Universidad Autónoma del Estado de Hidalgo
Instituto de Ciencias Básicas e Ingeniería
Centro de Investigación en Matemáticas

Desarrollo de Métodos en Diferencias Finitas de Alta Resolución para Ecuaciones Hiperbólicas con Términos Difusivos

Tesis que para obtener el título de

Licenciado en Matemáticas Aplicadas

presenta

Adán Uribe Bravo

bajo la dirección de

Dra. Silvia Jerez Galiano y Dr. Roberto Ávila Pozos

Pachuca, Hidalgo. Febrero de 2010.

Resumen

En este trabajo se presentan dos métodos numéricos para resolver problemas de convección-difusión no lineal con término convectivo dominante.

El primer método consiste en la descomposición de operadores, el cuál desacopla la ecuación en subproblemas más sencillos. El segundo método extiende el esquema TVD Flux-Limiter a ecuaciones hiperbólicas no lineales con términos difusivos.

Abstract

In this work we present two numerical methods to solve non-linear convection-diffusion equation with dominant convective term.

The first method consist in the operators decomposition, which decouples the partial differential equation in simpler subproblems. In the second method a recently TVD Flux-Limiter scheme is extended for numerical resolution of non-linear partial differential equations of hyperbolic type with diffusive terms.

Dedicatoria

A mis padres, hermanos
y amigos con todo cariño.

Agradecimientos

A Dios, a quien amo y le agradezco su constante cuidado, demostración de fidelidad y toda la paciencia que me ha tenido.

Quiero agradecer profundamente a mis papás, Sara Bravo Vargas y Francisco Uribe Zuñiga, quienes son el pilar de mi vida. Por ser las principales personas que inspiran en mí el deseo de superación y éxito. Por el gran cariño, amor que me tienen y todo el apoyo moral que me brindan. Esto me fortalece y hace que me sienta seguro y capaz de alcanzar mis metas. Sinceramente muchas gracias por todo.

A mi hermana, Elsa Uribe Bravo. Por su amistad y su cariño. Por estar siempre dispuesta a escuchar mis tristezas, frustraciones y alegrías. Por toda la motivación y consejos que me dio. ¡Gracias hermana!

A mi hermano, Iván Uribe Bravo por haber proporcionado los recursos económicos para la realización de mis estudios, los cuales culminan con la elaboración de esta tesis.

Al CIMAT A.C, en el cual tuve un espacio físico y la beca otorgada para el desarrollo de esta tesis. Deseo agradecer especialmente a mi directora de tesis, la Dra. Silvia Jerez Galiano por todo el apoyo brindado, paciencia, atención y esmero que puso en el desarrollo de esta tesis y que gracias a ella fue posible este trabajo.

Un agradecimiento a mis sinodales, por sus amables asesorías y brillantes consejos que fueron fundamentales para la conclusión de la presente tesis.

También quiero que aquí figuren los nombres de mis amigos Yaneli, Luis Ángel, Yolita, Edgar, Hortencia, Fernando, Fanny, Miguel Ángel, Ángeles, Deisi, Cecilia, Rubi, Virgilio, Chava, Selomit y Lety. Porque son mi fuente de energía e inspiración para seguir siempre adelante y porque le dan sentido a mi existir, por brindarme su amor y tiempo incondicionalmente, pero sobre todo por su amistad.

A todos mis profesores por los conocimientos que me han proporcionado, por transmitirme el gusto hacia las matemáticas y por alentarme en continuar mis estudios.

En fin, a todos, nombrados y no nombrados, que me han acompañado estos 7 largos y duros años, quisiera decirles, con todo mi cariño, gracias y brindar con ellos:*Para que desde las noches de un pasado imperfecto, pasando por el amanecer de un presente simple, lleguemos a la brillante melodía de un futuro perfecto.*

Índice general

Resumen I

Dedicatoria II

Agradecimientos III

1. Introducción 1

2. Métodos en Diferencias Finitas 5
- 2.1. Aproximación de derivadas mediante diferencias finitas 6
- 2.2. Propiedades Cualitativas de los Métodos en Diferencias Finitas 9
- 2.3. Ecuación de Difusión 13
  - 2.3.1. Método Euler Explícito 14
  - 2.3.2. Método de Euler Implícito 20
  - 2.3.3. Método de Crank-Nicholson 21
- 2.4. Ecuación de Convección No Lineal 22
  - 2.4.1. Estabilidad y Convergencia No Lineal 23
  - 2.4.2. Upwind 26
  - 2.4.3. Richtmyer Two-Step Lax-Wendroff 27
  - 2.4.4. TVD Flux-Limiter 30

3. Solución Numérica de la Ecuación Convección-Difusión Mediante un Método Splitting Tipo Strang 33
- 3.1. Método Splitting 33
- 3.2. Ecuación de Convección Difusión 36
- 3.3. Splitting Strang 37

4. Solución Numérica de la Ecuación Convección-Difusión Mediante un Método Viscous Flux Limiter **41**
4.1. Viscosidad numérica 43
4.2. Error de truncamiento local y consistencia 43
4.3. TVD-estabilidad 47

5. Resultados Numéricos **51**

A. Conceptos Básicos de EDP´s **59**

B. Teorema de Cauchy-Kovaleskaya **65**

C. Teorema de Harten **69**

D. Teorema de Lie **71**

Bibliografía **75**

CAPÍTULO 1

Introducción

Motivación

Diversos problemas de interés científico, tecnológico y social en diferentes campos de la ingeniería y las ciencias aplicadas tienen como modelo matemático la ecuación de convección-difusión. La ecuación de convección-difusión es una ecuación de evolución para una función $u(x,t)$ que puede representar la concentración de una especie química, la velocidad de un fluido, el precio de una acción o la temperatura de un gas. Esta ecuación es un modelo simple pero no obstante muy interesante en el que aparecen simultáneamente difusión y convección no lineal.

La presencia del término de difusión tiene un efecto regularizante sobre las soluciones para $t > 0$, en muchos casos el término difusivo es más pequeño que el término convectivo, dando lugar a problemas de convección dominante. Incluso en algunas situaciones el término difusivo se vuelve degenerado, como ocurre en algunos modelos de valoración de productos financieros (ver [26]).

Este hecho ha desembocado en que los algoritmos conocidos presenten dificultades importantes con respecto a la solución ya que producen oscilaciones cuando la solución es discontinua o posee cambios rápidos, otros métodos previenen la aparición de estas falsas oscilaciones numéricas pero contienen una excesiva disipación numérica en todos los puntos del mallado e incluso donde la solución es suave. Por lo tanto, es necesario utilizar esquemas no oscilatorios para la correcta resolución de problemas con discontinuidades, choques o cambios rápidos.

Objetivo

Es por ello, que el objetivo de esta tesis, es obtener métodos numéricos para resolver de manera eficiente la ecuación de problemas de convección-difusión no lineal como:

$$u_t - \nu u_{xx} + f(u)_x = 0, \qquad x \in [a, b], 0 \leq t \leq T, \tag{1.1}$$

donde ν representa el coeficiente de difusión que tomaremos constante, $u : \mathbb{R} \times \mathbb{R} \to \mathbb{R}^m$ es una función representativa del fenómeno velocidad, temperatura, concentración, etc. y $f : \mathbb{R}^m \to \mathbb{R}^m$ se le conoce como el vector de flujos de u.

Antecedentes

Es un hecho bien conocido que para problemas hiperbólicos es más complicado desarrollar métodos numéricos precisos y estables que para problemas elípticos o parabólicos. Por este motivo, en los últimos años se ha llevado a cabo una intensa actividad investigadora en el desarrollo de métodos numéricos que proporcionen aproximaciones más precisas y estables del término convectivo en ecuaciones hiperbólicas (ver por ejemplo [8], [15], [22]), los cuales utilizaremos en la solución de la ecuación de convección no lineal.

En este trabajo proponemos dos estrategias para aproximar la solución de problemas con convección dominante y/o discontinuidades: método Splitting Strang y método Viscous TVD Flux-Limiter. Estos métodos presentan buenas propiedades, lo que hacen de ellos una herramienta eficaz para la resolución numérica de problemas de convección.

El método Splitting Strang para la ecuación de convección-difusión no lineal combina una descomposición de operadores en la discretización temporal y diferencias finitas en la discretización espacial. Algunas de las ideas básicas de la descomposición de operadores fueron desarrolladas por Marchuk [23]. Como otros métodos de descomposición de operadores, nos permite desacoplar la ecuación de convección-difusión, en un problema de difusión y un problemas de convección. La solución de los subproblemas anteriores se realiza mediante métodos iterativos que conducen a la solución de problemas lineales parabólicos: Euler Explícito, Euler Implícito o Crank-Nicholson y TVD Flux-Limiter para el problema de tipo advección no lineal.

Los métodos de descomposición de operadores son muy versátiles y pueden aplicarse en muy diversas situaciones: *resolución de sistemas algebraicos* [14], *las ecuaciones de Navier-Stokes* [32], *el sistema de la termoelasticidad* [32] y *El método de la descomposición de dominios* para EDP´s [4].

El segundo método propuesto para resolver la ecuación (1.1) es un algoritmo Viscous Flux-Limiter. Este método es la extensión del algoritmo desarrollado por Jerez y Uh [15] para flujos convectivos-difusivos. La ecuación de convección-difusión no lineal está dominada por la convección, el término convectivo puede dar lugar a oscilaciones numéricas que desencadenen inestabilidades, provocando que las aproximaciones numéricas no converjan a la solución. Jerez y Uh describen un método TVD (Total Variation Diminishing) en [15] para la ecuación de convección no lineal. Este método conservativo hiperbólico da una buena aproximación numérica a las soluciones sin oscilaciones cerca de las discontinuidades. Las simulaciones realizadas en este artículo indican que el método TVD Flux-Limiter desarrollado por Jerez y Uh no genera oscilaciones para $CFL \leq 1$, y es óptimo en el sentido de que permite la mayor CFL para esquemas explícitos, $CFL = 1$ ([3], [6], [28], [18]). El método TVD-Flux-limiter consigue evitar la generación de oscilaciones espurias que se producen cerca de las discontinuidades.

Estas cualidades del método TVD-Flux-limiter hacen que sea plausible su extensión a problemas hiperbólicos con términos difusivos, es decir, a problemas de convección-difusión no lineal dominados por el término convectivo. Con este método se consigue evitar oscilaciones en cercanías de las discontinuidades y en presencia de términos convectivos fuertes.

Metodología

El presente trabajo está estructurado de la siguiente manera:

- En el Capítulo II se presentan los conceptos de esquemas en diferencias finitas y definiciones básicas que deben de cumplir dichos métodos tales como consistencia, estabilidad para que se tenga convergencia. Se hace una descripción de diferentes esquemas numéricos para resolver problemas de difusión y convección no lineal. Estos esquemas posteriormente se aplicaran en los siguientes capítulos al problema de convección y difusión que surgen en la descomposición de operadores de la ecuación de convección-difusión no lineal.

- En el Capítulo III se describe el esquema numérico de descomposición de operadores Splitting Strang aplicado a la ecuación de convección-difusión no lineal. El Splitting combina una descomposición de operadores en la discretización temporal de segundo orden y nos permite desacoplar la ecuación de convección-difusión obteniéndose problemas de tipo parabólico lineal e hiperbólico no lineal.

- En el Capítulo IV se analiza una extensión del método TVD Flux-Limiter, descrito en el Capítulo II para la ecuación de convección-difusión no lineal, al cual llamamos Viscous TVD Flux-Limiter y realizamos un análisis numérico completo para determinar propiedades importantes como error de truncamiento, consistencia, convergencia y TVD-estabilidad.

- En el Capítulo V se presentan resultados numéricos del método Splitting Strang y Viscous TVD Flux-Limiter para la ecuación de convección-difusión. En esta parte comparamos ambas aproximaciones con la solución exacta en problemas estándar donde la solución es conocida.

- Se completa la tesis con las secciones de conclusiones, bibliografía y apéndices, donde se desarrollan algunos conceptos básicos de EDP´s y teoremas usuales de métodos numéricos para EDP´s.

CAPÍTULO 2

Métodos en Diferencias Finitas

En general, encontrar la solución explícita de una ecuación diferencial parcial resulta difícil. Esta dificultad puede tener diversos orígenes, tales como la presencia de fronteras irregulares, hasta la existencia de términos no lineales en las ecuaciones mismas. Por lo tanto, en la mayoría de los casos la única opción es aproximar numéricamente a la solución a la ecuación.

Existen muchas formas distintas de resolver ecuaciones diferenciales parciales de forma numérica. Los métodos más populares son: las diferencias finitas, los elementos finitos, los volúmenes finitos y los métodos espectrales ([17], [19], [20]). Todos ellos tienen ventajas y desventajas. En realidad, cada uno de ellos puede ser el mejor según la aplicación que se considere. Nos limitaremos a estudiar el método de diferencias finitas (MDF) por ser el método conceptualmente más simple que no requiere transformar la EDP.

El MDF se basa en el desarrollo en series de Taylor. Requiere cierta regularidad en la malla de trabajo, por lo que resulta eficiente en geometrías especiales no muy irregulares. La discretización del problema, requiere reducir el dominio a un conjunto finito de puntos.

Como se dijo anteriormente el objetivo principal de esta tesis es obtener métodos para resolver numéricamente la ecuación diferencial parcial parabólica no lineal

$$u_t + f(u)_x = \nu u_{xx}, \tag{2.1}$$

donde ν es contante y f una función no lineal de u. Para asegurar la existencia y unicidad de la solución en (2.1) considérese condiciones iniciales suaves y condiciones

de contorno tipo Dirichlet, véase **Apéndice A**. Aunque también se estudian soluciones para problemas con condiciones iniciales discontinuos en los que se conoce su solución [34].

Otra manera de ver la ecuación (2.1) es como un sistema de dos ecuaciones. Una de ellas, la conocida ecuación del calor

$$u_t = \nu u_{xx}, \tag{2.2}$$

y la otra, una ecuación hiperbólica de primer orden no lineal

$$u_t + f(u)_x = 0. \tag{2.3}$$

En particular se demuestra en base al Teorema de Lie, véase **Apéndice D**, que la solución de la ecuación (2.1), se puede obtener resolviendo cada una de las ecuaciones (2.2) y (2.3), esto último se analiza en el siguiente capítulo.

Con la intención de resolver numéricamente los modelos (2.1) y el (2.2)-(2.3), en este capítulo iniciaremos revisando los métodos en diferencias finitas más conocidos para las ecuaciones (2.2) y (2.3), así como un estudio de sus propiedades cualitativas: consistencia, estabilidad y convergencia.

2.1. Aproximación de derivadas mediante diferencias finitas

El método de diferencias finitas se basa en el teorema de Taylor para aproximar las derivadas. El Teorema de Taylor que a continuación enunciaremos, nos dice que bajo ciertas condiciones, una función puede expresarse como un polinomio de Taylor mas un cierto error.

Teorema 2.1. ***Taylor*** *[2]. Sea f continua en [a,b] y con derivadas hasta de orden n continuas también en este intervalo cerrado; supóngase que $f^{(n+1)}(x)$ existe en (a,b), entonces para x y $x_o \in (a,b)$ se tiene*

$$f(x) = f(x_0) + f'(x_0)(x-x_0) + \frac{f^{(2)}(x_0)}{2!}(x-x_0)^2 + \ldots + \frac{f^{(n)}(x_0)}{n!}(x-x_0)^n + R_n(f),$$

donde $R_n(f) = \int_a^x \frac{f^{(n+1)}(t)}{n!}(x-t)^n dt$ *o* $R_n(f) = \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-x_0)^{n+1} dt$, $\xi \in (x_0, x)$.

Consideremos los desarrollos de Taylor de orden 4 para $u(x+h)$ y $u(x-h)$, entonces se tiene que

$$u(x+h) = u(x) + hu'(x) + \frac{1}{2}h^2u''(x) + \frac{1}{6}h^3u'''(x) + \frac{1}{24}h^4u''''(\xi_1) \text{ donde } \xi_1 \in (x, x+h) \tag{2.4}$$

y

$$u(x-h) = u(x) - hu'(x) + \frac{1}{2}h^2u''(x) - \frac{1}{6}h^3u'''(x) + \frac{1}{24}h^4u''''(\xi_2) \text{ donde } \xi_2 \in (x-h, x) \tag{2.5}$$

Al sumar (2.4) y (2.5) se obtiene

$$u(x+h) + u(x-h) = 2u(x) + h^2u''(x) + \frac{h^4}{24}(u''''(\xi_1) + u''''(\xi_2)), \tag{2.6}$$

que es igual a

$$u(x+h) + u(x-h) = 2u(x) + h^2u''(x) + \mathcal{O}(h^4), \tag{2.7}$$

donde $\mathcal{O}(h^4)$ denota los términos que contienen potencias de h de orden 4. Asumiendo que estos términos son pequeños en relación con las potencias menores de h, se sigue que

$$u''(x) \simeq \frac{1}{h^2}\{u(x+h) - 2u(x) + u(x-h)\}, \tag{2.8}$$

con un error de truncamiento de orden h^2. Al restar la ecuación (2.5) de la ecuación (2.4), y despreciar los términos de orden h^3 se obtiene

$$u'(x) \simeq \frac{1}{2h}\{u(x+h) - u(x-h)\}. \tag{2.9}$$

La aproximación (2.9) es de orden h^2 y aproxima la pendiente de la tangente en el punto $(x, u(x))$ mediante la pendiente de la recta que pasa por los puntos $(x-h, u(x-h))$ y $(x+h, u(x+h))$. Esta aproximación se conoce como aproximación por diferencias centradas. También se puede aproximar la pendiente de la tangente en $(x, u(x))$ por la pendiente de la recta que pasa por los puntos $(x, u(x))$ y $(x-h, u(x-h))$, obteniendo la aproximación por diferencias regresivas o de Euler Atrasadas de orden h

$$u'(x) \simeq \frac{1}{h}\{u(x) - u(x-h)\}, \tag{2.10}$$

o por la pendiente de la recta que pasa por los puntos $(x+h, u(x+h))$ y $(x, u(x))$, obteniendo la aproximación por diferencias progresistas o de Euler Adelantadas de orden h

$$u'(x) \simeq \frac{1}{h}\{u(x+h) - u(x)\}. \tag{2.11}$$

Ahora, tomemos una función u de las variables x y t y un rectángulo $\mathcal{R} = \{(x,t), 0 \leq x \leq L, 0 \leq t \leq b\}$. Dividamos $\mathcal{R}$ en m por N rectángulos de lados $\triangle x = h$, $\triangle t = k$ y definimos los puntos de la malla $x_j = jh$ y $t_n = nk$ donde $j = 0, 1, \ldots, m$ y para $n = 0, 1, \ldots, N$ (véase figura 2.1). A continuación desarrollamos las aproximaciones para las derivadas de $u(x,t)$ en los puntos de la malla como en (2.8). Entonces

$$\left(\frac{\partial^2 u}{\partial x^2}\right)_{j,n} \simeq \frac{u\{(j+1)h, nk\} - 2u\{jh, nk\} + u\{(j-1)h, nk\}}{h^2}, \tag{2.12}$$

que es equivalente a

$$\left(\frac{\partial^2 u}{\partial x^2}\right)_{j,n} \simeq \frac{U_{j+1}^n - 2U_j^n + U_{j-1}^n}{h^2}. \tag{2.13}$$

Donde $U_j^n \in \mathbb{R}^m$ es la aproximación de la solución exacta $u_j^n = u(x_j, t_n)$ en los puntos discretizados. De manera similar, obtenemos la aproximación

$$\left(\frac{\partial^2 u}{\partial t^2}\right)_{j,n} \simeq \frac{U_j^{n+1} - 2U_j^n + U_j^{n-1}}{k^2}.$$

De las ecuaciones (2.10) y (2.11) se siguen dos aproximaciones para $\partial u/\partial t$ en $u(x_j, t_n)$ son

$$\frac{\partial u}{\partial t} \simeq \frac{U_j^{n+1} - U_j^n}{k}, \tag{2.14}$$

y

$$\frac{\partial u}{\partial t} \simeq \frac{U_j^n - U_j^{n-1}}{k}. \tag{2.15}$$

Las aproximaciones (2.13), (2.14) y (2.15) se conocen como aproximaciones en diferencias finitas a las derivadas parciales de una función u.

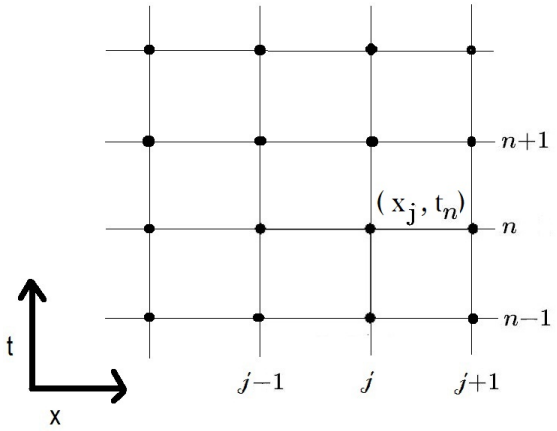

Figura 2.1: Discretización del dominio $\mathcal{R}$.

Por ejemplo si tomamos la ecuación diferencial de advección lineal

$$u_t + a u_x = 0 \tag{2.16}$$

donde $a \in \mathbb{R}$, un método en diferencias finitas será el obtenido de aproximar u_x por (2.11) y u_t por (2.14) obteniendo

$$\frac{U_j^{n+1} - U_j^n}{k} + a\frac{U_{j+1}^n - U_j^n}{h} = 0, \tag{2.17}$$

que es equivalente a

$$U_j^{n+1} = (1+\beta)U_j^n - \beta U_{j+1}^n, \tag{2.18}$$

donde $\beta = \dfrac{ak}{h}$.

