $I_j$ , sendo que  $r, s$  são

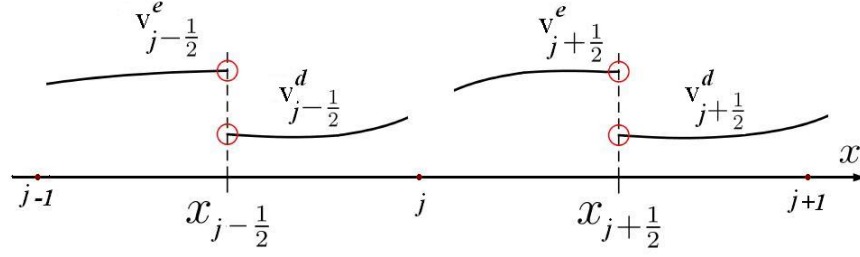


Figura 3.3: Aproximações à esquerda e à direita das fronteiras das células.

números que devem ser escolhidos de modo que  $r + s + 1 = k$ . Logo o estêncil é dado por

$$S(j) = \{I_{j-r}, I_{j-r+1}, \dots, I_j, \dots, I_{j+s}\}, \quad j = 1, \dots, N. \quad (3.28)$$

Observa-se que o estêncil  $S(j)$  em Eq. (3.28) é formado por  $k$  células e  $k$  pontos conhecidos.

Existe um único polinômio de grau no máximo  $k - 1 = r + s$  cujos valores médios para cada célula de  $S(j)$  é aproximado por:

$$\bar{v}_i = \frac{1}{\Delta x_j} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} p(\xi) d\xi, \quad i = j - r, \dots, j + s. \quad (3.29)$$

Observe a diferença dos índices: o índice  $j$  varia em toda a malha computacional,  $j = 1, \dots, N$ , e o índice  $i$  varia dentro de cada estêncil  $S(j)$ , para todo  $j$ . Para que fique mais claro segue um exemplo da malha computacional com poucos pontos, apenas para melhor visualização.

**Exemplo 3.1** Considere a malha computacional com  $N = 10$  na qual deve-se obter um polinômio interpolador com ordem de precisão  $k = 4$ .

Logo o grau do polinômio será  $k - 1 = 3$  e precisamos de 4 pontos conhecidos. Consideremos a célula  $I_5$ , uma possibilidade de estêncil é dada por  $r = 1$  e  $s = 2$ , ou seja  $S(5) = \{I_4, I_5, I_6, I_7\}$ . Logo  $j = 1, \dots, 10$  e  $i = 4, \dots, 7$  como mostra a Figura 3.4.

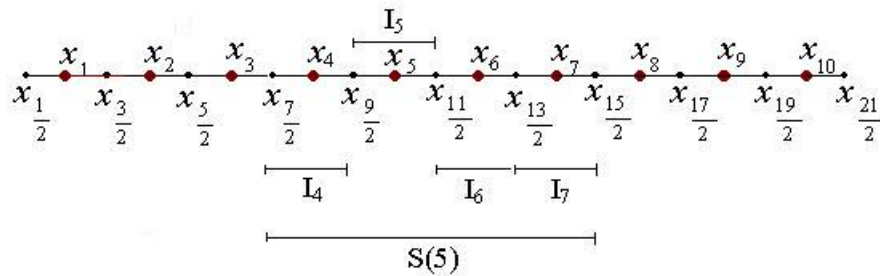


Figura 3.4: Ilustração da diferença entre o índice  $i$  e o índice  $j$ .

O estêncil  $S(5)$  possui  $k = 4$  células e  $k + 1 = 5$  pontos médios e  $k = 4$  pontos conhecidos.

O objetivo é utilizar os valores conhecidos  $\bar{v}_j$  nos centros das células,  $I_j$ , para encontrar uma aproximação nas fronteiras das células,  $v_{j+\frac{1}{2}}$ , para  $j = 0, \dots, N$ .

A questão é: dado o ponto de referência  $j$  como escolher o valor de  $r$ , a esquerda de  $j$ ? A resposta a essa questão será dada na seção 3.4.

Note que existem pontos que pertencem a duas células adjacentes, ou seja o ponto  $x_{j+\frac{1}{2}}$  pertence às células  $I_j$  e  $I_{j+1}$ . Analogamente tem-se que o ponto  $x_{j-\frac{1}{2}}$  pertence às células  $I_{j-1}$  e  $I_j$ , logo se identificarmos o valor de  $r$  como sendo  $r$  pontos à esquerda de  $x_{j+\frac{1}{2}}$  e não  $r$  células à esquerda de  $I_j$ , as constantes obtidas para a aproximação de  $v_{j+\frac{1}{2}}^e$  e  $v_{j-\frac{1}{2}}^d$  serão as mesmas.

Se  $r$  é identificado com a célula  $j$ , o estêncil  $S(j)$  é dado pela Eq. (3.28), mas como o ponto  $x_{j+\frac{1}{2}}$  pertence a célula  $I_{j+1}$  se identificarmos  $r$  com esta célula, o estêncil é dado por

$$S(j+1) = \{I_{j-r+1}, I_{j-r+1}, \dots, I_{j+1}, \dots, I_{j+s+1}\}, \quad j = 1, \dots, N,$$

obtendo duas possibilidades de polinômios. Porém se definirmos  $r$  como sendo a quantidade de pontos à esquerda de  $x_{j+\frac{1}{2}}$ , o estêncil dado por

$$\bar{S} = \{x_{j-r-1}, \dots, x_j, \dots, x_{j+s}\},$$

fornece o mesmo polinômio para aproximar  $v_{j+\frac{1}{2}}^e$  e  $v_{j-\frac{1}{2}}^d$ . Por isso é suficiente considerar, para o processo de reconstrução, apenas um valor. Ou seja, dado  $k$  aproximações médias

$$\bar{v}_{j-r}, \dots, \bar{v}_{j-r+k-1}$$

existem constantes  $c_{ri}$  tal que o valor de reconstrução na fronteira  $x_{j+\frac{1}{2}}$  da célula  $I_j$  é dado por

$$v_{j+\frac{1}{2}} = \sum_{i=0}^{k-1} c_{ri} \bar{v}_{j-r+i}. \quad (3.30)$$

Para se obter as constantes  $c_{ri}$  trabalha-se com a função primitiva de  $v(x)$  definida por

$$V(x) = \int_{-\infty}^x v(\xi) d\xi. \quad (3.31)$$

O valor de  $V\left(x_{j+\frac{1}{2}}\right)$  pode ser escrito como

$$V\left(x_{j+\frac{1}{2}}\right) = \int_{-\infty}^{x_{j-\frac{1}{2}}} v(\xi) d\xi + \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} v(\xi) d\xi.$$

Continuando com esse processo e usando Eq. (3.22) obtém-se

$$V\left(x_{j+\frac{1}{2}}\right) = \sum_{i=-\infty}^j \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} v(\xi) d\xi = \sum_{i=-\infty}^j \bar{v}_i \Delta x_i. \quad (3.32)$$

A Eq. (3.32) diz que conhecido as aproximações médias das células  $\bar{v}_i$  conhece-se também o valor de  $V(x)$  na fronteira da célula  $I_i$ . Denote por  $P(x)$  o único polinômio de grau no máximo  $k$ , o qual interpola  $V\left(x_{j+\frac{1}{2}}\right)$  nos  $k+1$  pontos

$$\left\{x_{j-r-\frac{1}{2}}, \dots, x_{j+s+\frac{1}{2}}\right\} \quad (3.33)$$

e cuja derivada é o polinômio  $p$

$$P'(x) \equiv p(x). \quad (3.34)$$

De acordo com o visto até o momento pode-se mostrar que realmente vale a Eq. (3.29)

$$\begin{aligned} \frac{1}{\Delta x} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} p(\xi) d\xi &= \frac{1}{\Delta x} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} P'(\xi) d\xi \\ &\stackrel{\text{TFC}}{=} \frac{1}{\Delta x} \left( P\left(x_{i+\frac{1}{2}}\right) - P\left(x_{i-\frac{1}{2}}\right) \right) \\ &= \frac{1}{\Delta x} \left( V\left(x_{i+\frac{1}{2}}\right) - V\left(x_{i-\frac{1}{2}}\right) \right) \\ &\stackrel{\text{def}}{=} \frac{1}{\Delta x} \left( \int_{-\infty}^{x_{i+\frac{1}{2}}} v(\xi) d\xi - \int_{-\infty}^{x_{i-\frac{1}{2}}} v(\xi) d\xi \right) \\ &= \frac{1}{\Delta x} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} v(\xi) d\xi \\ &= \bar{v}_i, \quad \text{para} \quad i = j-r, \dots, j+s. \end{aligned}$$

Note que a terceira igualdade é válida porque  $P(x)$  interpola  $V(x)$  nos pontos  $x_{i-\frac{1}{2}}$  e  $x_{i+\frac{1}{2}}$ , com  $i = j-r, \dots, j+s$ . Esta averiguação diz que  $p(x)$  realmente é o polinômio o qual estamos procurando.

Uma maneira de obter essas constantes  $c_{ri}$ , do polinômio  $p(x)$ , definidas em Eq. (3.30) é utilizar o polinômio interpolador de Lagrange:

$$P(x) = \sum_{m=0}^k V\left(x_{j-r+m-\frac{1}{2}}\right) \prod_{\substack{l=0 \\ l \neq m}}^k \frac{x - x_{j-r+l-\frac{1}{2}}}{x_{j-r+m-\frac{1}{2}} - x_{j-r+l-\frac{1}{2}}}, \quad (3.35)$$

subtraindo  $V\left(x_{j-r-\frac{1}{2}}\right)$  na Eq. (3.35) e observando que

$$\sum_{m=0}^k \prod_{\substack{l=0 \\ l \neq m}}^k \frac{x - x_{j-r+l-\frac{1}{2}}}{x_{j-r+m-\frac{1}{2}} - x_{j-r+l-\frac{1}{2}}} = 1, \quad (3.36)$$

obtém-se

$$P(x) - V\left(x_{j-r-\frac{1}{2}}\right) = \sum_{m=0}^k \left( V\left(x_{j-r+m-\frac{1}{2}}\right) - V\left(x_{j-r-\frac{1}{2}}\right) \right) \prod_{\substack{l=0 \\ l \neq m}}^k \frac{x - x_{j-r+l-\frac{1}{2}}}{x_{j-r+m-\frac{1}{2}} - x_{j-r+l-\frac{1}{2}}}. \quad (3.37)$$

Trabalhando com a Eq. (3.32), pode-se escrever

$$V\left(x_{j-r+m-\frac{1}{2}}\right) = V\left(x_{(j-r+m-1)+\frac{1}{2}}\right) = \sum_{i=-\infty}^{j-r+m-1} \bar{v}_i \Delta x \quad (3.38)$$

e

$$V\left(x_{j-r-\frac{1}{2}}\right) = V\left(x_{(j-r-1)+\frac{1}{2}}\right) = \sum_{i=-\infty}^{j-r-1} \bar{v}_i \Delta x \quad (3.39)$$

Como  $m = 0, \dots, k$ , tem-se que  $j-r-\frac{1}{2} \leq j-r+m-\frac{1}{2}$  e subtraindo a Eq. (3.39) da Eq. (3.38) obtemos

$$\begin{aligned} V\left(x_{j-r+m-\frac{1}{2}}\right) - V\left(x_{j-r-\frac{1}{2}}\right) &= \sum_{i=-\infty}^{j-r-1} \bar{v}_i \Delta x + \sum_{i=j-r}^m \bar{v}_i \Delta x - \sum_{i=-\infty}^{j-r-1} \bar{v}_i \Delta x \\ &= \sum_{i=0}^{m-1} \bar{v}_{j-r+i} \Delta x_{j-r+i}. \end{aligned} \quad (3.40)$$

Substituindo a Eq. (3.40) na Eq. (3.37), concluimos

$$P(x) - V\left(x_{j-r-\frac{1}{2}}\right) = \sum_{m=0}^k \sum_{i=0}^{m-1} \bar{v}_{j-r+i} \Delta x_{j-r+i} \prod_{\substack{l=0 \\ l \neq m}}^k \frac{x - x_{j-r+l-\frac{1}{2}}}{x_{j-r+m-\frac{1}{2}} - x_{j-r+l-\frac{1}{2}}}. \quad (3.41)$$

Agora, derivando a Eq. (3.41), obtém-se

$$p(x) = P'(x) = \sum_{m=0}^k \sum_{i=0}^{m-1} \bar{v}_{j-r+i} \Delta x_{j-r+i} \left( \frac{\sum_{\substack{l=0 \\ l \neq m}}^k \prod_{\substack{q=0 \\ q \neq l, m}}^k \frac{x - x_{j-r+q-\frac{1}{2}}}{x_{j-r+m-\frac{1}{2}} - x_{j-r+l-\frac{1}{2}}}}{\prod_{\substack{l=0 \\ l \neq m}}^k \left( x_{j-r+m-\frac{1}{2}} - x_{j-r+l-\frac{1}{2}} \right)} \right) \quad (3.42)$$

e avaliando Eq. (3.42) em  $x = x_{j+\frac{1}{2}}$  temos que

$$v_{j+\frac{1}{2}} = p\left(x_{j+\frac{1}{2}}\right) = \sum_{i=0}^{k-1} \left[ \sum_{m=i+1}^k \left( \frac{\sum_{\substack{l=0 \\ l \neq m}}^k \prod_{\substack{q=0 \\ q \neq l, m}}^k \frac{x_{j+\frac{1}{2}} - x_{j-r+q-\frac{1}{2}}}{x_{j-r+m-\frac{1}{2}} - x_{j-r+l-\frac{1}{2}}}}{\prod_{\substack{l=0 \\ l \neq m}}^k \left( x_{j-r+m-\frac{1}{2}} - x_{j-r+l-\frac{1}{2}} \right)} \right) \Delta x_{j-r+i} \right] \bar{v}_{j-r+i}. \quad (3.43)$$

Comparando a Eq. (3.30) com a Eq. (3.43), segue que as constantes  $c_{ri}$  são dadas por

$$c_{ri} = \sum_{m=i+1}^k \left( \frac{\sum_{\substack{l=0 \\ l \neq m}}^k \prod_{\substack{q=0 \\ q \neq l, m}}^k \frac{x_{j+\frac{1}{2}} - x_{j-r+q-\frac{1}{2}}}{x_{j-r+m-\frac{1}{2}} - x_{j-r+l-\frac{1}{2}}}}{\prod_{\substack{l=0 \\ l \neq m}}^k \left( x_{j-r+m-\frac{1}{2}} - x_{j-r+l-\frac{1}{2}} \right)} \right) \Delta x_{j-r+i}. \quad (3.44)$$

Observe que as constantes  $c_{ri}$  dependem do tamanho da célula  $\Delta x$ , da ordem de precisão requerida  $k$  e do valor de  $r$  escolhido para se definir o estêncil. As constantes  $c_{ri}$  são dadas pela Eq. (3.44) quando a malha não for uniforme. Para uma malha uniforme a Eq. (3.44) pode ser escrita como

$$c_{ri} = \sum_{m=i+1}^k \frac{\sum_{l=0}^k \prod_{\substack{q=0 \\ q \neq l, m}}^k \left[ x_{\frac{1}{2}} + j - \left( x_{\frac{1}{2}} + (j - r + q - 1) \right) \right] \Delta x}{\prod_{\substack{l=0 \\ l \neq m}}^k \left[ x_{\frac{1}{2}} + (j - r + m - 1) - \left( x_{\frac{1}{2}} + (j - r + l - 1) \right) \right] \Delta x} \quad (3.45)$$

e cancelando os termos comuns na Eq. (3.45), obtém-se

$$c_{ri} = \sum_{m=i+1}^k \frac{\sum_{l=0}^k \prod_{\substack{q=0 \\ q \neq l, m}}^k r - q + 1}{\prod_{\substack{l=0 \\ l \neq m}}^k (m - l)}. \quad (3.46)$$

De acordo com a Eq. (3.46), as constantes  $c_{ri}$  para uma malha uniforme, não dependem de  $j$  ou de  $\Delta x$ , apenas de  $r$  e  $k$ . Os valores das constantes  $c_{ri}$ , para uma malha uniforme, são fornecidos na tabela 4.1.

Analogamente se a Eq. (3.42) for avaliada em  $x = x_{j-\frac{1}{2}}$  obtém-se  $v_{j-\frac{1}{2}} = p\left(x_{j-\frac{1}{2}}\right)$ .

**Exemplo 3.2** Escolhido a ordem de precisão  $k = 4$ , pode-se tomar  $r = -1, 0, 1, 2, 3$  e temos  $j = 0, 1, 2, 3$ . De acordo com Eq. (3.43) temos

$$v_{j+\frac{1}{2}} = \sum_{i=0}^{k-1} c_{ri} \bar{v}_{j-r+i}.$$

Para  $r = 0, 1, 2, 3$  temos, respectivamente

$$v_{j+\frac{1}{2}} = c_{00} \bar{v}_j + c_{01} \bar{v}_{j+1} + c_{02} \bar{v}_{j+2} + c_{03} \bar{v}_{j+3} + O(\Delta x^4),$$

$$v_{j+\frac{1}{2}} = c_{10} \bar{v}_{j-1} + c_{11} \bar{v}_j + c_{12} \bar{v}_{j+1} + c_{13} \bar{v}_{j+2} + O(\Delta x^4),$$

$$v_{j+\frac{1}{2}} = c_{20} \bar{v}_{j-2} + c_{21} \bar{v}_{j-1} + c_{22} \bar{v}_j + c_{23} \bar{v}_{j+1} + O(\Delta x^4),$$

$$v_{j+\frac{1}{2}} = c_{30} \bar{v}_{j-3} + c_{31} \bar{v}_{j-2} + c_{32} \bar{v}_{j-1} + c_{33} \bar{v}_j + O(\Delta x^4).$$

Consultando a tabela 4.1 obtém-se, respectivamente

$$v_{j+\frac{1}{2}} = \frac{1}{4} \bar{v}_j + \frac{13}{12} \bar{v}_{j+1} - \frac{5}{12} \bar{v}_{j+2} + \frac{1}{12} \bar{v}_{j+3} + O(\Delta x^4),$$

$$v_{j+\frac{1}{2}} = -\frac{1}{12} \bar{v}_{j-1} + \frac{7}{12} \bar{v}_j + \frac{7}{12} \bar{v}_{j+1} - \frac{1}{12} \bar{v}_{j+2} + O(\Delta x^4).$$

$$v_{j+\frac{1}{2}} = \frac{1}{12}\bar{v}_{j-2} - \frac{5}{12}\bar{v}_{j-1} + \frac{13}{12}\bar{v}_j + \frac{1}{4}\bar{v}_{j+1} + O(\Delta x^4),$$

$$v_{j+\frac{1}{2}} = -\frac{1}{4}\bar{v}_{j-3} + \frac{13}{12}\bar{v}_{j-2} - \frac{23}{12}\bar{v}_{j-1} + \frac{25}{12}\bar{v}_j + O(\Delta x^4).$$

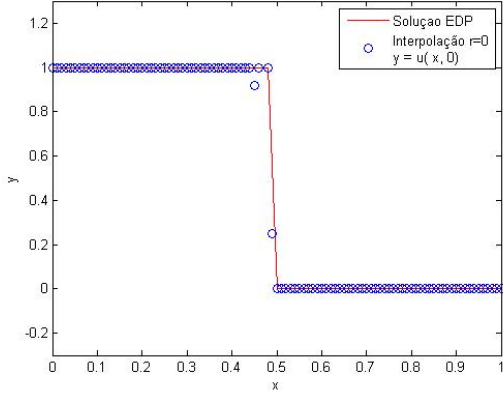


Figura 3.5: Interpolação de grau 3 com  $r = 0$  e  $\Delta x = 0.01$ , no intervalo  $[0, 1]$ .

