Luis Ferragut Canals

Análisis Numérico del Método de Diferencias Finitas para Ecuaciones en Derivadas Parciales

Monografías de Análisis Numérico y Matemática Aplicada
Universidad de Salamanca

Índice general

Prefacio

La aproximación mediante Diferencias Finitas de las derivadas fue ya utilizada por Euler en 1768. El procedimiento más sencillo para aproximar $\frac{du}{dt}$ consiste en reemplazarlo por $\frac{u^{n+1}-u^n}{\Delta t}$ lo que llevó en el caso de un problema de valor inicial al método de Euler que se estudia en los cursos de resolución numérica de Ecuaciones Diferenciales Ordinarias.

Para las Ecuaciones en Derivadas Parciales la primera aplicación del Método de Diferencias Finitas se atribuye a Runge en 1908 que estudió la ecuación de Poisson

$$\frac{\partial u}{\partial x^2} + \frac{\partial u}{\partial y^2} = \text{constante}$$

Aproximadamente al mismo tiempo, Richardson en Inglaterra realizó una investigación similar. Su artículo de 1910 fue el primer trabajo en el que se aplicaron métodos iterativos para resolver problemas en derivadas parciales mediante diferencias finitas.

Las primeras demostraciones de convergencia fueron realizadas por Le Roux, Phillips, Winer y Courant entre otros. Algunos consideran el artículo de Courant-Friedrichs y Lewy (1928) como el nacimiento de la moderna teoría de métodos numéricos para Ecuaciones en Derivadas Parciales.

La aparición de los primeros ordenadores (ENIAC) en los años 50 y 60 permiten un mayor desarrollo del método. Cabe destacar las aportaciones de Von Neumann (1951), John (1952) y Lax, Douglas, Kreiss, Lees, Samarskii, Widlund (1960's)

Estas notas se han agrupado en 5 capítulos de los cuales los capítulos 1 al 4 están dedicados a problemas lineales, parabólicos y elípticos. El capítulo 5 está dedicado integramente a problemas hiperbólicos. Por una parte problemas lineales de primer y segundo orden y en sección aparte problemas hiperbólicos no lineales de primer orden.

En el capítulo 1, se tratan problemas parabólicos en dimensión 1 espacial con coeficientes constantes y nos sirve de introducción a los conceptos básicos de consistencia, estabilidad y convergencia.

En el capítulo 2 se abordan ya los problemas parabólicos con coeficientes variables y con diversas condiciones de contorno. Introducimos métodos de análisis de estabilidad más generales, en particular el análisis de estabilidad en la norma de la energía y en la norma del máximo.

En el capítulo 3 extendemos los análisis anteriores a problemas en dimensión espacial mayor que 1. Introducimos también en este capítulo el método de direcciones alternadas.

El capítulo 4 está dedicado a problemas elípticos de segundo orden. Que aparecen típicamente en los problemas estacionarios de difusión. Realizamos el análisis de estabilidad en la norma de la energía y también utilizamos el principio del máximo para obtener la estabilidad en la norma del máximo. Finalmente vemos como el método de direcciones alternadas se puede considerar aquí. En definitiva en este contexto el método de direcciones alternadas es un método iterativo para resolver el sistema de ecuaciones algebraico correspondiente y estudiamos la convergencia.

En el capítulo 5 se estudia el Método de Diferencias Finitas para resolver problemas hiperbólicos. En una primera sección se estudian brevemente aspectos generales de las ecuaciones hiperbólicas lineales, resaltando la posibilidad de considerar soluciones no continuas, lo que lleva a la introducción de soluciones generalizadas. En una primera subsección dedicada a métodos numéricos para problemas de primer orden nos limitamos a métodos explícitos de los que damos varios ejemplos. Estos se pueden encuadran en un método general de $2l+1$ pasos. Analizanos la consistencia, estabilidad y convergencia de algunos de ellos. En una subsección aparte estudiamos la resolución numérica mediante el Método de Diferencias Finitas de la ecuación de ondas, que es un ejemplo de problema hiperbólico de segundo orden. Nos limitamos también a analizar un método explícito. La parte principal de este capítulo 5 se dedica a los problemas hiperbólicos de primer orden no lineales. En particular se requiere introducir la noción de solución débil y un análisis detallado de estas soluciones. Notablemente el problema de valor inicial asociado a una ecuación hiperbólica no lineal no tiene solución única requiriendo condiciones adicionales para asegurar la unicidad, como es la condición de entropía. Entre todas las soluciones matemáticamente posibles la solución entrópica será la físicamente aceptable. Los métodos numéricos tendrán que adaptarse a esta situación y asegurarse que las soluciones numéricas obtenidas convergen a la solución entrópica.

En cuanto a las referencias en las que están inspiradas buena parte de estas notas, nos hemos limitado a dar 3 referencias básicas, y otras 2 complementarias, pudiendo el lector acudir a las referencias citadas en estos libros para tener una bibliografía más extensa. Recomendamos [1] y [2] para la parte de problemas lineales y [3] para problemas hiperbólicos no lineales. A modo de complemento se encontrará parte del material relacionado en [4] y [5].

Salamanca, Mayo 2021 *Luis Ferragut*

Capítulo 1
Problemas parabólicos en dimensión 1 espacial. Introducción

Resumen
En primer lugar se expone la metodología general para abordar el análisis numérico del Método de Diferencias Finitas introduciendo los conceptos de consistencia, estabilidad y convergencia. A continuación en este capítulo se estudian métodos en Diferencias Finitas para resolver problemas parabólicos en dimensión 1 espacial con coeficientes constantes. Se analizará la convergencia a partir de la consistencia y estabilidad de los esquemas, utilizando diversas técnicas para el análisis de estabilidad según los casos como el análisis de la estabilidad en la norma del máximo y el método matricial

1.1. Aspectos generales del análisis numérico del Método de Diferencias Finitas

De manera general supongamos que tenemos un problema asociado a una ecuación diferencial que en gran parte de los casos será una ecuación en derivadas parciales. Para que el problema esté bien determinado la formulación del problema se completará con condiciones de contorno y eventualmente condiciones iniciales. Todo ello lo representamos aquí de la siguiente manera:

$$\mathscr{A}(u) = f \tag{1.1}$$

donde $\mathscr{A}$ es el operador diferencial y en su caso el operador en derivadas parciales y que podemos suponer que incluye las condiciones de contorno y eventualmente la condición o condiciones iniciales.

$$\mathscr{A} = \begin{bmatrix} \mathbf{A} \\ \mathbf{B} \end{bmatrix}$$

por ejemplo,

$\mathbf{A} = \dfrac{\partial}{\partial t} - \dfrac{\partial}{\partial x}(a\dfrac{\partial}{\partial x})$

$\mathbf{B}$ = operador definiendo las condiciones de contorno y eventualmente iniciales.