Definición de $\mathcal{O}(h)$

Usamos la notación $\mathcal{O}(h^p)$ para denotar el error de la aproximación y se lee "de orden p". Si $f(h)$ y $g(h)$ son dos funciones de h, entonces decimos que

$$f(h) = \mathcal{O}(g(h)) \quad h \to 0,$$

si existe una constante C tal que

$$\left|\frac{f(h)}{g(h)}\right| < C \qquad \text{para todo } h \text{ suficientemente pequeño,}$$

o, equivalentemente, si podemos acotar

$$|f(h)| < C|g(h)| \qquad \text{para todo } h \text{ suficientemente pequeño.}$$

Intuitivamente, esto significa que $f(h)$ decae a cero al menos tan rápido como la función $g(h)$ lo hace.

2.2. Propiedades Cualitativas de los Métodos en Diferencias Finitas

En este apartado se comentan algunas de las propiedades que puede tener, y en general es conveniente que tenga, un esquema numérico (véase [21]). Cuando en adelante se discutan los distintos esquemas, el hecho de poseer o no estas propiedades dará inmediatamente una idea de sus capacidades y limitaciones.

Los métodos basados en las diferencias finitas pueden ser de dos clases: **explícitos** e **implícitos**. Los esquemas explícitos son aquellos en los que el cálculo de las variables en un instante de tiempo se efectúa tan sólo con los valores que toman en el instante anterior. Por el contrario, un esquema implícito evalúa las variables en un punto del espacio en función de valores en otros puntos del espacio en el mismo instante, por lo que se debe resolver en cada paso de tiempo un sistema de ecuaciones que englobe todas las variables en todos los puntos del espacio en el instante t^{n+1}.

Los esquemas explícitos tienen un coste computacional pequeño en cada paso de tiempo, pero presentan el inconveniente de requerir pasos de tiempo muy pequeños durante el cálculo para que resulten estables. Los esquemas implícitos tienen la ventaja sobre los esquemas explícitos que son incondicionalmente estables, aunque en ocasiones el análisis de la convergencia de estos métodos resulta difícil de obtener y son más caros computacionalmente.

La multiplicidad de posibles aproximaciones nos lleva a la siguiente pregunta: ¿Cómo saber en qué caso usar una cierta aproximación y no otra? Desgraciadamente, no existe una respuesta general a esta pregunta. Sin embargo, sí existen guías que nos permiten escoger entre distintas aproximaciones en ciertos casos. Estas guías tienen que ver con los conceptos de consistencia, convergencia y estabilidad.

Estamos interesados en estudiar únicamente métodos explícitos de dos niveles por su sencillez en cálculos matemáticos y computacionales frente a los métodos implícitos. Estos métodos pueden ser escritos de la forma

$$U^{n+1} = \mathcal{H}_k(U^n),$$

donde U^{n+1} representa el vector de aproximación U_j^{n+1} en el tiempo t_{n+1} y $\mathcal{H}_k$ es un operador en diferencias. El valor U_j^{n+1} en un punto particular j típicamente depende de varios valores del vector U^n, entonces escribiremos

$$U_j^{n+1} = \mathcal{H}_k(U^n; j).$$

Por ejemplo para la ecuación de advección lineal (2.16),

$$U_j^{n+1} = (1+\beta)U_j^n - \beta U_{j+1}^n,$$

tenemos que el operador $\mathcal{H}_k$ toma la forma

$$\mathcal{H}_k(U^n; j) = (1+\beta)U_j^n - \beta U_{j+1}^n.$$

Error de truncamiento local y consistencia

Definición 2.1. Para un método cualquiera de dos niveles, definimos el ***error de truncamiento local*** [21] por

$$\tau_{h,k}(x,t) = \frac{1}{k}[u(x,t+k) - \mathcal{H}_k(u(\cdot,t);x)].$$

Típicamente tenemos que

$$\tau_{h,k}(x,t) = \mathcal{O}(h^p) + \mathcal{O}(k^q),$$

para algunos enteros p y q.

El ***error de truncamiento local***, $\tau_{h,k}(x,t)$, se puede entender como una medida local de qué tan bueno es el modelo discreto en diferencias finitas con respecto al modelo diferencial del que se parte.

El error se calcula reemplazando la aproximación U_j^n del método por la solución real $u(x_j, t_n)$. La solución real de la ecuación diferencial parcial es una aproximación de la solución de la ecuación en diferencias. Entonces la sustitución de la ecuación real en las ecuaciones en diferencias nos dará un indicador de que tan bién la solución de la ecuación en diferencias satisface la ecuación diferencial. Para ello el error se calcula reemplazando la aproximación del método por la solución real

Consideremos una cierta aproximación en diferencias finitas a la solución de una ecuación dife-rencial. Cuando la malla se refina (es decir, cuando Δt y Δx se hacen cada vez más pequeños), uno esperaría que la aproximación fuera cada vez mejor en el sentido de que los errores de truncamiento se hacen cada vez más pequeños. Buscamos entonces que en el límite continuo de nuestra aproximación se acerque a la ecuación diferencial original y no a otra. Cuando esto ocurre localmente se dice que nuestra aproximación es "consistente". En general, esta propiedad es muy fácil de ver de la estructura de las aproximaciones en diferencias finitas. La consistencia es fundamental en una aproximación en diferencias finitas. Si falla, aunque sea en un sólo punto, implica que no recuperaremos la solución correcta de la ecuación diferencial.

Definición 2.2. El método es ***consistente*** [21] si

$$\|\tau_{h,k}(x,t)\| \to 0, \quad \forall h, k \to 0.$$

Definición 2.3. Decimos que el método es de ***orden p en el tiempo*** y ***orden q en el espacio*** [21] si para datos iniciales suficientemente suaves con soporte compacto [1], existe constantes C_1 y C_2 tales que

$$\|\tau_{h,k}(x,t)\| \leq C_1 k^p + C_2 h^q, \quad \forall k \leq k_0, t \leq T.$$

Estamos interesados en conocer que tan bien U_j^n aproxima la solución real, entonces definimos el error global como la diferencia de la solución real y la solución calculada.

[1] Se dice que una función tiene soporte compacto si el conjunto donde no es nula conforma un conjunto cerrado y acotado

Definición 2.4. Definimos el ***error global*** [21] de un método como

$$E_j^n = U_j^n - u_j^n \quad \forall j, n.$$

La consistencia es solo una propiedad local: una aproximación consistente se reduce localmente a la ecuación diferencial en el límite continuo. En la práctica, estamos realmente interesados en una propiedad más global. Lo que realmente buscamos es que la aproximación mejore en un tiempo finito T cuando refinamos la malla. Es decir, la diferencia entre la solución exacta y la solución numérica en un tiempo fijo t debe tender a cero en el límite continuo. Esta condición se conoce como "convergencia".

La convergencia es diferente a la consistencia: esquemas consistentes pueden no ser convergentes. Esto no es difícil de entender si pensamos que en el límite cuando Δt tiende a cero, un tiempo finito T se puede alcanzar solo después de un número infinito de pasos. Esto implica que incluso si el error en cada paso es infinitesimal, su integral total puede ser finita. La solución numérica puede incluso divergir y el error resultar infinito.

Definición 2.5. Decimos que el método es ***convergente*** [21] para alguna norma en particular $\|\cdot\|$ si

$$\|E^n\| \to 0, \quad n \to \infty.$$

En general es muy difícil verificar analíticamente si un esquema de aproximación es convergente o no lo es. Numéricamente, por otro lado, es muy fácil ver si la solución aproximada converge a algo (es decir, no diverge). Lo difícil es saber si la solución numérica converge hacia la solución exacta y no a otra cosa. Por ello, la siguiente propiedad resulta importante para eliminar esta dificultad.

Independientemente del comportamiento de la solución a la ecuación diferencial, debemos pedir que las soluciones exactas de las ecuaciones en diferencias finitas permanezcan acotadas para cualquier tiempo finito T y cualquier intervalo de tiempo Δt (el error global esté acotado cuando crece n con h fijo). Este requisito se conoce como "estabilidad". En esta tesis tomamos la definición dada por Lax-Richtmyer.

Definición 2.6. : Estabilidad Lax-Richtmyer
Decimos que un método es ***estable*** [21] si para cada tiempo T existe una constante C y un valor $k_0 > 0$ tal que

$$\|\mathcal{H}_k^n\| \leq C, \quad \forall nk \leq T, k < k_0.$$

Notese que el método en particular es estable si $\|\mathcal{H}_k\| \leq 1$, dado que

$$\|\mathcal{H}_k^n\| \leq \|\mathcal{H}_k\|^n \leq 1, \quad \forall n, k.$$

Un resultado fundamental de la teoría de las aproximaciones en diferencias finitas es el teorema de Lax (para su demostración ver [31], ya que en esta Tesis solo utilizamos su resultado, ya que nos concentramos en su demostración), en el cual estos conceptos están relacionados:

Teorema de Equivalencia de Lax-Richtmyer

Teorema 2.2. *Un esquema que es consistente para un problema de valores iniciales lineal bien planteado para una ecuación diferencial parcial en diferencias finitas es convergente si y solo si es estable.* ***(ver apéndice A)***

Este teorema es de gran importancia pues relaciona el objetivo final de toda aproximación en diferencias finitas, es decir, la convergencia a la solución exacta, con una propiedad que es mucho más fácil de probar: la estabilidad. Lax estableció con el las condiciones bajo las cuales una implementación numérica da una aproximación válida de la solución a una ecuación diferencial.

2.3. Ecuación de Difusión

La Ecuación de Difusión constituye una herramienta de gran utilidad para dar solución a problemas de flujo de calor en cuerpos determinados. Si $u(x, y, z, t)$ es la temperatura en el punto (x, y, z), en un instante t, la ecuación es

$$u_t = \alpha^2 \nabla^2 u.$$

Esta ecuación aparece también en una gran variedad de problemas de la física matemática, por ejemplo, la concentración de material en difusión, la propagación de olas en canales de gran longitud, y la transmisión en cables eléctricos. En la termodinámica, la ecuación de difusión puede ser aplicada en tres situaciones: cuerpos sólidos (tres dimensiones), placas (dos dimensiones) y barras (una dimension).

La ecuación de difusión unidimensional $u_t = \alpha^2 \nabla^2 u$ aplicaría, por ejemplo, para el caso de una barra metálica larga y delgada, con aislamiento, ya que la temperatura de cualquier sección transversal será constante, debido a que el tiempo que tarda la temperatura en equilibrarse en distancias cortas se asume como despreciable.

En este caso, si asumimos que la barra tiene una longitud L, una temperatura inicial $f(x)$, y que los extremos se mantienen a temperatura cero, la distribución de temperatura en la barra está dada por la solución del problema de valores iniciales

$$u_t = \alpha^2 \nabla^2 u, 0 < x < L, t > 0 \quad (2.19)$$

$$u(x, 0) = f(x), 0 \leq x \leq L \quad (2.20)$$

con condiciones en la frontera tipo Dirichlet

$$\begin{aligned} u(0,t) &= 0 \\ u(L,t) &= 0. \end{aligned}$$

Este problema puede resolverse por medio del método analítico de separación de variables [24]. La solución general es

$$u(t,x) = \sum_{n=1}^{\infty} b_n e^{-\alpha^2 \left(\frac{n\pi}{L}\right)^2 t} \operatorname{sen}\left(\frac{nx\pi}{L}\right),$$

donde b_n es obtenido de la forma

$$b_n = \frac{2}{L}\int_0^L f(\tau) \operatorname{sen}\left(\frac{n\tau\pi}{L}\right) d\tau.$$

Para la ecuación de difusión se plantean tres variantes del método de diferencias finitas: Euler Explícito, Euler Implícito y de Crank-Nicholson.

Para el método Euler explícito se presenta su deducción y se estudian las propiedades de error de truncamiento local, consistencia, estabilidad y convergencia, para los demás métodos mencionamos su error de truncamiento local, consistencia y estabilidad, ya que el análisis es similar al realizado para el método Euler explícito.

2.3.1. Método Euler Explícito

Considere el problema planteado en las ecuaciones (2.19)-(2.20) de manera unidimensional. De las ecuaciones (2.13) y (2.14) se sigue que una aproximación por diferencias finitas para

$$\frac{\partial u}{\partial t} = \alpha^2 \frac{\partial^2 u}{\partial x^2},$$

es

$$\frac{U_j^{n+1} - U_j^n}{k} = \alpha^2 \frac{U_{j+1}^n - 2U_j^n + U_{j-1}^n}{h^2}, \tag{2.21}$$

donde U_j^n es la aproximación de la solución exacta $u_j^n = u(x_j; t_n)$ en los puntos de la malla, $x_j = jh$, $j = 0, 1, 2, \ldots, m-1, m$ y $t_n = nk, n = 0, 1, 2, ..., N$. La ecuación (2.21) puede reescribirse como

$$U_j^{n+1} = \lambda U_{j-1}^n + (1 - 2\lambda) U_j^n + \lambda U_{j+1}^n, \tag{2.22}$$

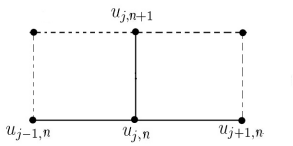

Figura 2.2: Esquema del Método Euler Explícito.

donde $\lambda = \alpha^2(k/h^2)$. La ecuación en diferencias se emplea para calcular las aproximaciones en la fila $(n+1)$-ésima de la malla a partir de las aproximaciones de la fila anterior, hagamos notar que esta fórmula proporciona explícitamente el valor U_j^{j+1} en función de U_{j-1}^n, U_j^n y U_{j+1}^n, véase figura 2.2.

Dado que la condición inicial $u(x,0)=f(x)$, para todo $0 \leq x \leq L$, implica que $U_j^0 = f(x_j)$, para toda $j = 0, 1, 2, \ldots, m$, podemos usar estos valores en la ecuación (2.22) para calcular el valor de U_j^1 para toda $j = 1, 2, \ldots, m-1$. Las condiciones de frontera $u(0,t) = 0$ y $u(L,t) = 0$ implican que $U_0^n = U_m^n = 0$, $(n = 0, 1, 2, \ldots, N)$ y por tanto, podemos determinar todos los elementos de la forma U_j^1. Ya conocidas todas las aproximaciones U_j^1 se pueden obtener, siguiendo un procedimiento semejante, los valores U_j^2, U_j^3,...,U_j^N.

Si hacemos $U^{(0)} = (f(x_1), f(x_2), \ldots, f(x_{m-1}))^T$ y $U^{(n)} = (U_1^n, U_2^n, \ldots, U_{m-1}^n)^T$, para todo $n = 1, 2, \ldots, i$, se puede plantear matricialmente este método de solución como

$$U^n = \mathrm{A}U^{n-1}, \quad \forall n = 1, 2, \ldots, N-1. \tag{2.23}$$

Donde $\mathrm{A} \in \mathbb{R}^{(N-1)\times(N-1)}$ es la siguiente matriz tridiagonal

$$\mathrm{A} = \begin{bmatrix} (1-2\lambda) & \lambda & 0 & \ldots & \ldots & 0 \\ \lambda & (1-2\lambda) & \lambda & \ddots & & \vdots \\ 0 & \lambda & (1-2\lambda) & & \ddots & \vdots \\ \vdots & \ddots & & & & 0 \\ \vdots & & \ddots & \lambda & (1-2\lambda) & \lambda \\ 0 & \ldots & \ldots & 0 & \lambda & (1-2\lambda) \end{bmatrix}.$$

Error de truncamiento local y Consistencia

Por definición el error de truncamiento local está dado por

$$\tau_{h,k} = \frac{1}{k}\left[U_j^{n+1} - U_j^n\right] - \frac{\alpha^2}{h^2}\left[U_{j+1}^n - 2U_j^n + U_{j-1}^n\right]. \tag{2.24}$$

Desarrollando en series de Taylor (tomaremos $u = u_j^n$ para facilitar la notación)

$$\begin{aligned} U_j^{n+1} &= u + ku_t + \frac{k^2}{2}u_{tt} + \frac{k^3}{6}u_{ttt} + \mathcal{O}(k^4), \\ U_{j+1}^n &= u + hu_x + \frac{h^2}{2}u_{xx} + \frac{h^3}{6}u_{xxx} + \frac{h^4}{24}u_{xxxx} + \mathcal{O}(h^5), \\ U_{j-1}^n &= u - hu_x + \frac{h^2}{2}u_{xx} - \frac{h^3}{6}u_{xxx} + \frac{h^4}{24}u_{xxxx} + \mathcal{O}(h^5). \end{aligned}$$

Obtenemos

$$\begin{aligned} \frac{1}{k}\left[U_j^{n+1} - U_j^n\right] &= u_t + \frac{k}{2}u_{tt} + \frac{k^2}{6}u_{ttt} + \mathcal{O}(k^3), \\ \frac{\alpha^2}{h^2}\left[U_{j+1}^n - 2U_j^n + U_{j-1}^n\right] &= \alpha^2 u_{xx} + \frac{\alpha^2 h^2}{12}u_{xxxx} + \mathcal{O}(h^4). \end{aligned}$$

Entonces

$$\begin{aligned} \tau_{h,k} &= u_t + \frac{k}{2}u_{tt} + \frac{k^2}{6}u_{ttt} + O(k^3) - \alpha^2 u_{xx} - \frac{\alpha^2 h^2}{12}u_{xxxx} + \mathcal{O}(h^4) \\ &= k\left(\frac{u_{tt}}{2}\right) + h^2\left(-\frac{\alpha^2 u_{xxxx}}{12}\right) + \mathcal{O}(k^2) + \mathcal{O}(h^4). \end{aligned}$$

Por lo tanto

$$\tau_{h,k} = \mathcal{O}(k) + \mathcal{O}(h^2), \tag{2.25}$$

es decir nuestro método es de primer orden en el tiempo y de segundo orden en el espacio.

Claramente si tomamos $h, k \to 0$ se tiene que $\tau_{h,k} \to 0$, entonces el método es *consistente*.

Estabilidad

Hay dos formas típicas de tratar la Estabilidad, es decir, acotar los errores en las aproximaciones de las soluciones de ecuaciones en diferencias:

- *Método Algebraico:* Expresa la ecuación en forma matricial y examina los valores propios de la matriz asociada.
- *Método de Fourier o Von Neumann:* Utiliza transformadas discretas de Fourier [27].

En esta sección utilizaremos el método algebraico.
Sean

$U \equiv$solución de la ecuación en diferencias con valores iniciales conocidos,

y

$U_* \equiv$solución de la ecuación correspondiente a valores iniciales perturbados (por ejemplo, valores iniciales con error de redondeo).

Por la fórmula (2.23) se tiene

$$U^n = AU^{n-1} = A(AU^{n-2}) = \ldots = A^nU^0,$$

donde U^0 es el vector de condiciones iniciales. Calculemos ahora los valores de la solución perturbada

$$U_*^1 = AU_*^0, \qquad U_*^2 = AU_*^1 = A^2U_*^0, \ldots, U_*^n = A^nU_*^0.$$

Supongamos que no hay errores posteriores, es decir, que los cálculos aritméticos posteriores son exactos. De la definición de error global (tomaremos $e_j = E_j^n$) tenemos

$$e_j = U^n - U_*^n = A^n(U^0 - U_*^0) = A^ne_0.$$

Entonces el esquema de Euler explícito será estable si e_j permanece acotado cuando $j \to \infty$. Esto puede calcularse expresando el vector de errores en términos de los vectores propios de A. Como $A \in \mathbb{R}^{(N-1)\times(N-1)}$ es una matriz simétrica, entonces A tiene $N-1$ vectores propios linealmente independientes v_s. Estos vectores propios forman una base en $\mathbb{R}^{(N-1)}$ y permiten expresar el vector de errores iniciales e_0 en la forma:

$$e_0 = \sum_{s=1}^{N-1} c_s v_s; \qquad c_s \in \mathbb{C}, s = 1, 2, \ldots, N-1.$$

Los errores a lo largo del mismo nivel de tiempo $t = k$ resultante de la propagación, vienen dados por

$$\begin{aligned}
e_1 &= Ae_0 = A\left(\sum_{s=1}^{N-1} c_s v_s\right) = \sum_{s=1}^{N-1} c_s A(v_s) = \sum_{s=1}^{N-1} c_s \lambda_s v_s, \\
e_2 &= Ae_1 = \sum_{s=1}^{N-1} c_s \lambda_s^2 v_s, \\
&\vdots \\
e_j &= \sum_{s=1}^{N-1} c_s \lambda_s^j v_s,
\end{aligned}$$

entonces $\|e_j\| \leq \sum_{s=1}^{N-1} |c_s||\lambda_s|^j \|v_s\|$, lo que demuestra que los errores no crecen exponencialmente con j siempre que

$$|\lambda_s| \leq 1, \qquad s = 1, 2, \ldots, N-1.$$

Para obtener información sobre $\sigma(A)$, donde $\sigma(A)$ es el espectro de A, necesitamos el siguiente Lema.

Lema 2.1. *[20]. Sean $d_0, d_1 \in \mathbb{R}$ y $M \in \mathbb{R}^{m\times m}$ una matriz tridiagonal simétrica de Toeplitz definida por*

$$\mathrm{M} = \begin{bmatrix} d_0 & d_1 & 0 & \dots & \dots & 0 \\ d_1 & d_0 & d_1 & \ddots & & \vdots \\ 0 & d_1 & d_0 & & \ddots & \vdots \\ \vdots & \ddots & & & & 0 \\ \vdots & & \ddots & d_1 & d_0 & d_1 \\ 0 & \dots & \dots & 0 & d_1 & d_0 \end{bmatrix}$$

Entonces los valores propios de M son

$$\lambda_s = d_0 + 2d_1 \cos(s\pi/(m+1)), \qquad s = 1, 2, \dots, m,$$

y el correspondiente vector propio r_s tiene en la j-ésima componente

$$r_{j,s} = \operatorname{sen}(s\pi j/(m+1)), \qquad j = 1, 2, \dots, m.$$

Una vez conocidos los valores propios de una matriz tridiagonal, estamos en condiciones de estudiar la condición de estabilidad del esquema:

$$|\lambda_s| \leq 1, \qquad \forall \lambda_s \in \sigma(A).$$

Por el Lema de los valores propios de una matriz tridiagonal, para la matriz A se tiene que $\sigma(A) = \{\lambda_s; 1 \leq s \leq N-1\}$ con

$$\begin{aligned} \lambda_s &= 1 - 2\lambda + 2\lambda \cos(s\pi/N), \\ &= 1 - 2(\lambda - \lambda \cos(s\pi/N)), \\ &= 1 - 4\lambda \operatorname{sen}^2\left(\frac{s\pi}{2N}\right), \end{aligned}$$

entonces $\sigma(A) = \{1 - 4\lambda \operatorname{sen}^2\left(\frac{s\pi}{2N}\right); 1 \leq s \leq N-1\}$.