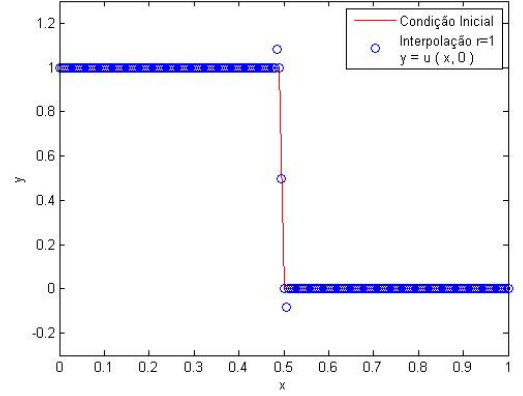


Figura 3.6: Interpolação de grau 3 com  $r = 1$  e  $\Delta x = 0.01$ , no intervalo  $[0, 1]$ .

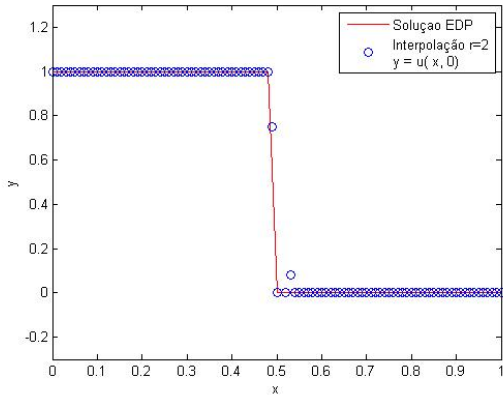


Figura 3.7: Interpolação de grau 3 com  $r = 2$  e  $\Delta x = 0.01$ , no intervalo  $[0, 1]$ .

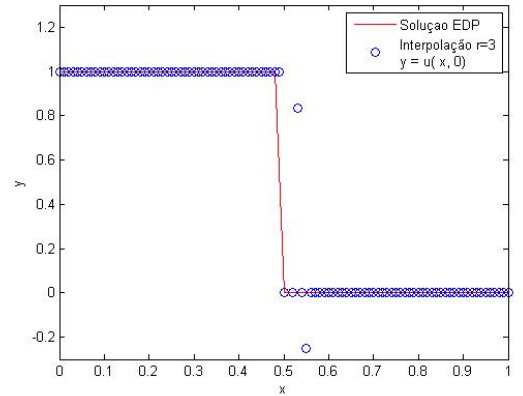


Figura 3.8: Interpolação de grau 3 com  $r = 3$  e  $\Delta x = 0.01$ , no intervalo  $[0, 1]$ .

As Figuras 3.5, 3.6, 3.7 e 3.8 ilustram as oscilações da aproximação numérica na região da descontinuidade, da condição inicial

$$u_0(x) = \begin{cases} 1, & \text{se } x < 0.5, \\ 0, & \text{se } x > 0.5, \end{cases} \quad (3.47)$$

para  $r = 0, 1, 2, 3$  respectivamente

Para melhorar a aproximação na região da descontinuidade da solução surgiu a idéia do estêncil adaptativo, na qual se baseia o esquema ENO.

| k | r  | i=0     | i=1      | i=2      | i=3       | i=4      | I=5      | i=6     |
|---|----|---------|----------|----------|-----------|----------|----------|---------|
| 1 | -1 | 1       |          |          |           |          |          |         |
|   | 0  | 1       |          |          |           |          |          |         |
| 2 | -1 | 3/2     | -1/2     |          |           |          |          |         |
|   | 0  | 1/2     | 1/2      |          |           |          |          |         |
|   | 1  | -1/2    | 3/2      |          |           |          |          |         |
| 3 | -1 | 11/6    | -7/6     | 1/3      |           |          |          |         |
|   | 0  | 1/3     | 5/6      | -1/6     |           |          |          |         |
|   | 1  | -1/6    | 5/6      | 1/3      |           |          |          |         |
|   | 2  | 1/3     | -7/6     | 11/6     |           |          |          |         |
| 4 | -1 | 25/12   | -23/12   | 13/12    | -1/4      |          |          |         |
|   | 0  | 1/4     | 13/12    | -5/12    | 1/12      |          |          |         |
|   | 1  | -1/12   | 7/12     | 7/12     | -1/12     |          |          |         |
|   | 2  | 1/12    | -5/12    | 13/12    | 1/4       |          |          |         |
|   | 3  | -1/4    | 13/12    | -23/12   | 25/12     |          |          |         |
| 5 | -1 | 137/60  | -163/60  | 137/60   | -21/20    | 1/5      |          |         |
|   | 0  | 1/5     | 77/60    | -43/60   | 17/60     | -1/20    |          |         |
|   | 1  | -1/20   | 9/20     | 47/60    | -13/60    | 1/30     |          |         |
|   | 2  | 1/30    | -13/60   | 47/60    | 9/20      | -1/20    |          |         |
|   | 3  | -1/20   | 17/60    | -43/60   | 77/60     | 1/5      |          |         |
|   | 4  | 1/5     | -21/20   | 137/60   | -163/60   | 137/60   |          |         |
| 6 | -1 | 49/20   | -71/20   | 79/20    | -163/60   | 31/30    | -1/6     |         |
|   | 0  | 1/6     | 29/20    | -21/20   | 37/60     | -13/60   | 1/30     |         |
|   | 1  | -1/30   | 11/30    | 19/20    | -23/60    | 7/60     | -1/60    |         |
|   | 2  | 1/60    | -2/15    | 37/60    | 37/60     | -2/15    | 1/60     |         |
|   | 3  | -1/60   | 7/60     | -23/60   | 19/20     | 11/30    | -1/30    |         |
|   | 4  | 1/30    | -13/60   | 37/60    | -21/20    | 29/20    | 1/6      |         |
|   | 5  | -1/6    | 31/30    | -163/60  | 79/20     | -71/20   | 49/20    |         |
| 7 | -1 | 363/140 | -617/140 | 853/140  | -2341/420 | 667/210  | -43/42   | 1/7     |
|   | 0  | 1/7     | 223/140  | -197/140 | 153/140   | -241/420 | 37/210   | -1/42   |
|   | 1  | -1/42   | 13/42    | 153/140  | -241/420  | 109/420  | -31/420  | 1/105   |
|   | 2  | 1/60    | -19/210  | 107/210  | 319/420   | -101/420 | 5/84     | -1/140  |
|   | 3  | -1/140  | 5/84     | -101/420 | 319/420   | 107/210  | -19/210  | 1/105   |
|   | 4  | 1/105   | -31/420  | 109/420  | -241/420  | 153/140  | 13/42    | -1/42   |
|   | 5  | -1/42   | 37/210   | -241/420 | 153/140   | -197/140 | 223/140  | 1/7     |
|   | 6  | 1/7     | -43/42   | 667/210  | -2341/420 | 853/140  | -617/140 | 363/140 |

Tabela 3.1: Constantes  $c_{ri}$  referente a Eq. (3.46)

### 3.3 Diferenças Divididas

Para resolução de leis de conservação há um interesse especial na classe de funções suaves por partes. Uma função suave por partes é uma função suave exceto para um número finito de pontos isolados. Assume-se que tais pontos possuem limites laterais para  $v(x)$  e suas derivadas.

A idéia básica do estêncil adaptativo é evitar introduzir no estêncil pontos nos quais a função possua variações bruscas, sempre que possível. Na prática o que se faz é procurar o melhor

estêncil possível para se trabalhar.

A escolha do melhor estêncil está relacionada com as diferenças divididas, portanto, primeiro será feito uma análise da teoria das diferenças divididas e do polinômio de Newton.

Dada a função  $V(x)$  definida em Eq. (3.31)-Eq. (3.32), a **diferença dividida de ordem zero** é definida por:

$$V \left[ x_{j+\frac{1}{2}} \right] \equiv V \left( x_{j+\frac{1}{2}} \right) \quad (3.48)$$

e para uma ordem qualquer  $i \geq 1$  as diferenças divididas são definidas por:

$$V \left[ x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}, \dots, x_{j+i-\frac{3}{2}}, x_{j+i-\frac{1}{2}} \right] \equiv \frac{V \left[ x_{j+\frac{1}{2}}, \dots, x_{j+i-\frac{1}{2}} \right] - V \left[ x_{j-\frac{1}{2}}, \dots, x_{j+i-\frac{3}{2}} \right]}{x_{j+i-\frac{1}{2}} - x_{j-\frac{1}{2}}}. \quad (3.49)$$

As diferenças divididas de  $V$  podem se escrever em função das aproximações médias  $\bar{v}_j$

$$V \left[ x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}} \right] \stackrel{\text{def}}{=} \frac{V \left( x_{j+\frac{1}{2}} \right) - V \left( x_{j-\frac{1}{2}} \right)}{x_{j+\frac{1}{2}} - x_{j-\frac{1}{2}}} \stackrel{3.32}{=} \bar{v}_j, \quad (3.50)$$

ou seja, a diferença dividida de primeira ordem de  $V(x)$  é igual a célula média  $\bar{v}$ .

O polinômio interpolador de Newton  $P(x)$ , de grau  $k$ , o qual interpola a função  $V(x)$  nos  $k+1$  pontos dados em Eq. (3.33) pode ser escrito como:

$$P(x) = \sum_{i=0}^k V \left[ x_{j-r-\frac{1}{2}}, \dots, x_{j-r+i-\frac{1}{2}} \right] \prod_{m=0}^{i-1} \left( x - x_{j-r+m-\frac{1}{2}} \right). \quad (3.51)$$

Derivando Eq. (3.51), obtém-se a equação

$$p(x) = \sum_{i=1}^k V \left[ x_{j-r-\frac{1}{2}}, \dots, x_{j-r+i-\frac{1}{2}} \right] \sum_{m=0}^{i-1} \prod_{\substack{l=0 \\ l \neq m}}^{i-1} \left( x - x_{j-r+l-\frac{1}{2}} \right), \quad (3.52)$$

a qual representa o polinômio  $p(x)$  dado na Eq. (3.34).

Note que na Eq. (3.52) apenas aparecem as diferenças divididas de ordem maior que um, pois a derivada da diferença dividida de ordem zero se anula. De acordo com este fato é possível expressar  $p(x)$  totalmente em função das diferenças divididas de  $\bar{v}$  usando Eq. (3.50), sem fazer nenhuma referência a  $V(x)$ . Ver Shu (1997) [29].

**Teorema 3.1** *Se a função  $V(x)$  é contínua no estêncil  $\{x_0, x_1, \dots, x_n, x\}$  então*

$$V[x_0, x_1, \dots, x_n, x] = \frac{V^{n+1}(\xi)}{(n+1)!}, \quad \xi \in (a, b). \quad (3.53)$$

**Demonstração:** Veja Ruggiero (1997) [27].

A Eq. (3.53) diz que em regiões de suavidade da função  $V$ , as diferenças divididas são proporcionais ao erro da interpolação.

De acordo com Krylov (1992) [16], pode-se mostrar que

$$V[x_0, x_1, \dots, x_n] = \sum_{J=0}^n \frac{V(x_J)}{(x_J - x_0)(x_J - x_1) \dots (x_J - x_{J-1})(x_J - x_{J+1}) \dots (x_J - x_n)}. \quad (3.54)$$

Se a função  $V(x)$  é descontínua em algum ponto do estêncil vale a propriedade de diferenças divididas no teorema 3.2, a qual será demonstrada para uma malha numérica igualmente espaçada.

**Teorema 3.2** *Se a função  $V(x)$  é descontínua no estêncil  $\{x_0, x_1, \dots, x_n, x\}$  temos que*

$$V[x_0, x_1, \dots, x_n] = O\left(\frac{1}{\Delta x^n}\right).$$

**Demonstração:**

Para  $n = 1$ , o resultado é facilmente comprovado, pois

$$V[x_0, x_1] = \frac{V(x_1) - V(x_0)}{\Delta x},$$

então

$$\lim_{\Delta x \rightarrow 0} \frac{V[x_0, x_1]}{\frac{1}{\Delta x}} = \lim_{\Delta x \rightarrow 0} \Delta x \frac{V(x_1) - V(x_0)}{\Delta x} = \text{constante},$$

ou seja,

$$V[x_0, x_1] = O\left(\frac{1}{\Delta x^1}\right).$$

O caso geral é uma consequência imediata da Eq. (3.54). Basta observar que o denominador de cada parcela que aparece em tal equação será um múltiplo de  $\Delta x^n$ . ■

Os teoremas 3.1 e 3.2 relacionam as diferenças divididas com a ordem do erro da aproximação de uma função. O teorema 3.1 diz que em regiões suaves a diferença dividida é proporcional a ordem do erro enquanto o teorema 3.2 diz que em regiões onde a função possui descontinuidades (choques) as diferenças divididas são inversamente proporcionais a ordem do erro. Ou seja, quanto menor a diferença dividida, mais suave será o estêncil e melhor será a aproximação. O esquema ENO se baseia nessas propriedades.

### 3.4 ENO

O esquema ENO será descrito usando o polinômio de Newton dado em Eq. (3.51), já que a escolha do estêncil se baseia nas diferenças divididas.

Seja  $k$  a ordem do polinômio interpolador (ENO), o objetivo é encontrar um estêncil com  $k+1$  pontos consecutivos nos quais estejam incluídos os pontos  $\left\{x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}\right\}$  e tal que a função  $V(x)$  é o mais suave possível dentro do estêncil.

**Exemplo 3.3** *Dado o grau do polinômio  $k = 3$  o estêncil deve ser composto por 4 pontos. Fixado o estêncil inicial  $\left\{x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}\right\}$  existem 3 possibilidades:*

$$\left\{x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}, x_{j+\frac{3}{2}}, x_{j+\frac{5}{2}}\right\}, \left\{x_{j-\frac{3}{2}}, x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}, x_{j+\frac{3}{2}}\right\} \text{ ou } \left\{x_{j-\frac{5}{2}}, x_{j-\frac{3}{2}}, x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}\right\}.$$

*O objetivo é encontrar qual é o melhor estêncil, dentre as três possibilidades.*

O processo se inicia com o estêncil composto por apenas dois pontos

$$\bar{S}_2(j) = \left\{x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}\right\}, \quad \forall j \quad (3.55)$$

e acrescenta-se os pontos, de um em um, de acordo com o menor grau de diferenças divididas.

Dado o estêncil inicial em Eq. (3.55), o segundo passo é escolher o próximo ponto: Se

$$\left|V\left[x_{j-\frac{3}{2}}, x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}\right]\right| \leq \left|V\left[x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}, x_{j+\frac{3}{2}}\right]\right|, \quad \forall j, \quad (3.56)$$

o estêncil  $\bar{S}_3(j)$  é dado por

$$\bar{S}_3(j) = \left\{x_{j-\frac{3}{2}}, x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}\right\}, \quad \forall j, \quad (3.57)$$

senão o estêncil  $\bar{S}_3(j)$  é dado por

$$\bar{S}_3(j) = \left\{x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}, x_{j+\frac{3}{2}}\right\}, \quad \forall j. \quad (3.58)$$

O terceiro passo é análogo ao anterior. Caso o estêncil escolhido,  $\bar{S}_3(j)$ , for dado pela Eq. (3.58), o próximo estêncil é obtido da seguinte forma:

- se

$$\left|V\left[x_{j-\frac{3}{2}}, x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}, x_{j+\frac{3}{2}}\right]\right| \leq \left|V\left[x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}, x_{j+\frac{3}{2}}, x_{j+\frac{5}{2}}\right]\right|, \quad \forall j, \quad (3.59)$$

o estêncil  $\overline{S}_4(j)$  é dado por

$$\overline{S}_4(j) = \left\{ x_{j-\frac{3}{2}}, x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}, x_{j+\frac{3}{2}} \right\}, \quad \forall j, \quad (3.60)$$

- senão o estêncil  $\overline{S}_4(j)$  é dado por

$$\overline{S}_4(j) = \left\{ x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}, x_{j+\frac{3}{2}}, x_{j+\frac{5}{2}} \right\}, \quad \forall j. \quad (3.61)$$

Se o estêncil escolhido  $\overline{S}_3(j)$  for dado pela Eq. (3.57) o processo é análogo, diferindo apenas pela inclusão dos pontos. E este processo se repete até que se obtenha o estêncil com a quantidade de pontos desejados.

Após a escolha do estêncil, o polinômio de Newton Eq. (3.51) é construído.

Considerando o exemplo 4.3 e o estêncil  $\overline{S}_4(j)$  dado na Eq. 3.61, o polinômio de Newton vem definido por

$$P(x) = V \left[ x_{j-\frac{1}{2}} \right] + V \left[ x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}} \right] \left( x - x_{j-\frac{1}{2}} \right) + V \left[ x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}, x_{j+\frac{3}{2}} \right] \left( x - x_{j-\frac{1}{2}} \right) \left( x - x_{j+\frac{1}{2}} \right) + V \left[ x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}, x_{j+\frac{3}{2}}, x_{j+\frac{5}{2}} \right] \left( x - x_{j-\frac{1}{2}} \right) \left( x - x_{j+\frac{1}{2}} \right) \left( x - x_{j+\frac{3}{2}} \right). \quad (3.62)$$

Note que cada estêncil  $\overline{S}$  de  $V$  possui um estêncil correspondente  $S$  de  $\overline{v}$  de acordo com Eq. (3.50). Por exemplo, para o estêncil Eq. (3.55) corresponde um estêncil  $S$  de  $\overline{v}$  composto por um único ponto

$$S(j) = \{x_j\}$$

e assim o polinômio Eq. (3.62) pode ser reescrito em função dos valores de  $\overline{v}$ . Na prática, para o esquema de diferenças finitas, o interpolador ENO é escrito em termos de  $\overline{v}$ .