El operador $\mathscr{A}$ actúa sobre funciones u definidas en un dominio de $\mathbb{R}^d$ o $\mathbb{R}^d \times R$. De modo que la solución u buscada es una función que toma valores en este dominio. La función u pertenece pues a un espacio funcional de dimensión infinita. De manera general los métodos numéricos para resolver el problema (1.1) consisten en sustituir éste por un problema algebraico. En particular si el operador $\mathscr{A}$ es lineal se sustituye por un problema (normalmente lineal) algebraico de ecuaciones. El Método de Diferencias Finitas es uno de esos métodos y consiste en buscar el valor de la solución u en un número finito de puntos del dominio en el que está definida. Sustituimos pues el problema (1.1) por un problema algebraico que representaremos así

$$\mathscr{A}_h(u_h) = f_h \tag{1.2}$$

Aquí h es un parámetro o parámetros que caracterizan la distribución de puntos del dominio elegidos donde aproximar (1.1). Nos referiremos a esta distribución de puntos como el mallado del dominio. En la práctica a medida que $h \to 0$ el número de puntos del mallado crece hacia ∞. Si el número de esos puntos es N entonces f_h está determinado por el valor de f en esos puntos. f_h es pues un vector de $\mathbb{R}^N$. Del mismo modo la solución u_h de (1.2) es a su vez un vector de $\mathbb{R}^N$. En consecuencia, tendremos que resolver un problema algebraico de N ecuaciones con N incógnitas. Si el operador $\mathscr{A}$ es lineal normalmente $\mathscr{A}_h$ será una matriz de N filas por N columnas y el sistema de ecuaciones es lineal. El propósito del análisis numérico será demostrar que la solución u_h aproxima la solución u en cierto sentido. Más precisamente, si $\{x_i;\ i = 1,...,N\}$ es el conjunto de puntos de la malla y u_i son la componentes de $u_h \in \mathbb{R}^N$ representando un valor aproximado de $u(x_i)$ en los puntos de la malla x_i, queremos evaluar el error $e_i = u(x_i) - u_i$ en todos los punto x_i. Llamando $e = (e_i)_i \in \mathbb{R}^N$ $i = 1,...,N$ evaluaremos la norma $||e||$. El procedimiento para evaluar este error se basa en dos conceptos que son la ***Consistencia*** y la ***Estabilidad***, que pasamos a definir a continuación, limitándonos por el momento a problemas lineales.

Definición 1.1 *El método (1.2) para resolver (1.1) es* ***Consistente*** *si se verifica*

$$\mathscr{A}_h\bar{u} - f_h = \tau_h \to 0 \text{ cuando } h \to 0 \tag{1.3}$$

donde $\bar{u}$ es el vector de $\mathbb{R}^N$ definido por $\bar{u}_i = u(x_i)$ para $i = 1,...,N$. *El vector $\tau_h = (\tau_i)_i \in \mathbb{R}^N$ recibe el nombre de error de consistencia. Si $\tau_i = \mathscr{O}(h^p)$ para todo $i = 1,...,N$ y por tanto $||\tau_h|| = \mathscr{O}(h^p)$, diremos que el error es de orden p.*

Definición 1.2 *El método (1.2) para resolver (1.1) es* ***Estable*** *si existe una constante C independiente de h tal que*

$$||v_h|| \leq C||\mathscr{A}_h v_h|| \quad \forall v_h \in \mathbb{R}^N \tag{1.4}$$

donde $||.||$ *es una norma en* $\mathbb{R}^N$

La consistencia y la estabilidad de un método implican la convergencia, como vemos en el siguiente

Teorema 1.1 *Si el método (1.2) para resolver (1.1) es* ***Consistente*** *y* ***Estable*** *el método es* ***Convergente****, en el sentido siguiente: Para* $e = (e_i)_i \in \mathbb{R}^N$ *definido por* $e_i = u(x_i) - u_i$ *tenemos*

$$||e|| \to 0 \quad \text{cuando } h \to 0 \tag{1.5}$$

Además si el error de consistencia τ_h *es de orden p, entonces* $||e|| = \mathscr{O}(h^p)$ *y diremos que el Método de Diferencias Finitas es de orden p.*

Demostración. Llamemos $\bar{u} \in \mathbb{R}^N$ al vector de componentes $\bar{u}_i = u(x_i)$ donde x_i son los puntos del mallado. Tendremos para el error $e = \bar{u} - u_h \in \mathbb{R}^N$

$$\begin{aligned} \mathscr{A}_h(\bar{u} - u_h) &= \mathscr{A}_h(\bar{u}) - \mathscr{A}_h(u_h) \\ &= \mathscr{A}_h(\bar{u}) - f_h = \tau_h \end{aligned}$$

de donde

$$||e|| = ||\bar{u} - u_h|| \leq C||\mathscr{A}_h(\bar{u} - u_h)|| = C||\tau_h|| \to 0 \text{ cuando } h \to 0$$

Naturalmente si $||\tau_h|| = \mathscr{O}(h^p)$ entonces $||e|| = \mathscr{O}(h^p)$ y el método es de orden p. ■

1.2. Métodos de Euler

1.2.1. Método de Euler explícito

En lo que sigue nos referiremos a t como la variable tiempo y mediante x, y variables que habitualmente en las aplicaciones representan las variables espaciales.

A modo introductorio consideraremos el método de Euler explícito para un problema parabólico en dimensión 1 espacial y coeficientes constantes. Sea un número real $D > 0$, y la función

$$f : [0,1] \times [0,T] \to \mathbb{R} \tag{1.6}$$

$$x, t \to f(x,t) \tag{1.7}$$

que supondremos al menos continua. Se quiere buscar la función

$$u : [0,1] \times [0,T] \to \mathbb{R} \tag{1.8}$$

$$x, t \to u(x,t) \tag{1.9}$$

verificando

$$\frac{\partial u}{\partial t} - D\frac{\partial^2 u}{\partial x^2} = f \quad \forall x \in (0,1),\ \forall t \in (0,T] \tag{1.10}$$

$$u(0,t) = u(1,t) = 0 \quad \forall t \in (0,T] \tag{1.11}$$

$$u(x,0) = v(x) \quad \forall x \in (0,1) \tag{1.12}$$

(1.10) es la ecuación en derivadas parciales, (1.11) son las condiciones de contorno y (1.12) es la condición inicial. El problema es un modelo simplificado de transmisión de calor en una barra de longitud unidad donde se busca la distribución de temperatura a lo largo de la barra. La función f representa una fuente de calor distribuida a lo largo de la barra. En los extremos se supone que se mantiene a la temperatura igual a cero. D representa el coeficiente de difusión térmica.