La condición de estabilidad es

$$\left|1 - 4\lambda \operatorname{sen}^2\left(\frac{s\pi}{2N}\right)\right| \leq 1, \qquad 1 \leq s \leq N-1,$$

que se puede escribir como

$$-1 \leq 1 - 4\lambda \operatorname{sen}^2\left(\frac{s\pi}{2N}\right) \leq 1.$$

La desigualdad de la derecha es evidente y la de la izquierda es equivalente a:

$$\lambda \leq \frac{1}{2\,\mathrm{sen}^2\left(\frac{s\pi}{2N}\right)}, \qquad 1 \leq s \leq N-1. \tag{2.26}$$

observemos que:

$$2\,\mathrm{sen}^2\left(\frac{s\pi}{2N}\right) \leq 2.$$

Así, la condición se satisface si

$$\lambda \leq \frac{1}{2}. \tag{2.27}$$

La formula (2.23) es estable si $0 \leq \lambda \leq \frac{1}{2}$. Esto significa que el tamaño de paso k debe cumplir $k \leq \frac{h^2}{2\alpha^2}$. Si esto no se cumple, entonces puede ocurrir que los errores introducidos en la fila $\{U_j^n\}$ se amplifiquen en alguna fila posterior $\{U_j^p\}$ para algún $p > n$.

Convergencia

Usando el Teorema de Equivalencia de Lax tenemos que el Método Euler Explícito es convergente bajo la misma condición con la cual es estable, es decir, cuando $0 \leq \lambda \leq \frac{1}{2}$ ver (2.27).

Como la precisión de la aproximación a la EDP (2.23) es de orden $\mathcal{O}(k) + \mathcal{O}(h^2)$ y como el término $\mathcal{O}(k)$ tiende a cero linealmente, no es sorprendente que k deba tomarse muy pequeño para obtener buenas aproximaciones. Aun así, la necesidad de que el método sea estable plantea consideraciones adicionales. Supongamos que las aproximaciones obtenidas en la malla no son suficientemente precisas y que debemos reducir los tamaños de paso $\triangle x = h_0$, $\triangle t = k_0$. Si tomamos como nuevo tamaño de paso para la coordenada x, simplemente $\triangle x = h_1 = \frac{h_0}{2}$, y queremos mantener el mismo valor del cociente λ, entonces k_1, debe cumplir

$$k_1 = \frac{r(h_1)^2}{\alpha^2} = \frac{r(h_0)^2}{4\alpha^2} = \frac{k_0}{4}.$$

En consecuencia: Hay que doblar el numero de nodos de la malla en el eje de la variable x y cuadruplicarlo en el eje de la variable t, con lo cual el esfuerzo de ordenador se multiplica por ocho. Este esfuerzo extra, nos obliga a buscar métodos mas eficaces que no estén sujetos a restricciones de estabilidad tan exigente.

2.3.2. Método de Euler Implícito

Si en la ecuación (2.21) reemplazamos el lado izquierdo por (2.15) se obtiene la siguiente aproximación en diferencias finitas a la ecuación de calor

$$\frac{U_j^n - U_j^{n-1}}{k} = \alpha^2 \frac{U_{j+1}^n - 2U_j^n + U_{j-1}^n}{h^2}.$$

Se puede reescribir como

$$(1+2\lambda)U_j^n - \lambda U_{j+1}^n - \lambda U_{j-1}^n = U_j^{n-1}. \tag{2.28}$$

Gráficamente en la figura 2.3 se encuentra el método (2.28).

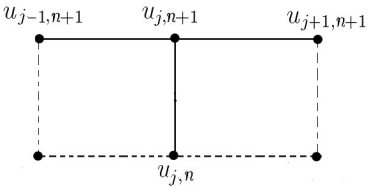

Figura 2.3: Esquema del Método Euler Implícito.

Aplicando el hecho de que $U_j^0 = f(x_j)$ para toda $j = 1, 2, ..., m-1$ y $U_0^n = U_m^n = 0$ para toda $n = 1, 2, ..., N$, este método de diferencias tiene la representación matricial

$$\mathrm{A} = \begin{bmatrix} (1+2\lambda) & -\lambda & 0 & \dots & \dots & 0 \\ -\lambda & (1+2\lambda) & -\lambda & \ddots & & \vdots \\ 0 & -\lambda & (1+2\lambda) & & \ddots & \vdots \\ \vdots & \ddots & & & & 0 \\ \vdots & & \ddots & -\lambda & (1+2\lambda) & -\lambda \\ 0 & \dots & \dots & 0 & -\lambda & (1+2\lambda) \end{bmatrix} \begin{bmatrix} U_1^n \\ U_2^n \\ \vdots \\ U_{m-1}^n \end{bmatrix} = \begin{bmatrix} U_1^{n-1} \\ U_2^{n-1} \\ \vdots \\ U_{m-1}^{n-1} \end{bmatrix}$$

o

$$\mathrm{A}U^n = U^{n-1}, \text{ para todo } n = 1, 2, \dots, N-1.$$

Así debemos resolver ahora un sistema lineal para obtener $U^{(n)}$ a partir de $U^{(n-1)}$. La precision de esta método es de orden $\mathcal{O}(k)+\mathcal{O}(h^2)$ y es estable para cualquier λ, por lo tanto es convergente debido al Teorema de Equivalencia de Lax. Dado que $\lambda > 0$, la matriz A es definida positiva y tridiagonal. Para resolver este sistema, se puede emplear la factorización LU de Crout para sistemas lineales tridiagonales. En este algoritmo suponemos, para propósitos de detención o paro, que se da una cota para t.

2.3.3. Método de Crank-Nicholson

Este método es un método implícito. Con el método de diferencias regresivas se solucionó el problema de estabilidad. Sin embargo, para evitar la falta de precisión, generada por el error de truncamiento, se requiere que los intervalos de tiempo sean mucho mas pequeños que los de espacio ($h >> k$), y esto reduce su eficiencia. Por tanto, se hace necesario un método que permita tomar valores similares para h y k, y que además sea estable para todo λ. Este método se puede obtener al promediar el método de diferencias progresivas en el n-ésimo paso en t,

$$\frac{U_j^{n+1} - U_j^n}{k} = \alpha^2 \frac{U_{j+1}^n - 2U_j^n + U_{j-1}^n}{h^2},$$

y el método de diferencias regresivas en el $(n+1)$-ésimo paso en t

$$\frac{U_j^{n+1} - U_j^n}{k} = \alpha^2 \frac{U_{j+1}^{n+1} - 2U_j^{n+1} + U_{j-1}^{n+1}}{h^2}.$$

Entonces se tiene que una aproximación por diferencias para la ecuación de calor como

$$\frac{U_j^{n+1} - U_j^n}{k} = \frac{\alpha^2}{2}\left[\frac{U_{j+1}^n - 2U_j^n + U_{j-1}^n}{h^2} + \frac{U_{j+1}^{n+1} - 2U_j^{n+1} + U_{j-1}^{n+1}}{h^2}\right], \tag{2.29}$$

que se puede reescribir como

$$-\frac{\lambda}{2}U_{j-1}^{n+1} + (1+\lambda)U_j^{n+1} - \frac{\lambda}{2}U_{j+1}^{n+1} = \frac{\lambda}{2}U_{j-1}^n + (1-\lambda)U_j^n + \frac{\lambda}{2}U_{j+1}^n. \tag{2.30}$$

Observamos el grafo del método (2.30) en la figura 2.4.

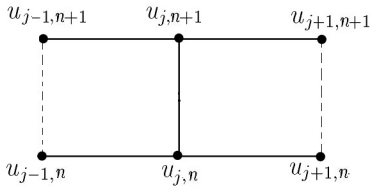

Figura 2.4: Esquema del Método Crank-Nicholson.

esta representado de forma matricial como

$$\mathrm{A}U^{n+1} = \mathrm{B}U^n, \tag{2.31}$$

donde

$$\mathrm{A} = \begin{bmatrix} (1+\lambda) & \frac{-\lambda}{2} & 0 & \dots & \dots & 0 \\ \frac{-\lambda}{2} & (1+\lambda) & \frac{-\lambda}{2} & \ddots & & \vdots \\ 0 & \frac{-\lambda}{2} & (1+\lambda) & & \ddots & \vdots \\ \vdots & \ddots & & & & 0 \\ \vdots & & \ddots & \frac{-\lambda}{2} & (1+\lambda) & \frac{-\lambda}{2} \\ 0 & \dots & \dots & 0 & \frac{-\lambda}{2} & (1+\lambda) \end{bmatrix}$$

y

$$\mathrm{B} = \begin{bmatrix} (1-\lambda) & \frac{\lambda}{2} & 0 & \dots & \dots & 0 \\ \frac{\lambda}{2} & (1-\lambda) & \frac{\lambda}{2} & \ddots & & \vdots \\ 0 & \frac{\lambda}{2} & (1-\lambda) & & \ddots & \vdots \\ \vdots & \ddots & & & & 0 \\ \vdots & & \ddots & \frac{\lambda}{2} & (1-\lambda) & \frac{\lambda}{2} \\ 0 & \dots & \dots & 0 & \frac{\lambda}{2} & (1-\lambda) \end{bmatrix}$$

La precisión de esta método es de orden $\mathcal{O}(k^2)+\mathcal{O}(h^2)$ y es estable para cualquier λ. Usando el Teorema de Equivalencia de Lax concluimos que el método de Crank-Nicholson es convergente. Para resolver este sistema, se puede usar la factorización LU de Crout para sistemas lineales tridiagonales. Al igual que en el método de diferencias regresivas, se da una cota para t, para propósitos de detención o paro.

2.4. Ecuación de Convección No Lineal

La ecuación de convección no lineal es una ecuación en derivadas parciales de primer orden hiperbólica que habitualmente aparece en las aplicaciones que están relacionadas con leyes de conservación de cantidades físicas.

Comenzaremos con el problema de Cauchy

$$u_t + f(u)_x = 0, \quad x \in [a,b], 0 \le t \le T, \tag{2.32}$$
$$u(x,0) = u_0(x),$$

donde u recibe el nombre de variable conservativa y es una función vectorial de m componentes y $f : \mathbb{R}^m \to \mathbb{R}^m$ se conoce como función flujo de u, y escrito de esa forma se conoce como sistema de leyes de conservación de la variable vectorial u.

El sistema es no lineal por lo tanto su solución desarrolla discontinuidades mejor conocidas como ondas de choque, ver figura 2.5. La aparición de estas discontinuidades en las soluciones hace que su aproximación por métodos numéricos

comunes sea engorrosa y no satisfactoria. Por esta razón se han desarrollado esquemas numéricos para tratar de resolver este tipo de ecuaciones llamados métodos de alta resolución.

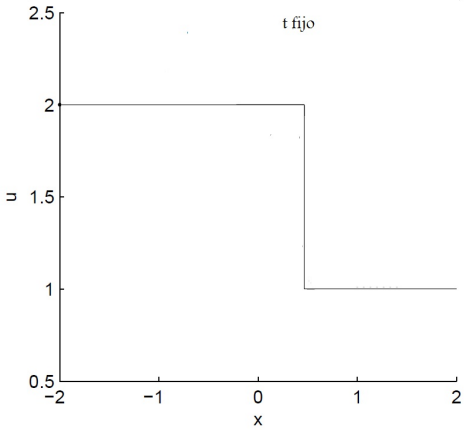

Figura 2.5: Ejemplo de onda de choque

Los métodos de alta resolución son esquemas numéricos especialmente diseñados para resolver leyes de conservación de tipo hiperbólico. Los métodos de alta resolución son esquemas que unifican en un solo algoritmo los métodos de primer y segundo orden, de tal manera que los de primer orden sean los que aproximen mejor las soluciones de leyes de conservación sin oscilaciones alrededor de las discontinuidades, mientras que en las regiones suaves la solución sea aproximada por los métodos de segundo orden.

Formalmente los métodos de alta resolución son esquemas conservativos de la forma :

$$U_j^{n+1} = U_j^n - \frac{k}{h}(F(U_j^n, U_{j+1}^n) - F(U_{j-1}^n, U_j^n)),$$

donde F es llamado función flujo numérico y en este caso sólo depende de dos variables.

Para establecer la convergencia de los métodos de alta resolución es necesario introducir la noción de esquemas TVD (Total Variation Diminishing).

2.4.1. Estabilidad y Convergencia No Lineal

Desafortunadamente el Teorema de Equivalencia de Lax-Richtmyer teorema (2.2) solo puede ser aplicado a problemas lineales. En esta sección introducimos conceptos de estabilidad no lineal que nos pemitirán garantizar la convergencia de esquemas numéricos para problemas no lineales.

El modo mas simple de medir cuan oscilatoria es una función o una aproximación numérica es analizando su variación total.

Definición 2.7. [13]. Definimos la **variación total** de U^n como

$$TV(U^n) := \sum_{i=-\infty}^{\infty} \mid U_{i+1}^n - U_i^n \mid .$$

Sea $\mathcal{K}$ el siguiente conjunto compacto

$$\mathcal{K} = \{U^n \in \mathcal{L}_1 : TV(U^n) \leq R \text{ y } \mathrm{Sop}^2(u(\cdot,t) \subset [-M,M] \ \ \forall t \in [0,T]\}. \tag{2.33}$$

Definición 2.8. [21]. Un método numérico es **variación total estable** o simplemente **TV-estable**, si todas las aproximaciones U^n para $k < k_0$ están en algún conjunto de la forma (2.33) (donde R y M pueden depender de los datos iniciales u_0 y de la función flujo $f(u)$ pero no de k).

Los siguientes teoremas nos garantizan la TV-estabilidad y la convergencia de esquemas numéricos.

Teorema 2.3. *Considere un método conservativo con flujo numérico continuo Lipschitz $F(U;j)$ y supóngase que para cada condición inicial u_0 existe algún $k_0, R > 0$ tales que*

$$TV(U^n) \leq R \quad \forall n,k \ \ con \ \ k < k_0, nk \leq T, \tag{2.34}$$

entonces el método es TV-estable.

Demostración.Véase [21]. ∎

Teorema 2.4. *Supóngase que U^n es generada por un método conservativo con flujo numérico continuo Lipschitz y es consistente con alguna ley escalar de conservación. Si el método es TV-estable, es decir, si $TV(U^n)$ es uniformemente acotada para todo n,k con $k < k_0, nk \leq T$, entonces el método es convergente.*

Demostración.Véase [21]. ∎

[2]Supongase que $\mathcal{X}$ es un espacio topologico y $f : \mathcal{X} \to \mathbb{C}$ una función. Es soporte de f es el conjunto $sop(f) = \overline{\{x \in \mathcal{X} | f(x) = 0\}}$, es decir, $Sop(f) \subset [-M,M]$ significa que $f(x) \equiv 0$ para $|x| > M$.

Variación Total Decreciente

Harten introdujo el siguiente importante concepto:

Definición 2.9. *[13]. Un sistema numérico es* ***Variación Total Decreciente*** *(TVD) si*

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

para todo n.

Harten propuso un criterio simple para determinar si un esquema numérico cumple la propiedad TVD, en el teorema conocido por su nombre.

Teorema 2.5. ***(Harten)*** *Si un esquema numérico explícito de diferencias finitas es de la forma:*

$$U_i^{n+1} = U_i^n - C_{i-1}^n(U_i^n - U_{i-1}^n) + D_i^n(U_{i+1}^n - U_i^n), \tag{2.35}$$

donde los coeficientes C_{i-1}^n y D_i^n son valores arbitrarios (que pueden depender de U^n de un modo no lineal), entonces es

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

si se satisfacen las siguientes condiciones sobre los coeficientes:

$$\begin{aligned} C_i^n &\geq 0, \\ D_i^n &\geq 0, \\ 0 \leq C_i^n + D_i^n &\leq 1. \end{aligned}$$

Para la demostración del teorema ver **Apéndice C**.

Los esquemas TVD o de alta resolución consiguen evitar oscilaciones espurias en el contorno de las discontinuidades manteniendo el segundo orden en el dominio.

El estudio de esquemas numéricos en diferencias finitas clásicas para la solución de sistemas hiperbólicos en forma de ley de conservación, son los métodos conocidos como los del tipo Lax-Wendroff y los del tipo Upwind. Los métodos numéricos tipo Lax-Wendroff constituyen un grupo de métodos centrados o simétricos, de segundo orden en espacio. Para problemas no lineales existen algunas variantes del clásico esquema Lax-Wendroff, esquemas del tipo predictor-corrector, tales como los métodos Richtmyer o Lax-Wendroff de dos pasos y el de McCormack.

Para la solución numérica de la ecuación de convección no lineal describiremos dos métodos explícitos clásicos y uno de alta resolución combinación de ambos. En primer lugar, un esquema Upwind de primer orden espacial y temporal, en segundo

lugar un esquema Lax Wendroff de dos pasos y por último un esquema con corrección TVD que le confiere propiedades no oscilatorias.

Para el método Lax-Wendroff se presenta su deducción, error de truncamiento local y consistencia, para los métodos Upwind y TVD Flux-Limiter error de truncamiento local, consistencia, y TV-estabilidad.

2.4.2. Upwind

Para poder incorporar en el esquema numérico las propiedades del fenómeno físico se desarrollaron los esquemas upwind (contraviento o contraflujo), esquemas descentrados que utilizan derivadas espaciales hacia delante ó hacia atrás dependiendo del sentido de propagación de la onda, es decir, decide cual es la mejor opción entre Euler Atrasadas y Euler Adelantadas.

El esquema conservativo *Upwind* más simple para ecuaciones hiperbólicas no lineales con soluciones de ondas de choque está dado por

$$U_j^{n+1} = U_j^n - \frac{k}{h}\Big[\lambda^+(f(U_j^n) - f(U_{j-1}^n)) - \lambda^-(f(U_{j+1}^n) - f(U_j^n))\Big], \tag{2.36}$$

con

$$\lambda^+ = max\left(\frac{f_u(U_j^n)}{|f_u(U_j^n)|}, 0\right) \qquad \lambda^- = min\left(\frac{f_u(U_j^n)}{|f_u(U_j^n)|}, 0\right)$$

donde f_u denota la derivada parcial de f respecto de u.

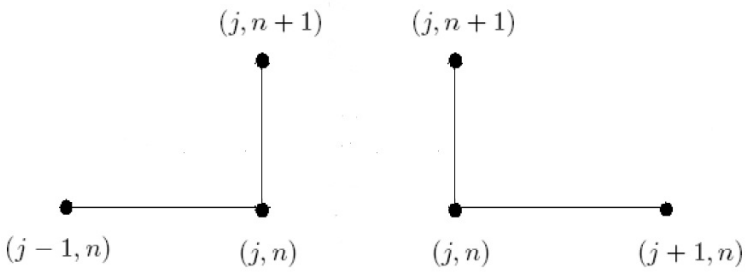

Figura 2.6: Esquema del Método Upwind

Proposición 2.1. *El esquema Upwind (2.36) es* ***consistente****, con error de truncamiento de primer orden en el espacio y el tiempo [21].*

Nota 2.1. *El esquema Upwind (2.36) es* ***TV-estable****, véase [21].*

Nota 2.2. *Por los teoremas (2.3) y (2.4) se tiene que el esquema Upwind (2.36) es* ***convergente****, véase [21].*

Este esquema Upwind, basado en la dirección de los flujos, previene la aparición de falsas oscilaciones numéricas, pero contiene una excesiva disipación numérica en todos los puntos del mallado puesto, que es un esquema de primer orden y por consiguiente poco preciso incluso donde la solución es suave.

2.4.3. Richtmyer Two-Step Lax-Wendroff

La principal característica de este método radica en la combinación de discretizaciones centradas de tiempo y espacio. El esquema de Lax-Wendroff, es el esquema más sencillo explícito de precisión de segundo orden, pero tiene deficiencias, tales como la generación de oscilaciones cerca de las discontinuidades y requiere la evaluación de las matrices Jacobianas que puede ser una operación costosa en la práctica. Por lo tanto Richtmyer y Morton [25] desarrollaron un procedimiento en dos etapas conocidas como *Richtmyer two-step Lax-Wendroff*. Este esquema evita la estimación de matrices Jacobianas, y es capaz de manejar la no linealidad de una manera directa.

El método *Richtmyer Two-Step Lax-Wendroff* esta dado por

Paso predictor

$$U_{j+1/2}^{n+1/2} = \frac{1}{2}\left[(U_j^n + U_{j+1}^n) - \frac{k}{h}\left(f(U_{j+1}^n) + f(U_j^n)\right)\right], \tag{2.37}$$

$$U_{j-1/2}^{n+1/2} = \frac{1}{2}\left[(U_j^n + U_{j-1}^n) - \frac{k}{h}\left(f(U_j^n) + f(U_{j-1}^n)\right)\right]. \tag{2.38}$$

Paso corrector

$$U_j^{n+1} = U_j^n - \frac{k}{h}\left[f(U_{j+1/2}^{n+1/2}) - f(U_{j-1/2}^{n+1/2})\right]. \tag{2.39}$$

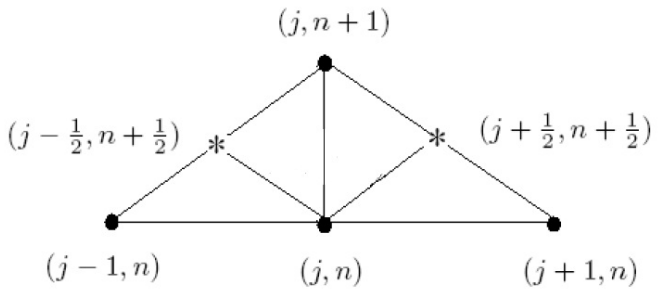

Figura 2.6: Esquema del Método Richtmyer-Lax-Wendroff.