Para a condição inicial Eq. (3.47), página 71, o interpolador ENO de grau 3 é ilustrado na Figura 3.9. Pode-se observar que, de fato, o interpolador ENO não apresenta oscilação na região da descontinuidade da condição inicial. Comparando a Figura 3.9 com as Figuras 3.5, 3.6, 3.7 e 3.8 podemos ver o ganho, em relação as oscilações, quando utilizado o interpolador ENO.

**Observação 3.4** *Para um polinômio de grau 3 é preciso adicionar ao estêncil inicial 3 pontos para que cada estêncil seja composto por  $k + 1$  pontos e a escolha do melhor ponto para ser adicionado se inicia na coluna da tabela de diferenças divididas de ordem 1.*

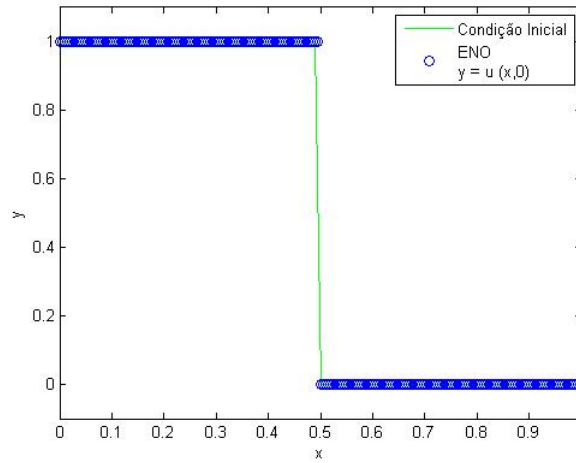


Figura 3.9: Aproximação ENO de grau 3 com  $\Delta x = 0.01$ , no intervalo  $[0, 1]$ .

### Roteiro para o código do ENO

%  $k$  - grau do polinômio

%  $is$  - o ponto mais à esquerda do estêncil

%  $m$  - é a coluna, da tabela das diferenças divididas, na qual se encontra a diferença dividida de ordem 1

1. discretização do domínio: construção da malha inteira ( $j$ ), para todo  $j$ ;
2. construção da malha metade ( $j + \frac{1}{2}$ ), para todo  $j$ , a partir da malha inteira;

$$x_{j+\frac{1}{2}} = \frac{x_j + x_{j+1}}{2};$$

3. dadas as aproximações médias  $\bar{v}_j$  da função  $v(x)$ , constrói-se a tabela de diferenças divididas fazendo  $\bar{v} = u_0(x_j)$ ;
4. escolha do melhor estêncil utilizando o processo Eq. (3.55)-Eq. (3.61)

- $is=j$
- for  $m=2:k$ 
  - if ( $abs(V(is-1, m)) \leq abs(V(is, m))$ )
  - $is = is - 1$
  - else
  - end

5. Usando a forma polinomial de Lagrange Eq. (3.42) ou a forma polinomial de Newton Eq. (3.52), construir o polinômio de grau no máximo  $k-1$  em  $S_j$  para todo  $j$ , com ordem

de precisão desejada na Eq. (3.25) quando a função for suave em  $I_j$ . Utilizar  $p_j(x)$  para aproximar a função  $v(x)$  nas fronteiras das células

$$\begin{aligned} v_{j+\frac{1}{2}}^e &= p_j\left(x_{j+\frac{1}{2}}\right) & \text{e} & & v_{j+\frac{1}{2}}^d &= p_{j+1}\left(x_{j+\frac{1}{2}}\right), \\ v_{j-\frac{1}{2}}^d &= p_j\left(x_{j-\frac{1}{2}}\right) & \text{e} & & v_{j-\frac{1}{2}}^e &= p_{j-1}\left(x_{j-\frac{1}{2}}\right). \end{aligned} \quad (3.63)$$

**Observação 3.5** *Na prática é mais conveniente contruir as aproximações Eq. (3.63) utilizando o polinômio de Newton, visto que a escolha do estêncil é feita através das diferenças divididas.*

**Observação 3.6** *O caminho mais natural para o tratamento das condições de fronteira para o esquema ENO é avaliar somente valores dentro do domínio computacional onde se dá a escolha do melhor estêncil. Na prática, para que o código decida qual estêncil está totalmente contido no domínio computacional cria-se um conjunto de células fantasmas fora do domínio computacional com valores 'grandes' e grandes variações (isto é,  $u_{-j} = (10^j)^{10}$  se  $x_{-j}$  para  $j = 1, 2, \dots$ ). Dessa maneira teremos altos valores para as diferenças divididas, forçando o esquema à ficar dentro do domínio desejado.*

*Outra opção é construir o polinômio de Newton convencional nas fronteiras.*

### Propriedades do Interpolador ENO

Segue algumas propriedades esquema ENO. Essas propriedades serão apenas comentadas, para informações adicionais consultar Shu (1997) [29].

1. A ordem de precisão,  $x \in I_j$ , é válida para toda célula  $I_j$  a qual não contém a descontinuidade. Isso significa que o esquema ENO pode recuperar a alta precisão da interpolação até para a célula a qual contém a descontinuidade.
2. O polinômio  $P_j(x)$  é monótono em qualquer célula  $I_j$  a qual contém a descontinuidade de  $V(x)$ .
3. O processo de reconstrução obedece a propriedade TVD. Esta última propriedade é consequência das duas primeiras.

O interesse maior é conseguir uma interpolação de alta ordem no intervalo de choque (na descontinuidade) a qual não seja oscilatória (não possua oscilações na interpolação próxima à descontinuidade). De acordo com estas propriedades, é o que ocorre com o interpolador ENO.

Segue um esboço da propriedade 2, para o problema de Riemann Eq. (1.55)

$$V(x) = \begin{cases} u_l, & \text{se } x < 0, \\ u_r, & \text{se } x > 0. \end{cases} \quad (3.64)$$

Seja  $P_j(x)$  o polinômio de grau  $k$  que interpola  $V(x)$  nos  $k+1$  pontos

$$x_{\frac{1}{2}} < x_{\frac{3}{2}} < \dots < x_{k+\frac{1}{2}}$$

e seja

$$x_{i_0-\frac{1}{2}} < 0 < x_{i_0+\frac{1}{2}},$$

a célula que contém a descontinuidade, para algum  $i_0 \in (0, k)$ . Para toda célula que não contém a descontinuidade  $[x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}]$ ,  $i \neq i_0$ , tem-se

$$P\left(x_{i-\frac{1}{2}}\right) = V\left(x_{i-\frac{1}{2}}\right) = V\left(x_{i+\frac{1}{2}}\right) = P\left(x_{i+\frac{1}{2}}\right),$$

considerando  $i < i_0$  e  $i > i_0$ , separadamente. Pelo teorema do valor médio existe  $\xi \in (x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}})$  tal que

$$P'(\xi_i) = 0, \quad \forall i \neq i_0.$$

como a grelha possui  $k-1$  células, as quais não possuem a descontinuidade, então existem  $k-1$  pontos  $\xi_i$  tais que  $P'(\xi_i) = 0$ . Como  $P(x)$  é um polinômio de grau no máximo  $k-1$ , então ele pode conter no máximo  $k-1$  zeros distintos, logo  $P'(\xi_i) \neq 0 \forall x \in (x_{i_0-\frac{1}{2}}, x_{i_0+\frac{1}{2}})$ . Ou seja,  $P(x)$  é monótono na célula que contém a descontinuidade  $(x_{i_0-\frac{1}{2}}, x_{i_0+\frac{1}{2}})$ , pois  $P'(\xi_i) < 0$  ou  $P'(\xi_i) > 0$  nesta célula.

## 3.5 Esquemas ENO para Leis de Conservação Unidimensionais

Nesta seção descrevemos a discretização espacial através do esquema ENO (ou do esquema WENO (apêndice A) o qual se encontra no apêndice A) para leis de conservação unidimensionais, definidas em Eq. (1.12) com condições iniciais  $u_0$ . O domínio será discretizado como na seção 3.2.

### 3.5.1 Reconstrução da Função Fluxo Numérica

Dados os valores discretizados da função  $v(x)$

$$v_j = v(x_j), \quad j = 0, 1, \dots, N, \quad (3.65)$$

deseja-se encontrar uma função de fluxo numérico  $F_{j+\frac{1}{2}}$ , relativa ao estêncil Eq. (3.28), ou seja,

$$F_{j+\frac{1}{2}} \equiv F(v_{j-r}, \dots, v_{j+s}), \quad j = 0, 1, \dots, N, \quad (3.66)$$

de forma que a aproximação da derivada  $v'(x)$ , da ordem de  $\Delta x^k$ , seja dada por

$$\frac{1}{\Delta x} (F_{j+\frac{1}{2}} - F_{j-\frac{1}{2}}) = v'(x_j) + O(\Delta x^k), \quad j = 0, 1, \dots, N, \quad (3.67)$$

de acordo com Eq. (2.44).

A solução deste problema é essencial para um esquema conservativo de alta ordem, baseado nos valores dos pontos (diferenças finitas) ou nas aproximações médias (volumes finitos).

Assumimos que a malha numérica é uniforme  $\Delta x_j = \Delta x$ , para que se possa obter uma aproximação conservativa de qualquer grau. Os esquemas ENO e WENO podem se basear em uma malha numérica variável, porém não se consegue uma aproximação conservativa de ordem maior que dois, de acordo com Shu (1997) [29].

Se for possível encontrar uma função  $h(x)$ , a qual depende do tamanho da grelha, definida implicitamente por

$$v(x) = \frac{1}{\Delta x} \int_{x-\frac{\Delta x}{2}}^{x+\frac{\Delta x}{2}} h(\xi) d\xi, \quad (3.68)$$

então

$$v'(x) = \frac{1}{\Delta x} \left[ h\left(x + \frac{\Delta x}{2}\right) - h\left(x - \frac{\Delta x}{2}\right) \right]. \quad (3.69)$$

Portanto, deve-se ter

$$F_{j+\frac{1}{2}} = h\left(x_{j+\frac{1}{2}}\right) + O(\Delta x^k), \quad (3.70)$$

para obter Eq. (3.67).

Na prática não é fácil aproximar  $h(x)$  através de Eq. (3.68). Pela definição de  $h(x)$  notamos que a função **conhecida**  $v(x)$  é a aproximação média da função **desconhecida**  $h(x)$ ; assim para encontrar a função  $h(x)$  basta usar o *processo de reconstrução* exibido na seção 3.2.1.

Seja a primitiva de  $h$ , definida por

$$H(x) = \int_{-\infty}^x h(\xi) d\xi, \quad (3.71)$$

usando a Eq. (3.68), tem-se claramente que

$$H\left(x_{j+\frac{1}{2}}\right)=\sum_{i=-\infty}^j\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}}h\left(\xi\right)d\xi=\sum_{i=-\infty}^j\bar{v}_i\Delta x. \quad (3.72)$$

Assim, dado os valores nos pontos  $v_i$ , identifica-se as aproximações médias da função  $h(x)$  em Eq. (3.68), então a função primitiva  $H(x)$  é conhecida nas interfaces das células  $x = x_{j+\frac{1}{2}}$ . Usando o processo de reconstrução, seção 3.2.1, encontra-se uma aproximação para  $h\left(x_{j+\frac{1}{2}}\right)$  a qual forma o fluxo numérico  $F_{j+\frac{1}{2}}$  em Eq. (3.67).

Em outras palavras, se o estêncil para a função  $F_{j+\frac{1}{2}}$ , em Eq. (3.67), é dado por

$$\bar{S}=\left\{x_{j-r},\ldots,x_j,\ldots,x_{j+s}\right\}, \quad (3.73)$$

sendo  $r+s=k-1$ , então o fluxo numérico é dado por

$$F_{j+\frac{1}{2}}=\sum_{i=0}^{k-1}c_{ri}v_{j-r+i},$$

sendo que as constantes  $c_{ri}$  são dadas na tabela 4.1.

## 3.6 Métodos Runge-Kutta TVD

Até o momento foi considerado apenas a discretização espacial. Esta seção trata da discretização temporal. Com o processo descrito nas seções anteriores, a EDP (1.12) se transforma em um sistema de equações diferenciais ordinárias. Para resolver este sistemas de EDO's usa-se métodos numéricos padrões, porém para se evitar resultados numéricos pobres e oscilações não-físicas do problema estudado, é importante que o método utilizado seja um método pertencente à classe TVD. Para maiores detalhes sobre o assunto consultar as referências Butcher (1987), Gottlieb e Shu (1998), Shu e Osher (1988) e Suresh e Huynh (1997) [1], [10], [30] e [31].

A classe dos métodos Runge-kutta TVD de alta ordem foi desenvolvido em Shu e Osher (1988) [30]. Estes métodos são usados para resolver um sistema de EDO's, com condição inicial escrito da forma:

$$u_t=L(u) \quad (3.74)$$

e

$$u_0=u(0), \quad (3.75)$$

resultante do método de discretização espacial da EDP

$$u_t = -f(u)_x.$$

Vale ressaltar que este método se aplica a qualquer EDP e não apenas para leis de conservação.

Note que  $L(u)$  na Eq. (3.74) é uma aproximação para a derivada  $-f(u)_x$ , usando o esquema ENO ou o esquema WENO.

De acordo com Shu e Osher (1988) [30] e Butcher (1987) [1], um método Runge-kutta explícito, para Eq. (3.74), pode ser escrito na forma:

$$\begin{aligned} u^{(i)} &= u^{(0)} + \Delta t \sum_{k=0}^{i-1} c_{ik} L(u^{(k)}), & i = 1, \dots, m \\ u^{(0)} &= u^n, & u^{(m)} = u^{n+1}, \end{aligned} \quad (3.76)$$

para problemas não dependentes do tempo, sendo  $m$  o número de estágios do método.

**Observação 3.7** Comparando as duas maneiras de se escrever os métodos de Runge-Kutta, pode-se verificar que a Eq. (3.76) é equivalente aos métodos Runge-kutta tradicionais.

Quando o operador  $L(u)$  depende explicitamente de  $t$  ou quando as condições de fronteira dependem do tempo, a fórmula geral dos métodos explícitos de Runge-kutta Eq. (3.76) se escreve como

$$\begin{aligned} u^{(i)} &= u^{(0)} + \Delta t \sum_{k=0}^{i-1} c_{ik} L(u^{(k)}, t^{(0)} + \lambda_k \Delta t), & i = 1, \dots, m \\ u^{(0)} &= u^n, & u^{(m)} = u^{n+1}, \end{aligned} \quad (3.77)$$

tal que

$$\lambda_k = \sum_{l=0}^{k-1} c_{kl}. \quad (3.78)$$

**Observação 3.8** Quando se fala em problemas que não dependem do tempo, temos equações para as quais o operador  $L(u)$  não possui a variável  $t$  explícita em sua expressão, ou seja, são da forma

- $u_t = u$ ,

ao passo que equações dependentes do tempo são da forma

- $u_t = u + t$

Para facilidade de notação, serão consideradas as equações para problemas não dependentes do tempo, o outro caso é inteiramente análogo. Com o objetivo de encontrar o método Runge-Kutta TVD, reescreve-se a Eq. (3.76) como segue

$$u^{(i)} = \sum_{k=0}^{i-1} \alpha_{ik} u^{(0)} + \Delta t \sum_{k=0}^{i-1} c_{ik} L(u^{(k)}), \quad (3.79)$$

sendo  $\alpha_{ik} \geq 0$  e  $\sum_{k=0}^{i-1} \alpha_{ik} = 1$ . Ou seja ,

$$u^{(i)} = \alpha_{i0} u^{(0)} + \sum_{k=1}^{i-1} \alpha_{ik} u^{(0)} + \Delta t \sum_{k=0}^{i-1} c_{ik} L(u^{(k)}). \quad (3.80)$$

Da Eq. (3.76) temos que

$$u^{(0)} = u^{(k)} - \Delta t \sum_{l=0}^{k-1} c_{kl} L(u^{(l)}) \quad (3.81)$$

e substituindo a Eq. (3.81) na Eq. (3.80) temos

$$\begin{aligned} u^{(i)} &= \alpha_{i0} u^{(0)} + \sum_{k=1}^{i-1} \alpha_{ik} \left( u^{(k)} - \Delta t \sum_{l=0}^{k-1} c_{kl} L(u^{(l)}) \right) + \Delta t \sum_{k=0}^{i-1} c_{ik} L(u^{(k)}) \\ &= \sum_{k=0}^{i-1} \alpha_{ik} u^{(k)} - \Delta t \sum_{k=1}^{i-1} \alpha_{ik} \sum_{l=0}^{k-1} c_{kl} L(u^{(l)}) + \Delta t \sum_{k=0}^{i-1} c_{ik} L(u^{(k)}) \\ &= \sum_{k=0}^{i-1} [\alpha_{ik} u^{(k)} + \Delta t c_{ik} L(u^{(k)})] - \Delta t \sum_{k=1}^{i-1} \alpha_{ik} \sum_{l=0}^{k-1} c_{kl} L(u^{(l)}). \end{aligned} \quad (3.82)$$

O somatório  $\sum_{k=1}^{i-1} \alpha_{ik} \sum_{l=0}^{k-1} c_{kl} L(u^{(l)})$  será reescrito para que a Eq. (3.82) fique mais concisa. Abrindo este somatório e colocando  $L(u^{(k)})$  em evidência, concluímos

$$\sum_{k=1}^{i-1} \alpha_{ik} \sum_{l=0}^{k-1} c_{kl} L(u^{(l)}) = \sum_{k=0}^{i-2} \left[ \sum_{l=k+1}^{i-1} \alpha_{il} c_{lk} \right] L(u^{(k)}) = \sum_{k=0}^{i-1} \left[ \sum_{l=k+1}^{i-1} \alpha_{il} c_{lk} \right] L(u^{(k)}) \quad (3.83)$$

onde a última igualdade é válida, pois para  $k = i-1$  se obtém uma parcela nula. Substituindo a Eq. (3.83) na Eq. (3.82) obtém-se

$$u^{(i)} = \sum_{k=0}^{i-1} [\alpha_{ik} u^{(k)} + \Delta t c_{ik} L(u^{(k)})] - \Delta t \sum_{k=0}^{i-1} \left[ \sum_{l=k+1}^{i-1} \alpha_{il} c_{lk} \right] L(u^{(k)}),$$

ou seja

$$u^{(i)} = \sum_{k=0}^{i-1} \left\{ \alpha_{ik} u^{(k)} + \Delta t L(u^{(k)}) \left[ c_{ik} - \sum_{l=k+1}^{i-1} \alpha_{il} c_{lk} \right] \right\}. \quad (3.84)$$

Definindo

$$\beta_{ik} = c_{ik} - \sum_{l=k+1}^{i-1} \alpha_{il} c_{lk}, \quad (3.85)$$

obtemos

$$u^{(i)} = \sum_{k=0}^{i-1} \left( \alpha_{ik} u^{(k)} + \beta_{ik} \Delta t L(u^{(k)}) \right), \quad i = 1, 2, \dots, m \quad (3.86)$$

De acordo com Gottlieb e Shu (1998) [10], sob certas condições para as constantes  $\alpha_{ik}, \beta_{ik}$ , o esquema Eq. (3.86) é TVD, como diz o teorema 3.3.