Vamos a construir un método de diferencias finitas para resolver de forma aproximada el problema anterior. Consideramos en el plano x,t el rectángulo $[0,1]\times[0,T]$. Sea $x_j,\ j = 0,1,...,M+1$ una partición de $[0,1]$ en subintervalos iguales de tamaño h de manera que $x_0 = 0$ y $x_{M+1} = 1$ y una partición $t_n,\ n = 0,1,...,N$ del intervalo $[0,T]$ con $t_0 = 0$ y $t_{N+1} = T$ en subintervalos de tamaño k dando lugar a una partición del rectángulo en subrectángulos según se muestra en la figura (1.1). Nos referiremos a esta partición del rectángulo como la malla o el mallado del mismo. En el mallado anterior, asociado a cada punto (x_j,t_n) del mismo consideraremos el

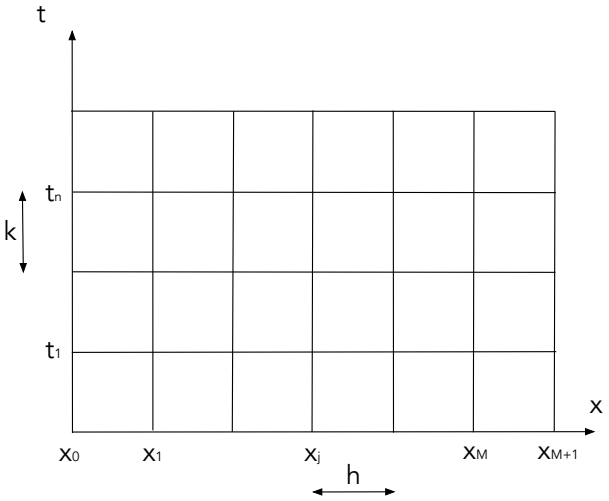


Figura 1.1 Malla de diferencias finitas

número u_j^n que representará una aproximación del valor exacto de la solución u del problema (1.10)-(1.11)-1.12) en este punto, es decir $u_j^n \approx u(x_j,t_n)$ con $x_j = jh$ y $t_n = nk$. Para cada valor de n tenemos definido un vector $u^n = (u_1^n, u_2^n, ..., u_M^n)^t \in \mathbb{R}^M$

del espacio euclídeo M-dimensional. En $\mathbb{R}^M$ podemos considerar distintas normas, p.e.,

$$||v||_\infty = \max_{1\le j\le M} |v_j| \tag{1.13}$$

$$||v|| = (h\Sigma(v_j)^2)^{1/2} \tag{1.14}$$

Operadores en diferencias

A continuación introducimos algunas notaciones para los operadores en diferencias finitas.

Asociado a un mallado como el descrito anteriormente sea

$$v^n = (v_1^n, v_2^n, ..., v_M^n)^t \in \mathbb{R}^M$$

Diferencia progresiva (respecto a x)

$$\partial_x v_j^n = \frac{v_{j+1}^n - v_j^n}{h} \tag{1.15}$$

Diferencia regresiva (respecto a x)

$$\bar{\partial}_x v_j^n = \frac{v_j^n - v_{j-1}^n}{h} \tag{1.16}$$

Diferencia progresiva (respecto a t)

$$\partial_t v_j^n = \frac{v_j^{n+1} - v_j^n}{k} \tag{1.17}$$

Diferencia regresiva (respecto a t)

$$\bar{\partial}_t v_j^n = \frac{v_j^n - v_j^{n-1}}{k} \tag{1.18}$$

El esquema de Euler explícito consiste en aproximar los términos de la ecuación en los puntos (x_j, t_n) mediante diferencias finitas y se escribe

$$\partial_t u_j^n - D\bar{\partial}_x(\partial_x u_j^n) = f_j^n = f(x_j, t_n) \quad j = 1, ..., M \; n = 1, ..., N \tag{1.19}$$

$$u_0^n = u_{M+1}^n = 0 \quad n = 1, ..., N \tag{1.20}$$

$$u_j^0 = v_j = v(x_j) \quad \text{para } j = 1 \text{ y } j = M \tag{1.21}$$

es decir

$$\frac{u_j^{n+1}-u_j^n}{k}-D\frac{u_{j-1}^n-2u_j^n+u_{j+1}^n}{h^2}=f_j^n \quad j=1,...,M;\ n=1,...,N \tag{1.22}$$

$$u_0^n=u_{M+1}^n=0 \quad n=1,...,N \tag{1.23}$$

$$u_j^0=v_j=v(x_j) \quad \text{para } j=1 \text{ y } j=M \tag{1.24}$$

la resolución del problema anterior es inmediata, pues dado $u_j^0=v_j$ obtenemos u_j^{n+1} para $j=1,...,M$

$$u_j^{n+1}=u_j^n+\frac{kD}{h^2}(u_{j-1}^n-2u_j^n+u_{j+1}^n)+kf_j^n \ j=1,...M$$
$$u_0^n=u_{M+1}^n=0$$

Pasamos ahora e estudiar la convergencia del método analizando la consistencia y la estabilidad.

Consistencia

Teorema 1.2 *El método (1.19)-(1.20)-(1.21) es consistente, con error de consistencia de orden* 2 *en la variable x y de orden* 1 *en la variable t, es decir,*

$$\tau_j^n=\mathcal{O}(h^2)+\mathcal{O}(k)$$

Demostración. Para probar la consistencia evaluamos la expresión que resulta al sustituir en la ecuación en diferencias la solución exacta, es decir, utilizando los desarrollos de Taylor en un entorno de (x_j,t_n):

$$u(x_j,t_n+k)=u(x_j,t_n)+k\frac{\partial u}{\partial t}(x_j,t_n)+\mathcal{O}(k^2)$$
$$\frac{u(x_j,t_n+k)-u(x_j,t_n)}{k}=\frac{\partial u}{\partial t}(x_j,t_n)+\mathcal{O}(k)$$
$$u(x_j+h,t_n)=u(x_j,t_n)+h\frac{\partial u}{\partial t}(x_j,t_n)+\frac{h^2}{2}\frac{\partial^2 u}{\partial x^2}(x_j,t_n)+\frac{h^3}{6}\frac{\partial^3 u}{\partial x^3}(x_j,t_n)+\mathcal{O}(h^4)$$
$$u(x_j-h,t_n)=u(x_j,t_n)-h\frac{\partial u}{\partial t}(x_j,t_n)+\frac{h^2}{2}\frac{\partial^2 u}{\partial x^2}(x_j,t_n)-\frac{h^3}{6}\frac{\partial^3 u}{\partial x^3}(x_j,t_n)+\mathcal{O}(h^4)$$
$$\frac{u(x_j+h,t_n)-2u(x_j,t_n)+u(x_j-h,t_n)}{h^2}=\frac{\partial^2 u}{\partial t}(u_j,t_n)+\mathcal{O}(h^2)$$

de donde el error de consistencia τ_j^n es:

$$\frac{u(x_j,t_n+k)-u(x_j,t_n)}{k}-D\frac{u(x_j-h,t_n)-2u(x_j,t_n)+u(x_j+h,t_n)}{h^2}-f(x_j,t_n)$$
$$=\tau_j^n(h,k)=\mathcal{O}(h^2)+\mathcal{O}(k)\to 0 \text{ cuando } h,k\to 0 \tag{1.25}$$

■

Estabilidad

Teorema 1.3 *Sea* $\lambda = \frac{kD}{h^2}$. *Bajo la condición* $\lambda \leq 1/2$ *el método (1.19)-(1.20)-(1.21) verifica*

$$||u^N||_\infty \leq |||u^0||_\infty + k\sum_{n=0}^{N-1} ||f^n||_\infty \leq |||u^0||_\infty + T \max_{n=0}^{N-1} ||f^n||_\infty \tag{1.26}$$

Diremos que el método es condicionalmente estable.