Deducción del método

La deducción de este método es por medio de desarrollos de Taylor, primero calculamos aproximaciones en el centro de los rectángulos de la malla formada por la discretización y finalmente usamos estas aproximaciones para calcular la solución en el punto deseado.

Deducción del Paso Predictor: desarrollando en series de Taylor se tiene (denotando $u = u(x_j, t_n)$)

$$u\left(x_j + \frac{h}{2}, t_n + \frac{k}{2}\right) = u + \frac{h}{2}u_x + \frac{k}{2}u_t + \mathcal{O}(h^2, k^2).$$

Pero por (2.32) tenemos que

$$u_t = -f(u)_x,$$

entonces

$$u\left(x_j + \frac{h}{2}, t_n + \frac{k}{2}\right) = u + \frac{h}{2}u_x - \frac{k}{2}f(u)_x + \mathcal{O}(h^2, k^2), \tag{2.40}$$

además tenemos que

$$u_x = \frac{u(x_j + h, t_n) - u(x_j, t_n)}{h} + \mathcal{O}(h), \tag{2.41}$$

$$f(u)_x = \frac{f(u)(x_j + h, t_n) - f(u)(x_j, t_n)}{h} + \mathcal{O}(h), \tag{2.42}$$

sustituyendo (2.41) y (2.42) en (2.40) obtenemos

$$\begin{aligned} u\left(x_j + \frac{h}{2}, t_n + \frac{k}{2}\right) &= u(x_j, t_n) + \frac{h}{2}\left[\frac{1}{h}(u(x_j + h, t_n) - u(x_j, t_n)) + \mathcal{O}(h)\right] \\ &\quad -\frac{k}{2}\left[\frac{1}{h}(f(u)(x_j + h, t_n) - f(u)(x_j, t_n)) + \mathcal{O}(h)\right] \\ &\quad +\mathcal{O}(h^2, k^2), \end{aligned}$$

entonces

$$\begin{aligned} u\left(x_j + \frac{h}{2}, t_n + \frac{k}{2}\right) &= u(x_j, t_n) + \frac{1}{2}[u(x_j + h, t_n) - u(x_j, t_n) + \mathcal{O}(h^2)] \\ &\quad -\frac{k}{2h}[f(u)(x_j + h, t_n) - f(u)(x_j, t_n) + \mathcal{O}(h^2)] \\ &\quad +O(h^2, k^2). \end{aligned}$$

Por lo tanto el método para calcular la aproximación en el punto central del rectángulo esta dado por

$$U_{j+1/2}^{n+1/2} = U_j^n + \frac{1}{2}(U_{j+1}^n - u_j^n) - \frac{k}{2h}(f(U_{j+1}^n) - f(U_j^n)),$$

es decir

$$U_{j+1/2}^{n+1/2} = \frac{1}{2}(U_{j+1}^n + U_j^n) - \frac{k}{2h}(f(U_{j+1}^n) - f(U_j^n)),$$

y análogamente obtenemos

$$U_{j-1/2}^{n+1/2} = \frac{1}{2}(U_j^n + U_{j-1}^n) - \frac{k}{2h}(f(U_j^n) - f(U_{j-1}^n)).$$

Deducción del Paso Corrector usamos el desarrollo de Taylor y denotando $u = u(x_j, t_n)$ se obtiene

$$\begin{aligned} u(x_j, t_n + k) &= u + ku_t + \mathcal{O}(k^2) \\ &= u - kf(u)_x + \mathcal{O}(k^2), \end{aligned}$$

entonces para aproximar la derivada $f(u(x_j, t_n))_x$ usamos

$$\begin{aligned} f(u)\left(x_j + \frac{h}{2}, t_n + \frac{k}{2}\right) &= f(u) + \frac{h}{2}f(u)_x + \frac{k}{2}f(u)_t + \mathcal{O}(h^2, k^2), \\ f(u)\left(x_j - \frac{h}{2}, t_n + \frac{k}{2}\right) &= f(u) - \frac{h}{2}f(u)_x + \frac{k}{2}f(u)_t + \mathcal{O}(h^2, k^2), \end{aligned}$$

luego

$$f(u)_x = \frac{1}{h}\left(f(u)\left(x_j + h2, t_n + k2\right) - f(u)\left(x_j - \frac{h}{2}, t_n + \frac{k}{2}\right) + \mathcal{O}(h^2, k^2)\right),$$

así obtenemos

$$\begin{aligned} u(x_j, t_n + k) &= u(x_j, t_n) - \frac{k}{h}\left[f(u)\left(x_j + \frac{h}{2}, t_n + \frac{k}{2}\right) - f(u)\left(x_j - \frac{h}{2}, t_n + \frac{k}{2}\right)\right] \\ &\quad + \frac{k}{h}\mathcal{O}(h^2, k^2) + \mathcal{O}(k^2). \end{aligned}$$

Por lo tanto la aproximación de $u(x_j, t_{n+1})$ esta dada por

$$U_j^{n+1} = U_j^n - \frac{k}{h}\left(f\left(U_{j+1/2}^{n+1/2}\right) - f\left(U_{j-1/2}^{n+1/2}\right)\right). \tag{2.43}$$

Proposición 2.2. *El método Richtmyer two-step Lax-Wendroff es* ***consistente****, de primer orden en el paso predictor y de segundo orden en el paso corrector tanto espacial como temporalmente [34].*

2.4.4. TVD Flux-Limiter

En este sección utilizaremos el método numérico *TVD Flux-Limiter* [15] que incorpora al método Richtmyer two-step Lax-Wendroff (R2LW) un esquema conservativo Upwind en el paso predictor y una función no convencional flux-limiter en el paso corrector.

Una importante característica de este método radica en la medición de la suavidad de una función a través de diferencias finitas no estándar, el cual asigna θ para determinar la mejor aproximación

$$f(u)_x(x,t) \approx \frac{1}{h}\left[\theta\Big(f(u(x+h,t))-f(u(x,t))\Big)+(1-\theta)\Big(f(u(x,h))-f(u(x-h,t))\Big)\right], \tag{2.44}$$

donde

$$\theta = \frac{|f(u(x,t)) - f(u(x-h,t))|}{|f(u(x+h,t)) - f(u(x,t))| + |f(u(x,h)) - f(u(x-h,t))|}. \tag{2.45}$$

Notese que el valor de θ siempre se encuentra en el intervalo [0,1].

Si la distancia $|f(u(x,t))-f(u(x-h,t))|$ es mayor que $|f(u(x+h,t))-f(u(x,t))|$ entonces $\theta > 1/2$ y consecuentemente el método no estándar proporciona un mayor peso a las aproximaciones adelantadas. Si la distancia $|f(u(x,t)) - f(u(x-h,t))|$ es menor que $|f(u(x+h,t)) - f(u(x,t))|$ entonces $\theta < 1/2$, entonces el método no estándar proporciona un mayor peso a las aproximaciones atrasadas. Si las distancias $|f(u(x+h,t)) - f(u(x,t))|$ y $|f(u(x,t)) - f(u(x-h,t))|$ son muy parecidas el valor de θ será muy cercano a $1/2$ y entonces nuestro método adquiere la forma de diferencias centradas.

Resumiendo, el método TVD flux-limiter es dado por:
Paso predictor

$$U_{j+\frac{1}{2}}^{n+1/2} = \frac{1}{2}(U_j^n + U_{j+1}^n) - \frac{k}{2h}(f(U_{j+1}^n) + f(U_j^n)), \tag{2.46}$$

$$U_{j-\frac{1}{2}}^{n+1/2} = \frac{1}{2}(U_j^n + U_{j-1}^n) - \frac{k}{2h}(f(U_j^n) + f(U_{j-1}^n)), \tag{2.47}$$

$$U_j^{n+1/2} = U_j^n - \frac{k}{2h}\left[\lambda^+(f(U_j^n) - f(U_{j-1}^n)) - \lambda^-(f(U_{j+1}^n) - f(U_j^n))\right]. \tag{2.48}$$

Paso corrector

$$\begin{aligned} U_j^{n+1} &= U_j^{n+1/2} - \frac{k}{h}\left[\phi(\theta_j^{n+1/2})\Big(f(U_{j+1/2}^{n+1/2}) - f(U_j^{n+1/2})\Big)\right. \\ &\quad \left. +\phi(\theta_{j-1/2}^{n+1/2})\Big(f(U_j^{n+1/2}) - f(U_{j-1/2}^{n+1/2})\Big)\right], \end{aligned} \tag{2.49}$$

donde

$$\theta_j^{n+1/2} = \frac{|f(U_j^{n+1/2}) - f(U_{j-1/2}^{n+1/2})|}{|f(U_{j+1/2}^{n+1/2}) - f(U_j^{n+1/2})| + |f(U_j^{n+1/2}) - f(U_{j-1/2}^{n+1/2})|}, \tag{2.50}$$

$$\theta_{j-1/2}^{n+1/2} = 1 - \theta_j^{n+1/2}, \tag{2.51}$$

y ϕ es la función flux-limiter definida en la siguiente proposición.

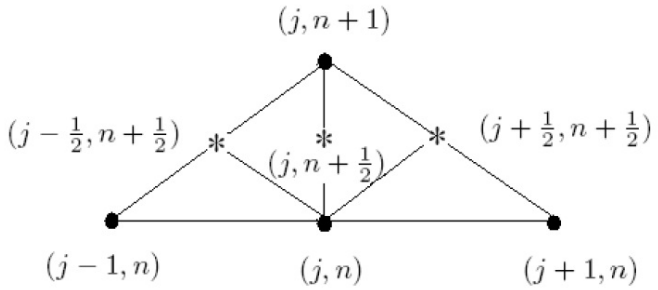

Figura 2.8: Esquema del Método TVD flux-limiter

Proposición 2.3. *El esquema TVD flux-limiter (2.47)-(2.50) es* ***consistente*** *de primer orden temporalmente como espacialmente en el paso predictor y de primer orden de precisión en el tiempo y segundo o primer orden de precisión en el espacio dependiendo del valor de ϕ en el paso corrector [15].*

La siguiente proposición nos garantiza la TVD-estabilidad.

Proposición 2.4. *El esquema (2.47)-(2.50) para (2.32) es TVD-estable si la condición CFL se satisface*

$$|c_j^n| \leq 1 \quad \forall j, n.$$

donde $c_j^n = \frac{k}{h} f_u(U_j^n)$ y si $|c_j^n| \leq \frac{1}{2}$ la función flux-limiter es dada por

$$\phi(\theta_j^{n+1/2}) = \begin{cases} 0 & si\ b_j^n \leq 0, \\ (\theta_j^{n+1/2})^\alpha & si\ 0 < b_j^n < 5, \\ \theta_j^{n+1/2} & si\ b_j^n \geq 5, \end{cases}$$

y si $|c_j^n| > \frac{1}{2}$ entonces

$$\phi(\theta_j^{n+1/2}) = \begin{cases} 0 & si\ b_j^n \leq 0, \\ \theta_j^{n+1/2} & si\ 0 < b_j^n, \end{cases}$$

donde el parámetro de flujo local es determinado por

$$b_j^n = \begin{cases} \dfrac{U_{j+1}^n - U_j^n}{U_j^n - U_{j-1}^n} & si\ c_j^n \geq 0, \\ \dfrac{U_j^n - U_{j-1}^n}{U_{j+1}^n - U_j^n} & si\ c_j^n < 0, \end{cases}$$

$\theta_j^{n+1/2}$ *es el parámetro definido por (2.45) y* $\alpha \in [0{,}7, 1]$ *[15].*

En [15] se muestra la efectividad de este método TVD Flux-Limiter para resolver numéricamente ecuaciones hiperbólicas conservativas no lineales con funciones de flujo convexo y no convexo.

No se presentan gráficamente resultados de estos esquemas numéricos pero si el lector esta interesado en verlos, véase:

- Método *Euler Explícito* [2].
- Método *Euler Implícito* [2].
- Método *Crank-Nicholson* [2].
- Método *Upwind* [21].
- Método *Richtmyer Two-Step Lax-Wendroff* [34].
- Método *TVD Flux-Limiter* [15].

CAPÍTULO 3

Solución Numérica de la Ecuación Convección-Difusión Mediante un Método Splitting Tipo Strang

En este capítulo se presenta un método para resolver la ecuación convección-difusión no lineal que describe el flujo de un fluido viscoso. El método consiste en la aplicación de la descomposición de operadores a la discretización temporal del problema original. Con esto se obtienen subproblemas más sencillos que son resueltos mediante técnicas iterativas, entonces se puede reconstruir una aproximación para el problema completo. Este esquema de descomposición de operadores es el esquema clásico de descomposición de operadores simetrizado Strang.

3.1. Método Splitting

La complejidad de los modelos matemáticos para analizar problemas que se plantean de un determinado fenómeno aumenta cada vez más al tratar de describir adecuadamente la realidad. Frecuentemente, en dichos problemas podemos encontrar varios sistemas de naturaleza distinta acoplados: Ecuaciones en Derivadas Parciales (EDP) de diferentes tipos, es decir, modelos heterogéneos o de carácter multi-físico.

Es por eso que, a menudo, el análisis de los modelos no se puede realizar mediante el uso de uno de los métodos matemáticos o numéricos específicos de determinado tipo de sistemas, sino que es preciso combinar varios de ellos. Resulta por tanto natural desarrollar y utilizar métodos de descomposición que permitan dividir el sistema en subsistemas más simples con características precisas e identificadas en los que podamos aplicar un método específico.

Estos métodos recibieron diversos nombres como métodos de factorización, de barrido, de descomposición (Splitting) y de Direcciones Alternadas. Los primeros métodos Splitting aparecen a mediados de los 50 en los trabajos de Douglas, Peaceman, Ratchford, Yanenko y Glowinski (ver [12], [23]).

El método Splitting resuelve de manera iterada los diversos subsistemas en los que el sistema global se ha descompuesto. Este método iterativo permite aplicar en cada subsistema un método específico pero al mismo tiempo recuperar las propiedades globales del sistema.

Como un primer paso para entender el método Splitting, consideremos el sistema de EDO

$$\begin{aligned} \dot{x}(t) &= (A+B)x(t), \\ x(0) &= x_0, \end{aligned} \tag{3.1}$$

donde $x = x(t)$ es una incógnita vectorial en $\mathbb{R}^N$ dependiente del parámetro temporal $t \in \mathbb{R}$, A y B son matrices cuadradas $N \times N$ con coeficientes constantes independientes de t.

Para cada dato inicial $x_0 \in \mathbb{R}^N$ el sistema (3.1) admite una única solución global $x_0 \in C^n(\mathbb{R}, \mathbb{R}^N)$, donde mediante C^n denotamos la clase de funciones analíticas. La solución viene dada por la fórmula de representación:

$$x(t) = e^{(A+B)t}x_0. \tag{3.2}$$

Ahora consideremos el siguiente resultado

Teorema 3.1. ***(Lie)****Dadas dos matrices cuadradas $N \times N$ A y B se tiene*

$$e^{(A+B)} = \lim_{n\to\infty} \left(e^{\frac{A}{n}} e^{\frac{B}{n}} \right)^n \tag{3.3}$$

Demostración: Véase **Apéndice D**.

El Teorema nos dice lo siguiente: en virtud de (3.3) se tiene que

$$e^{(A+B)} \sim \left(e^{\frac{A}{n}} e^{\frac{B}{n}} \right)^n,$$

además, para n fijo

$$\left(e^{\frac{A}{n}} e^{\frac{B}{n}} \right)^n = e^{\frac{A}{n}} e^{\frac{B}{n}} \cdots e^{\frac{A}{n}} e^{\frac{B}{n}},$$

es decir se trata de un producto iterado, n veces, del operador $e^{A/n}e^{B/n}$.

Vemos que al aplicar $e^{A/n}e^{B/n}$ a un elemento $x_0 \in \mathbb{R}^N$ lo que se obtiene es

$$\left[e^{A/n}e^{B/n}\right]x_0 = e^{A/n}\left[e^{B/n}x_0\right].$$

Por otra parte $e^{B/n}x_0 = y_0$ es el valor en el instante $t = 1/n$ de la solución de

$$\dot{y} = By; \quad y(0) = x_0,$$

mientras que $e^{B/n}y_0 = z_0$ es el valor en el instante $t = 1/n$ de la solución de

$$\dot{z} = Az; \quad z(0) = y_0.$$

Al aplicar el Teorema 3.1 a la solución de (3.1) se obtiene

$$x(t) = e^{(A+B)t}x_0 = \lim_{n\to\infty}\left(e^{\frac{At}{n}}e^{\frac{Bt}{n}}\right)^n x_0,$$

lo cual significa que $x(t)$ se aproxima mediante la expresión

$$x_n(t) = \left(e^{\frac{At}{n}}e^{\frac{Bt}{n}}\right)^n x_0,$$

la cual se encuentra del modo siguiente:

- Se itera n veces un procedimiento en el que, arrancando del valor del paso anterior, se avanza un paso temporal del tamaño t/n en la resolución del sistema (3.1) .
- En cada paso lo que se hace es resolver de manera consecutiva cada uno de los dos sistemas involucrados en (3.1)

$$x' = Ax, \tag{3.4}$$

y

$$x' = Bx. \tag{3.5}$$

Conviene observar que, en ningún caso se resuelve el sistema completo (3.1) sino que siempre se resuelvan los subsistemas (3.4) y(3.5). En cada intervalo temporal de longitud t/n se resuelven ambos sistemas (3.4) y(3.5), lo cual indica que se recorre el intervalo temporal en dos ocasiones, una por cada subsistema.

Estas características del método iterativo hacen que se denomine método Splitting. Es importante señalar, y esta es una de las virtudes del método Splitting, que cada uno de los sistemas (3.4) y(3.5) pueden resolverse mediante métodos distintos, de manera que se pueden utilizar métodos mejor adaptados a las características de cada subsistema.

En el caso de problemas no lineales, se trabaja con su sistema discreto asociado, linealizando y aplicando el teorema de Lie en cada subintervalo del dominio discreto.

3.2. Ecuación de Convección Difusión

En esta sección se analiza el método Splitting explícito para la solución de la ecuación de convección-difusión no lineal

$$u_t - \nu u_{xx} + f(u)_x = 0, \quad x \in [a,b], 0 \leq t \leq T. \tag{3.6}$$

En vista de la propia estructura de la ecuación (3.6) es obvio que en ella subyacen dos modelos. Por una parte la ecuación del calor

$$u_t - \nu u_{xx} = 0, \tag{3.7}$$

y por otra, la ecuación de convección no lineal

$$u_t + f(u)_x = 0. \tag{3.8}$$

En este tipo de situaciones, en los que el modelo en consideración incorpora dos subsistemas bien reconocibles, es natural utilizar métodos de descomposición o splitting que permiten obtener la solución del sistema global a partir de la solución de cada subsistema y por otra parte aplicar a cada uno de los subsistemas un método numérico específico.

El Teorema de Lie demuestra que para dos matrices $n \times n$ A_1 y A_2 cualesquiera se tiene

$$e^{A_1+A_2} = \lim_{j\to\infty} (e^{\frac{A_1}{j}} e^{\frac{A_2}{j}})^j,$$

con independencia de que A_1 y A_2 conmuten o no. Resulta de utilidad a la hora de resolver la ecuación diferencial

$$x' = (A_1 + A_2)x, \quad t \in \mathbb{R} \tag{3.9}$$

$$x(0) = x_0,$$

puesto que su solución es de la forma

$$x(t) = e^{(A_1+A_2)t} x_0. \tag{3.10}$$

La ecuación convección-difusión no lineal (3.6) puede también entenderse como un problema de la forma (3.9), si bien en este caso la incognita $u = u(x,t)$ es, para cada $t > 0$, una función que depende de x y que por tanto pertenece a un espacio de dimension infinita, A_1 es el operador diferencial $A_1 u = \partial_x^2 u$ y A_2 es el operador no-lineal $A_2 u = -(\frac{u^2}{2})_x$.

3.3. Splitting Strang

Los sistemas de ecuaciones de convección-difusión están dominados por convección, que es el caso más complejo desde el punto de vista numérico: si bien la solución de (3.6) suele ser suave para $t > 0$, sus gradientes pueden ser muy grandes y una capa de solución de ondas de choque viscosas puede estar fuera del alcance práctico. Por lo tanto, la aplicación del método de captura de ondas de choques, desarrollado originalmente para los sistemas de leyes de conservación hiperbólica puede ser ventajoso. Al mismo tiempo, aun cuando el impacto de la difusión no es demasiado significativa, su presencia suele reducir la eficiencia de los esquemas numéricos explícitos.

Una forma de superar esta dificultad es utilizar un algoritmo de descomposición de operadores, que puede describirse brevemente como sigue. Denotando por A_1 con el operador de la solución exacta asociado con el sistema (lineal) parabólico (3.7) y por A_2 el operador de la solución exacta asociado con el sistema hiperbólico (3.8). Entonces, la introducción de un paso de tiempo Δt, la solución original del sistema convección-difusión (que se supone que estará disponible en el momento t) se desarrolla en el tiempo en tres subetapas o subpasos:

$$u(x, t + \Delta t) = A_1 \left(\frac{\Delta t}{2} \right) A_2(\Delta t) A_1 \left(\frac{\Delta t}{2} \right) u(x, t). \tag{3.11}$$

La idea es resolver el primer subproblema $u_t = A_1 u$ en sólo un paso de tiempo de longitud $\Delta t/2$, luego utilizar el resultado como datos para un paso de tiempo en el segundo subproblema $u_t = A_2 u$ y finalmente dar medio paso de tiempo en $u_t = A_1 u$.

Este algoritmo de la división en tres etapas, es llamado ***Splitting Strang*** (ver [30]) y es conocido (por razones obvias) como un método splitting simétrico.

Proposición 3.1. *El método Splitting Strang (3.11) es un esquema de segundo orden y estable [30].*

Al analizar el Splitting Strang, vemos que ahora se esta aproximando la solución $e^{\Delta t(A_1+A_2)}$ por $e^{\frac{\Delta t}{2}A_1} e^{\Delta t A_2} e^{\frac{\Delta t}{2}A_1}$ en (3.10).