**Teorema 3.3** *O método Runge-kutta Eq. (3.86) é TVD se o coeficiente CFL  $c$  satisfaz*

$$c = \min_{i,k} \frac{\alpha_{ik}}{|\beta_{ik}|}, \quad (3.87)$$

para  $\alpha_{ik} \geq 0$ ,  $\beta_{ik}$  negativo. Temos  $c$  é o coeficiente CFL para discretização temporal de alta ordem.

**Demonstração:** *Veja Gottlieb e Shu (1998) [10].*

De forma análoga tem-se, para as equações dependentes do tempo, que a Eq. (3.77) pode ser reescrita como

$$u^{(i)} = \sum_{k=0}^{i-1} \left( \alpha_{ik} u^{(k)} + \beta_{ik} \Delta t L(u^{(k)}, t^{(0)} + \lambda_k \Delta t) \right), \quad i = 1, 2, \dots, m. \quad (3.88)$$

Neste trabalho será usado Runge-Kutta de terceira ordem:

$$u^{(1)} = \alpha_{10} u^{(0)} + \beta_{10} \Delta t L(u^{(0)}), \quad (3.89)$$

$$u^{(2)} = \alpha_{20} u^{(0)} + \beta_{20} \Delta t L(u^{(0)}) + \alpha_{21} u^{(1)} + \beta_{21} \Delta t L(u^{(1)}) \quad \text{e} \quad (3.90)$$

$$u^{(3)} = \alpha_{30} u^{(0)} + \beta_{30} \Delta t L(u^{(0)}) + \alpha_{31} u^{(1)} + \beta_{31} \Delta t L(u^{(1)}) + \alpha_{32} u^{(2)} + \beta_{32} \Delta t L(u^{(2)}). \quad (3.91)$$

Para menor armazenamento computacional, pode-se reescrever as Eqs. (3.89) - (3.91) de forma que seja preciso calcular o operador  $L$  apenas do passo anterior. Da Eq. (3.89), tem-se

$$L(u^{(0)}) = \frac{u^{(1)} - \alpha_{10} u^{(0)}}{\beta_{10} \Delta t}. \quad (3.92)$$

Substituindo a Eq. (3.92) na Eq. (3.90), obtemos

$$u^{(2)} = \left( \alpha_{20} - \frac{\beta_{20}}{\beta_{10}} \alpha_{10} \right) u^{(0)} + \left( \alpha_{21} - \frac{\beta_{20}}{\beta_{10}} \right) u^{(1)} + \beta_{21} \Delta t L(u^{(1)}) \quad (3.93)$$

e, consequentemente,

$$L(u^{(1)}) = \left[ u^{(2)} - \left( \alpha_{20} - \frac{\beta_{20}}{\beta_{10}} \alpha_{10} \right) u^{(0)} + \left( \alpha_{21} - \frac{\beta_{20}}{\beta_{10}} \right) u^{(1)} \right] \frac{1}{\beta_{21} \Delta t}. \quad (3.94)$$

Substituindo as Eqs. (3.92) e (3.94) na Eq. (3.91), chegamos a que

$$\begin{aligned} u^{(3)} = & \left[ \alpha_{30} - \frac{\beta_{30}}{\beta_{10}} \alpha_{10} - \frac{\beta_{31}}{\beta_{21}} \left( \alpha_{20} - \frac{\beta_{20}}{\beta_{10}} \alpha_{10} \right) \right] u^{(0)} + \left[ \frac{\beta_{30}}{\beta_{10}} + \alpha_{31} + \frac{\beta_{31}}{\beta_{21}} \left( \alpha_{21} + \frac{\beta_{20}}{\beta_{10}} \right) \right] u^{(1)} \\ & + \left( \frac{\beta_{31}}{\beta_{21}} + \alpha_{32} \right) u^{(2)} + \beta_{32} \Delta t L(u^{(2)}), \end{aligned} \quad (3.95)$$

o método Runge-Kutta com menor armazenamento é composto pelas Eqs. (3.89), (3.93) e (3.95). Para obter as constantes do método Runge-Kutta seja TVD é preciso comparar as Eqs. (3.89), (3.93) e (3.95) com a expansão em série de Taylor, segundo Gottlieb e Shu (1998) [10], resultando no sistema.

$$\begin{cases} \alpha_{32} = 1 - \alpha_{31} - \alpha_{30} \\ \beta_{32} = \frac{3\beta_{10}-2}{6P(\beta_{10}-P)} \\ \beta_{21} = \frac{1}{6\beta_{10}-\beta_{32}} \\ \beta_{31} = \frac{\frac{1}{2}-\alpha_{32}\beta_{10}\beta_{21}-P\beta_{32}}{\beta_{10}} \\ \beta_{30} = 1 - \alpha_{31}\beta_{10} - \alpha_{32}P - \beta_{31} - \alpha_{32}\beta_{32} \\ \beta_{20} = P - \alpha_{21}\beta_{10} - \beta_{21}, \end{cases} \quad (3.96)$$

sendo que  $\alpha_{21}, \alpha_{30}, \alpha_{31}, \beta_{10}$  e  $P = \beta_{20} + \alpha_{21}\beta_{10} + \beta_{21}$  são parâmetros livres. A análise para se obter os valores das constantes não é objeto de estudo deste trabalho. Segundo Gottlieb e Shu (1998) [10], após extensiva busca para se resolver o sistema Eq. (3.96), de acordo com as exigências dos possíveis parâmetros, se obtém

$$\begin{aligned} u^{(1)} &= u^{(0)} + \Delta t L(u^{(0)}), \\ u^{(2)} &= \frac{3}{4}u^{(0)} + \frac{1}{4}u^{(1)} + \frac{1}{4}\Delta t L(u^{(1)}) \quad \text{e} \\ u^{(3)} &= \frac{1}{3}u^{(0)} + \frac{2}{3}u^{(2)} + \frac{2}{3}\Delta t L(u^{(2)}). \end{aligned} \quad (3.97)$$

As Eqs. (3.97) compõem o método Runge-Kutta TVD com menor armazenamento computacional do que o método Runge-Kutta TVD escrito na sua forma original

$$\begin{aligned}
u^{(1)} &= u^{(0)} + \Delta t L(u^{(0)}), \\
u^{(2)} &= u^{(0)} + \frac{1}{4}\Delta t L(u^{(0)}) + \frac{1}{4}\Delta t L(u^{(1)}) \quad \text{e} \\
u^{(3)} &= u^{(0)} + \frac{1}{6}L(u^{(0)}) + \frac{1}{6}\Delta t L(u^{(1)}) + \frac{2}{3}\Delta t L(u^{(2)}), \tag{3.98}
\end{aligned}$$

pois requer o cálculo do operador  $L$  apenas do passo anterior.

Comparando Eq. (3.96) e Eq. (3.98), pode-se recuperar (obter) os valores para as constantes  $\alpha, \beta$ , expostos nas tabelas 3.2 e 3.2, respectivamente.

|                             |                   |                             |
|-----------------------------|-------------------|-----------------------------|
| $\alpha_{10} = 1$           |                   |                             |
| $\alpha_{20} = 1$           | $\alpha_{21} = 0$ |                             |
| $\alpha_{30} = \frac{1}{3}$ | $\alpha_{31} = 0$ | $\alpha_{32} = \frac{2}{3}$ |

Tabela 3.2: Constantes  $\alpha_{ik}$  para Runge-Kutta de 3ª ordem

|                            |                            |                            |
|----------------------------|----------------------------|----------------------------|
| $\beta_{10} = 1$           |                            |                            |
| $\beta_{20} = \frac{1}{4}$ | $\beta_{21} = \frac{1}{4}$ |                            |
| $\beta_{30} = 0$           | $\beta_{31} = 0$           | $\beta_{32} = \frac{2}{3}$ |

Tabela 3.3: Constantes  $\beta_{ik}$  para Runge-Kutta de 3ª ordem

Para os problemas dependentes do tempo, falta apenas obter as constantes  $\lambda_k$ . De posse das constantes  $\alpha_k, \beta_k$  pode-se obter as constantes  $\lambda_k$ .

Da Eq. (3.85) tem-se

$$c_{ik} = \beta_{ik} + \sum_{l=k+1}^{i-1} \alpha_{il} c_{lk} \quad \text{e}$$

$$\sum_{k=0}^{i-1} c_{ik} = \sum_{k=0}^{i-1} \beta_{ik} + \sum_{k=0}^{i-1} \sum_{l=k+1}^{i-1} \alpha_{il} c_{lk}.$$

Da Eq. (3.78), obtemos

$$\lambda_i = \sum_{k=0}^{i-1} \beta_{ik} + \sum_{l=1}^{i-1} \alpha_{il} \sum_{k=0}^{l-1} c_{lk},$$

logo

$$\lambda_i = \sum_{k=0}^{i-1} \beta_{ik} + \sum_{l=1}^{i-1} \alpha_{il} d_l, \quad 1 \leq l \leq m-1. \tag{3.99}$$

Para o método Runge-Kutta TVD de três estágios, dependentes do tempo, precisa-se de duas constantes  $\lambda_1, \lambda_2$ . De acordo com a Eq. (3.99) e com as tabelas 3.2-3.3, obtém-se

$$\boxed{\lambda_1 = 1 \quad \lambda_2 = \frac{1}{2}}$$

Tabela 3.4: Constantes  $\lambda_i$  para Runge-Kutta de 3ª ordem

Assim como para os problemas não dependentes do tempo, pode-se reescrever o método Eq. (3.88) de forma que requiera menor armazenamento computacional, como já obtido nas Eqs. (3.98) para problemas não dependentes do tempo. Como os cálculos são inteiramente análogos será apresentado apenas o resultado:

$$\begin{aligned} u^{(1)} &= u^{(0)} + \Delta t L(u^{(0)}, t^{(0)}), \\ u^{(2)} &= \frac{3}{4}u^{(0)} + \frac{1}{4}u^{(1)} + \frac{1}{4}\Delta t L(u^{(1)}, t^{(0)} + \Delta t) \quad \text{e} \\ u^{(3)} &= \frac{1}{3}u^{(0)} + \frac{2}{3}u^{(2)} + \frac{2}{3}\Delta t L\left(u^{(2)}, t^{(0)} + \frac{1}{2}\Delta t\right). \end{aligned} \quad (3.100)$$

Como ilustração da eficiência dos métodos Runge-Kutta TVD será apresentado o exemplo 3.4.

**Exemplo 3.4** Considere o problema de valor inicial

$$\begin{cases} u_t = u + t, \\ u(1) = 2. \end{cases} \quad (3.101)$$

Como a equação diferencial ordinária depende explicitamente de  $t$ , o método utilizado para aproximá-la é dado pela Eq. (3.100). Na Figura 3.10 comparou-se o método de Runge-

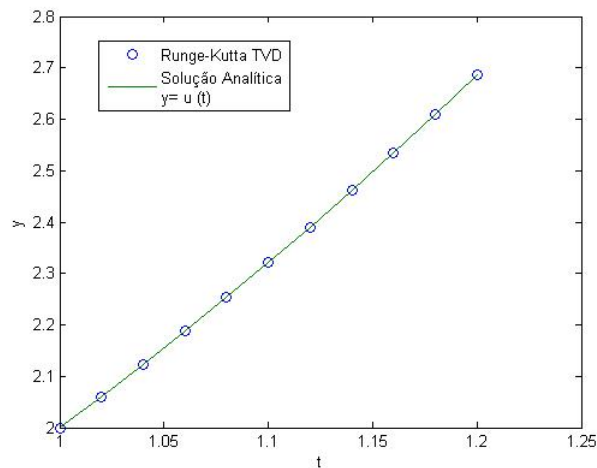


Figura 3.10: Runge-Kutta TVD de 3ª ordem para EDO's

Kutta TVD de 3ª ordem com a solução analítica  $u(t) = 4e^{(t-1)} - t - 1$ . A Figura 3.11 ilustra o comportamento do erro pontual, para o método Runge-Kutta TVD de 3ª ordem, da aproximação numérica exibida na Figura 3.10 e pode-se notar que este método apresenta uma aproximação de alta ordem, ou seja, da ordem de  $10^{-7}$ .

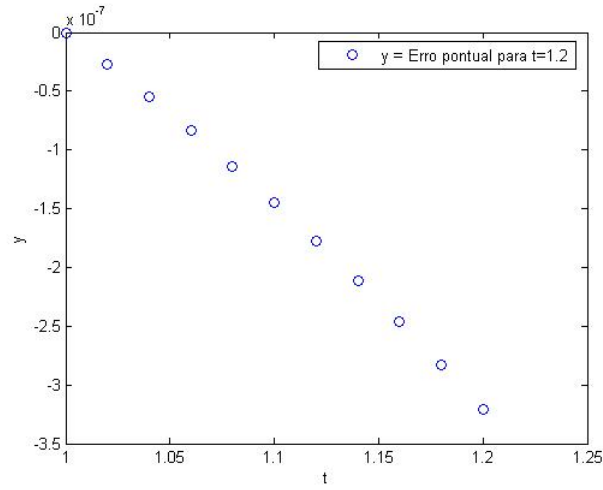


Figura 3.11: Ilustração do erro pontual do método de Runge-Kutta de 3ª ordem para  $N = 10$  no intervalo  $[0, 1.2]$ .

É importante observar que, mesmo trabalhando com métodos não-oscilatórios na discretização espacial, ENO-WENO, se o método da aproximação temporal, Runge-Kutta, não pertencer a classe dos métodos TVD, os resultados na região do choque (descontinuidade da solução) podem apresentar oscilações. As Figuras 4.6 e 4.9 trazem uma comparação entre os métodos de Runge-Kutta TVD de segunda ordem

$$\begin{aligned} u^{(1)} &= u^{(0)} + \Delta t L(u^{(0)}) \\ u^{(2)} &= \frac{1}{2}u^{(0)} + \frac{1}{2}u^{(1)} + \frac{1}{2}\Delta t L(u^{(1)}) \end{aligned} \quad (3.102)$$

e os métodos de Runge-Kutta de segunda ordem não-TVD

$$\begin{aligned} u^{(1)} &= u^{(0)} - 20\Delta t L(u^{(0)}) \\ u^{(2)} &= u^{(0)} + \frac{41}{40}\Delta t L(u^{(0)}) - \frac{1}{40}\Delta t L(u^{(1)}) . \end{aligned} \quad (3.103)$$

# Capítulo 4

## Resultados Numéricos e Conclusões

### 4.1 Resultados Numéricos

Os resultados numéricos obtidos neste capítulo utilizaram, para a aproximação no tempo os métodos Runge-Kutta de terceira ordem dado na Eq. (3.97) e de segunda ordem dado na Eq. (3.102) (páginas 85 e 88) e, para a aproximação no espaço, o esquema das diferenças finitas na seção 3.1 (página 58) baseado no esquema ENO de quarta ordem.

**Exemplo 4.1** Para a equação de advecção o fluxo 'splitting' Eq. (3.6), página 59, é dado por

$$f^+(u) = \frac{1}{2}(a.u + \alpha u) \text{ e } f^-(u) = \frac{1}{2}(a.u - \alpha u), \quad (4.1)$$

onde  $\alpha = |a|$ . Considere o problema de Riemann, com  $0.1 \leq x \leq 1$  e  $t \geq 0$ , composto pela equação de advecção com  $a = 1$ :

$$u_t + u_x = 0,$$

e condição inicial dada por:

$$u_0(x) = \begin{cases} 1, & \text{se } x < 0, \\ 0, & \text{se } x > 0. \end{cases}$$

Para este caso, tem-se que  $\alpha = 1$ . Os resultados numéricos, para este problema, foram obtidos para o domínio  $[-0.1, 1]$ , com  $t = 0.5$  e  $CFL = 0.5$ . A Figura 4.1 exibe a solução obtida com o método Runge-Kutta TVD de 3ª ordem,  $\Delta t = 0.00275$  e  $\Delta x = 0.0055$ . Na Figura 4.2 estão ilustrados os resultados numéricos obtidos com o método Runge-Kutta TVD de 2ª,  $\Delta t = 0.00275$  e  $\Delta x = 0.0055$ .

Comparando os resultados obtidos através do método numérico semi-discreto ilustrado nas Figuras 4.1 e 4.2, para a equação de advecção, com o método numérico de Lax-Friedrichs e o método numérico 'upwind' dados pelas Eqs. 2.26 e 2.36 e ilustrados nas Figuras 2.5 e 2.6, respectivamente, observa-se um ganho significativo em relação à suavização da solução na

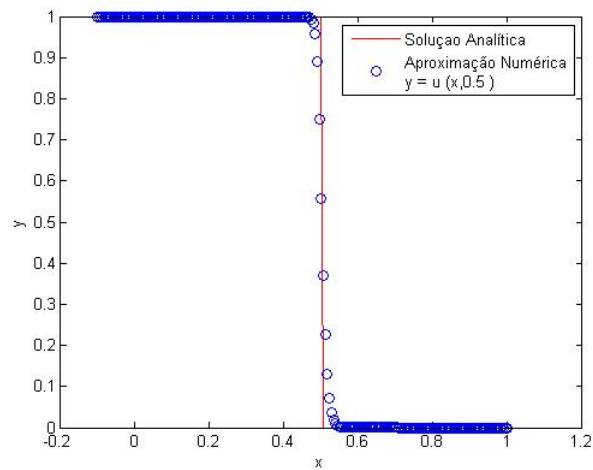


Figura 4.1: Aproximação do exemplo 4.1: Runge-Kutta de 3ª ordem no domínio  $[-0.1, 1]$  com  $\Delta x = 0.0055$ ,  $t = 0.5$ ,  $\Delta t = 0.00275$  e  $CFL = \Delta t / \Delta x = 0.5$ .