Demostración. Sea $\lambda = \frac{kD}{h^2} \leq 1/2$. El esquema anterior se escribe

$$\begin{aligned} u_j^{n+1} &= \lambda(u_{j-1}^n + u_{j+1}^n) + (1-2\lambda)u_j^n + kf_j^n \; j = 1, ...M \\ u_0^n &= u_{M+1}^n = 0 \\ u_j^0 &= v_j \end{aligned}$$

Tomando valores absolutos, puesto que $\lambda > 0$, $1-2\lambda \geq 0$ y la suma de coeficientes $\lambda + \lambda + (1-2\lambda) = 1$ resulta

$$|u_j^{n+1}| \leq \lambda|u_{j-1}^n| + \lambda|u_{j+1}^n| + (1-2\lambda)|u_j^n| + k|f_j^n|$$

y tomando el máximo para $j = 1, ..., M$, teniendo en cuenta las condiciones de contorno (1.20)

$$||u^{n+1}||_\infty \leq |||u^n||_\infty + k||f^n||_\infty$$

aplicando la anterior relación para $n = 0, 1, ..., N-1$

$$||u^N||_\infty \leq |||u^0||_\infty + k\sum_{n=0}^{N-1} ||f^n||_\infty \leq |||u^0||_\infty + T \max_{n=0}^{N-1} ||f^n||_\infty$$

■

Convergencia

La Consistencia y la estabilidad implican la convergencia.

Teorema 1.4 *Con la condición* $\lambda = \frac{kD}{h^2} \leq 1/2$ *que asegura la estabilidad el método (1.19)-(1.20)-(1.21) es convergente de orden* 2 *en h y de orden* 1 *en k.*

Demostración. Consideremos ahora el error

$$e_j^n = u(x_j,t_n) - u_j^n \; j = 0,...,M+1;\; n = 0,...,N$$

El correspondiente vector de $\mathbb{R}^M$, $e^n = (e_1^n, e_2^n, ..., e_M^n)^t \in \mathbb{R}^M$ verifica las ecuaciones en diferencias

$$\frac{e_j^{n+1} - u_j^n}{k} - D\frac{e_{j-1}^n - 2u_j^n + e_{j+1}^n}{h^2} = \tau_j^n \quad j = 1,...,M;\; n = 1,...,N \tag{1.27}$$

$$e_0^n = e_{M+1}^n = 0 \quad n = 1,...,N \tag{1.28}$$

$$e_j^0 = 0 \quad \text{para } j = 1 \text{ y } j = M \tag{1.29}$$

que se obtienen restando las ecuaciones en diferencias (1.22), y (1.25) y teniendo en cuenta que $u_j^0 = v_j = v(x_j)\; \forall j = 0,1,...,M+1$. Finalmente aplicando la propiedad de estabilidad (1.26) a las ecuaciones del error (1.27)-(1.28)-(1.29) obtenemos

$$||e^n||_\infty \leq T \max_{n=0}^{n-1} ||\tau^n||_\infty \; \forall n = 0,1,...,N \tag{1.30}$$

gracias a la propiedad de consistencia (1.25). Tenemos pues que el método es de orden 2 en h y de orden 1 en k, bajo la condición $\lambda \leq 1/2$, es decir, $\frac{kD}{h^2} \leq \frac{1}{2}$, o bien, $k \leq \frac{h^2}{2D}$ lo que obliga a tomar valores de k muy pequeños. ■

1.2.2. Método de Euler implícito

En el esquema de Euler implícito para resolver el problema (1.10-(1.11)-(1.12) aproximamos los términos de la ecuación en los puntos (x_j, t_{n+1}) utilizando diferencias finitas.

$$\partial_t u_j^n - D\bar{\partial}_x(\partial_x u_j^{n+1}) = f_j^{n+1} = f(x_j,t_{n+1})\; j = 1,...,M\,;n = 0,...,N \tag{1.31}$$

$$u_0^n = u_{M+1}^n = 0\; n = 1,...,N \tag{1.32}$$

$$u_j^0 = v_j = v(x_j) \quad \text{para } j = 1 \text{ y } j = M \tag{1.33}$$

es decir,

$$\frac{u_j^{n+1} - u_j^n}{k} - D\frac{u_{j-1}^{n+1} - 2u_j^{n+1} + u_{j+1}^{n+1}}{h^2} = f_j^{n+1} \quad j = 1,...,M\,;n = 0,...,N \tag{1.34}$$

$$u_0^n = u_{M+1}^n = 0 \quad n = 1,...,N \tag{1.35}$$

$$u_j^0 = v_j = v(x_j) \quad \text{para } j = 1 \text{ y } j = M \tag{1.36}$$

Consistencia

Análogamente al caso de Euler explícito obtenemos que el error de consistencia es $\tau_j^{n+1}(h,k) = \mathscr{O}(h^2) + \mathscr{O}(k) \to 0$ cuando $h,k \to 0$.

Estabilidad

Teorema 1.5 *El método (1.31)-(1.32)-(1.33) verifica*

$$||u^N||_\infty \leq ||u^0||_\infty + k\sum_{n=0}^{N-1} ||f^n||_\infty \leq ||u^0||_\infty + T \max_{n=0}^{N-1} ||f^n||_\infty \tag{1.37}$$

Diremos que el método es incondicionalmente estable.