Por lo tanto el método de Splitting Strang queda de la siguiente manera

$$\frac{\partial u}{\partial t} + \mathcal{A}_1 u = 0 \quad \text{en} \quad (t^n, t^{n+1/2}) \qquad u(t^n) = u^n \rightarrow \tilde{u}^{n+1/2} = u(t^{n+1/2}) \tag{3.12}$$

$$\frac{\partial u}{\partial t} + \mathcal{A}_2 u = 0 \quad \text{en} \quad (t^n, t^{n+1}) \qquad u(t^n) = \tilde{u}^{n+1/2} \rightarrow u^{n+1/2} = u(t^{n+1}) \tag{3.13}$$

$$\frac{\partial u}{\partial t} + \mathcal{A}_1 u = 0 \quad \text{en} \quad (t^{n+1/2}, t^{n+1}) \qquad u(t^{n+1/2}) = u^{n+1/2} \rightarrow u^{n+1} = u(t^{n+1}) \tag{3.14}$$

El método Splitting Strang es un esquema bastante robusto y relativamente fácil de programar siguiendo el procedimiento descrito arriba. Gráficamente se puede visualizar como esta trabajando el método Splitting Strang simétrico

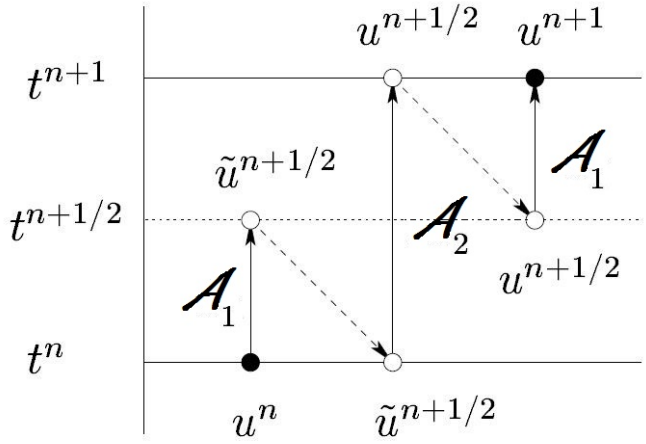

Figura 3.2: Splitting Strang Simétrico

Solución de los Subproblemas

En la práctica, la solución de los operadores A_1 y A_2 son reemplazados por sus aproximaciones numéricas. Teniendo en cuenta que la principal ventaja de la técnica de descomposición de operadores es el hecho de que el hiperbólico (3.8) y el parabólico (3.7) subproblemas de diferente naturaleza, se pueden resolver numéricamente por diferentes métodos. Por lo que, para resolver (3.12) y (3.14) (problemas de difusión) usamos los esquemas (2.22), (2.28) y (2.30). Para resolver el (3.13) (problema de advección) usamos el método (2.47)-(2.50) un método eficiente en problemas convectivos no lineales [15].

Hemos presentado un esquema numérico para fluidos viscosos, el cual descompone el problema en una sucesión de subproblemas más simples en cada nivel de tiempo, a través de la discretización temporal. La solución de dichos subproblemas, se lleva a cabo mediante métodos iterativos que, en cada iteración, requieren de la solución de un problema parabólico lineal y un problema hiperbólico no lineal.

Los resultados numéricos del método Splitting Strang para la ecuación convección-difusión no lineal se presentan en el **capítulo 5**.

CAPÍTULO 4

Solución Numérica de la Ecuación Convección-Difusión Mediante un Método Viscous Flux Limiter

Los métodos Viscous flux-Limiter (VFL) se usan en problemas de flujos viscosos. Los métodos VFL reducen la difusión numérica en las discontinuidades y combinan la viscosidad física, la cual es maximizada, con viscosidad numérica, para capturar y dejar libre de oscilaciones las soluciones de la ecuación de convección-difusión.

En este capítulo, se presenta la extensión del método TVD flux-Limiter (2.47)-(2.50) para la ecuación convección-difusión no lineal. Una vez obtenido el método, al cual llamamos Viscous TVD Flux-Limiter, se realiza un análisis para determinar propiedades importantes como error de truncamiento, consistencia, estabilidad y convergencia.

Consideramos el problema conservativo no lineal de convección-difusión

$$u_t + f(u)_x = \nu u_{xx}, \qquad x \in [a,b], 0 \leq t \leq T, \tag{4.1}$$

$$u(x,0) = u_0(x),$$

donde ν es el coeficiente constante de viscosidad y f es la función de flujo.

Discretizamos la ecuación (4.1), utilizando el esquema conservativo explícito desarrollado en [15] para el término de convección y discretizamos el término difusivo de (4.1) utilizando las diferencias centradas de segundo orden. El nuevo método Viscous TVD Flux-limiter, tiene dos pasos, para el primer paso se deja igual que el paso predictor del método TVD Flux-Limiter [15], para el segundo paso, es decir,

para el paso corrector [15], agregamos la aproximación del termino difusivo o viscoso.

Resumiendo, el nuevo método Viscous TVD flux-limiter es
Paso Predictor

$$U_{j+\frac{1}{2}}^{n+1/2} = \frac{1}{2}(U_j^n + U_{j+1}^n) - \frac{k}{2h}(f(U_{j+1}^n) + f(U_j^n)), \tag{4.2}$$

$$U_{j-\frac{1}{2}}^{n+1/2} = \frac{1}{2}(U_j^n + U_{j-1}^n) - \frac{k}{2h}(f(U_j^n) + f(U_{j-1}^n)), \tag{4.3}$$

$$U_j^{n+1/2} = U_j^n - \frac{k}{2h}\left[\lambda^+(f(U_j^n) - f(U_{j-1}^n)) - \lambda^-(f(U_{j+1}^n) - f(U_j^n))\right]. \tag{4.4}$$

Paso corrector si $\mathcal{C} > 0$ entonces

$$U_j^{n+1} = \begin{cases} U_j^{n+1/2} - \frac{k}{h}\left[\phi(\theta_j^{n+1/2})\left(f(U_{j+1/2}^{n+1/2}) - f(U_j^{n+1/2})\right)\right. & \text{si } \phi(\theta_j^{n+1/2}) \geq 2/Re_c(1-|\mathcal{C}|), \\ \left.+\phi(\theta_{j-1/2}^{n+1/2})\left(f(U_j^{n+1/2}) - f(U_{j-1/2}^{n+1/2})\right)\right] & \\ +\frac{\nu k}{h^2}(U_{j+1}^n - 2U_j^n + U_{j-1}^n) & \\ U_j^n - \frac{k}{h}\left(f(U_{j+1/2}^{n+1/2}) - f(U_{j-1/2}^{n+1/2})\right) & \text{en otro caso,} \\ +\frac{\nu k}{h^2}(U_{j+1}^n - 2U_j^n + U_{j-1}^n) & \end{cases} \tag{4.5}$$

y si $\mathcal{C} < 0$ entonces

$$U_j^{n+1} = \begin{cases} U_j^{n+1/2} - \frac{k}{h}\left[\phi(\theta_j^{n+1/2})\left(f(U_{j+1/2}^{n+1/2}) - f(U_j^{n+1/2})\right)\right. & \text{si } \phi(\theta_{j-1/2}^{n+1/2}) \geq 2/Re_c(1-|\mathcal{C}|), \\ \left.+\phi(\theta_{j-1/2}^{n+1/2})\left(f(U_j^{n+1/2}) - f(U_{j-1/2}^{n+1/2})\right)\right] & \\ +\frac{\nu k}{h^2}(U_{j+1}^n - 2U_j^n + U_{j-1}^n) & \\ U_j^n - \frac{k}{h}\left(f(U_{j+1/2}^{n+1/2}) - f(U_{j-1/2}^{n+1/2})\right) & \text{en otro caso,} \\ +\frac{\nu k}{h^2}(U_{j+1}^n - 2U_j^n + U_{j-1}^n) & \end{cases} \tag{4.6}$$

donde

$$\theta_j^{n+1/2} = \frac{|f(U_j^{n+1/2}) - f(U_{j-1/2}^{n+1/2})|}{|f(U_{j+1/2}^{n+1/2}) - f(U_j^{n+1/2})| + |f(U_j^{n+1/2}) - f(U_{j-1/2}^{n+1/2})|}, \tag{4.7}$$

$$\theta_{j-1/2}^{n+1/2} = 1 - \theta_j^{n+1/2},$$

$$\mathcal{C} = \frac{k}{h} f_u(U_j^{n+1/2}),$$

y ϕ es la función flux-limiter definida en [15].

4.1. Viscosidad numérica

En la ecuación de convección-difusión no lineal al tener en consideración que está dominada por el término convectivo, se tiene que este término es fuente de dispersión numérica que produce en el fluido oscilaciones en la simulación, especialmente en presencia de cambios abruptos en las variables que surgen de discontinuidades en el campo de la solución del problema ya que la viscosidad física del problema no es suficiente para remover las oscilaciones que produce el termino convectivo dominante. Los métodos Viscous Flux-Limiter son una forma simple de compensar este comportamiento agregando disipación numérica o *viscosidad artificial* al esquema. La viscosidad artificial permite conseguir un esquema estable frente a una solución discontinua [5], [9]. Sin ella, los métodos producen oscilaciones espurias en el entorno de las discontinuidades, que pueden llegar a dar problemas de estabilidad del cálculo y discontinuidades no reales (fruto del proceso de cálculo) [1].

Los métodos Viscous Flux-Limiter son métodos que permiten detectar las zonas donde la viscosidad artificial es necesaria, es decir, añada viscosidad en aquellas zonas donde puedan desarrollarse inestabilidades , de forma que los choques son capturados de forma precisa, sin dañar la calidad de la solución en aquellas zonas en que el flujo es suave.

Toro en [33] encontró que los esquemas TVD no requieren viscosidad artificial si

$$Re_c \leq \frac{2}{1-|\mathcal{C}|}. \tag{4.8}$$

4.2. Error de truncamiento local y consistencia

Para el primer paso, el error de truncamiento esta definido por

$$\begin{aligned} \tau_{h,k} &= \frac{1}{k}\left[2u\left(x_j+\frac{h}{2},t_n+\frac{k}{2}\right)-u(x_j+h,t_n)-u(x_j,t_n)\right] \\ &\quad +\frac{1}{h}\left[f(u)(x_j+h,t_n)-f(u)(x_j,t_n)\right]. \end{aligned} \tag{4.9}$$

Tomando en cuenta los desarrollos de Taylor

$$\begin{aligned} u\left(x_j+\frac{h}{2},t_n+\frac{k}{2}\right) =& u+\frac{h}{2}u_x+\frac{k}{2}u_t+\frac{k^2}{8}u_{tt}+2\frac{kh}{8}u_{tx}+\frac{h^2}{8}u_{xx}+\mathcal{O}(h^3,k^3), \\ u(x_j+h,t_n) =& u+hu_x+\frac{h^2}{2}u_{xx}+\mathcal{O}(h^3), \\ f(u)(x_j+h,t_n) =& f(u)+hf(u)_x+\frac{h^2}{2}f(u)_{xx}+\mathcal{O}(h^3). \end{aligned}$$

Tenemos que el primer término de (4.9) es

$$\frac{1}{k}\left[2u\left(x_j+\frac{h}{2},t_n+\frac{k}{2}\right)-u(x_j+h,t_n)-u(x_j,t_n)\right]=u_t+\frac{k}{4}u_{tt}+\frac{h}{2}u_{xx}-\frac{h^2}{4k}u_{xx}+\mathcal{O}(h^2,k^2),$$

y el segundo término de (4.9) es

$$\frac{1}{h}\left[f(u)(x_j+h,t_n)-f(u)(x_j,t_n)\right]=f(u)_x+\frac{h}{2}f(u)_{xx}+\mathcal{O}(h^2),$$

de manera que

$$\tau_{h,k}=k\left(\frac{u_{tt}}{4}\right)+h\left(\frac{f(u)_{xx}}{2}+\frac{u_{tx}}{2}\right)+h^2\left(\frac{\nu u_{xx}}{h^2}+\frac{u_{xx}}{4k}\right)+\mathcal{O}(h^2,k^2),$$

y

$$\tau_{h,k}=\mathcal{O}(h)+\mathcal{O}(k). \tag{4.10}$$

Por lo tanto el primer paso es de *primer orden* tanto espacial como temporalmente.

Si $\phi(\theta_j^{n+1/2})\geq 2/Re_c(1-|\mathcal{C}|)$ y $\phi(\theta_{j-1/2}^{n+1/2})\geq 2/Re_c(1-|\mathcal{C}|)$ el orden de truncamiento para el segundo paso esta definido por

$$\begin{aligned}\tau_{h,k}=&\frac{1}{k}\left[u(x_j,t_n+k)-u\left(x_j,t_n+\frac{k}{2}\right)\right]\\&+\frac{1}{h}\left[\phi(\theta_j^{n+1/2})\left(f(u)\left(x_j+\frac{h}{2},t_n+\frac{k}{2}\right)-f(u)\left(x_j,t_n+\frac{k}{2}\right)\right)\right.\\&\left.+\phi(\theta_{j-1/2}^{n+1/2})\left(f(u)\left(x_j,t_n+\frac{k}{2}\right)-f(u)\left(x_j-\frac{h}{2},t_n+\frac{k}{2}\right)\right)\right]\\&-\frac{\nu}{h^2}\left[u(x_j+h,t_n)-2u(x_j,t_n)+u(x_j-h,t_n)\right].\end{aligned} \tag{4.11}$$

Tomando en cuenta los desarrollos de Taylor y denotando $u=u(x_j,t_n+\frac{k}{2})$,

$\phi(\theta_{j-1/2}^{n+1/2}) = \theta$ y $\phi(\theta_{j-1/2}^{n+1/2}) = (1-\theta)$ tenemos:

$$\begin{aligned}
u(x_j+h,t_n) =& u + hu_x - \frac{k}{2}u_t + \frac{k^2}{8}u_{tt} - 2\frac{kh}{4}u_{tx} + \frac{h^2}{2}u_{xx} \\
&- \frac{k^3}{48}u_{ttt} + 3\frac{k^2h}{24}u_{ttx} - 3\frac{kh^2}{12}u_{txx} + \frac{h^3}{6}u_{xxx} + \mathcal{O}(h^4,k^4), \\
u(x_j,t_n) =& u - \frac{k}{2}u_t + \frac{k^2}{8}u_{tt} - \frac{k^3}{48}u_{ttt} + \mathcal{O}(k^4), \\
u(x_j-h,t_n) =& u - hu_x - \frac{k}{2}u_t + \frac{k^2}{8}u_{tt} + 2\frac{kh}{4}u_{tx} + \frac{h^2}{2}u_{xx} \\
&- \frac{k^3}{48}u_{ttt} - 3\frac{k^2h}{24}u_{ttx} - 3\frac{kh^2}{12}u_{txx} - \frac{h^3}{6}u_{xxx} + \mathcal{O}(h^4,k^4), \\
u(x_j,t_n+k) =& u + \frac{k}{2}u_t + \frac{k^2}{8}u_{tt} + \frac{k^3}{48}u_{ttt} + \mathcal{O}(k^4), \\
f(u)\left(x_j+\frac{h}{2},t_n+\frac{k}{2}\right) =& f(u) + \frac{h}{2}f(u)_x + \frac{h^2}{8}f(u)_{xx} + \frac{h^3}{48}f(u)_{xxx} + \mathcal{O}(h^4), \\
f(u)\left(x_j-\frac{h}{2},t_n+\frac{k}{2}\right) =& f(u) - \frac{h}{2}f(u)_x + \frac{h^2}{8}f(u)_{xx} - \frac{h^3}{48}f(u)_{xxx} + \mathcal{O}(h^4).
\end{aligned}$$

Sustituyendo las anteriores expresiones en obtenemos que el primer término de (4.11) es

$$\frac{1}{k}\left[u(x_j,t_n+k) - u\right] = \frac{1}{2}u_t + \frac{k}{8}u_{tt} + \frac{k^2}{48}f(u)xtt + \mathcal{O}(k^3),$$

el segundo término de (4.11) es

$$\begin{aligned}
&\frac{1}{h}\left[\theta\left(f(u)\left(x_j+\frac{h}{2},t_n+\frac{k}{2}\right) - f(u)\left(x_j,t_n+\frac{k}{2}\right)\right)\right. \\
&\left.+ (1-\theta)\left(f(u)\left(x_j,t_n+\frac{k}{2}\right) - f(u)\left(x_j-\frac{h}{2},t_n+\frac{k}{2}\right)\right)\right] \\
&= \frac{1}{2}f(u)_x + (2\theta-1)\frac{h}{8}f(u)_{xx} + \frac{h^2}{48}f(u)_{xxx} + (2\theta-1)\mathcal{O}(h^3),
\end{aligned}$$

y el tercer término de (4.11) es

$$-\frac{\nu}{h^2}\left[u(x_j+h,t_n) - 2u(x_j,t_n) + u(x_j-h,t_n)\right] = -\nu u_{xx} + \frac{\nu k}{2}u_{txx} + \mathcal{O}(h^2,k^4),$$

tenemos que

$$\tau_{h,k} = k\left(\frac{-\nu u_{txx}}{2} + \frac{u_{tt}}{8}\right) + (2\theta-1)h\left(\frac{f(u)_{xx}}{8}\right) + \frac{k^2u_{ttt}}{48} + \frac{h^2f(u)_{xxx}}{48} + \mathcal{O}(h^2,k^4) + \mathcal{O}(k^3),$$

entonces

$$\tau_{h,k} = (2\theta-1)\mathcal{O}(h) + \mathcal{O}(h^2) + \mathcal{O}(k). \tag{4.12}$$

Por lo tanto el segundo paso es de *primer orden* temporal y de *primer orden o segundo orden* espacial dependiendo del valor de θ. Mientras θ sea más cercano a $1/2$ el método se comportará más como uno de segundo orden.

Si $\phi(\theta_j^{n+1/2}) \leq 2/Re_c(1-|\mathcal{C}|)$ y $\phi(\theta_{j-1/2}^{n+1/2}) \leq 2/Re_c(1-|\mathcal{C}|)$ el orden de truncamiento para el segundo paso esta definido por

$$\begin{aligned}\tau_{h,k} =&\frac{1}{k}\Big[u(x_j,t_n+k)-u(x_j,t_n)\Big]+\frac{1}{h}\Big[f(u)\Big(x_j+\frac{h}{2},t_n+\frac{k}{2}\Big)-f(u)\Big(x_j-\frac{h}{2},t_n+\frac{k}{2}\Big)\Big]\\ &-\frac{\nu}{h^2}\Big[u(x_j+h,t_n)-2u(x_j,t_n)+u(x_j-h,t_n)\Big].\end{aligned} \tag{4.13}$$

Tomando en cuenta los desarrollos de Taylor y denotando $u=u(x_j,t_n)$

$$u(x_j+h,t_n) =u+hu_x+\frac{h^2}{2}u_{xx}+\frac{h^3}{6}u_{xxx}+\frac{h^4}{24}u_{xxxx}+\mathcal{O}(h^5), \tag{4.14}$$

$$u(x_j-h,t_n) =u-hu_x+\frac{h^2}{2}u_{xx}-\frac{h^3}{6}u_{xxx}+\frac{h^4}{24}u_{xxxx}+\mathcal{O}(h^5), \tag{4.15}$$

$$u(x_j,t_n+k) =u+ku_t+\frac{k^2}{2}u_{tt}+\frac{k^3}{6}u_{ttt}+\mathcal{O}(k^4), \tag{4.16}$$

$$\begin{aligned}f(u)(x_j+\frac{h}{2},t_n+\frac{k}{2}) =&f(u)+\frac{h}{2}f(u)_x+\frac{k}{2}f(u)_t+\frac{k^2}{8}f(u)_{tt}+2\frac{kh}{8}f(u)_{tx}\\ &+\frac{h^2}{8}f(u)_{xx}+\frac{k^3}{48}f(u)_{ttt}+3\frac{k^2h}{48}f(u)_{ttx}\\ &+3\frac{kh^2}{48}f(u)_{txx}+\frac{h^3}{48}f(u)_{xxx}+\mathcal{O}(h^4,k^4),\end{aligned} \tag{4.17}$$

$$\begin{aligned}f(u)(x_j-\frac{h}{2},t_n+\frac{k}{2}) =&f(u)-\frac{h}{2}f(u)_x+\frac{k}{2}f(u)_t+\frac{k^2}{8}f(u)_{tt}-2\frac{kh}{8}f(u)_{tx}\\ &+\frac{h^2}{8}f(u)_{xx}+\frac{k^3}{48}f(u)_{ttt}-3\frac{k^2h}{48}f(u)_{ttx}\\ &+3\frac{kh^2}{48}f(u)_{txx}-\frac{h^3}{48}f(u)_{xxx}+\mathcal{O}(h^4,k^4),\end{aligned} \tag{4.18}$$

sustituyendo (4.14), (4.15), (4.16), (4.17) y (4.18) en (4.13) tenemos

$$\begin{aligned}\tau_{h,k} =&u_t+\frac{k}{2}u_{tt}+\frac{k^2}{6}u_{ttt}+\mathcal{O}(k^3)+f(u)_x+\frac{k}{2}f(u)_{tx}+\frac{k^2}{48}f(u)_{ttx}+\frac{h^2}{24}f(u)xxx+\mathcal{O}(h^3,k^3)\\ &-\nu u_{xx}-\frac{\nu h^2}{12}u_{xxxx}+\mathcal{O}(h^4),\\ =&k\Big(\frac{u_{tt}}{2}+\frac{f(u)_{tx}}{2}\Big)+k^2\Big(\frac{u_{ttt}}{6}+\frac{f(u)_{ttx}}{48}\Big)+h^2\Big(\frac{f(u)_{xxx}}{24}-\frac{\nu u_{xxxx}}{12}\Big)+\mathcal{O}(h^3,k^3),\end{aligned}$$

entonces

$$\tau_{h,k}=\mathcal{O}(h^2)+\mathcal{O}(k). \tag{4.19}$$

Por lo tanto el segundo paso es de *primer orden* temporal y de *segundo orden* espacial.