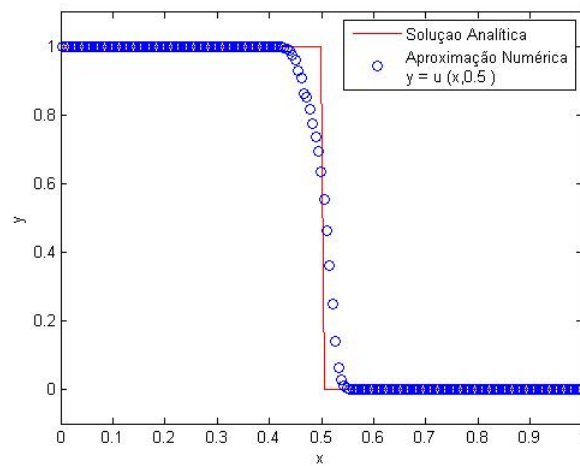


Figura 4.2: Aproximação do exemplo 4.1: Runge-Kutta de 2ª ordem no domínio  $[-0.1, 1]$  com  $\Delta x = 0.0055$ ,  $t = 0.5$ ,  $\Delta t = 0.00275$  e  $CFL = \Delta t / \Delta x = 0.5$ .

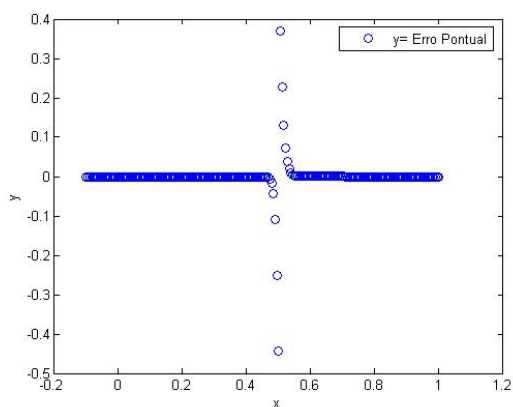


Figura 4.3: Ilustração do erro pontual para a aproximação ilustrada na figura 4.1.

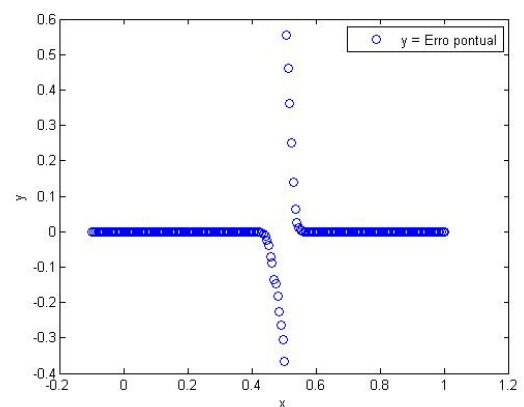


Figura 4.4: Ilustração do erro pontual para a aproximação ilustrada na figura 4.2.

| Método       | N   | Erro     |
|--------------|-----|----------|
| R-k 2ª ordem | 10  | 1.062e-1 |
|              | 50  | 3.643e-2 |
|              | 100 | 3.131e-2 |
|              | 200 | 2.062e-2 |
| R-k 3ª ordem | 10  | 1.21e-1  |
|              | 50  | 2.673e-2 |
|              | 100 | 1.627e-2 |
|              | 200 | 9.602e-3 |

Tabela 4.1: Tabela de erros para equação de advecção, na norma  $L_1$ , com ENO de 4ª ordem.

região da formação de choque. De acordo com a tabela 4.1 os métodos Runge-Kutta TVD para a equação de advecção, de segunda e de terceira ordem, apresentam resultados com ordem de precisão semelhante para uma malha computacional de até 100 subintervalos, com o mesmo valor da condição CFL. Quando se refina a malha computacional o método Runge-Kutta TVD de terceira ordem apresentou maior ordem precisão do que o método Runge-Kutta TVD de segunda ordem e também apresentou menor suavização, na região da formação do choque. Como pode ser visto nas Figuras 4.1 e 4.2 e na tabela 4.1.

As Figuras 4.3 e 4.4 ilustram o comportamento do erro pontual da aproximação numérica, para equação de advecção, ilustradas nas Figuras 4.1 e 4.2 respectivamente. Podemos ver que em regiões de suavidade o erro pontual é praticamente zero; já na região do choque, há um aumento significativo do mesmo. Ainda assim, estes erros são menores do que os erros cometidos pelos métodos numéricos tradicionais.

**Exemplo 4.2** Para a equação de Burgers

$$u_t + \left( \frac{u^2}{2} \right)_x = 0,$$

o fluxo 'splitting' Eq. (3.6), página 59, é dado por

$$f^+(u) = \frac{1}{2} \left( \frac{u^2}{2} + \alpha u \right) \quad e \quad f^-(u) = \frac{1}{2} \left( \frac{u^2}{2} - \alpha u \right), \quad (4.2)$$

onde  $\alpha = \max |u|$ , o qual deve ser atualizado a cada passo de tempo real do Runge-Kutta.

Considere o problema de Riemann composto pela equação de Burgers, na qual a função fluxo é dada por  $f(u) = \frac{u^2}{2}$  e pela condição inicial:

$$u_0(x) = \begin{cases} 1.2, & \text{se } x < 0, \\ 0.4, & \text{se } x > 0, \end{cases}$$

no domínio  $-0.1 \leq x \leq 1.6$ ,  $t \geq 0$ . Os resultados numéricos obtidos através do método Runge-Kutta de 3ª ordem são ilustrados na Figura 4.5 e na Figura 4.6 está ilustrado os resultados numéricos obtidos através do método Runge-Kutta TVD de 2ª ordem. Os resultados numéricos,

para ambos os métodos, considerou o domínio  $[-0.1, 1.6]$  com  $t = 1$ ,  $\Delta t = 0.00425$ ,  $CFL = 0.6$  e  $\Delta x = 0.0085$ .

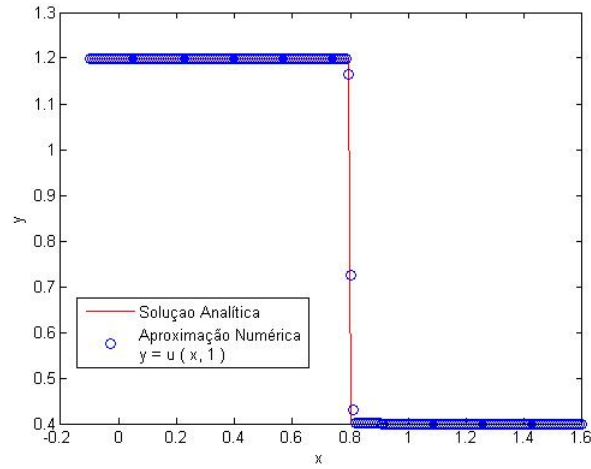


Figura 4.5: Aproximação do exemplo 4.2: Runge-Kutta de 3ª ordem no domínio  $[-0.1, 1.6]$ ,  $\Delta x = 0.0085$  e  $t = 1$ ,  $\Delta t = 0.00425$ ,  $CFL = 0.6$ .

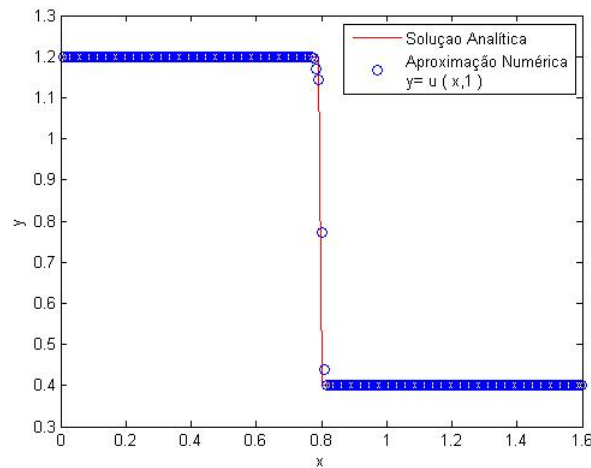


Figura 4.6: Aproximação do exemplo 4.2: Runge-Kutta de 2ª ordem no domínio  $[-0.1, 1.6]$ ,  $\Delta x = 0.0085$  e  $t = 1$ ,  $\Delta t = 0.00425$ ,  $CFL = 0.6$ .

| Método       | N   | Erro      |
|--------------|-----|-----------|
| R-k 2ª ordem | 10  | —         |
|              | 50  | —         |
|              | 100 | 8.629e-3  |
|              | 200 | 4.299e-3  |
| R-k 3ª ordem | 10  | 1.7494e-1 |
|              | 50  | 3.003e-2  |
|              | 100 | 5.486e-3  |
|              | 200 | 3.337e-3  |

Tabela 4.2: Tabela de erros para equação de Burgers, na norma  $L_1$ , com ENO de 4ª ordem.

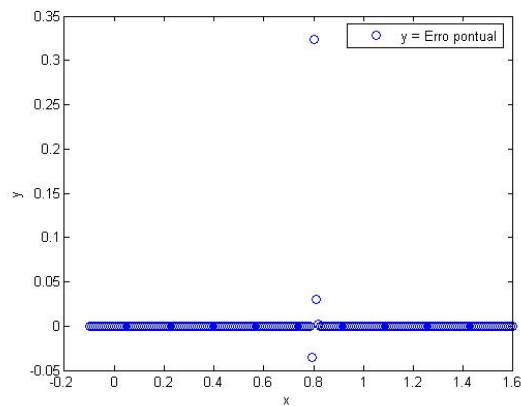


Figura 4.7: Erro pontual da aproximação ilustrada na figura 4.5.

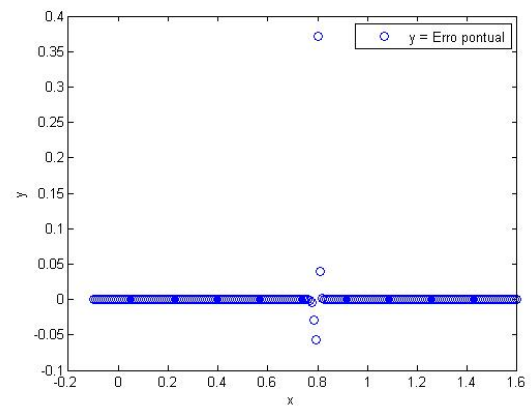


Figura 4.8: Erro pontual da aproximação ilustrada na figura 4.6.

Comparando os resultados obtidos, através do método numérico semi-discreto ilustrados nas Figuras 4.5 e 4.6, para a equação de Burgers, com o método numérico ‘upwind’ dado pela Eq. (2.46) e ilustrado na Figura 2.10 observa-se o ganho em relação a aproximação obtida na região da descontinuidade da solução. As Figuras 4.5 e 4.6 ilustram que a aproximação numérica conseguiu reproduzir a descontinuidade da solução sem suavizá-la e de acordo com a tabela 4.2 temos que esta aproximação é de alta ordem quando a malha computacional é refinada.

As Figuras 4.7 e 4.8 ilustram o comportamento do erro pontual da aproximação numérica, para equação de Burgers, ilustradas nas Figuras 4.5 e 4.6 respectivamente. Podemos ver que os métodos Runge-Kutta TVD de segunda e de terceira ordem apresentam resultados numéricos semelhantes. Podemos ver que em regiões de suavidade da solução o erro pontual, da aproximação numérica, é praticamente zero aumentando significativamente na região do choque. Ainda assim, apresentam erros menores do que os métodos numéricos tradicionais. De acordo com a tabela 4.2 os métodos Runge-Kutta TVD de terceira ordem e de segunda ordem apresentaram ordem de precisão semelhantes para uma malha computacional de 100 e de 200 subintervalos e esta relação se mantém quando a malha é refinada (ao menos para os valores testados 300 e 400 subintervalos). Para uma malha computacional de 10 e de 50 subintervalos o método Runge-Kutta TVD de segunda ordem não conseguiu aproximar a solução.

Para exibir a importância dos métodos Runge-Kutta TVD é ilustrado, na Figura 4.9, os resultados numéricos obtidos para o problema do exemplo 4.2, através do método numérico Runge-Kutta **não** TVD de segunda ordem Eq. (3.103).

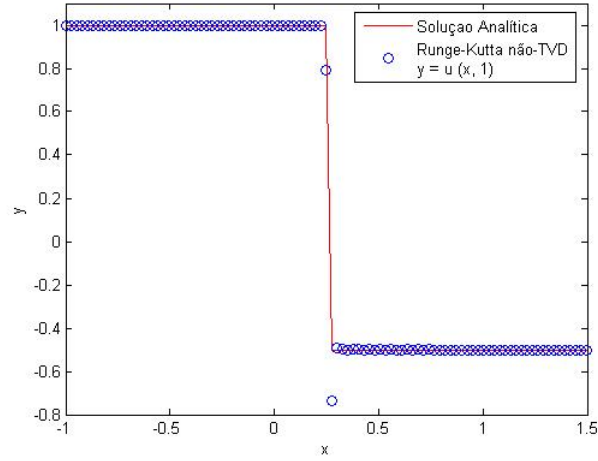


Figura 4.9: Aproximação Runge-Kutta não TVD de 2ª ordem:  $[-1, 1.6]$ ,  $\Delta x = 0.026$  e  $t = 1$ ,  $\Delta t = 0.01$ ,  $CFL = 0.3846$ .

**Exemplo 4.3** *Este exemplo traz o estudo numérico do problema de Riemann, estudado no capítulo 1, composto pela equação de Burgers*

$$u_t + \left( \frac{u^2}{2} \right)_x = 0,$$

e pela condição inicial

$$u_0(x) = \begin{cases} 0, & \text{se } x \leq 0, \\ 1, & \text{se } x > 0, \end{cases}$$

no domínio  $-0.2 \leq x \leq 1.5$ ,  $t \geq 0$ . Neste caso temos a formação do leque de rarefação (como visto na página 26).

Os resultados foram obtidos no domínio  $[-0.2, 1.5]$ , com  $\Delta t = 0.0085$ ,  $CFL = 0.5$  e  $\Delta x = 0.017$  e são ilustrados nas Figuras 4.10, 4.11 e 4.12, para  $t = 0$ ,  $t = 0.5$  e  $t = 1$  respectivamente. Podemos ver que o método numérico conseguiu aproximar a solução de entropia (Eq. (1.72)) da lei de conservação quando ocorre a formação do leque de rarefação. Na Figura 4.13 está ilustrado a comparação entre a solução exata nos tempos citados. Podemos ver, nesta figura, a formação do leque de rarefação. Nas Figuras 4.14 e 4.14 está ilustrado o erro pontual, para as aproximação obtidas nas Figuras 4.10 e 4.12 respectivamente, o qual é da ordem de  $10^{-2}$ . A tabela 4.3 mostra o erro na norma  $L_1$ , deste exemplo, para  $t = 0, 1$ .

| Tempo | Erro norma $L_1$ |
|-------|------------------|
| 0     | 4.25e-3          |
| 1     | 7.179e-3         |

Tabela 4.3: Tabela de erros na norma  $L_1$ , para a onda de rarefação, com ENO de 4ª ordem e Runge-Kutta de 3ª ordem.

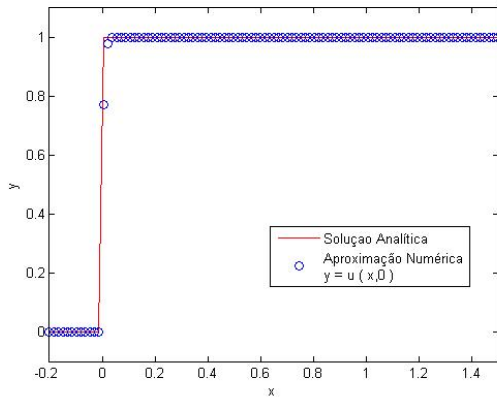


Figura 4.10: Aproximação Runge-Kutta de terceira ordem para rarefação :  $[-0.2, 1.5]$ ,  $\Delta x = 0.017$  e  $t = 0$ ,  $CFL = 0.5$ .

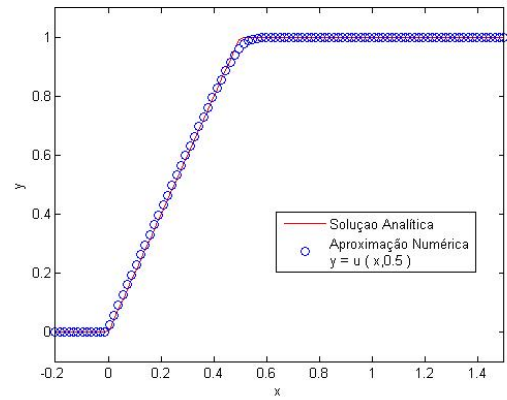


Figura 4.11: Solução do exemplo 4.3: no domínio  $[-0.2, 1.5]$ ,  $\Delta x = 0.017$  e  $t = 0.5$ ,  $\Delta t = 0.0085$ ,  $CFL = 0.5$ .

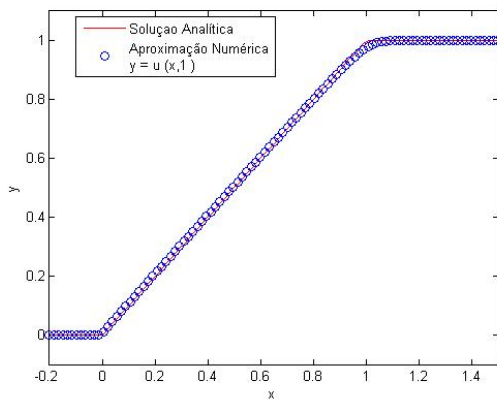


Figura 4.12: Aproximação Runge-Kutta de terceira ordem para rarefação:  $[-0.2, 1.5]$ ,  $\Delta x = 0.017$  e  $t = 1$ ,  $\Delta t = 0.0085$ ,  $CFL = 0.5$ .