Demostración. Con las mismas notaciones que en el caso explícito, tenemos

$$\begin{aligned} &-\lambda u_{j+1}^{n+1} + (1+2\lambda)u_j^{n+1} - \lambda u_{j-1}^{n+1} = u_j^n + kf_j^{n+1} \\ &u_0^n = u_{M+1}^n = 0 \\ &u_j^0 = v_j \end{aligned}$$

que podemos escribir así

$$(1+2\lambda)u_j^{n+1} = \lambda u_{j+1}^{n+1} + \lambda u_{j-1}^{n+1} + u_j^n + kf_j^{n+1}$$

tomando valores absolutos

$$(1+2\lambda)|u_j^{n+1}| \leq \lambda |u_{j+1}^{n+1}| + \lambda |u_{j-1}^n| + |u_j^n| + k|f_j^{n+1}|$$

y tomando el máximo para todo j

$$\begin{aligned} (1+2\lambda)||u^{n+1}||_\infty &\leq 2\lambda ||u^{n+1}||_\infty + ||u^n||_\infty + k||f^{n+1}||_\infty \\ ||u^{n+1}||_\infty &\leq ||u^n||_\infty + k||f^{n+1}||_\infty \end{aligned}$$

y finalmente aplicando la desigualdad anterior recursivamente tomando sucesivamente $n = 0,1,2,...,N-1$

$$||u^N||_\infty \leq |||u^0||_\infty + k\sum_{n=0}^{N-1} ||f^n||_\infty \leq |||u^0||_\infty + T \max_{n=0}^{N-1} ||f^n||_\infty$$

■

Convergencia

La consistencia y la estabilidad nos dan la convergencia. En efecto tendremos, razonando como en el caso explícito, para el error $e^n = (e_1^n, e_2^n, ..., e_M^n)^t \in \mathbb{R}^M$

$$||e^n||_\infty \leq T \max_{n=0}^{n-1} ||\tau^n||_\infty \; \forall n = 0, 1, ..., N \tag{1.38}$$

Tenemos pues que el método es de orden 2 en h y de orden 1 en k.

Ejercicio

Considerar el problema

$$\frac{\partial u}{\partial t} - \frac{\partial^2 u}{\partial x^2} = 0 \quad \forall x \in (0,1), \forall t \in (0,T] \tag{1.39}$$

$$u(0,t) = u(1,t) = 0 \tag{1.40}$$

$$u(x,0) = v(x) \tag{1.41}$$

y el correspondiente método de Euler explícito

$$\partial_t u_j^n - \bar{\partial}_x(\partial_x u_j^n) = 0 \quad j = 1, ..., M; \; n = 1, ..., N \tag{1.42}$$

$$u_0^n = u_{M+1}^n = 0 \quad n = 1, ..., N \tag{1.43}$$

$$u_j^0 = v_j = v(jh) \quad \text{para } j = 1 \text{ y } j = M \tag{1.44}$$

Demostrar que la condición $\lambda \leq 1/2$ es necesaria para la estabilidad, considerando el caso particular

$$u_j^0 = v_j = (-1)^j \sin(\pi jh) = (-1)^j Im(e^{i(\pi jh)})$$

Solución

$$\begin{aligned}
u_j^1 &= \lambda(u_{j+1}^0 + u_{j-1}^0) + (1-2\lambda)u_j^0 \\
&= Im\Big(\lambda(-1)^{j+1} e^{i(\pi(j+1)h)} + \lambda(-1)^{j-1} e^{i(\pi(j-1)h)} + (1-2\lambda)(-1)^j e^{i(\pi jh)}\Big) \\
&= Im\left(e^{i(\pi jh)}\Big((-1)^{j+1}\big(2\lambda(\frac{e^{i\pi h} + e^{-i\pi h}}{2}) + (1-2\lambda)(-1)^j\big)\Big)\right) \\
&= Im(-1)^j e^{i(\pi jh)}\Big(1 - 2\lambda - 2\lambda\cos(\pi h)\Big)
\end{aligned}$$

tenemos pues,

$$u_j^1 = \big(1 - 2\lambda - 2\lambda\cos(\pi h)\big)u_j^0$$

y aplicando recursivamente este procedimiento

$$u_j^n = \big(1 - 2\lambda - 2\lambda\cos(\pi h)\big)^n u_j^0$$

Si $h \to 0$ y $\lambda > 1/2$

$$|1 - 2\lambda - 2\lambda\cos(\pi h)| = |2\lambda + 2\lambda\cos(\pi h) - 1| \geq \gamma > 1$$

lo que implica $||u^n||_\infty \geq \gamma^n ||u^0||_\infty \to \infty$ cuando $k, h \to 0$

1.3. Método matricial de análisis de la estabilidad

Consideraremos aquí un método general de un paso para resolver el problema (1.10)-(1.11)-(1.12).

Dado θ, $0 \leq \theta \leq 1$ y $t_{n+\theta} = (1-\theta)t_n + \theta t_{n+1}$ el método numérico general de un paso se escribe así:

$$\partial_t u_j^n - D\bar{\partial}_x(\partial_x u_j^{n+\theta}) = f_j^{n+\theta} = f(x_j, t_{n+\theta}) \quad n \geq 0,\ j = 1, ..., M \tag{1.45}$$

$$u_0^n = u_M^n = 0 \quad n > 0 \tag{1.46}$$

$$u_j^0 = v_j = v(x_j) \quad j = 1, ..., M \tag{1.47}$$

donde $u_j^{n+\theta} = (1-\theta)u_j^n + \theta u_j^{n+1}$. Denotando, como en la sección anterior $\lambda = \frac{Dk}{h^2}$ el esquema anterior se puede escribir con notación matricial así,

$$(\mathbf{I} + \theta\lambda\mathbf{A})u^{n+1} = \big(\mathbf{I} - (1-\theta)\lambda\mathbf{A}\big)u^n + f^{n+\theta}$$
$$u^0 = v$$

donde $\mathbf{A}$ es la matriz tridiagonal

$$\mathbf{A} = \begin{bmatrix} 2 & -1 & \dots & \dots & 0 \\ -1 & 2 & \dots & \dots & 0 \\ & \ddots & \ddots & \ddots & \\ & & & & -1 \\ 0 & 0 & \dots & -1 & 2 \end{bmatrix}$$

de modo que en cada paso hay que resolver un sistema lineal de ecuaciones, con matriz tridiagonal en todos los casos, salvo en el caso $\theta = 0$ en cuyo caso la matriz del sistema es la identidad. El caso $\theta = 0$ es el método de Euler explícito mientras que el caso $\theta = 1$ corresponde al método de Euler implícito. Para todos los valores

de θ el método es consistente y se tiene

$$\tau_j^{n+\theta} = \begin{cases} \mathscr{O}(h^2) + \mathscr{O}(k) & \text{si } \theta \neq \frac{1}{2} \\ \mathscr{O}(h^2) + \mathscr{O}(k^2) & \text{si } \theta = \frac{1}{2} \end{cases}$$

El caso $\theta = 1/2$ se conoce como el método de Crank-Nicolson. Veamos que efectivamente es un método de orden dos tanto en la discretización con respecto a la variable x como con respecto a la variable t. La consistencia de los casos de Euler implícito y explícito ya ha sido estudiada en la sección anterior. Para el resto de valores de θ el estudio de la consistencia se deja como ejercicio.