Proposición 4.1. *El método Viscous TVD Flux-Limiter definido en (4.2)-(4.7) es* ***consistente****.*

Demostración.
Por la definición de consistencia dada en [21] y por (4.10), (4.12) y (4.19). ∎

4.3. TVD-estabilidad

En esta sección presentaremos la TVD-estabilidad para el método Viscous TVD Flux-limiter, haciendo uso del teorema de Harten.

Proposición 4.2. *El método Viscous TVD Flux-Limiter definido en (4.2)-(4.7) es TVD estable si*

$$\phi_j^{n+1/2} \geq \frac{2}{Re_c(1-|\mathcal{C}|)}, \tag{4.20}$$

para $\mathcal{C} \geq 0$ *y*

$$\phi_{j-1/2}^{n+1/2} \geq \frac{2}{Re_c(1-|\mathcal{C}|)}, \tag{4.21}$$

si $\mathcal{C} < 0$*, donde* $\mathcal{C} = \frac{k}{h} f_u(U_j^{n+1/2})$ *y* $Re_c = |\mathcal{C}|/d$*.*

Demostración.

Discretizando el lado convectivo de (4.1), utilizando el esquema conservativo explícito desarrollado en [15] para el término de convección y discretizando el término difusivo de (4.1) utilizando las diferencias centradas de segundo orden tenemos

$$\begin{aligned} U_j^{n+1} =& U_j^n - \frac{k}{2h}\Big[\lambda^+(f(U_j^n) - f(U_{j-1}^n)) - \lambda^-(f(U_{j+1}^n) - f(U_j^n))\Big] \\ & - \frac{k}{h}\Big[\phi(\theta_j^{n+1/2})\Big(f(U_{j+1/2}^{n+1/2}) - f(U_j^{n+1/2})\Big) + \phi(\theta_{j-1/2}^{n+1/2})\Big(f(U_j^{n+1/2}) - f(U_{j-1/2}^{n+1/2})\Big)\Big] \\ & + \frac{\nu k}{h^2}(U_{j+1}^n - 2U_j^n + U_{j-1}^n), \end{aligned} \tag{4.22}$$

asumiendo que f es una función suave, aplicamos el teorema del valor medio y sustituyendo $U_{j+1/2}^{n+1/2}$, $U_j^{n+1/2}$ y $U_{j-1/2}^{n+1/2}$ definidos en (4.2),(4.3) y (4.4) en (4.22) tenemos

$$\begin{aligned} U_j^{n+1} =& U_j^n - \frac{1}{2}\Bigg[\lambda^+\sigma_{j-1}^n(1+\phi_j^{n+1/2}\sigma_j^{n+1/2} - \phi_{j-1/2}^{n+1/2}\sigma_{j-1/2}^{n+1/2}) \\ &+ \phi_{j-1/2}^{n+1/2}\sigma_{j-1/2}^{n+1/2}(1-\sigma_j^n) + 2d\Bigg](U_j^n - U_{j-1}^n) \\ &+ \frac{1}{2}\Bigg[\lambda^-\sigma_j^n(1+\phi_j^{n+1/2}\sigma_j^{n+1/2} - \phi_{j-1/2}^{n+1/2}\sigma_{j-1/2}^{n+1/2}) \\ &+ \phi_j^{n+1/2}\sigma_j^{n+1/2}(\sigma_j^n - 1) + 2d\Bigg](U_{j+1}^n - U_j^n), \end{aligned} \tag{4.23}$$

donde $\sigma_m^n = \dfrac{k}{h} f u(\xi_m^n)$ y $d = \dfrac{\nu k}{h^2}$. Procediendo de la misma manera que en [15], supongamos que todos los coeficientes σ_m^n son positivos, entonces $\lambda^+ = 1$, $\lambda^- = 0$ y los coeficientes de Harten son

$$C_{j-1} = \frac{1}{2}\Bigg[\sigma_{j-1}^{n+1/2}(\sigma_{j-1}^n + (1-\sigma_j^n)b_j^n)\phi_j^{n+1/2} + \sigma_{j-1/2}^{n+1/2}\phi_{j-1/2}^{n+1/2} + \sigma_j^n - 2d(b_j^n - 1)\Bigg], \tag{4.24}$$

$$D_j = 0, \tag{4.25}$$

donde $b_j^n = \dfrac{U_{j+1}^n - U_j^n}{U_j^n - U_{j-1}^n}$. Sea $\mathcal{C} = max\{\sigma_{j-1}^{n+1/2}, \sigma_{j-1/2}^{n+1/2}, \sigma_{j-1}^n, \sigma_j^n\}$ entonces

$$\begin{aligned} C_{j-1} &\leq \frac{1}{2}\Bigg[\mathcal{C}(\mathcal{C} + (1-\mathcal{C})b_j^n)\phi_j^{n+1/2} + \mathcal{C}\phi_{j-1/2}^{n+1/2} + \mathcal{C} - 2d(b_j^n - 1)\Bigg] \\ &\leq \frac{\mathcal{C}}{2}\Bigg[(\mathcal{C} + (1-\mathcal{C})b_j^n)\phi_j^{n+1/2} + \phi_{j-1/2}^{n+1/2} + 1 - \frac{2d}{\mathcal{C}}(b_j^n - 1)\Bigg] \\ &= \frac{\mathcal{C}}{2}\Bigg[1 + \mathcal{C}\phi_j^{n+1/2} + \phi_{j-1/2}^{n+1/2} + (1-\mathcal{C})b_j^n\phi_j^{n+1/2} - \frac{2d}{\mathcal{C}}(b_j^n - 1)\Bigg] \\ &= \mathcal{C}\Bigg[\frac{1 + \mathcal{C}\phi_j^{n+1/2} + \phi_{j-1/2}^{n+1/2} + 2d/\mathcal{C}}{2} + \frac{(1-\mathcal{C})}{2}\left(b_j^n\phi_j^{n+1/2} - \frac{2db_j^n}{\mathcal{C}(1-\mathcal{C})}\right)\Bigg], \end{aligned}$$

sea $Re_c = \mathcal{C}/d$ (número de Reynolds por celda) y reescribiendo la ecuación anterior en función de Re_c, obtenemos

$$C_{j-1} \leq \mathcal{C}\Bigg[\frac{1 + \mathcal{C}\phi_j^{n+1/2} + \phi_{j-1/2}^{n+1/2} + 2d/\mathcal{C}}{2} + \frac{(1-\mathcal{C})}{2}\left(b_j^n\phi_j^{n+1/2} - \frac{2b_j^n}{Re_c(1-\mathcal{C})}\right)\Bigg],$$

por Leveque [21], teniendo en cuenta que si $|\mathcal{C}| \leq 1$ y

$$\left| b_j^n \phi_j^{n+1/2} - \frac{2b_j^n}{Re_c(1-\mathcal{C})} \right| \leq 2, \tag{4.26}$$

entonces $C_{j-1} \leq 1$. Por [15] tenemos que la función Flux-Limiter definida en la proposición (2.4) satisface la desigualdad

$$\left| b_j^n \phi_j^{n+1/2} \right| \leq 2.$$

Observemos que ahora tenemos

$$b_j^n \left(\phi_j^{n+1/2} - \frac{2}{Re_c(1-\mathcal{C})} \right),$$

entonces

$$\frac{(1-\mathcal{C})}{2} b_j^n \phi_j^{n+1/2} \geq \frac{(1-\mathcal{C})}{2} b_j^n \left(\phi_j^{n+1/2} - \frac{2}{Re_c(1-\mathcal{C})} \right),$$

así que si

$$\phi_j^{n+1/2} \geq \frac{2}{Re_c(1-\mathcal{C})}, \tag{4.27}$$

aplicamos el Viscous Flux-Limiter, ya que Toro [33] nos dice que el método TVD requiere viscosidad artificial. Ahora si

$$\phi_j^{n+1/2} < \frac{2}{Re_c(1-\mathcal{C})}, \tag{4.28}$$

por (4.8) se tiene que es suficiente la viscosidad física del problema para remover las oscilaciones espurias de las soluciones numéricas, por lo que no es necesario la función Flux-limiter, entonces

$$\lambda^+ = \lambda^- = 0, \tag{4.29}$$

$$\phi_{j+1/2}^{n+1/2} = \phi_{j-1/2}^{n+1/2} = 1/2. \tag{4.30}$$

Ahora supongamos que todos los coeficientes σ_m^n son negativos , entonces los coeficientes de Harten son

$$C_{j-1} = 0, \tag{4.31}$$

$$D_j = \frac{1}{2}\left[-\sigma_{j-1/2}^{n+1/2}(-\sigma_j^n + (1+\sigma_{j-1}^n) b_j^n) \phi_{j-1/2}^{n+1/2} - \sigma_j^{n+1/2} \phi_{j-1/2}^{n+1/2} - \sigma_j^n - 2d(b_j^n + 1) \right], \tag{4.32}$$

donde $b_j^n = \dfrac{U_j^n - U_{j-1}^n}{U_{j+1}^n - U_j^n}$. Análogamente, obtenemos que si

$$\phi_{j-1/2}^{n+1/2} \geq \frac{2}{Re_c(1+\mathcal{C})}, \tag{4.33}$$

aplicamos el Viscous Flux-Limiter, en caso contrario hacemos

$$\lambda^+ = \lambda^- = 0, \tag{4.34}$$

$$\phi_{j+1/2}^{n+1/2} = \phi_{j-1/2}^{n+1/2} = 1/2. \tag{4.35}$$

■

Proposición 4.3. *El método Viscous TVD Flux-Limiter definido en (4.2)-(4.7) es convergente.*

Demostración.

Por la Proposición (4.2) y el teorema (2.4). ■

Las aproximaciones numéricas para la ecuación de convección-difusión se presentan en el capítulo 5.

CAPÍTULO 5

Resultados Numéricos

En este capítulo se exponen los resultados numéricos obtenidos mediante el método Splitting Strang y el método Viscous TVD Flux-limiter y se ven algunas simulaciones numéricas para ejemplificar las ventajas y desventajas que poseen estos métodos. Las simulaciones se realizaron en el programa **MatLab**.

Para el método Splitting Strang se implementa para el operador hiperbólico no lineal el método explícito *TVD Flux-Limiter* (2.47)-(2.50). Los ejemplos numéricos presentados en [15] nos muestran la gran ventaja del método, al combinar Upwind con Lax-Wendroff, dado que proporciona una buena aproximación a soluciones suaves, oscilatorias o discontinuas de problemas hiperbólicos no lineales en forma conservativa. El operador lineal parabólico es resuelto por el método *Euler Explícito* (2.22), la elección para su implementación se debe a que el método *Splitting Strang* es un método de descomposición de operadores explícito, aunque el método *Euler Explícito* no sea muy preciso.

En los siguientes ejemplos se observa la gran ventaja del esquema Viscous TVD Flux-Limiter sobre el método Splitting Strang, dado que nos proporciona buenas aproximaciones a las soluciones.

Ejemplo 1

Para este ejemplo utilizaremos la ecuación de convección-difusión

$$u_t + au_x = \nu u_{xx},$$

con condición inicial $u(x,0) = \text{sen}(\pi x)$ y las condiciones iniciales $u(0,t) = 0$ y

$u(2,t) = 0$ para $t \in [0,T]$, $a = 1$ y $\nu = 1$.

En las siguientes gráficas podemos observar las correspondientes soluciones de los métodos Splitting Strang y Viscous TVD Flux-Limiter en el tiempo $T = 0{,}01$.

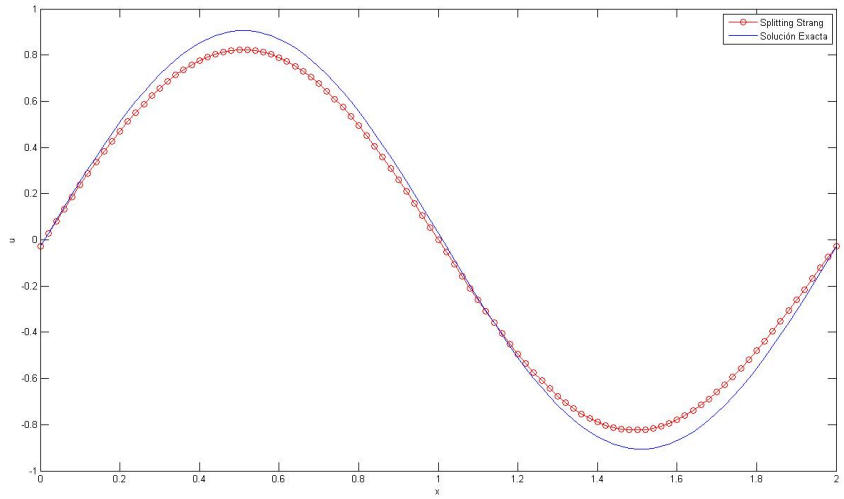

Figura 5.1: Simulación numérica del método Splitting Strang para el ejemplo 1 para T=0.01.

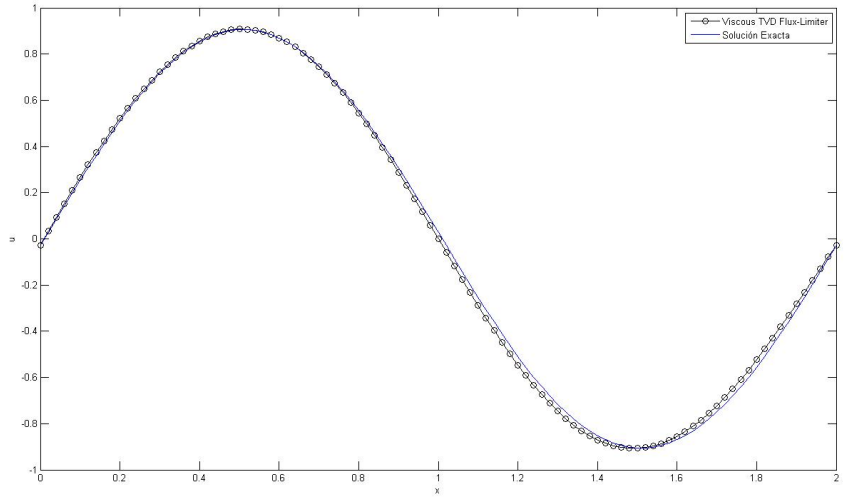

Figura 5.2: Simulación numérica del método Viscous TVD Flux-Limiter para el ejemplo 1 para T=0.01.

En la *figura* **5.3** están las soluciones de todos los métodos numéricos aplicados a este problema.

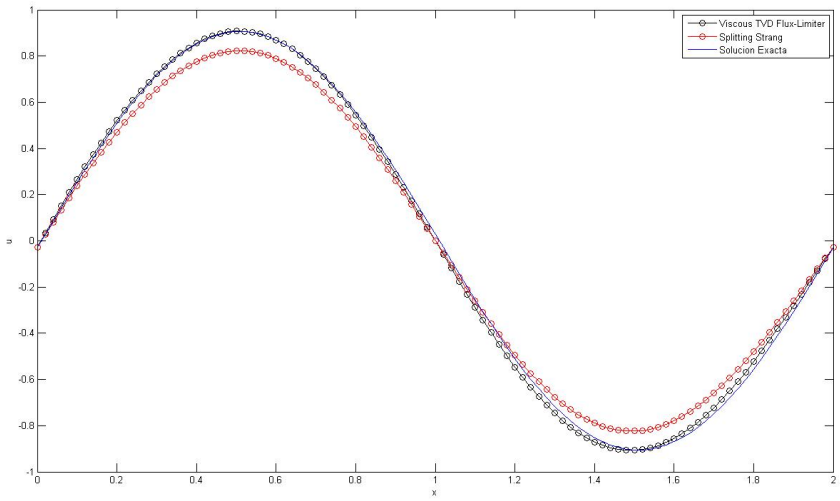

Figura 5.3: Simulaciones numéricas de los dos métodos aplicados al ejemplo 1 para T=0.01.

En este ejemplo se puede apreciar que el método *Splitting Strang* va por debajo de la solución exacta, mientras que el método *Viscous TVD Flux-Limiter* si proporciona una buena aproximación.

Ejemplo 2

Para nuestro primer ejemplo no lineal consideremos la ecuación conservativa de Burgers viscosa con $f(u) = \dfrac{u^2}{2}$

$$u_t + f(u)_x = \nu u_{xx},$$

con $\nu = ,001$ y los intervalos temporal [0,1] y espacial [0,1], con un tamaño de paso $\Delta x = 0{,}01$, $\Delta t = 0{,}005$ y la condición inicial discontinua

$$u(x,0) \begin{cases} 1 & \text{si } x \leq 0.1, \\ 0 & \text{si } x > 0.1. \end{cases}$$

La solución exacta para $t = T$ esta dado por

$$u(x,T) \begin{cases} 1 & \text{si } x \leq \frac{T}{2} + 0.1, \\ 0 & \text{si } x > \frac{T}{2} + 0.1. \end{cases}$$

En las siguientes gráficas podemos observar las correspondientes soluciones de los métodos Splitting Strang y Viscous TVD Flux-Limiter en el tiempo $T = 0{,}5$.

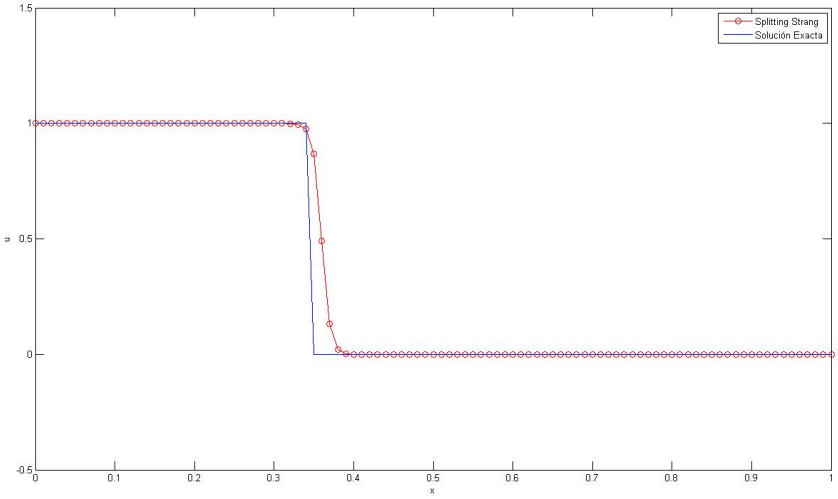

Figura 5.4: Método Splitting Strang para el ejemplo 2, en el tiempo T=0.5.

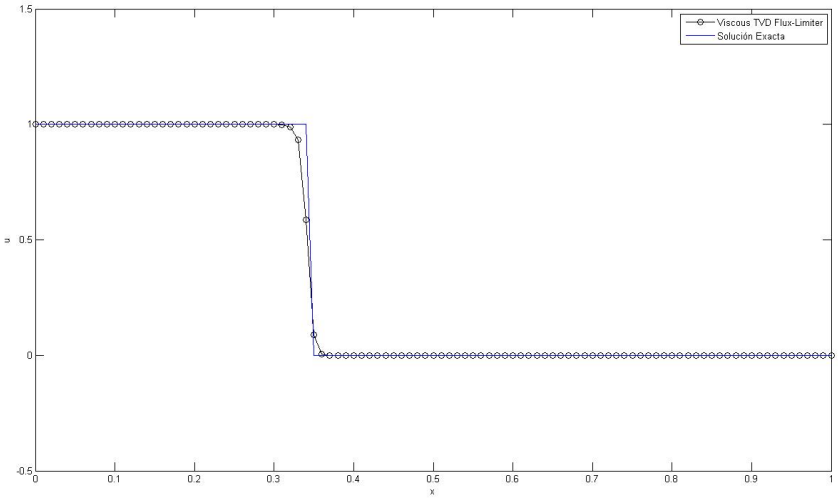

Figura 5.5: Método Viscous TVD Flux-Limiter para el ejemplo 2, en el tiempo T=0.5.

En la *figura* **5.6** están las soluciones de los métodos numéricos aplicados a este problema.

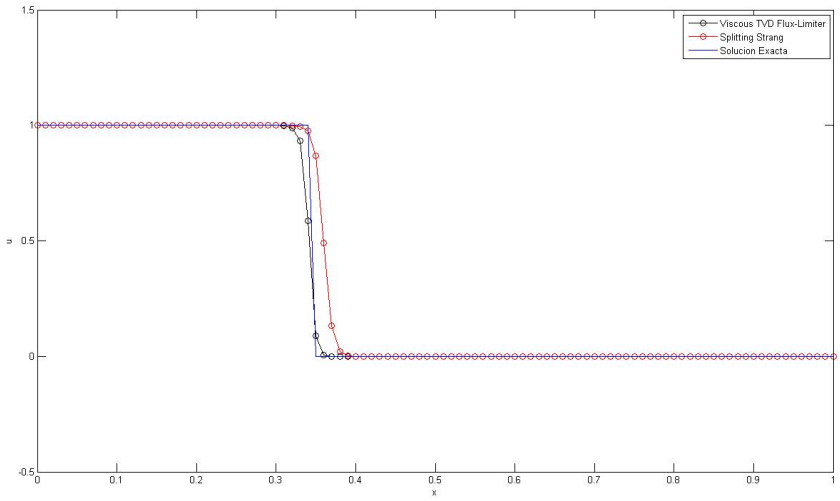

Figura 5.6: Simulaciones numéricas de los dos métodos aplicados al ejemplo 2 para T=0.5.

Se puede apreciar que el método *Splitting Strang* simula bien la solución, pero contiene una disipación numérica entorno a la discontinuidad, llevando a la pérdida de precision numérica en esta región. Por otro lado el método *Viscous TVD Flux-Limiter* es más preciso reduciendo la difusión numérica evitando las oscilaciones espurias en la onda de choque.

Ejemplo 3

Para este ejemplo consideraremos nuevamente la ecuación de Burgers viscosa con solución

$$u(x,t) = -\frac{2\,\mathrm{senh}(x)}{\cosh(x) - e^{-t}}$$

para $\nu = 1$ y condiciones de frontera $u(-9,t) = 2$, $u(9,t) = -2$ para $t \in [0,T]$. Para las simulaciones de este ejemplos se utilizó un tamaño de paso $\Delta x = 0{,}2$ y $\Delta t = 0{,}0083$.