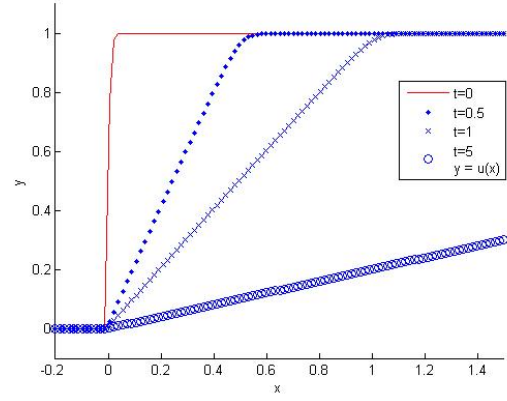


Figura 4.13: Comparação da aproximação numérica, formando o leque de rarefação para  $t = 0, 0.5, 1, 5$ :  $[-0.2, 1.5]$ ,  $\Delta x = 0.017$ ,  $\Delta t = 0.0085$ ,  $CFL = 0.5$ .

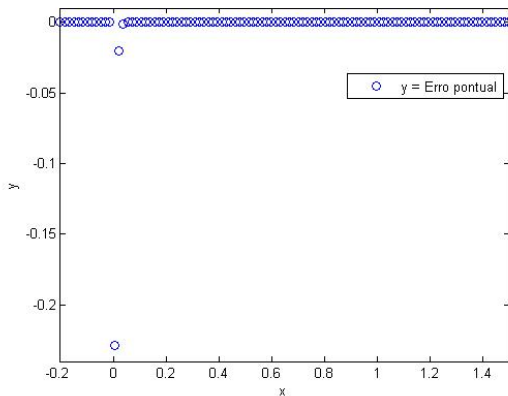


Figura 4.14: Erro pontual da aproximação ilustrada na figura 4.10.

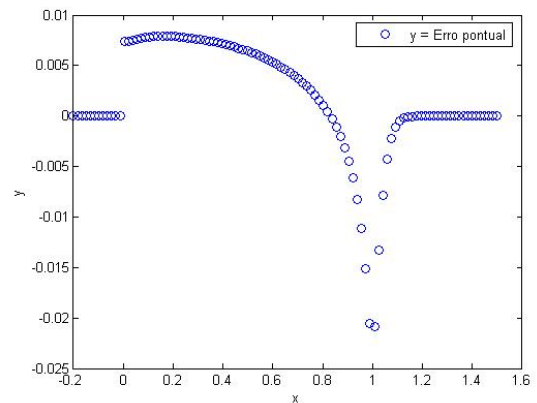


Figura 4.15: Erro pontual da aproximação ilustrada na figura 4.12.

## 4.2 Conclusões

Os esquemas Runge-Kutta TVD apresentam alta ordem de precisão em regiões de suavidade da solução, decaindo significativamente na região de choque, porém a aproximação numérica é ainda melhor do que a aproximação numérica dos métodos tradicionais. Para a equação de Burgers as aproximações numéricas obtidas pelos métodos Runge-Kutta TVD de segunda e de terceira ordem são muito próximos, ao passo que, para a equação de advecção as aproximações numéricas obtidas apresentam uma diferença considerável.

É importante a utilização do método Runge-Kutta TVD, pois mesmo tendo utilizado um método não oscilatório para aproximação espacial (ENO) quando o método utilizado para a discretização temporal foi um método Runge-Kutta não TVD o resultado numérico apresentou oscilação.

De acordo com Shu (1997) [29] existem poucos resultados teóricos sobre os esquemas ENO e WENO, mas na prática estes métodos são muito robustos e estáveis.

De acordo com Jiang e Shu (1996) [15] o esquema WENO possui alta ordem de precisão ( $O(\Delta x^{2k-1})$ ) em regiões de suavidade da função fluxo, onde a ordem do esquema ENO é de ( $O(\Delta x^k)$ ) ao menos para  $k = 2, 3$ . Contudo em regiões de descontinuidades a ordem do interpolador WENO é a mesma do interpolador ENO.

A continuidade natural deste trabalho se resume em:

- o estudo de funções de fluxos descontínuas;
- o estudo de aplicação dos métodos numéricos semi-discretos para diferentes equações;
- o estudo destes métodos em malhas não-estruturadas;
- o estudo destes métodos em domínios de dimensões mais elevadas.

# Apêndice A

## WENO

O método numérico Essencialmente Não-Oscilatório Ponderado (WENO) é a combinação convexa de todos os possíveis polinômios dentro do estêncil. Logo o esquema WENO trabalha com todos os estêncis possíveis e resulta no decaimento das oscilações na região da descontinuidade, com maior precisão do que a apresentada pelo ENO. O esquema WENO foi proposto por Liu, Osher e Chan (1994) em [22]. Supondo que  $k$  é a quantidade de pontos escolhidos para o estêncil veremos que a ordem do esquema WENO será  $O(\Delta x^{2k-1})$ .

### Deficiências do ENO

1. Quando ambos os lados da Eq. (3.59) for zero ocorre a livre escolha do estêncil. Isto acontece nas células próximas aos zeros da solução e de suas derivadas, ocasionando a escolha errada do estêncil e, conseqüentemente, prejuízos na aproximação.
2. O processo de escolha do estêncil envolve estrutura 'if', o que pode não ser muito conveniente computacionalmente.
3. Para um polinômio de grau  $k$  precisa-se de  $k + 1$  pontos, obtendo  $k$  células. No total de possibilidades dos  $k$  estêncis tem-se um total de  $2k - 1$  células. No esquema ENO é usado apenas um estêncil de  $k$  células resultando na  $k$ -ésima ordem de precisão e se todos os possíveis estêncis forem usados, obtém-se a ordem de precisão  $2k - 1$ .

**Exemplo A.1** Para um polinômio de ordem  $k = 3$ , tem-se 3 possíveis estêncis. Considere a célula inicial  $\{x_3, x_4\}$ , as possibilidades de estêncis são :

$$S_0 = \{x_3, x_4, x_5\}, \quad S_1 = \{x_2, x_3, x_4\} \quad e \quad S_2 = \{x_1, x_2, x_3\}.$$

resultando no estêncil total

$$\{x_1, x_2, x_3, x_4, x_5\},$$

composto por  $2k - 1$  pontos da malha inteira ( $2k - 1$  células) logo a ordem de precisão do polinômio é  $2k - 1 = 5$ .

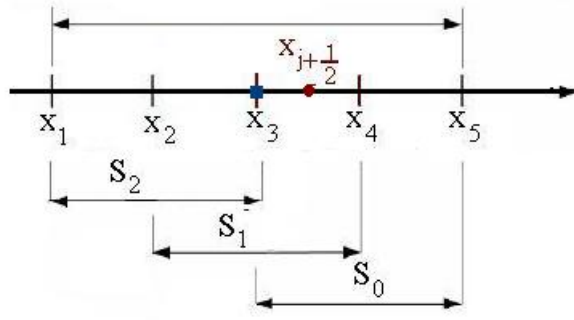


Figura A.1: Possibilidades de estênceis para o interpolador ENO de ordem 3.

De acordo com Shu (1997) [29], na literatura encontram-se algumas propostas para resolver estas questões, com alguns resultados positivos. O ganho realmente significativo ocorreu com o esquema WENO desenvolvido em Liu; Osher e Chan (1994) [22]. Para simplicidade será considerada uma grelha uniforme  $\Delta x_j = \Delta x$ .

A idéia básica do WENO, baseado em um estêncil fixo, é usar a combinação convexa de todos os polinômios gerados nos possíveis estênceis, ao invés de usar apenas um candidato, como no esquema ENO.

Sejam os  $k$  estênceis

$$\bar{S}_r(j) = \{x_{j-r}, x_{j-r+1}, \dots, x_{j-r+k-1}\} \quad \text{com} \quad r = 0, 1, \dots, k-1, \quad (\text{A.1})$$

sendo que cada estêncil possui  $k+1$  pontos e  $2k-1$  células.

**Observação A.1** Temos que

$$\begin{aligned} \bar{S}_0(j) &= \{x_j, x_{j+1}, \dots, x_{j+k-1}\}, \\ \bar{S}_1(j) &= \{x_{j+1}, x_{j+2}, \dots, x_{j+k}\}, \\ &\vdots \\ \bar{S}_{k-1}(j) &= \{x_{j-k+1}, x_{j-k+2}, \dots, x_j\} \quad \text{com} \quad j = 1, \dots, N. \end{aligned}$$

Os  $k$  estênceis produzem  $k$  diferentes valores de reconstrução para  $v_{j+\frac{1}{2}}$ . De acordo com Eq. 3.30 tem-se as aproximações dadas por

$$v_{j+\frac{1}{2}}^{(r)} = \sum_{i=0}^{k-1} c_{ri} \bar{v}_{j-r+i} \quad \text{com} \quad r = 0, 1, \dots, k-1. \quad (\text{A.2})$$

O processo de reconstrução WENO trabalha com todos os possíveis valores  $v_{j+\frac{1}{2}}^{(r)}$ , para isto faz-se a combinação convexa de todos esses valores, para estimar uma melhor aproximação para  $v$

nas fronteiras das células  $v_{j+\frac{1}{2}}$   $j = 0, \dots, N$ .

Sejam as constantes  $w_r$  tais que

$$\sum_{r=0}^{k-1} w_r = 1 \quad (\text{A.3})$$

e  $w_r \geq 0$ , a aproximação WENO é dada por

$$v_{j+\frac{1}{2}} = \sum_{i=0}^{k-1} w_r v_{j+\frac{1}{2}}^{(r)} \quad \text{com} \quad r = 0, 1, \dots, k-1. \quad (\text{A.4})$$

Aparentemente a chave para o sucesso do WENO está na escolha das constantes  $w_r$ , então o cálculo do método WENO se resume em calcular essas constantes.

Se a função  $v(x)$  é suave em todos os estênceis da Eq. A.1, a teoria de interpolação polinomial tradicional garante que existem constantes  $d_r$  tais que

$$v_{j+\frac{1}{2}} = \sum_{i=0}^{k-1} d_r v_{j+\frac{1}{2}}^{(r)} = v\left(x_{j+\frac{1}{2}}\right) + O\left(\Delta x^{2k-1}\right) \quad \text{para} \quad r = 0, 1, \dots, k-1. \quad (\text{A.5})$$

Para encontrar os valores das constantes  $d_r$  deve-se igualar os polinômios de Lagrange construídos para cada um dos  $k$  estênceis, de modo que cada estêncil possui  $k+1$  pontos (veja observação A.1) e para o estêncil composto pelos  $2k$  pontos. Na Eq. A.5 é dada a combinação convexa dos valores  $v_{j+\frac{1}{2}}$  nos  $k$  estênceis de Eq. A.1. Por outro lado deseja-se aproximar  $v_{j+\frac{1}{2}}$  utilizando o estêncil composto por  $2k-1$  células.

O caso geral não será exposto aqui, como ilustração será feito o caso  $k = 3$ .

Considerando os 3 possíveis estênceis  $\bar{S}_r(j)$ , em Eq. A.1,

$$\bar{S}_0(j) = \{x_j, x_{j+1}, x_{j+2}\}, \quad \bar{S}_1(j) = \{x_{j-1}, x_j, x_{j+1}\} \quad e \quad \bar{S}_2(j) = \{x_{j-2}, x_{j-1}, x_j\}, \quad (\text{A.6})$$

as respectivas aproximações para  $v_{j+\frac{1}{2}}$  são dadas por

$$\begin{aligned} v_{j+\frac{1}{2}}^{(0)} &= c_{00}\bar{v}_j + c_{01}\bar{v}_{j+1} + c_{02}\bar{v}_{j+2}, \\ v_{j+\frac{1}{2}}^{(1)} &= c_{10}\bar{v}_{j-1} + c_{11}\bar{v}_j + c_{12}\bar{v}_{j+1}, \\ v_{j+\frac{1}{2}}^{(2)} &= c_{20}\bar{v}_{j-2} + c_{21}\bar{v}_{j-1} + c_{22}\bar{v}_j. \end{aligned} \quad (\text{A.7})$$

Consultando a tabela 4.1 para  $k = 3$ ,  $r = 0, 1, 2$  e  $i = 0, 1, 2$ , as Eqs. (A.7) se reduzem à

$$\begin{aligned} v_{j+\frac{1}{2}}^{(0)} &= \frac{1}{3}\bar{v}_j + \frac{5}{6}\bar{v}_{j+1} - \frac{1}{6}\bar{v}_{j+2}, \\ v_{j+\frac{1}{2}}^{(1)} &= -\frac{1}{6}\bar{v}_{j-1} + \frac{5}{6}\bar{v}_j + \frac{1}{3}\bar{v}_{j+1}, \\ v_{j+\frac{1}{2}}^{(2)} &= \frac{1}{3}\bar{v}_{j-2} + -\frac{7}{6}\bar{v}_{j-1} + \frac{11}{6}\bar{v}_j. \end{aligned} \quad (\text{A.8})$$

Sustituindo as Eqs. A.8 na Eq. (A.5) tem-se

$$\begin{aligned} v_{j+\frac{1}{2}} &= d_0 v_{j+\frac{1}{2}}^{(0)} + d_1 v_{j+\frac{1}{2}}^{(1)} + d_2 v_{j+\frac{1}{2}}^{(2)} \\ v_{j+\frac{1}{2}} &= \frac{1}{3}d_2 \bar{v}_{j-2} - \left(\frac{1}{6}d_1 - \frac{7}{6}d_2\right) \bar{v}_{j-1} + \left(\frac{11}{6}d_2 + \frac{5}{6}d_1 + \frac{1}{3}d_0\right) \bar{v}_j + \left(\frac{1}{3}d_1 + \frac{5}{6}d_0\right) \bar{v}_{j+1} - \frac{1}{6}d_0 \bar{v}_{j+2} \end{aligned} \quad (\text{A.9})$$

Por outro lado, desenvolvendo o polinômio de Lagrange para  $2k-1 = 5$  células, cujo estêncil é dado por  $\bar{S}(j) = \{x_{j-2}, x_{j-1}, x_j, x_{j+1}, x_{j+2}\}$ , obtem-se

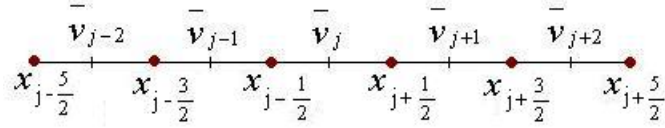


Figura A.2: Estêncil do polinômio interpolador de Lagrange para cálculo das constantes  $d_r$ ;  $r = 0, 1, 2$ .

$$v_{j+\frac{1}{2}} = \sum_{i=0}^{(2k-1)-1} \tilde{c}_{ri} \bar{v}_{j+i-2} = \tilde{c}_{20} \bar{v}_{j-2} + \tilde{c}_{21} \bar{v}_{j-1} + \tilde{c}_{22} \bar{v}_j + \tilde{c}_{23} \bar{v}_{j+1} + \tilde{c}_{24} \bar{v}_{j+2},$$

e de acordo com a tabela 4.1 o polinômio de Lagrange se reduz a

$$v_{j+\frac{1}{2}} = \frac{1}{30} \bar{v}_{j-2} - \frac{13}{60} \bar{v}_{j-1} + \frac{47}{60} \bar{v}_j + \frac{9}{20} \bar{v}_{j+1} - \frac{1}{20} \bar{v}_{j+2}. \quad (\text{A.10})$$

Igualando as Eqs. (A.9) e (A.10) obtém-se o sistema

$$\begin{cases} \frac{1}{3}d_2 = \frac{1}{30} \\ -\frac{1}{6}d_0 = -\frac{1}{20} \\ \frac{1}{3}d_1 + \frac{5}{6}d_0 = \frac{9}{20} \\ -\frac{1}{6}d_1 + \frac{7}{6}d_2 = -\frac{13}{60} \\ \frac{11}{6}d_2 + \frac{5}{6}d_1 + \frac{1}{3}d_0 = \frac{47}{60}, \end{cases} \quad (\text{A.11})$$

de onde temos

$$\begin{cases} d_0 = \frac{3}{10} \\ d_1 = \frac{3}{5} \\ d_2 = \frac{1}{10} \end{cases} \quad (\text{A.12})$$

De maneira análoga pode-se calcular os valores de  $d_r$ , para qualquer valor de  $r$ . Na tabela A.1 encontram-se os valores para  $k = 1, 2, 3, 4$ . Verifica-se que  $\sum_{r=0}^{k-1} d_r = 1$ .

| k | i=0            | i=1             | i=2             | i=3            |
|---|----------------|-----------------|-----------------|----------------|
| 1 | 1              |                 |                 |                |
| 2 | $\frac{2}{3}$  | $\frac{1}{3}$   |                 |                |
| 3 | $\frac{3}{10}$ | $\frac{3}{5}$   | $\frac{1}{10}$  |                |
| 4 | $\frac{4}{35}$ | $\frac{18}{35}$ | $\frac{12}{35}$ | $\frac{1}{35}$ |

Tabela A.1: Constantes  $d_r$  referentes a Eq. (A.5)

Realizando cálculos análogos aos anteriores é possível verificar que o sistema resultante, quando se iguala as Eqs. (A.9) e (A.10) para obter as constantes  $d_r$  nos  $k$  estênceis, possui  $2k - 1$  equações sendo que  $k - 1$  equações serão repetições das outras (veja sistema formado na Eq. (A.11)). Sempre ocorrerá um sistema compatível determinado e a combinação sempre será convexa.