Consistencia del método de Crank-Nicolson

Teorema 1.6 *El método (1.45)-(1.46)-(1.47) para $\theta = 1/2$ es consistente de orden 2 en h y de orden 2 en k.*

Demostración. Las equaciones (1.45)-(1.46)-(1.47) para $\theta = 1/2$ se escriben de forma desarrollada así

$$\frac{u_j^{n+1} - u_j^n}{k} - \frac{1}{2} D\Big(\frac{u_{j-1}^{n+1} - 2u_j^{n+1} + u_{j+1}^{n+1}}{h^2} + \frac{u_{j-1}^n - 2u_j^n + u_{j+1}^n}{h^2} \Big) = f_j^{n+1/2} \quad (1.48)$$

$$u_0^n = u_{M+1}^n = 0 \quad n = 1, ..., N \quad (1.49)$$

$$u_j^0 = v_j = v(x_j) \quad \text{para } j = 1 \text{ y } j = M \quad (1.50)$$

Consideremos primero los siguientes desarrollos de Taylor para una función v dependiente de una variable t

$$v(t+k) = v(t+\frac{k}{2}) + \frac{\partial v}{\partial t}(t+\frac{k}{2})\frac{k}{2} + \frac{1}{2}\frac{\partial^2 v}{\partial t^2}(t+\frac{k}{2})(\frac{k}{2})^2 + \mathscr{O}(k^3)$$

$$v(t) = v(t+\frac{k}{2}) - \frac{\partial v}{\partial t}(t+\frac{k}{2})\frac{k}{2} + \frac{1}{2}\frac{\partial^2 v}{\partial t^2}(t+\frac{k}{2})(\frac{k}{2})^2 + \mathscr{O}(k^3)$$

restando y dividiendo por k tenemos

$$\frac{v(t+k) - v(t)}{k} = \frac{\partial v}{\partial t}(t+\frac{k}{2}) + \mathscr{O}(k^2) \quad (1.51)$$

y sumando y dividendo por 2

$$\frac{v(t+k) + v(t)}{2} = v(t+\frac{k}{2}) + \mathscr{O}(k^2) \quad (1.52)$$

Para estudiar el error de consistencia sustituimos en la expresión (1.48) la solución exacta y evaluamos la diferencia con la ecuación exacta (1.10). Utilizando

las expresiones (1.51) y (1.52) obtenemos para la solución exacta de (1.10)-(1.11)-(1.12), evaluando todo sus términos para el valor de $t = t_n + k/2$

$$\begin{aligned}
&\frac{u(x_j,t_{n+1}) - u(x_j,t_n)}{k} = \frac{\partial u}{\partial t}(x_j,t_n + k/2) + \mathscr{O}(k^2)\\
&\frac{u(x_{j+1},t_{n+1}) - 2u(x_j,t_{n+1}) + u(x_{j-1},t_{n+1})}{h^2} = \frac{\partial^2 u}{\partial x^2}(x_j,t_{n+1}) + \mathscr{O}(h^2)\\
&\frac{u(x_{j+1},t_n) - 2u(x_j,t_n) + u(x_{j-1},t_n)}{h^2} = \frac{\partial^2 u}{\partial x^2}(x_j,t_n) + \mathscr{O}(h^2)\\
&\frac{1}{2}\big(f(x_j,t_{n+1} + f(x_j,t_n)\big) = f(x_j,t_n + k/2) + \mathscr{O}(k^2)
\end{aligned}$$

juntando estas estimaciones obtenemos que el error de consistencia en $(x_j, t_n + k/2)$ es

$$\begin{aligned}
\tau_j^{n+1/2} &= \frac{u(x_j,t_{n+1}) - u(x_j,t_n)}{k}\\
-\frac{1}{2}D\Big(&\frac{u(x_{j+1},t_{n+1}) - 2u(x_j,t_{n+1}) + u(x_{j-1},t_{n+1})}{h^2} + \frac{u(x_{j+1},t_n) - 2u(x_j,t_n) + u(x_{j-1},t_n)}{h^2}\Big)\\
&- \frac{1}{2}\Big(f(x_j,t_{n+1} + f(x_j,t_n\Big)\\
&= \frac{\partial u}{\partial t}(x_j,t_n + k/2) - D\frac{\partial^2 u}{\partial x^2}(x_j,t_n + k/2) - f(x_j,t_n + k/2) + \mathscr{O}(k^2) + \mathscr{O}(h^2)\\
&= \mathscr{O}(k^2) + \mathscr{O}(h^2)
\end{aligned}$$

■

Estudio de la estabilidad

El esquema general de un paso descrito anteriormente se puede escribir

$$u^{n+1} = \big(\mathbf{I} + \theta\lambda\mathbf{A}\big)^{-1}\big(\mathbf{I} - (1-\theta)\lambda\mathbf{A}\big)u^n + k\big(\mathbf{I} + \theta\lambda\mathbf{A}\big)^{-1} f^{n+\theta}$$

o bien escribiendo

$$\mathbf{T} = \big(\mathbf{I} + \theta\lambda\mathbf{A}\big)^{-1}\big(\mathbf{I} - (1-\theta)\lambda\mathbf{A}\big)$$

y

$$\mathbf{S} = \big(\mathbf{I} + \theta\lambda\mathbf{A}\big)^{-1}$$

$$u^{n+1} = \mathbf{T}u^n + k\mathbf{S}f^{n+\theta} \tag{1.53}$$

Teorema 1.7 *Si* $||\mathbf{T}|| \leq 1$ *el método (1.53) es estable y se verifica*

$$||u^n|| \leq ||u^0|| + T||\mathbf{S}|| \max_l ||f^{l+\theta}||$$

donde T es el valor máximo de t.

Demostración. Sea $||.||$ una norma vectorial en $\mathbb{R}^M$ y para una matriz $\mathbf{B}$ sea

$$||\mathbf{B}|| = \sup_{v \neq 0} \frac{||\mathbf{B}v||}{||v||}$$

la correspondiente norma matricial inducida por la anterior norma vectorial. Aplicando recursivamente la relación (1.53) tenemos

$$u^n = \mathbf{T}^n u^0 + k \sum_{l=0}^{n-1} \mathbf{T}^{n-1-l} \mathbf{S} f^{n+\theta}$$

y tomando normas

$$||u^n|| \leq ||\mathbf{T}||^n ||u^0|| + k \sum_{l=0}^{n-1} ||\mathbf{T}||^{n-1-l} ||\mathbf{S}||.||f^{l+\theta}||$$

Si $||\mathbf{T}|| \leq 1$ el esquema anterior es estable pues

$$||u^n|| \leq ||u^0|| + k \sum_{l=0}^{n-1} ||\mathbf{S}||.||f^{l+\theta}||$$

$$||u^n|| \leq ||u^0|| + kn||\mathbf{S}|| \max_l ||f^{l+\theta}||$$

y finalmente

$$||u^n|| \leq ||u^0|| + T||\mathbf{S}|| \max_l ||f^{l+\theta}||$$

donde T es el valor máximo de t ■

Condiciones suficiente de estabilidad

La condición suficiente de estabilidad es $||\mathbf{T}|| \leq 1$. Si $\mathbf{T}$ es simétrica y $||.||$ es la norma euclídea entonces $||\mathbf{T}|| = \rho(\mathbf{T})$ donde $\rho(\mathbf{T})$ es el radio espectral de $\mathbf{T}$, es decir el máximo del conjunto de los módulos de los valores propios de $\mathbf{T}$. Los valores propios de $\mathbf{T}$ son a su vez función de los valores propios de $\mathbf{A}$.