En las siguientes gráficas podemos observar las correspondientes soluciones de los métodos Splitting Strang y Viscous TVD Flux-Limiter en el tiempo $T = 0{,}83$.

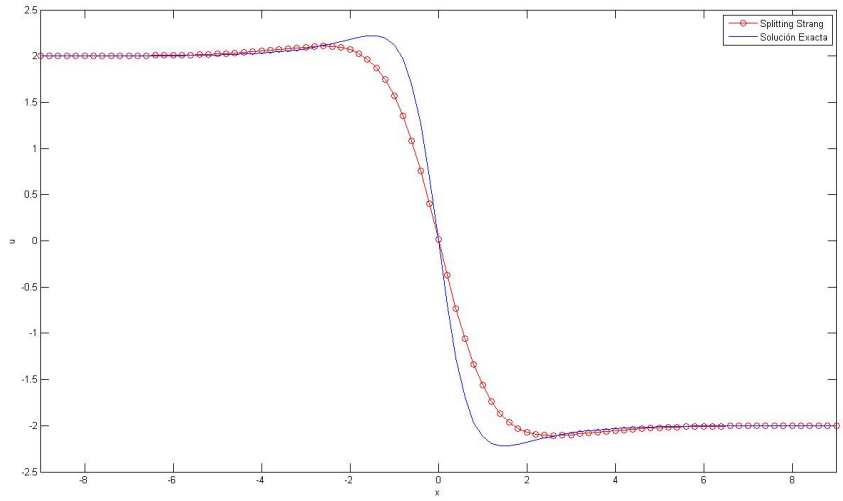

Figura 5.7: Método Splitting Strang para el ejemplo 3, en el tiempo T=0.83.

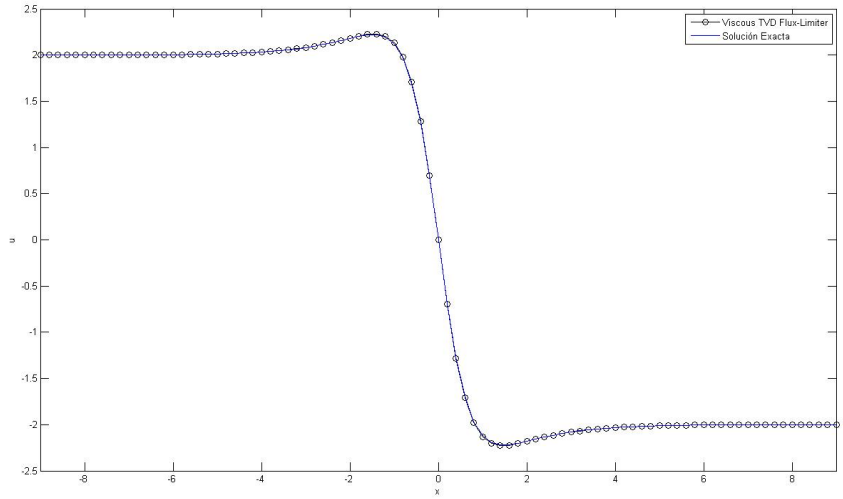

Figura 5.8: Método Viscous TVD Flux-Limiter para el ejemplo 3, en el tiempo T=0.83.

En la *figura* **5.9** están las soluciones de los métodos numéricos aplicados a este problema.

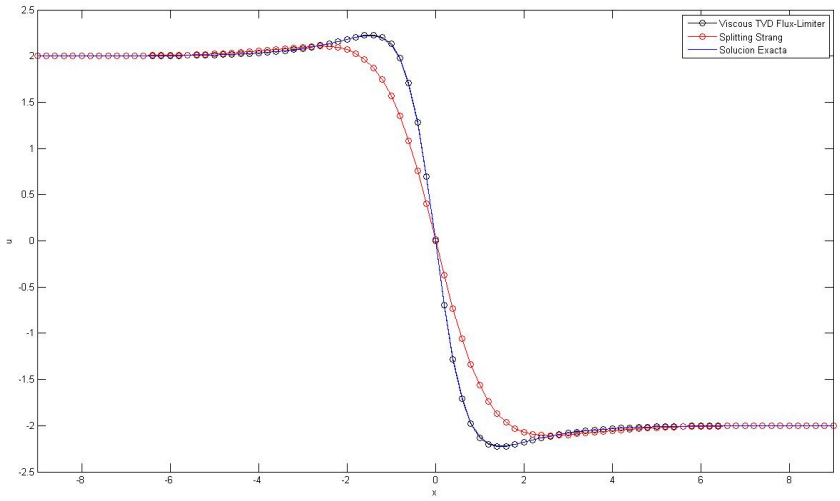

Figura 5.9: Simulaciones numéricas de los dos métodos aplicados al ejemplo 3 para T=0.83.

En este ejemplo, donde tenemos cambios rápidos de la solución, se puede apreciar que el método *Splitting Strang* no es tan preciso como quisiéramos, mientras que el método *Viscous TVD Flux-Limiter* sí proporciona una buena aproximación.

CONCLUSIONES

El objetivo de este trabajo ha sido obtener métodos numéricos precisos y estables para problemas de convección-difusión no lineal. Estos métodos numéricos resuelven un sistema general de convección-difusión no lineal en forma conservativa. Se establecieron propiedades importantes que un método eficiente debe de verificar, como lo son el error de truncamiento, la consistencia, estabilidad y consecuentemente la convergencia.

El primero método presentado es un esquema numérico, en el cual se descompone el problema en una sucesión de subproblemas más simples en cada nivel de tiempo, a través de la discretización temporal. La solución de dichos subproblemas estacionarios, se lleva a cabo mediante métodos iterativos que en cada iteración, requieren de la solución de problemas hiperbólicos no lineales y parabólicos lineales. El segundo esquema es una extensión del método TVD Flux-Limiter a sistemas hiperbólicos con términos difusivos, ya que la ecuación de convección-difusión no lineal en nuestro caso esta dominada por el término convectivo.

Los ejemplos numéricos presentados mostraron la gran ventaja del método propuesto Viscous TVD Flux-Limiter sobre el método Splitting Strang, dado que proporcionó una buena aproximación a soluciones suaves y oscilatorias.

APÉNDICE A

Conceptos Básicos de EDP´s

En este apéndice presentamos una breve introducción de las ecuaciones diferenciales parciales de segundo orden. El objetivo es dar las definiciones referentes a las EDP´s. Las siguientes definiciones se encuentran en [11] y [24]. Considérese una variable dependiente u que depende de un numero de variables independientes x_i, $i = 1, 2, ..., n$. Expresado en términos de una relación funcional, es posible escribir:

$$u = u(x_1, x_2, ..., x_n)$$

La relación funcional anterior es analítica si admite una expansion en serie de Taylor en todas sus variables independientes. Esto es equivalente a decir que u tiene derivadas de cualquier orden en cualquiera de las n variables. En este caso, una ecuación diferencial a derivadas parciales (EDP) es cualquier relación entre las derivadas de u. La forma general para u escalar es:

$$F(u, x_1, x_2, ..., x_n, \frac{\partial u}{\partial x_1}, \frac{\partial u}{\partial x_2}, ..., \frac{\partial u}{\partial x_n}, \frac{\partial^2 u}{\partial x_1^2}, \frac{\partial^2 u}{\partial x_2^2}, ..., \frac{\partial^2 u}{\partial x_n^2}, \frac{\partial^2 u}{\partial x_1 \partial x_2}, ..., \frac{\partial^m u}{\partial x_n^m}, ...) = 0 \qquad \text{(A.1)}$$

El orden de una ecuación diferencial parcial es el de la derivada de mayor orden que aparezca en dicha ecuación. La ecuación diferencial (A.1) es de orden m. Se dirá que la ecuación (A.1) es lineal si la función F es lineal respecto de u y de todas las derivadas parciales de u que aparecen en ella. De manera análoga que en EDO, la ecuación (A.1) es no lineal si posee una dependencia no lineal en las derivadas de mayor orden de u que aparecen en la ecuación.

Los tres ejemplos de problemas clásicos se estudio de las ecuaciones diferenciales en derivadas parciales son:

- Ecuaciones de tipo hiperbólico (problemas que requieren fenómenos oscilatorios: vibraciones de cuerda, membranas, oscilaciones electromagnéticas).

- Ecuaciones de tipo parabólico (problemas que se presentan al estudiar los procesos de conductibilidad térmica y difusión).

- Ecuaciones de tipo elíptico (problemas que aparecen al estudiar procesos estacionarios, o sea que no cambian con el tiempo).

Consideremos ahora una ecuación general de derivadas parciales en dos variables independientes, x e y, lineal, de segundo orden, con coeficientes variables:

$$A(x,y)u_{xx} + B(x,y)u_{xy} + C(x,y)u_{yy} + D(x,y)u_x + E(x,y)u_y + F(x,y)u = G(x,y) \qquad \text{(A.2)}$$

donde $u = u(x,y)$ es la función incógnita y $A(x,y), B(x,y), ..., G(x,y)$ funciones de las variables x e y en una región $\Omega \subset \mathbb{R}^2$, a $G(x,y)$ se le denominada término independiente. Al igual que ocurría con las ecuaciones diferenciales ordinarias, si $G(x,y)$ es idénticamente nula, la ecuación diferencial en derivadas parciales se denomina homogénea; en caso contrario se hablará de ecuación no homogénea.

Hay tres tipos de EDP representadas por la forma general anterior, la clasificación se realiza en términos del signo del discriminante $(B^2 - 4AC)$, a saber:

(1) *Elípticas* en Ω, si $\triangle = B^2 - 4AC < 0$ en Ω.

(2) *Parabólicas* en Ω, si $\triangle = B^2 - 4AC = 0$ en Ω .

(3) *Hiperbólicas* en Ω, si $\triangle = B^2 - 4AC > 0$ en Ω .

Es importante mencionar que si los coeficientes A, B, y C dependen de las coordenadas, entonces la clasificación de las ecuaciones es local únicamente, esto se debe a que el tipo de la ecuación puede cambiar con las coordenadas.

Definición A.1. Sea la EDP definida en (A.1) de orden m, se llama solución de dicha EDP en cierta region Ω, de variación de las x_i; $\forall i = 1, 2, ..., n$ a una función cualquiera $u = u(x_1, x_2, ..., x_n) \in \mathcal{C}^m(\Omega)$, tal que al sustituir u y sus derivadas en (A.1), se convierte en la identidad respecto a x_i; $\forall i = 1, 2, ..., n$ en la región Ω.

La solución de ecuaciones diferenciales en derivadas parciales lineales y no lineales, resulta mas difícil que la solución de ecuaciones diferenciales ordinarias debido a que no existen métodos generales de resolución efectivos. Estamos interesados en hallar la solución más general posible de una EDP. La solución general para una ecuación diferencial ordinaria lineal de orden $n \in \mathbb{N}$, es un conjunto de funciones dependientes de n constantes arbitrarias. En lo que a las ecuaciones diferenciales parciales se refiere, en lugar de constantes, la solución general depende de constantes

arbitrarias.

Por ejemplo, la solución general de la ecuación $\dfrac{\partial^2 u}{\partial x \partial y} = 0$ tiene la forma

$$u(x, y) = f(x) + g(y). \tag{A.3}$$

Donde f y g son las funciones derivables arbitrarias. Para obtener de (A.3) una solución particular, será necesario añadir ciertas condiciones adicionales que permitan determinar las funciones f y g explícitamente, lo que puede resultar incluso más difícil que la obtención de la propia solución general. Esta dificultad no aparece en el caso ordinario, en el que la obtención de soluciones particulares requiere solamente la determinación de constantes arbitrarias. Por este motivo, en lugar de obtener soluciones generales y a partir de ellas las correspondientes particulares que verifiquen ciertas condiciones prescritas, consideremos solamente procedimientos que permitan obtener directamente estas soluciones particulares.

Otra diferencia entre soluciones de una EDO Y EDP se refiere al conjunto de soluciones de la ecuación homogénea. Para las ecuaciones ordinarias lineales de orden $n \in \mathbb{N}$, constituye un espacio vectorial de dimensión n, igual al orden de la ecuación. Esto viene a significar que toda solución se puede expresar como una combinación lineal de soluciones linealmente independientes. Desafortunadamente esto no es cierto, en general, para las ecuaciones diferenciales en derivadas parciales lineales homogéneas, debido a que, aunque el conjunto formado por sus soluciones tiene también estructuras de espacio vectorial, se tiene que, su dimensión es infinita.

Para determinar la solución completamente en un proceso físico es insuficiente solo la ecuación diferencial del proceso, hacen falta dos tipos de condiciones auxiliares.

- **Condiciones iniciales**: al igual que en el caso de las ecuaciones ordinarias, prescriben el valor de la función y/o sus derivadas en cierto instante . Supondremos siempre que, lo que no restará generalidad a los resultados que se obtengan. El número de condiciones iniciales a considerar depende del orden de la derivada parcial con respecto al tiempo que contenga la ecuación.

$$u(x, 0) = f_1(x), u_t(x, 0) = \left.\frac{\partial u}{\partial t}\right|_{t=0} = f_2(x)$$

- **Condiciones de frontera**: están asociadas a variables que representan alguna dimensión espacial y que, por tanto se hallan restringidas a una cierta región Ω finita o semiinfinita en el espacio. Si la ecuación diferencial en derivadas parciales es de segundo orden, va ser necesario conocer el valor de la solución, de su derivada o de una combinación lineal de ellas, en la frontera $\partial\Omega$ de la región considerada. Así pues trataremos tres tipos de condiciones de frontera:

1) *Condiciones de Dirichlet:* consisten en prescribir el valor de la solución en la frontera y pueden representarse por

$$u(x,t)|_{\partial\Omega} = g(t).$$

2) *Condiciones de Neumann:* consisten en prescribir el valor de la derivada según la dirección normal (gradiente) de la solución en la frontera y pueden representarse por

$$\frac{\partial u}{\partial n}(x,t)\bigg|_{\partial\Omega} = g(t).$$

3) *Condiciones de Cauchy ó Robin:* son de carácter mixto pues prescriben el valor de una combinación lineal de la solución y su derivada según la dirección normal en la frontera. Se pueden representar por:

$$u(x,t)|_{\partial\Omega} + k\frac{\partial u}{\partial n}(x,t)\bigg|_{\partial\Omega} = g(t).$$

El concepto de condición de frontera incluye como caso especial el concepto de condición inicial.

Problema inverso bien planteado

Con mucha frecuencia las EDP surgen al confeccionar modelos de un determinado fenómeno, usando diferentes hipótesis. Evidentemente nos interesan aquellos esquemas que describen adecuadamente la realidad y permiten además efectuar predicciones. Si nuestro modelo representa adecuadamente el fenómenos físico que se esta estudiando, cabe esperar que posea una solución única. A priori no es evidente si un problema de ecuaciones en derivadas parciales esta *bien planteado*: aunque físicamente parezca razonable que la solución haya de ser única, puede ocurrir que no tenga solución o que, de tener, no sea única. Aunque se este seguro de que la solución existe y es única, esto no basta para afirmar que el problema está bien planteado. En efecto, en un problema realista los datos se obtienen a partir de experimentos, así que forzosamente no serán exactos, de manera que, para que el problema tenga sentido, ha de ocurrir también que un pequeño cambio en esos datos no influya de manera notable en la solución. Pero aun hay más: como en la mayor parte de las ocasiones estas ecuaciones tienen que resolverse numéricamente, lo que siempre implica aproximaciones y errores de redondeo, es preciso que estas aproximaciones no produzcan grandes desviaciones del resultado exacto, ya que en caso contrario el modelo usado sería de muy poca utilidad.

El estudio de los problemas bien planteados es lo que se suele denominar "teoría clásica de las ecuaciones en derivadas parciales". Comenzando el siglo XX, Jaques Hadamard definió un problema como mal planteado (*ill posed*) si la solución del problema no existe o no es única o si no es una función continua de los datos, es decir, si no es una función continua de u, tales problemas son extremadamente sensibles a las perturbaciones en los datos que pueden conducir a perturbaciones arbitrariamente grandes en la respuesta. La teoría clásica de ecuaciones diferenciales parciales trata casi exclusivamente con problemas bien planteados.

Definición A.2. Sea X y Y espacios normados, $K : X \to Y$ una transformación lineal o no lineal. La ecuación $Kx = y$ es llamada *bien planteada* [16] si cumple las siguientes propiedades siguientes.

1 *Existencia:* Para cada $y \in Y$ existe $x \in X$ tal que $Kx = y$.

2 *Unicidad:* Para cada $y \in Y$ existe a lo mas $x \in X$ tal que $Kx = y$.

3 *Estabilidad:* La solución depende continuamente de y, es decir, para cada sucesión $\{x_n\} \subset X$ con $Kx_n \to Kx$ $(n \to \infty)$, entonces $x_n \to x$ $(n \to \infty)$.

Ecuaciones para las cuales al menos una de las tres propiedades no se cumple es llamada *mal planteada*.

La existencia se demuestra usualmente encontrando una solución que satisfaga la EDP y las condiciones iniciales y/o de frontera. La unicidad implica que la solución encontrada es la única solución posible para el problema considerado. La dependencia continua de los datos iniciales y de frontera es una propiedad que resulta de observaciones de sistemas físicos.

Las dos primeras exigencias son razonables, pero la tercera pudiera parecer caprichosa. Sin embargo es especialmente importante para problemas que aparecen en aplicaciones físicas, ya que, como se ha comentado, es deseable que la solución que buscamos, además de ser única, cambie muy poco cuando se modifican ligeramente las condiciones que especifican el problema. Esta condición implica que pequeñas perturbaciones o errores en las condiciones iniciales o de borde resultan en pequeñas variaciones de la solución de la EDP.

APÉNDICE B

Teorema de Cauchy-Kovaleskaya

Uno de los resultados generales de la teoría de EDP, es un teorema de existencia y unicidad de soluciones de una ecuación en derivadas parciales de orden k con condiciones iniciales para funciones analíticas, que se aplica tanto a los casos lineales como no lineales, se debe a Cauchy y Kovaleskaya (Sofia Kowaleski).

Un aspecto importante a considerar es el tipo de soluciones que se buscan, es decir, las propiedades que se exigen a $u = u(x)$. La decisión en este aspecto está condicionada por las propiedades de la propia función que define la EDP en (A.1). En este sentido se concentrará en una clase de problemas en que es posible deducir un importante resultado sobre la existencia de soluciones analíticas. En este punto es importante comentar la noción de EDP en forma normal o de Kovaleskaya.

Definición B.1. Sea una EDP (A.1) con variables independientes

$$x = (x_0 = t, \mathbf{x}), \mathbf{x} := (x_1, \ldots, x_{n-1}).$$

Decimos que la EDP posee forma normal (*o de Kovaleskaya*) de orden $r > 0$ respecto a la variable t si puede escribirse como:

$$\frac{\partial^r u}{\partial t^r} = G(x, D^\alpha u), \quad r > 0,$$

siendo G una función que depende polinomicamente de un número finito de derivadas

$$D^\alpha u = \frac{\partial^{\alpha_0} u}{\partial t^{\alpha_0}} \frac{\partial^{\alpha_1} u}{\partial x_1^{\alpha_1}} \cdots \frac{\partial^{\alpha_{n-1}} u}{\partial x_{n-1}^{\alpha_{n-1}}}$$

pero debe ser independiente de las siguientes

$$\frac{\partial^r u}{\partial t^r};\ D^\alpha u \quad \text{con } |\alpha| > r.$$

De acuerdo con la definición anterior r es el orden de (A.1). Para analizar si una EDP posee la forma normal respecto de una de sus variables independientes t, lo primero que hay que hacer es despejar la derivada respecto de t de orden más alto y después comprobar que el segundo miembro no aparezcan derivadas de orden estrictamente superior.

Ejemplo

La EDP

$$u_x - u_y = \log(xy),$$

posee la forma normal respecto de cualquier de sus variables x ó y, pues puede escribirse como

$$u_x = u_y + \log(xy),$$

o bien

$$u_y = u_x - \log(xy)$$

y en ambos casos se verifica la condición de normalidad respectiva[1].

Definición B.2. Dada una EDP normal de orden r respecto de una variable t

$$\frac{\partial^r u}{\partial t^r} = G(x, D^\alpha u), \quad x \in \Omega, \tag{B.1}$$

sobre un dominio

$$\Omega = I \times \Lambda,$$

siendo I un intervalo abierto de $\mathbb{R}$ y Λ un abierto de $\mathbb{R}^{n-1}$, un problema de Cauchy con valores iniciales consiste en determinar una solución $u = u(x)$ de (B.1) que satisfaga las r condiciones iniciales definida ($\mathbf{x} \in \Lambda$)

$$\begin{aligned} u(t_0, \mathbf{x}) &= \Phi_0(\mathbf{x}), \\ \frac{\partial u}{\partial t}(t_0, \mathbf{x}) &= \Phi_1(\mathbf{x}), \\ &\vdots \\ \frac{\partial^{r-1} u}{\partial t^{r-1}}(t_0, \mathbf{x}) &= \Phi_{r-1}(\mathbf{x}), \end{aligned}$$

[1] sin embargo la función $\log(xy)$ sólo es analítica cuando $xy > 0$, luego la EDP es normal-analítica en x e y en los dominios $\Omega_1 = \{(x, y) \in \mathbb{R}^2, \text{con } x > 0, y > 0\}$ y $\Omega_2 = \{(x, y) \in \mathbb{R}^2, \text{con } x < 0, y < 0\}$

donde $t_0 \in I$ y $\Phi_i = \Phi_i(\mathbf{x})$ son una serie de funciones dadas, que reciben el nombre de valores iniciales del problema.

El siguiente teorema nos garantiza la existencia y unicidad locales de una solución analítica de un problema de Cauchy con datos iniciales, siempre que se verifique que la EDP es normal y que tanto la EDP como los datos iniciales dependen analíticamente de las variables independientes.