Para o caso suave deseja-se que

$$w_r = d_r + O(\Delta x^{k-1}) \quad \text{para} \quad r = 0, 1, \dots, k-1. \quad (\text{A.13})$$

O que implicará uma ordem de precisão igual a  $(2k - 1)$

$$v_{j+\frac{1}{2}} = \sum_{i=0}^{k-1} w_r v_{j+\frac{1}{2}}^{(r)} = v\left(x_{j+\frac{1}{2}}\right) + O(\Delta x^{2k-1}) \quad \text{para} \quad r = 0, 1, \dots, k-1, \quad (\text{A.14})$$

pois

$$\begin{aligned}
\sum_{i=0}^{k-1} w_r v_{j+\frac{1}{2}}^{(r)} - \sum_{i=0}^{k-1} d_r v_{j+\frac{1}{2}}^{(r)} &= \sum_{i=0}^{k-1} w_r v_{j+\frac{1}{2}}^{(r)} - \sum_{i=0}^{k-1} d_r v_{j+\frac{1}{2}}^{(r)} + \sum_{i=0}^{k-1} (w_r - d_r) v \left( x_{j+\frac{1}{2}} \right) \\
&= \sum_{i=0}^{k-1} w_r \left( v_{j+\frac{1}{2}}^{(r)} - v \left( x_{j+\frac{1}{2}} \right) \right) + \sum_{i=0}^{k-1} d_r \left( -v_{j+\frac{1}{2}}^{(r)} + v \left( x_{j+\frac{1}{2}} \right) \right) \\
&= \sum_{i=0}^{k-1} \left( v_{j+\frac{1}{2}}^{(r)} - v \left( x_{j+\frac{1}{2}} \right) \right) (w_r - d_r) \\
&= \sum_{i=0}^{k-1} O(\Delta x^k) \cdot O(\Delta x^{k-1}) \\
&= O(\Delta x^{2k-1}).
\end{aligned}$$

Na primeira igualdade usou-se a Eq. (A.3) e a Eq. (A.13), pois

$$\sum_{i=0}^{k-1} (d_r - w_r) v \left( x_{j+\frac{1}{2}} \right) = v \left( x_{j+\frac{1}{2}} \right) \left( \sum_{i=0}^{k-1} d_r - \sum_{i=0}^{k-1} w_r \right) = 0. \quad (\text{A.15})$$

Na quarta igualdade tem-se que  $v_{j+\frac{1}{2}}^{(r)}$  aproxima  $v \left( x_{j+\frac{1}{2}} \right)$  com a ordem de precisão  $O(\Delta x^k)$ . Isso mostra que a aproximação para  $v_{j+\frac{1}{2}}$  está bem definida.

Quando a função  $v(x)$  possui descontinuidade(s) dentro do estêncil Eq. (A.1), pode-se escolher como peso(s) correspondente(s)  $w_r$  sendo zero. Outra consideração é que os pesos devem ser funções suaves dentro do estêncil envolvido. De acordo com Jiang e Shu (1996) [15], os pesos dados pela Eq. (A.16) são de classe  $C^\infty$  e eficientes computacionalmente.

Levando em conta essas considerações chega-se aos pesos da forma:

$$w_r = \frac{\alpha_r}{\sum_{s=0}^{k-1} \alpha_s}, \quad r = 0, 1, \dots, k-1, \quad (\text{A.16})$$

com

$$\alpha_r = \frac{d_r}{(\epsilon + \beta_r)^2}. \quad (\text{A.17})$$

O termo  $\epsilon > 0$  foi introduzido para evitar que o denominador de Eq. (A.17) seja zero. E em Jiang e Shu (1996) [15], foi utilizado em todos os testes  $\epsilon = 10^{-6}$ . Os coeficientes  $\beta_r$  são chamados de indicadores de suavidade do estêncil  $S_r$ .

No primeiro trabalho onde o esquema WENO foi construído (veja Liu; Osher e Chan (1994) [22]), os coeficientes  $\beta_r$  foram determinados utilizando diferenças divididas, de acordo com a filosofia do ENO, considerando a Eq. (3.53) e o teorema 3.2. Em Jiang e Shu (1996) [15], após intenso experimentos, os coeficientes  $\beta_r$  são escolhidos de forma diferente, obtendo melhorias

no método.

Seja  $p_r(x)$  o polinômio de reconstrução do estêncil  $S_r(j)$ ; define-se

$$\beta_r = \sum_{l=0}^{k-1} \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} \Delta x^{2l-1} \left( \frac{\partial p_r(x)}{\partial^l x} \right)^2 dx. \quad (\text{A.18})$$

O lado direito da Eq. (A.18) é a soma dos quadrados da norma  $L_2$  para todas as derivadas do polinômio interpolador  $p_r(x)$  sobre o intervalo  $(x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}})$ . O fator  $\Delta x^{2l-1}$  foi introduzido para remover toda dependência de  $\Delta x$  nas derivadas de  $p_r(x)$ .

Os coeficientes  $\beta_r$  serão calculados para  $k = 2, 3$ , de acordo com Jiang e Shu (1996) [15].

Para  $k = 2$  tem-se

$$\beta_0 = \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} \Delta x^{-1} (p_0(x))^2 dx + \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} \Delta x^1 \left( \frac{\partial p_0(x)}{\partial x} \right)^2 dx \quad (\text{A.19})$$

e

$$\beta_1 = \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} \Delta x^{-1} (p_1(x))^2 dx + \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} \Delta x^1 \left( \frac{\partial p_1(x)}{\partial x} \right)^2 dx, \quad (\text{A.20})$$

sendo  $p_0$  o polinômio de Lagrange, dado na Eq. (3.42), que interpola  $v(x)$  no estêncil  $\bar{S}_0(j) = \{x_j, x_{j+1}\}$  e  $p_1$  é o polinômio de Lagrange, que interpola  $v(x)$  no estêncil  $\bar{S}_1(j) = \{x_{j-1}, x_j\}$ .

Para  $k = 2$ , com a grelha uniforme e de acordo com a Eq. (3.43), o polinômio  $p(x)$  é dado por

$$p(x) = \sum_{i=0}^1 \left( \sum_{m=j+1}^2 \frac{\sum_{l=0}^2 \prod_{\substack{q=0 \\ q \neq l, m}}^2 (x - x_{j-r+q-\frac{1}{2}})}{\prod_{l=0}^2 (x_{j-r+m-\frac{1}{2}} - x_{j-r+l-\frac{1}{2}})} \right), \quad (\text{A.21})$$

que é equivalente à

$$p(x) = \left[ \frac{(x - x_{j-r+\frac{3}{2}}) + (x - x_{j-r-\frac{1}{2}})}{(x_{j-r+\frac{1}{2}} - x_{j-r-\frac{1}{2}})(x_{j-r+\frac{1}{2}} - x_{j-r+\frac{3}{2}})} + \frac{(x - x_{j-r+\frac{1}{2}}) + (x - x_{j-r-\frac{1}{2}})}{(x_{j-r+\frac{3}{2}} - x_{j-r-\frac{1}{2}})(x_{j-r+\frac{3}{2}} - x_{j-r+\frac{1}{2}})} \right] \Delta x \bar{v}_{j-r} + \left[ \frac{x - x_{j-r+\frac{1}{2}} + x - x_{j-r-\frac{1}{2}}}{(x_{j-r+\frac{3}{2}} - x_{j-r-\frac{1}{2}})(x_{j-r+\frac{3}{2}} - x_{j-r+\frac{1}{2}})} \right] \Delta x \bar{v}_{j-r+1}. \quad (\text{A.22})$$

Derivando a Eq. (A.22), obtém-se

$$p'(x) = \left( \frac{2}{-\Delta x^2} + \frac{2}{2\Delta x^2} \right) \Delta x \bar{v}_{j-r} + \left( \frac{2}{2\Delta x^2} \right) \Delta x \bar{v}_{j-r+1} = \frac{1}{\Delta x} (\bar{v}_{j-r+1} - \bar{v}_{j-r}) \quad (\text{A.23})$$

e para  $r = 0$  e  $r = 1$  tem-se, respectivamente

$$\begin{aligned} p'_0(x) &= \frac{1}{\Delta x} (\bar{v}_{j+1} - \bar{v}_j) & \text{e} \\ p'_1(x) &= \frac{1}{\Delta x} (\bar{v}_j - \bar{v}_{j-1}). \end{aligned} \quad (\text{A.24})$$

Substituindo as Eqs. (A.24) nas Eqs. (A.20), tem-se

$$\begin{aligned} \beta_0 &= \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} \Delta x \left( \frac{1}{\Delta x} (\bar{v}_{j+1} - \bar{v}_j) \right)^2 dx & \text{e} \\ \beta_1 &= \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} \Delta x \left( \frac{1}{\Delta x} (\bar{v}_j - \bar{v}_{j-1}) \right)^2 dx. \end{aligned} \quad (\text{A.25})$$

Como  $(\bar{v}_{j+1} - \bar{v}_j)$  e  $(\bar{v}_j - \bar{v}_{j-1})$  são constantes, calculando as integrais em Eq. (A.25) e eliminando os termos possíveis em  $\Delta x$ , tem-se as constantes  $\beta_0, \beta_1$  para  $k = 2$  dadas por

$$\begin{aligned} \beta_0 &= (\bar{v}_{j+1} - \bar{v}_j)^2 & \text{e} \\ \beta_1 &= (\bar{v}_j - \bar{v}_{j-1})^2 \end{aligned} \quad (\text{A.26})$$

Analogamente tem-se para  $k = 3$ . Como os cálculos da derivada são um pouco mais complicados, será feito aqui um esboço dos cálculos. De acordo com Eq. (3.43), o polinômio de Lagrange para  $k = 3$  em uma grelha uniforme é dado por:

$$\begin{aligned} p(x) &= \left[ \left( x - x_{j-r+\frac{3}{2}} \right) \left( x - x_{j-r+\frac{5}{2}} \right) + \left( x - x_{j-r-\frac{1}{2}} \right) \left( x - x_{j-r+\frac{5}{2}} \right) + \left( x - x_{j-r-\frac{1}{2}} \right) \left( x - x_{j-r+\frac{3}{2}} \right) \right] \\ &\quad \frac{\bar{v}_{j-r}}{2\Delta x^2} \\ &+ \left[ \left( x - x_{j-r+\frac{1}{2}} \right) \left( x - x_{j-r+\frac{5}{2}} \right) + \left( x - x_{j-r-\frac{1}{2}} \right) \left( x - x_{j-r+\frac{5}{2}} \right) + \left( x - x_{j-r-\frac{1}{2}} \right) \left( x - x_{j-r+\frac{1}{2}} \right) \right] \\ &\quad \frac{\bar{v}_{j-r} + \bar{v}_{j-r+1}}{-2\Delta x^2} \\ &+ \left[ \left( x - x_{j-r+\frac{1}{2}} \right) \left( x - x_{j-r+\frac{3}{2}} \right) + \left( x - x_{j-r-\frac{1}{2}} \right) \left( x - x_{j-r+\frac{3}{2}} \right) + \left( x - x_{j-r-\frac{1}{2}} \right) \left( x - x_{j-r+\frac{1}{2}} \right) \right] \\ &\quad \frac{\bar{v}_{j-r} + \bar{v}_{j-r+1} + \bar{v}_{j-r+2}}{6\Delta x^2}. \end{aligned}$$

a derivada de  $p(x)$  é dada por

$$\begin{aligned} p'(x) &= \left[ \left( x - x_{j-r+\frac{5}{2}} \right) + \left( x - x_{j-r+\frac{3}{2}} \right) + \left( x - x_{j-r-\frac{1}{2}} \right) \right] \frac{\bar{v}_{j-r}}{\Delta x^2} + \\ &\quad \left[ \left( x - x_{j-r+\frac{5}{2}} \right) + \left( x - x_{j-r+\frac{1}{2}} \right) + \left( x - x_{j-r-\frac{1}{2}} \right) \right] \frac{\bar{v}_{j-r} + \bar{v}_{j-r+1}}{-\Delta x^2} + \\ &\quad \left[ \left( x - x_{j-r+\frac{3}{2}} \right) + \left( x - x_{j-r+\frac{1}{2}} \right) + \left( x - x_{j-r-\frac{1}{2}} \right) \right] \frac{\bar{v}_{j-r} + \bar{v}_{j-r+1} + \bar{v}_{j-r+2}}{3\Delta x^2} \end{aligned} \quad (\text{A.27})$$

e fazendo uma mudança de variável

$$x = x_{j-r-\frac{1}{2}} + y\Delta x \quad (\text{A.28})$$

na Eq. (A.27), obtém-se

$$\begin{aligned}
 p' \left( x_{j-r-\frac{1}{2}} + y\Delta x \right) &= [(-3\Delta x + y\Delta x) + (2\Delta x + y\Delta x) + (y\Delta x)] \frac{\bar{v}_{j-r}}{\Delta x^2} \\
 &\quad - [(-3\Delta x + y\Delta x) + (-\Delta x + y\Delta x) + (y\Delta x)] \frac{\bar{v}_{j-r} + \bar{v}_{j-r+1}}{-\Delta x^2} + \\
 &\quad \left[ (-2\Delta x + y\Delta x) + (-\Delta x + y\Delta x) + (y\Delta x) \frac{\bar{v}_{j-r} + \bar{v}_{j-r+1}}{-\Delta x^2} \right] \frac{\bar{v}_{j-r} + \bar{v}_{j-r+1} + \bar{v}_{j-r+2}}{3\Delta x^2}.
 \end{aligned} \tag{A.29}$$

Reordenando os termos, tem-se

$$p' \left( x_{j-r-\frac{1}{2}} + y\Delta x \right) = \frac{1}{\Delta x} [y(\bar{v}_{j-r} - 2\bar{v}_{j-r+1} + \bar{v}_{j-r+2}) + (-2\bar{v}_{j-r} + 3\bar{v}_{j-r+1} - \bar{v}_{j-r+2})] \tag{A.30}$$

de onde, para  $r = 0, 1, 2$ , tem-se respectivamente

$$\begin{aligned}
 p'_0 \left( x - x_{j-\frac{1}{2}} + y\Delta x \right) &= \frac{1}{\Delta x} [y(\bar{v}_j - 2\bar{v}_{j+1} + \bar{v}_{j+2}) + (-2\bar{v}_j + 3\bar{v}_{j+1} - \bar{v}_{j+2})] \\
 p'_1 \left( x - x_{j-\frac{3}{2}} + y\Delta x \right) &= \frac{1}{\Delta x} [y(\bar{v}_{j-1} - 2\bar{v}_j + \bar{v}_{j+1}) + (-2\bar{v}_{j-1} + 3\bar{v}_j - \bar{v}_{j+1})] \\
 p'_2 \left( x - x_{j-\frac{5}{2}} + y\Delta x \right) &= \frac{1}{\Delta x} [y(\bar{v}_{j-2} - 2\bar{v}_{j-1} + \bar{v}_j) + (-2\bar{v}_{j-2} + 3\bar{v}_{j-1} - \bar{v}_j)].
 \end{aligned} \tag{A.31}$$

Derivando a Eq. (A.30), obtém-se a derivada segunda do polinômio  $p(x)$

$$p'' \left( x_{j-r-\frac{1}{2}} + y\Delta x \right) = \frac{1}{\Delta x^2} [(\bar{v}_{j-r} - 2\bar{v}_{j-r+1} + \bar{v}_{j-r+2})], \tag{A.32}$$

de onde, para  $r = 0, 1, 2$  tem-se, respectivamente

$$\begin{aligned}
 p''_0 \left( x_{j-\frac{1}{2}} + y\Delta x \right) &= \frac{1}{\Delta x^2} [(\bar{v}_j - 2\bar{v}_{j+1} + \bar{v}_{j+2})], \\
 p''_1 \left( x_{j-\frac{3}{2}} + y\Delta x \right) &= \frac{1}{\Delta x^2} [(\bar{v}_{j-1} - 2\bar{v}_j + \bar{v}_{j+1})] \quad \text{e} \\
 p''_2 \left( x_{j-\frac{5}{2}} + y\Delta x \right) &= \frac{1}{\Delta x^2} [(\bar{v}_{j-2} - 2\bar{v}_{j-1} + \bar{v}_j)].
 \end{aligned} \tag{A.33}$$

Para  $k = 3$ ,  $r = 0$  a Eq. (A.18) se reduz a

$$\beta_0 = \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} \Delta x (p'_0(x))^2 dx + \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} \Delta x^3 (p''_0(x))^2 dx.$$

Considerando a troca de variável Eq. (A.28) e substituindo  $p'_0$  e  $p''_0$  na equação acima obtém-se

$$\beta_0 = \int_0^1 \Delta x^2 \left( \frac{yR + L}{\Delta x} \right)^2 dy + \int_0^1 \Delta x^4 (\bar{v}_j - 2\bar{v}_{j+1} + \bar{v}_{j+2})^2 dy, \tag{A.34}$$

sendo  $R = (\bar{v}_j - 2\bar{v}_{j+1} + \bar{v}_{j+2})$  e  $L = (-2\bar{v}_j + 3\bar{v}_{j+1} - \bar{v}_{j+2})$ . Calculando as integrais na Eq. A.34, tem-se

$$\beta_0 = \frac{R^2}{3} + RL + L^2 + (\bar{v}_j - 2\bar{v}_{j+1} + \bar{v}_{j+2})^2.$$

Substituindo as variáveis  $R, L$ , obtém-se

$$\beta_0 = \frac{4}{3} (\bar{v}_j - 2\bar{v}_{j+1} + \bar{v}_{j+2})^2 + (\bar{v}_j - 2\bar{v}_{j+1} + \bar{v}_{j+2}) (-2\bar{v}_j + 3\bar{v}_{j+1} - \bar{v}_{j+2}) + (-2\bar{v}_j + 3\bar{v}_{j+1} - \bar{v}_{j+2})^2$$

e

$$\beta_0 = \frac{13}{12} (\bar{v}_j - 2\bar{v}_{j+1} + \bar{v}_{j+2})^2 + \frac{3}{12} (\bar{v}_j - 2\bar{v}_{j+1} + \bar{v}_{j+2})^2 + (2\bar{v}_j^2 + 3\bar{v}_{j+1}^2 - 5\bar{v}_j\bar{v}_{j+1} + \bar{v}_j\bar{v}_{j+2} - \bar{v}_{j+1}\bar{v}_{j+2}).$$

Desenvolvendo o quadrado da segunda parcela e tirando mínimo múltiplo comum com a terceira, parcela concluímos que

$$\beta_0 = \frac{13}{12} (\bar{v}_j - 2\bar{v}_{j+1} + \bar{v}_{j+2})^2 + \frac{1}{4} (3\bar{v}_j - 4\bar{v}_{j+1} + \bar{v}_{j+2})^2. \quad (\text{A.35})$$

Agora, realizando cálculos análogos aos descritos para  $\beta_0$ , obtém-se os valores de  $\beta_1$  e  $\beta_2$  são dados por

$$\begin{aligned} \beta_1 &= \frac{13}{12} (\bar{v}_{j-1} - 2\bar{v}_j + \bar{v}_{j+1})^2 + \frac{1}{4} (\bar{v}_{j-1} - \bar{v}_{j+1})^2 \\ \beta_2 &= \frac{13}{12} (\bar{v}_{j-2} - 2\bar{v}_{j-1} + \bar{v}_j)^2 + \frac{1}{4} (\bar{v}_{j-2} - 4\bar{v}_{j-1} + 3\bar{v}_j)^2. \end{aligned} \quad (\text{A.36})$$