Si β_p son los valores propios de $\mathbf{A}$, entonces los valores propios de $\mathbf{T}$ son:

$$\frac{1-(1-\theta)\lambda\beta_p}{1+\theta\lambda\beta_p} \tag{1.54}$$

La condición de estabilidad $\rho(\mathbf{T}) = ||\mathbf{T}|| \leq 1$ es entonces

$$\left| \frac{1-(1-\theta)\lambda\beta_p}{1+\theta\lambda\beta_p} \right| \leq 1 \ \forall p = 1, ..., M \tag{1.55}$$

Como **A** es simétrica definida positiva, los valores propios $\beta_b > 0$ y en

$$-1-\theta\lambda\beta_p \leq 1-(1-\theta)\lambda\beta_p \leq 1+\theta\lambda\beta_b$$

la segunda desigualdad se cumple siempre. De la primera desigualdad obtenemos

$$(1-2\theta)\lambda\beta_p \leq 2$$

Si $\theta \geq 1/2$ la desigualdad anterior se cumple siempre, y si $\theta < 1/2$ la condición de estabilidad se cumplirá si

$$\lambda \leq \frac{2}{(1-2\theta)\beta_p} \quad \forall p$$

es decir, la condición de estabilidad será para $\theta < 1/2$

$$k \leq \frac{2h^2}{(1-2\theta)D\beta_{max}}$$

donde β_{max} es el valor propio máximo.

Cálculo de los valores propios de A

Para completar el estudio vamos a calcular los valores propios de **A**. Para ello hay que resolver el problema algebraico de valores y vectores propios

$$\mathbf{A}w = \beta w \tag{1.56}$$

Si **A** es simétrica definida positiva, existen soluciones (w_p, β_p) $p = 1, ..., M$ donde $(w_p)_p$ son los vectores propios, y β_p son los valores propios.

La ecuación j-ésima de (1.56) es

$$-w_{j+1} + w_j - w_{j+1} = \beta w_j \tag{1.57}$$

Por analogía con las ecuaciones diferenciales de segundo orden ($y'' = \beta y$ cuyas soluciones son del tipo $y = e^{ir\pi x}$), probaremos soluciones del tipo

$$w_p = (w_{p,j})_{j=1}^M = (e^{i(jh\pi p)})_{j=1}^M$$

Sustituyendo en (1.57) se obtiene

$$\begin{aligned}
&-e^{i(j+1)h\pi p} + 2e^{ijh\pi p} - e^{i(j-1)h\pi p} = \beta_p e^{ijh\pi p} \\
&(-e^{ih\pi p} + 2 - e^{-ih\pi p})e^{i(jh\pi p)} = \beta_p e^{ijh\pi p} \\
&2(1-cos(h\pi p))e^{i(jh\pi p)} = \beta_p e^{ijh\pi p}
\end{aligned}$$

es decir $\beta_p = 2(1-cos(h\pi p))$. Si tenemos en cuenta las condiciones de contorno tomaremos la parte imaginaria de w_p,

$$w_p = (w_p^j)_{j=1}^M = (\sin(jh\pi p))_{j=1}^M$$

de esta forma w_p verifica las condiciones de contorno, en efecto

$$w_p^0 = 0 \quad w_p^{M+1} = \sin((M+1)h\pi p) = 0$$

Los M vectores propios son $w_p,\ p = 1,...,M$ y los M valores propios son

$$\beta_p = 2(1-\cos(h\pi p))\ p = 1,...,M$$

Observemos que para $p \geq M+1$ se repiten los valores y los vectores propios. Por otra parte, $\beta_{\text{máx}} = \beta_M = 2(1-\cos(h\pi M)) < 4$ y concluimos que una condición suficiente de estabilidad para $\theta < 1/2$ es

$$k \leq \frac{h^2}{2(1-2\theta)D}$$

Convergencia

La consistencia y la estabilidad implican la convergencia, en efecto, para el error tenemos

$$\begin{aligned} ||e^n|| &\leq ||e^0|| + ||\mathbf{S}||T \max_l ||\tau^{l+\theta}|| \\ &\leq ||\mathbf{S}||TC(k^r + h^2) \to 0 \text{ cuando } k,h \to 0 \end{aligned}$$

donde hemos tenido en cuenta que

$$\tau_j^{n+\theta} = \mathcal{O}(k^r + h^2) \text{ donde} \begin{cases} r = 1 \text{ si } \theta \neq \dfrac{1}{2} \\ r = 2 \text{ si } \theta = \dfrac{1}{2} \end{cases}$$

La constante C depende de las derivadas de la solución exacta u, según se desprende de los desarrollos de Taylor efectuados al estudiar el error de consistencia.

Observemos también que

$$||S|| = \rho(S) = \frac{1}{1+\theta\lambda\beta_1} < 1$$

y por lo tanto,

$$\begin{aligned} ||e^n|| &\leq ||e^0|| + T \max_l ||\tau^{l+\theta}|| \\ &\leq TC(k^r + h^2) \to 0 \text{ cuando } k,h \to 0 \end{aligned}$$

Ejercicio

Considerar el esquema de Euler explícito para el problema homogéneo, es decir

$$\frac{u_j^{n+1}-u_j^n}{k}-D\frac{u_{j-1}^n-2u_j^n+u_{j+1}^n}{h^2}=0 \quad j=1,...,M;\ n\geq 1$$
$$u_0^n=u_{M+1}^n=0 \quad n\geq 1$$
$$u_j^0=v_j=v(x_j) \quad j=1,...,M$$

1. Hallar la solución del esquema en diferencias utilizando la versión discreta del método de variables separadas, es decir, ensayar soluciones de la forma

$$u_j^n=\gamma^n w_j$$

2. Comparar la solución obtenida con la solución exacta.
3. Deducir la condición de estabilidad por comparación con la solución exacta, es decir, elegir k de modo que $\lim_{n\to\infty} u_j^n=0$

Solución de la parte 1

Sustiyuyendo $u_j^n=\gamma^n w_j$ en las ecuaciones

$$\frac{\gamma^{n+1}w_j-\gamma^n w_j}{k}-D\frac{\gamma^n w_{j-1}-2\gamma^n w_j+\gamma^n w_{j+1}}{h^2}=0 \quad j=1,...,M;\ n\geq 1 \tag{1.58}$$
$$\gamma^n w_0=\gamma^n w_{M+1}=0 \quad n\geq 1 \tag{1.59}$$
$$\gamma^0 w_j=v_j \quad j=1,...,M \tag{1.60}$$