Teorema B.1. *Sea un problema normal de Cauchy con valores iniciales para una EDP normal y $x_0 = (t_0, \mathbf{x}_0) \in \Lambda$ un punto de un dominio tal que*

i) Condición de analiticidad de la EDP

Como función de x la función $G(x, D^\alpha u)$ del segundo miembro es analítica en x_0.

ii) Condición de analiticidad de los valores iniciales

Los valores iniciales $\Phi_i(\mathbf{x})$, $(i = 0, ..., r-1)$ son funciones analíticas en $\boldsymbol{x}_0$.

Entonces existe una función $u = u(x)$ definida sobre un abierto $\Omega_0 \subset \Omega$ que contiene a $\boldsymbol{x}_0$ tal que:

i) La función $u = u(x)$ satisface la EDP en Ω_0 y las condiciones iniciales en todo punto $(t_0, \mathbf{x}) \in \Omega_0$.

ii) La función $u = u(x)$ es la única función analítica en Ω_0 que satisface tales propiedades.

Consultar la demostración de este teorema en [7] y [10].

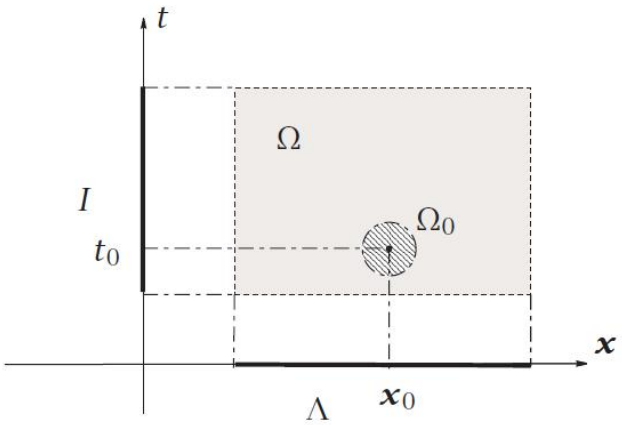

Figura B.1: Interpretación Geométrica del Teorema de Cauchy-Kovaleskaya

Hay una clara analogía entre este resultado y los teoremas de existencia básicos de la teoría de ecuaciones diferenciales ordinarias (EDO). Diremos que una EDP es normal-analítica si es normal y satisface la condición de analiticidad del Teorema de Cauchy-Kovaleskaya, entonces, bajo condiciones apropiadas se verifica:

La solución local de una EDP de orden r depende de r contantes arbitrarias.

La solución analítica local de una EDP normal-analítica de orden r depende de r funciones analíticas arbitrarias.

APÉNDICE C

Teorema de Harten

A continuación enunciamos y demostramos el Teorema de Harten que es de gran utilidad para probar la TVD estabilidad de esquemas de alta resolución para ecuaciones diferenciales parciales hiperbólicas no lineales.

Teorema C.1. *Si un esquema numérico explícito de diferencias finitas es de la forma:*

$$U_i^{n+1} = U_i^n - C_{i-1}^n(U_i^n - U_{i-1}^n) + D_i^n(U_{i+1}^n - U_i^n), \tag{C.1}$$

donde los coeficientes C_{i-1}^n y D_i^n son valores arbitrarios (que pueden depender de U^n de un modo no lineal), entonces

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

si se satisfacen las siguientes condiciones sobre los coeficientes:

$$\begin{aligned} C_i^n &\geq 0 \\ D_i^n &\geq 0 \\ 0 \leq C_i^n + D_i^n &\leq 1, \end{aligned}$$

donde TV (TV por sus siglas en Inglés) es la Variación Total de la función.

Demostración. Incrementado el espacio en (C.1) se tiene

$$U_{i+1}^{n+1} = U_{i+1}^n - C_i^n(U_{i+1}^n - U_i^n) + D_{i+1}^n(U_{i+2}^n - U_{i+1}^n), \tag{C.2}$$

restando (C.2) de (C.1) se obtiene

$$\begin{aligned} U_{i+1}^{n+1} - U_i^{n+1} =& U_{i+1}^n - U_i^n - C_i^n(U_{i+1}^n - U_i^n) + C_{i-1}^n(U_i^n - U_{i-1}^n) \\ &+ D_{i+1}^n(U_{i+2}^n - U_{i+1}^n) - D_i^n(U_{i+1}^n - U_i^n), \end{aligned}$$

entonces

$$U_{i+1}^{n+1} - U_i^{n+1} = (1 - C_i^n - D_i^n)(U_{i+1}^n - U_i^n) + C_{i-1}^n(U_i^n - U_{i-1}^n) + D_{i+1}^n(U_{i+2}^n - U_{i+1}^n),$$

tomando valor absoluto y por hipótesis se tiene $1 - C_i^n - D_i^n \geq 0$

$$\mid U_{i+1}^{n+1} - U_i^{n+1} \mid \leq (1 - C_i^n - D_i^n) \mid U_{i+1}^n - U_i^n \mid + C_{i-1}^n \mid U_i^n - U_{i-1}^n \mid + D_{i+1}^n \mid U_{i+2}^n - U_{i+1}^n \mid,$$

entonces

$$\begin{aligned} \mid U_{i+1}^{n+1} - U_i^{n+1} \mid \leq & \mid U_{i+1}^n - U_i^n \mid - C_i^n \mid U_{i+1}^n - U_i^n \mid + C_{i-1}^n \mid U_i^n - U_{i-1}^n \mid \\ & - D_i^n \mid U_{i+1}^n - U_i^n \mid + D_{i+1}^n \mid U_{i+2}^n - U_{i+1}^n \mid . \end{aligned}$$

Sumando para $-\infty < i < \infty$ se tiene

$$\begin{aligned} \sum_{i=-\infty}^{\infty} \mid U_{i+1}^{n+1} - U_i^{n+1} \mid \leq & \sum_{i=-\infty}^{\infty} \mid U_{i+1}^n - U_i^n \mid - \sum_{i=-\infty}^{\infty} C_i^n \mid U_{i+1}^n - U_i^n \mid \\ & + \sum_{i=-\infty}^{\infty} C_{i-1}^n \mid U_i^n - U_{i-1}^n \mid - \sum_{i=-\infty}^{\infty} D_i^n \mid U_{i+1}^n - U_i^n \mid \\ & + \sum_{i=-\infty}^{\infty} D_{i+1}^n \mid U_{i+2}^n - U_{i+1}^n \mid, \end{aligned}$$

cambiando el orden de la suma del término tercero y quinto del lado derecho tenemos

$$\begin{aligned} \sum_{i=-\infty}^{\infty} \mid U_{i+1}^{n+1} - U_i^{n+1} \mid \leq & \sum_{i=-\infty}^{\infty} \mid U_{i+1}^n - U_i^n \mid - \sum_{i=-\infty}^{\infty} C_i^n \mid U_{i+1}^n - U_i^n \mid \\ & + \sum_{k=+\infty}^{\infty} C_k^n \mid U_{k+1}^n - U_k^n \mid - \sum_{i=-\infty}^{\infty} D_i^n \mid U_{i+1}^n - U_i^n \mid \\ & + \sum_{k=-\infty}^{\infty} D_k^n \mid U_{k+1}^n - U_k^n \mid, \end{aligned}$$

entonces

$$\sum_{i=-\infty}^{\infty} \mid U_{i+1}^{n+1} - U_i^{n+1} \mid \leq \sum_{i=-\infty}^{\infty} \mid U_{i+1}^n - U_i^n \mid .$$

Por lo tanto

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

durante la prueba del teorema se asumió que $C_{i-1}^n \geq 0$, $D_{i+1}^n \geq 0$ ∎

APÉNDICE D

Teorema de Lie

A continuación enunciamos y demostramos el Teorema de Lie que se usa en los métodos de descomposición.

Es bien sabido que, en general, salvo que las matrices A y B conmuten, no es cierto que e^{A+B} coincida con el producto $e^A e^B$. Analizamos este hecho calculando el error global.

El error global se define de la siguiente forma

$$le(x;t) = e^{(A+B)t} - e^{At}e^{Bt},$$

entonces desarrollando en series de Taylor se tiene

$$\begin{aligned} le(x;t) =& \left(I + (A+B)t + \frac{1}{2}(A+B)^2t^2\right) - \left(I + t(A+B) + t^2(AB + \frac{1}{2}A^2 + \frac{1}{2}B^2)\right) + \mathcal{O}(t^3) \\ =& t^2\left(\frac{1}{2}(A+B)^2 - (AB + \frac{1}{2}A^2 + \frac{1}{2}B^2)\right) + \mathcal{O}(t^3) \\ =& t^2\left[\frac{1}{2}A^2 + \frac{1}{2}AB + \frac{1}{2}BA + \frac{1}{2}B^2 - AB - \frac{1}{2}A^2 - \frac{1}{2}B^2\right] + \mathcal{O}(t^3) \\ =& \frac{t^2}{2}(BA - AB) + \mathcal{O}(t^3). \end{aligned}$$

El término $BA - AB$ que interviene en esta diferencia se denomina *conmutador* y se denota de la siguiente manera

$$[A, B] = AB - BA,$$

donde el conmutador es la medida del grado de conmutatividad de las matrices A y B.

Por lo tanto

$$le(x;t) = \frac{t^2}{2}[A,B] + \mathcal{O}(h^3), \tag{D.1}$$

es decir, el error global es de orden dos.

Teorema D.1. *Dadas dos matrices cuadradas de orden $N \times N$ A y B se tiene*

$$e^{(A+B)} = \lim_{n\to\infty} (e^{\frac{A}{n}} e^{\frac{B}{n}})^n.$$

Demostración. En virtud de (D.1) se tiene

$$e^{(\frac{A}{n}+\frac{B}{n})t} = e^{\frac{A}{n}t} e^{\frac{B}{n}t} + \mathcal{O}\left(\frac{t^2}{n^2}\right). \tag{D.2}$$

En (D.2) el término principal del resto es

$$[A,B]/2n^2. \tag{D.3}$$

La fórmula (D.3) es rigurosa en el sentido que

$$\left\| e^{\frac{At}{n}+\frac{Bt}{n}} - e^{\frac{A}{n}t} e^{\frac{B}{n}t} \right\| \leq \frac{C}{n^2}|t^2|, \tag{D.4}$$

donde $C > 0$ es una constante independiente de n. La prueba de (D.4) puede realizarse utilizando el criterio de la mayorante de Weierstrass. Obviamente, la constante C depende de A y B pero conviene subrayar que es independiente del parámetro n.

Por (D.2) tenemos

$$e^{(A+B)t} = \left[e^{(A+B)\frac{t}{n}}\right]^n = \left[e^{\frac{A}{n}t} e^{\frac{B}{n}t} + \mathcal{O}\left(\frac{t^2}{n^2}\right)\right]^n.$$

Por otra parte

$$\left[e^{\frac{A}{n}t} e^{\frac{B}{n}t} + \mathcal{O}\left(\frac{t^2}{n^2}\right)\right]^n = \left\{e^{\frac{At}{n}} e^{\frac{Bt}{n}} \left[1 + e^{-\frac{A}{n}t} e^{-\frac{B}{n}t} \mathcal{O}\left(\frac{t^2}{n^2}\right)\right]\right\}^n.$$

Basta por lo tanto concluir que

$$\lim_{n\to\infty} \left(e^{\frac{A}{n}t} e^{\frac{B}{n}t}\right)^n = \lim_{n\to\infty} \left\{e^{\frac{A}{n}t} e^{\frac{B}{n}t} \left[1 + e^{-\frac{A}{n}t} e^{-\frac{B}{n}t} \mathcal{O}\left(\frac{t^2}{n^2}\right)\right]\right\}^n. \tag{D.5}$$

Denotando

$$C_n = e^{\frac{At}{n}} e^{\frac{Bt}{n}},$$

se trata de probar que

$$\lim_{n\to\infty} C_n^n = \lim_{n\to\infty} \left[C_n \left(1 + C_n^{-1} \mathcal{O}\left(\frac{t^2}{n^2}\right)\right)\right]^n. \tag{D.6}$$

Tenemos que

$$\left[C_n \left(1 + C_n^{-1} \mathcal{O}\left(\frac{t^2}{n^2}\right)\right)\right]^n = C_n^n \left(1 + C_n^{-1} \mathcal{O}\left(\frac{t^2}{n^2}\right)\right)^n,$$

aplicando la expansión de series de Taylor a la función $(1+x)^n$ en el punto $x = 0$, se tiene

$$(1+x)^n = 1 + nx,$$

sea $x = C_n^{-1}\mathcal{O}\left(\frac{t^2}{n^2}\right)$, entonces

$$\left(1 + C_n^{-1} \mathcal{O}\left(\frac{t^2}{n^2}\right)\right)^n = 1 + nC_n^{-1}\mathcal{O}\left(\frac{t^2}{n^2}\right),$$

por lo tanto

$$C_n^n \left(1 + C_n^{-1} \mathcal{O}\left(\frac{t^2}{n^2}\right)\right)^n = C_n^n + nC_n^{n-1}\mathcal{O}\left(\frac{t^2}{n^2}\right),$$

utilizando el teorema del Valor Medio tenemos

$$\begin{aligned}\left[C_n \left(1 + C_n^{-1} \mathcal{O}\left(\frac{t^2}{n^2}\right)\right)\right]^n - C_n^n =& nC_n^{n-1}\mathcal{O}\left(\frac{t^2}{n^2}\right)(1+\xi_n)^{n-1} \\ =& t^2\mathcal{O}\left(\frac{1}{n}\right) C_n^{n-1}(1+\xi_n)^{n-1},\end{aligned}$$

donde ξ_n es un elemento del segmento que une 0 con $C_n^{-1}\mathcal{O}\left(\frac{t^2}{n^2}\right)$.

Por tanto

$$\begin{aligned}\left\| \left[C_n \left(1 + C_n^{-1} \mathcal{O}\left(\frac{t^2}{n^2}\right)\right)\right]^n - C_n^n \right\| \leq& \frac{Ct^2}{n} \|C_n^{n-1}\| \|1+\xi_n\|^{n-1} \\ \leq& \frac{Ct^2}{n} \|C_n^{n-1}\| (1+\|\xi_n\|)^{n-1} \\ \leq& \frac{Ct^2}{n} \|C_n\|^{n-1} (1+\|\xi_n\|)^{n-1} \\ \leq& \frac{Ct^2}{n} \|e^{\frac{At}{n}} e^{\frac{Bt}{n}}\|^{n-1} (1+\|\xi_n\|)^{n-1} \\ \leq& \frac{Ct^2}{n} \|e^{\frac{At}{n}}\|^{n-1} \|e^{\frac{Bt}{n}}\|^{n-1} (1+\|\xi_n\|)^{n-1},\end{aligned}$$

entonces

$$\left\| \left[C_n \left(1 + C_n^{-1} \mathcal{O}\left(\frac{t^2}{n^2} \right) \right) \right]^n - C_n^n \right\| \leq \frac{Ct^2}{n} e^{\|A\| \frac{(n-1)t}{n}} e^{\|B\| \frac{(n-1)t}{n}} (1 + \|\xi_n\|)^{n-1}. \quad \text{(D.7)}$$

Ahora bien como

$$\begin{aligned} \|\xi_n\| \leq & \left\| C_n^{-1} \mathcal{O}\left(\frac{t^2}{n^2} \right) \right\| \\ \leq & \left\| C_n^{-1} \right\| \mathcal{O}\left(\frac{t^2}{n^2} \right) \\ \leq & \left\| e^{\frac{At}{n}} e^{\frac{Bt}{n}} \right\| \mathcal{O}\left(\frac{t^2}{n^2} \right) \\ \leq & e^{\|A\| \frac{t}{n}} e^{\|B\| \frac{t}{n}} \mathcal{O}\left(\frac{t^2}{n^2} \right) \\ = & \mathcal{O}\left(\frac{t^2}{n^2} \right), \end{aligned}$$

entonces

$$(1 + \|\xi_n\|)^{n-1} = \left[\left(1 + 1/\|\xi_n\|^{-1} \right)^{\|\xi_n\|^{-1}} \right]^{\|\xi_n\|(n-1)} \to e^0 = 1,$$

y por lo tanto, volviendo a (D.7) se deduce que

$$\left\| \left[C_n \left(1 + C_n^{-1} \mathcal{O}\left(\frac{t^2}{n^2} \right) \right) \right]^n - C_n^n \right\| = \mathcal{O}\left(\frac{1}{n} \right), \quad 0 \leq t \leq T.$$

Se obtiene por tanto (D.5)y (D.6) que es el resultado que se precisaba probar. ∎

Bibliografía

[1] ALCRUDO F., *Esquemas de alta resolución de variación total decreciente para el estudio de flujos discontinuos de superficie libre*, Tesis Doctoral. Facultad de Ciencias de la Universidad de Zaragoza. 1992.

[2] BURDEN R. Y FAIRES J., *Análisis Númerico*, Thomson Learning, 2002.

[3] CARRILLO J. A. AND VAZQUEZ J. L., *Fine asymptotics for fast diffusion equations,* Comm. Partial Differential Equations **Vol. 28** (2003), 1023-1056.

[4] COURANT R. and HILBERT D., *Methods of Mathematical Physics*, John Wiley and Sons. 1989.

[5] CHAUDHRY M. H., *Open Channel Flow*, Prentice Hall, New Jersey, 1993.

[6] CHAN T. F. AND ZHOU H. M., *Adaptive ENO-Wavelet Transforms for Discontinuous Functions,* 12th Int. Conf. on Domain Dec. Methods, No. 9, (2001).

[7] DIBENEDETTO EMMANUELE, *Partial Difference Equations*, Boston-Birkhäuser, 1997.

[8] DOUGLAS J. AND RUSSELL T. F., *Numerical methods for convection-dominated diffusion problems based on combining the method of characteristics with finite element or finite difference procedures*, SIAM Journal on Numerical Analysis **Vol. 19** (1982) No. 5, 871-885.

[9] DRIKAKIS DIMITRIS AND RIDER WILLIAM, *High-resolution methods for incompressible and low-speed flows*, Springer, 2005.

[10] FRITZ JOHN, *Partial Difference Equations*, Spring-Verlag: Applied Mathematical Sciences, 1980.

[11] GUENTHER R. AND LEE JOHN, *Parcial Difference Equations*, Dover, 1995.

[12] GLOWINSKI R., *Handbook of Numerical Analysis, Volume IX: Numerical Methods for Fluids*, North-Holland, 2003.

[13] HARTEN AMI, *High Resolution Schemes for Hyperbolic Conservation Laws*, Journal of Computational Physics **Vol. 49** (1983) No. 3, Pages 357-393.

[14] ISERLES A., *A First Course in the Numerical Analysis of Differential Equations*, Cambridge Texts in Applied Mathematics, Cambridge University Press. 1996.

[15] JEREZ GALIANO SILVIA AND UH ZAPATA MIGUEL, *A new TVD flux-limiter method for solving nonlinear hyperbolic equations*, Journal of Computational and Applied Mathematics **Vol. 234** (2010) , No. 5, 1395-1403.

[16] KIRSCH ANDREAS, *An Introduction to the Mathematical Theory of Inverse Problems*, Springer-Verlag, 1996.

[17] KOPRIVA DAVID A., *Implementing Spectral Methods for Partial Differential Equations: Algorithms for Scientists and Engineers*, Springer, 2009.

[18] KOZAKEVICIUS J. AND SANTOS L. C. C., *High resolution method for solving 1D Euler equation in wavelet*, Preprint, UFSM, IME, USP, Brazil (2003).

[19] KNABNER P. AND ANGERMANN L.,*Numerical Methods for Elliptic and Parabolic Partial Differential Equations*, Springer, 2003.

[20] LEVEQUE RANDALL J., *Finite difference methods for ordinary and partial differential equations: steady-state and time-dependent problems,* Philadelphia: SIAM 2007.

[21] LEVEQUE RANDALL J., *Numerical Methods for Conservation Laws*, Second Edition, Birkhäuser Verlag, Basel-Boston-Berlin, 1992.

[22] LIU Y., BANK R. E., DUPONT T. F., GARCIA S. AND SANTOS R. F., *Symmetric error estimates for moving mesh mixed methods for advection-diffusion equations*, SIAM Journal on Numerical Analysis **Vol. 40** (2003) No. 6, 2270-2291.

[23] MARCHUK G. I., *Handbook of numerical analysis, Vol. 1: Splitting and alternating direction methods*, North-Holland, 1990.

[24] PINSKY M., *Introduction to Parcial Difference equations*, McGraw Hill, 1984.

[25] RICHTMYER R. D. AND MORTON K. W., *Difference methods for initial value problems*, Second Edition, John Wiley and Sons, 1967.

[26] RODRIGUEZ M., *Numerical analysis of second order Lagrange-Galerkin schemes. Application to option pricing problems,* Tesis. Universidad de Santiago de Compostela 2005.

[27] SHOICHIRO NAKAMURA, *Métodos Numéricos Aplicados con Software*, Prentice-Hall, 1992.

[28] SHU C. W., *Essentially Non-Oscillatory and Weighted Essentially Non-Oscillatory Schemes for Hyperbolic Conservation Laws*, ICASE Report 97-65, November (1997).

[29] SMITH G., *Numerical Solution of Partial Differential Equations: Finite Difference Methods*, Oxford University Press, 1999.

[30] STRANG GILBERT, *On the Construction and Comparison of Difference Schemes*, SIAM Journal on Numerical Analysis, **Vol. 5** (1968) No. 3, 506-517.

[31] STRIKWERDA JOHN C., *Finite Difference Schemes and Partial Differential Equations*, Second Edition, Wadsworth&Brooks/Cole, 1989.

[32] TEMAN R., *Theory and Numerical Analysis of the Navier-Stokes equations*, North-Holland. 1977.

[33] TORO ELEUTERIO F., *Viscous flux limiters*, In Notes on Numerical Fluid Mechanics. Vieweg, Vos, Rizzi and Rhyming (Editors). **Vol. 35** (1992), 592-600.

[34] TORO ELEUTERIO F., *Riemann Solvers and Numerical Methods for Fluid Dynamics: A Practical Introduction,* Third Edition, Springer Verlag, Berlin, 2009.