Segue a verificação da ordem de precisão dessas constantes. Se o estêncil é suave, pela série de Taylor, tem-se

$$\bar{v}_{j-2} = \bar{v}(x_j - 2\Delta x) = \bar{v}_j - 2\Delta x \bar{v}'_j + \frac{4\Delta x^2}{2} \bar{v}''_j - \frac{8\Delta x^3}{6} \bar{v}'''_j,$$

$$\bar{v}_{j-1} = \bar{v}(x_j - \Delta x) = \bar{v}_j - \Delta x \bar{v}'_j + \frac{\Delta x^2}{2} \bar{v}''_j - \frac{\Delta x^3}{6} \bar{v}'''_j,$$

$$\bar{v}_{j+1} = \bar{v}(x_j + \Delta x) = \bar{v}_j + \Delta x \bar{v}'_j + \frac{\Delta x^2}{2} \bar{v}''_j + \frac{\Delta x^3}{6} \bar{v}'''_j,$$

$$\bar{v}_{j+2} = \bar{v}(x_j + 2\Delta x) = \bar{v}_j + 2\Delta x \bar{v}'_j + \frac{4\Delta x^2}{2} \bar{v}''_j + \frac{8\Delta x^3}{6} \bar{v}'''_j.$$

Substituindo o desenvolvimento de Taylor nas Eqs. (A.35) - (A.36), tem-se

$$\beta_0 = (\Delta x \bar{v}'_j)^2 + \left[ 1 + \Delta x^2 \left( \frac{-2}{3} \frac{\bar{v}'''_j}{\bar{v}'_j} + \frac{1}{9} \Delta x^2 \frac{(\bar{v}'''_j)^2}{\bar{v}'_j} + \frac{13}{12} \frac{\bar{v}''_j}{\bar{v}'_j} + \frac{13}{6} \Delta x \frac{\bar{v}'''_j \bar{v}''_j}{\bar{v}'_j} + \frac{13}{12} \Delta x^2 \frac{\bar{v}'''_j}{\bar{v}'_j} \right) \right],$$

$$\beta_1 = (\Delta x \bar{v}'_j)^2 + \left[ 1 + \Delta x^2 \left( \frac{1}{3} \frac{\bar{v}'''}{\bar{v}'_j} + \frac{1}{36} \Delta x^2 \left( \frac{\bar{v}'''}{\bar{v}'_j} \right)^2 + \frac{13}{12} \left( \frac{\bar{v}'''}{\bar{v}'_j} \right)^2 \right) \right] \quad \text{e}$$

$$\beta_2 = (\Delta x \bar{v}'_j)^2 + \left[ 1 + \Delta x^2 \left( \frac{-2}{3} \frac{\bar{v}'''}{(\bar{v}'_j)^2} + \frac{2}{9} \Delta x^2 \left( \frac{\bar{v}'''}{\bar{v}'_j} \right)^2 + \frac{13}{12} \left( \frac{\bar{v}''}{\bar{v}'_j} \right)^2 - 2 \Delta x \frac{\bar{v}''' \bar{v}''}{\bar{v}'_j} + \Delta x^2 \left( \frac{\bar{v}'''}{\bar{v}'_j} \right)^2 \right) \right].$$

ou seja

$$\beta_r = (\Delta x \bar{v}'_j)^2 + (1 + O(\Delta x^2)) \quad \text{para} \quad r = 0, 1, 2. \quad (\text{A.37})$$

Note que as constantes  $\beta_r$  em Eq. (A.35) - Eq. (A.36) foram encontradas considerando a direção 'upwind', a movimentação do fluxo da esquerda para a direita. Se a direção do fluxo se inverter, o procedimento para obter as constantes pode ser modificado simetricamente usando  $x_{j+\frac{1}{2}}$  na mudança de variável Eq. (A.28).

De acordo com Shu (1997) [29], a ordem de precisão das constantes  $\beta_r$  é dada por:

- Se a função  $v(x)$  é suave no estêncil  $S_r(j)$ , tem-se:  $\beta_r = O(\Delta x^2)$ .
- Se a função  $v(x)$  possui uma descontinuidade no estêncil  $S_r(j)$ , tem-se:  $\beta_r = O(1)$ .

Utilizando a Eq. (A.16) e a Eq. (A.17), tem-se:

- Se a função  $v(x)$  é suave no estêncil  $S_r(j)$ , tem-se:  $w_r = O(1)$ .
- Se a função  $v(x)$  possui uma descontinuidade no estêncil  $S_r(j)$ , tem-se:  $\beta_r = O(\Delta x^2)$ .

### Roteiro para o código WENO

Dado as aproximações médias  $\bar{v}_j$  da função  $v(x)$ , para cada célula  $I_j$ , obtém-se a aproximação 'upwind' para a função  $v(x)$  nas fronteiras das células  $v_{j-\frac{1}{2}}^d$  e  $v_{j+\frac{1}{2}}^e$  da seguinte maneira:

1. Obtenha os  $k$  valores de reconstrução  $v_{j+\frac{1}{2}}^{(r)}$  e  $v_{j-\frac{1}{2}}^{(r)}$  de acordo com a Eq. (A.2), usando Eq. (3.30), para  $r = 0, 1, \dots, k-1$ .
2. Obtidas as constantes  $d_r$ , tabela A.1, obter as constantes  $\beta_r$  de acordo com Eq. (A.18), para  $r = 0, 1, \dots, k-1$ . Para  $k = 2, 3$  as constantes são dadas pelas Eq. (A.27), Eq. (A.35) - Eq. (A.36).
3. Obter as constantes  $w_r$ , de acordo com Eq. (A.16) e Eq. (A.17).
4. Encontre a aproximação nas fronteiras das células, de acordo com a Eq. (A.4).

# Apêndice B

## Esquemas ENO e WENO via Volume Finito para Leis de Conservação Escalares

Para o método dos volumes finitos, ou esquemas baseados na aproximação dos valores médios das células, não se resolve a lei de conservação na sua forma diferencial como na Eq. (1.12) mas sim sua versão na forma integral

$$\frac{d}{dt} \int_{I_j} u(x, t) dx = - \left[ f \left( u \left( x_{j+\frac{1}{2}}, t \right) \right) - f \left( u \left( x_{j-\frac{1}{2}}, t \right) \right) \right],$$

sobre o intervalo  $I_j = \left[ x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}} \right]$ . Fazendo

$$\bar{u}(x_j, t) \equiv \frac{1}{\Delta x} \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} u(\xi, t) d\xi, \quad (\text{B.1})$$

obtemos

$$\frac{d}{dt} [\bar{u}(x_j, t) \Delta x_j] = - \left[ f \left( u \left( x_{j+\frac{1}{2}}, t \right) \right) - f \left( u \left( x_{j-\frac{1}{2}}, t \right) \right) \right].$$

Assim

$$\frac{d}{dt} \bar{u}_j(x, t) = - \frac{1}{\Delta x_j} \left[ F_{j+\frac{1}{2}} - F_{j-\frac{1}{2}} \right] \quad (\text{B.2})$$

onde  $F_{j+\frac{1}{2}}, F_{j-\frac{1}{2}}$  são as funções de fluxo numéricas as quais aproximam, respectivamente,  $f \left( u \left( x_{j+\frac{1}{2}}, t \right) \right), f \left( u \left( x_{j-\frac{1}{2}}, t \right) \right)$ .

A função de fluxo numérica  $F_{j+\frac{1}{2}}$  é definida por

$$\begin{aligned} F_{j+\frac{1}{2}} &= h \left( u_{j+\frac{1}{2}}^e, u_{j+\frac{1}{2}}^d \right) & e \\ F_{j-\frac{1}{2}} &= h \left( u_{j-\frac{1}{2}}^e, u_{j-\frac{1}{2}}^d \right). \end{aligned} \quad (\text{B.3})$$

Para que o método obedeça ao teorema 2.5, se deve obter uma função de fluxo monótona.

**Definição B.1** Uma função de fluxo  $h$ , de dois argumentos  $h(y, z)$ , é dita um *fluxo monótono* se:

- se  $h(y, z)$  é Lipschitz contínua em ambos os argumentos;
- se  $h(y, z)$  é uma função não-decrescente em  $y$  e não-crescente em  $z$ ;
- se  $h(y, z)$  é consistente com a função fluxo  $f$ , ou seja,  $h(y, y) = f(y)$ .

De acordo com o exemplo 2.1, visto na seção 2.3.2, tem-se que o método de Lax-Friedrichs é monótono. Além disso a função de fluxo numérica de Lax-Friedrichs também é uma função de fluxo monótono. Fazendo na Eq. (2.77)  $y = U_j$  e  $z = U_{j+1}$  obtém-se

$$h(y, z) = \frac{1}{2} \left[ f(y) + f(z) - \frac{\Delta x}{\Delta t} (z - y) \right]. \quad (\text{B.4})$$

Novamente, tem-se que  $\alpha = \max |f'(u)|$  e a Eq. (B.4) se reduz à

$$h(y, z) = \frac{1}{2} [f(y) + f(z) - \alpha (z - y)]. \quad (\text{B.5})$$

É importante lembrar que o esquema conservativo Eq. (B.2), segundo o teorema de Lax-Wendroff, se convergir deverá ser para uma solução fraca do problema.

### Procedimento para o esquema de volume finito no caso escalar unidimensional

1. Obtenha as aproximações para os valores

$$\begin{aligned} v_{j+\frac{1}{2}}^e &= p_j \left( x_{j+\frac{1}{2}} \right) & \text{e} & & v_{j+\frac{1}{2}}^d &= p_{j+1} \left( x_{j+\frac{1}{2}} \right), \\ v_{j-\frac{1}{2}}^d &= p_j \left( x_{j-\frac{1}{2}} \right) & \text{e} & & v_{j-\frac{1}{2}}^e &= p_{j-1} \left( x_{j-\frac{1}{2}} \right). \end{aligned} \quad (\text{B.6})$$

usando o esquema ENO ou o esquema WENO.

2. Calcule o fluxo monótono, Eq. (B.5).
3. Forme o esquema, Eq. (B.2).

# Referências Bibliográficas

- [1] BUTCHER, J. C., *The numerical analysis of ordinary differential equations; Runge-Kutta and general linear methods*. John Wiley, 1987.
- [2] CHENG, J. E SHU, C.W., *High Order Schemes for CFD: A Review*.
- [3] COLELLA, P. E WOODWARD, P., *The Piecewise Parabolic Method (PPM) for Gas-Dynamical Simulations*. Journal of Computational Physics, Vol. 54 (1984), pp.174-201.
- [4] CONWAY, J. B., *A course in functional analysis*. Springer, 1990 2ª edição.
- [5] COURANT, R.; FRIEDRICHS, K. O. E LEWY, H., *Über die partiellen Differenzengleichungen der mathematischen Physik*. Math. Ann, Vol. 100 (1928), pp. 32-74.
- [6] COURANT, R.; FRIEDRICHS, K. O. E LEWY, H., *On the partial difference equations of mathematical physics*. IBM Journal, Vol. 11 (1967), pp. 215-234.
- [7] DE AZEVEDO, A. V. F. E ESCHENAZI, C. S., *Leis de conservação com aplicações ao tráfego nas cidades*. II Bienal da Sociedade Brasileira de Matemática, Salvador 2004.
- [8] DiPERNA, R. J., *Finite difference schemes for conservation laws*. Communications on Pure Applied Mathematics, Vol. 25 (1982), pp. 379-450.
- [9] EVANS, L. C., *Partial Differential Equations*. Graduate Studies in mathematics, Vol. 19, American Mathematical Society 1998.
- [10] GOTTLIEB, S. E SHU, C.-W., *Total variation diminishing Runge-Kutta schemes*. Mathematics of Computation, Vol. 67 (1998), N° 221, pp. 73-85.
- [11] GRANDALL, M. G. E MAJDA, A., *Monotone difference approximations for scalar conservation laws*. Mathematics of Computation, Vol. 34 (1980), pp. 1-21.
- [12] HARTEN, A.; HYMAN, J. M. E LAX, P. D., *On Finite-difference approximation an entropy conditions for shocks*. Communications on Pure Applied Mathematics, Vol. 29 (1976), pp. 297-322 (com apêndice por Barbara Keyfitz).
- [13] HARTEN, A. E OSHER, S., *Uniformly high-order accurate non-oscillatory schemes I*. SIAM Journal on Numerical Analysis, Vol. 24 (1987), pp.279-309.

- [14] HARTEN, A.; OSHER, S.; ENGQUIST, B. E CHAKRAVARTHY, S., *Uniformly high-order essentially non-oscillatory schemes III*. Journal of Computational Physics, Vol. 71 (1987), pp.231-303.
- [15] JIANG, G. E SHU, C.W., *Efficient implementation of weighted ENO schemes*. Journal of Computational Physics, Vol. 126 (1996), pp.202-228.
- [16] KRYLOV, V. I., *Approximate calculation of integrals*. Dover Publications, INC., 2005.
- [17] LEVEQUE, R. J., *Numerical Methods for Conservation Laws*. Lectures in Mathematics, ETH Zurich, Segunda edição 1992.
- [18] LEVEQUE, R. J., *Finite-Volume Methods for Hiperbolic Problems*. Cambridge Texts in Applied Mathematics, UNIVERSITY OF CAMBRIDGE, 2004.
- [19] LEVEQUE, R. J. E TEMPLE, B., *Stability of Godunov's method for a class of  $2 \times 2$  systems of conservation laws*. Trans. Amer. Math. Soc., Vol. 288 (1985), pp. 115-123.
- [20] LIMA, E. L., *Curso de Análise-Volume 1*. Projeto Euclides, IMPA, Rio de Janeiro, Décima Edição 2000.
- [21] LIMA, E. L., *Curso de Análise-Volume 2*. Projeto Euclides, IMPA, Rio de Janeiro, Nona Edição 2006.
- [22] LIU, X.D; OSHER, S. E CHAN, T., *Weighted essentially non-oscillatory schemes*. Journal of Computational Physics, Vol. 115 (1994), pp. 200-212.
- [23] LIU, T. P., *The deterministic version of the Glimm scheme*. Comm. Math. Phys., Vol. 57 (1977), pp. 135-148.
- [24] LOGAN, J. D., *An Introduction to Nonlinear Partial Differential Equations*. Pure and Applied Mathematics, Wiley-Interscience series of Texts, Monographs and Tracts, 1994.
- [25] MAROTTI, L. P., *Aspectos teóricos e numéricos das leis de conservação: um estudo visando a equação de Buckley-Leverett*. UFRJ-Instituto Politécnico, Graduação em Engenharia Mecânica, Nova Friburgo, 2008.
- [26] REZZOLLA, L., *Numerical Methods for the solution of Hyperbolic Partial Differential Equations*. Lectures Notes, SISSA, International School for advanced studies, Abril 2004. [www.sissa.it/rezzolla](http://www.sissa.it/rezzolla)
- [27] RUGGIERO, M. A. G. E LOPES, V. L. R., *Cálculo Numérico: aspectos teóricos e computacionais*. MAKRON., São Paulo. Segunda edição. 1997.
- [28] SMOLLER, J., *Shock Waves and Reaction-Diffusion Equation*. Springer-Verlag, Segunda edição 1994.

- [29] SHU, C.W., *Essentially Non-Oscillatory and Weighted Essentially Non-Oscillatory Schemes for Hyperbolic Conservation Laws*. ICASE Report N° 97-65, Novembre 1997.
- [30] SHU, C.W. E OSHER, S., *Efficient implementaton of essentially non-oscillatory shock capturing schemes*. Journal of Computational Physics, Vol. 77 (1988), pp.439-471.
- [31] SURESH, A. E HUYNH, H. T., *Accuracy monotonicity preserving schemes with Runge-Kutta time stepping*. Journal of Computational Physics, Vol. 136 (1997), pp.83-99.
- [32] THOE, D. W. E ZACHMANOGLU, E. C., *Introduction to Partial Differential Equation with aplications*. Dover, 1986.

# Índice Remissivo

- Aproximação média, 30
- Choque, 14
- Condição
  - CFL, 31
  - de Entropia, 21
  - de fronteira, 78
  - de Rankine-Hugoniot, 20
- Conjunto compacto, 47
- Constantes  $c_{ri}$ , 70
- Curvas Características, 11
- Deficiências do Interpolador ENO, 97
- Definição
  - Consistência, 36
  - Convergência, 35
  - Ordem do método numérico, 36
- Domínio de Dependência
  - EDP, 32
  - Numérico, 32
- Equação
  - de Advecção, 7
  - de Burgers, 8
  - lei de conservação, 5
- Erro
  - de truncamento local, 35
  - pontual, 34
- Espaço
  - $L_{1,T}$ , 48
  - $L_1$ , 48
- Estêncil, 32
- Estabilidade, 36
- Estabilidade não-linear, 46
- Fluxo ‘splitting’, 59
- Formação de Choque, 23
- Função de fluxo monótona, 108
- Lax-Friedrichs
  - equação não-linear, 55
- Lei de conservação
  - Forma diferencial conservativa, 7
  - Forma integral, 7
  - Forma não-conservativa, 7
- Leque de rarefação, 26
- Lipschitz, 46
  - contínua, 46
- Método Numérico
  - Semi-discreto, 58
  - Conservativo, 42
  - equação não-linear, 46
  - Lax-Friedrichs equação advecção, 36
  - Upwind equação advecção, 40
- Método numérico
  - monótonos, 55
  - preservam monotonicidade, 55
  - TVD, 54
- Número CFL, 34
- Normas
  - $L_1$ , 35
  - $L_1$ , dados discretos, 35
  - $L_2$ , 35
  - T, 47
- Polinômio
  - Lagrange, 68
  - Lagrange para  $v_{j+1/2}$ , 69
  - Newton, 73
- Problema

- de Cauchy, 8
- Problema de Riemann, 21
  - Equação de Advecção, 21
  - Equação de Burgers, 23
- Propriedades ENO, 78
- Roteiro
  - Diferenças finitas-ENO, 60
  - ENO, 77
- Runge-Kutta não-TVD 2ª ordem, 88
- Runge-Kutta TVD
  - 2ª ordem, 88
  - 3ª ordem, 85
  - equações dependentes do tempo, 87
- Solução analítica da equação de Advecção, 11
- Solução fraca, 15
- Teorema
  - convergência métodos lineares, 31
  - Convergência métodos não-lineares, 53
  - Lax-Wendroff, 52
  - solução de entropia, 56
- TV-estável, 49
- Unicidade da condição de entropia, 21
- Variação Total, 48
- WENO
  - constantes  $w_r$ , 102