Las ecuaciones (1.58) se pueden escribir de la forma

$$\frac{\gamma^{n+1}-\gamma^n}{\lambda\gamma^n}=\frac{w_{j-1}-2w_j+w_{j+1}}{w_j} \quad j=1,...,M;n\geq 1$$

donde $\lambda=\frac{Dk}{h^2}$. Puesto que el primer miembro de las ecuaciones solo depende de n y el segundo solo depende de j ambos miembros tienen un valor constante que denotamos $-\beta$, de modo que

$$\gamma^{n+1}-\gamma^n=-\beta\lambda\gamma^n \quad n=1,2,...$$
$$-w_{j-1}+2w_j-w_{j+1}=\beta w_j \quad j=1,...,M$$

el segundo bloque de ecuaciones es el problema de valores y vectores propios de la matriz

$$\mathbf{A} = \begin{bmatrix} 2 & -1 & \dots & \dots & 0 \\ -1 & 2 & \dots & \dots & 0 \\ & \ddots & \ddots & \ddots & \\ & & & & -1 \\ 0 & 0 & \dots & -1 & 2 \end{bmatrix}$$

cuyos valores propios son, según hemos visto anteriormente

$$\beta_p = 2(1-\cos(h\pi p)) \; p = 1, ..., M$$

y sus correspondiente vectores propios

$$w_p = (w_p^j)_{j=1}^M = (\sin(jh\pi p))_{j=1}^M$$

que se han escogido de modo que el vector ampliado

$$\tilde{w}_p = (\tilde{w}_p^j)_{j=0}^{M+1} = (\sin(jh\pi p))_{j=0}^{M+1}$$

verifique las condiciones de contorno (1.59).

Resolvamos ahora la parte temporal.

$$\begin{aligned} \gamma^{n+1} - \gamma^n &= -\beta_p \lambda \gamma^n \\ \gamma^{n+1} - (1-\beta_p\lambda)\gamma^n &= 0 \\ \gamma^n(\gamma - (1-\beta_p\lambda)) &= 0 \\ \gamma &= 1-\beta_p\lambda \end{aligned}$$

Finalmente, una expresión de la forma

$$(1-\beta_p\lambda)^n \tilde{w}_{p,j} = (1-\beta_p\lambda)^n \sin(j\pi hp) \quad j = 0, ..., M+1$$

verifica la ecuación en diferencias y también las condiciones de contorno. Sin embargo esta solución no verifica la condición inicial. Para obtener una solución que verifique la condición inicial tomamos una combinación lineal

$$u_j^n = \sum_{p=1}^M C_p (1-\beta_p\lambda)^n w_p^j$$

de manera que para $n = 0$ sea igual a v_j, es decir,

$$v_j = \sum_{p=1}^M C_p w_{p,j} = \sum_{p=1}^M C_p \sin(j\pi hp)$$

Para calcular C_p utilizamos la ortogonalidad de los vetores w_p,

$$(w_p, w_q) = h\sum_{j=1}^{M} \sin(j\pi hp)\sin(j\pi hq) = \begin{cases} \frac{1}{2} & \text{si } p = q \\ 0 & \text{si } p \neq q \end{cases}$$

Para $v = (v_j)_{j=1}^{M}$ tendremos

$$\begin{aligned}(v, w_q) &= h\sum_{j}^{M} v_j \sin(j\pi hq) \\ &= h\sum_{j}\sum_{p} C_p \sin(j\pi hp)\sin(j\pi hq) = C_q\frac{1}{2}\end{aligned}$$

de donde

$$C_q = 2h\sum_{j} v_j \sin(j\pi hq)$$

Finalmente la solución es

$$u_j^n = 2h\Big(\sum_{p=1}^{M}\big(\sum_{l} v_l \sin(l\pi hp)\big)\big(1-2(1-\cos(h\pi p))\lambda\big)^n \sin(j\pi hp)\Big) \tag{1.61}$$

Ejercicio

Considerar el siguiente método de Richardson para resolver

$$\frac{\partial u}{\partial t} - \frac{\partial^2 u}{\partial x^2} = 0 \tag{1.62}$$
$$u(0,t) = u(1,t) = 0 \tag{1.63}$$
$$u(x,0) = v(x) \tag{1.64}$$

El método de Richardson es

$$\hat{\partial}_t u_j^n \bar{\partial}_x \partial_x u_j^n = 0 \tag{1.65}$$
$$u_0^n = u_{M+1}^n = 0 \tag{1.66}$$
$$u_j^0 = v_j \tag{1.67}$$

donde

$$\hat{\partial}_t u_j^n = \frac{1}{2}(\partial_t u_j^n + \bar{\partial}_t u_j^n) = \frac{u_j^{n+1} - u_j^{n-1}}{2k}$$

1. Demostrar que el método es consistente de orden $h^2 + k^2$
2. Demostrar que el método es inestable.

Solución de la parte 2

Utilizando el método de separación de variables, ensayamos una solución de la forma

$$u_j^n = \gamma^n w_j$$

sustituyendo en la ecuación en diferencias (1.65) resulta

$$\frac{\gamma^{n+1}-\gamma^{n-1}}{2k} w_j - \gamma^n \frac{-w_{j-1}+2w_j-w_{j+1}}{h^2} = 0$$
$$\frac{\gamma^{n+1}-\gamma^{n-1}}{2\lambda\gamma^n} = \frac{-w_{j-1}+2w_j-w_{j+1}}{w_j} = -\beta$$

de donde razonando como en el ejercicio anterior, poniendo $\lambda = \frac{k}{h^2}$

$$\gamma^{n+1} + 2\lambda\beta\gamma^n - \gamma^{n-1} = 0$$
$$w_{j-1} - 2w_j + w_{j+1} = -\beta w_j$$

de la primera ecuación obtenemos que γ es solución de

$$\gamma^2 + 2\lambda\beta\gamma - 1 = 0$$

cuyas soluciones son

$$\gamma_\pm = -\lambda\beta \pm \sqrt{1+\lambda^2\beta^2} \tag{1.68}$$

y la solución $|\gamma_-| > 1$. Por lo tanto la solución verifica

$$\lim_{n\to\infty} u_j^n \to \infty$$

y el método es inestable.

Aclaración:

El método inicialmente necesita dos valores iniciales u^0 y u^1 para poder calcular $u^n, n = 2,3,....$ La solución general se obtiene mediante una combinación lineal de las 2 soluciones $\gamma_\pm$. Las constantes de esta combinación lineal vendrán determinadas por los dos valores iniciales u^0 y u^1. En teoría solo en casos muy particulares (en los que interviniese únicamente γ_+ tendriamos estabilidad. En cualquier caso los errores de redondeo terminarán por afectar a las dos raíces $\gamma_\pm$ y en consecuencia la solución que obtendremos será inestable en cualquier caso.

Capítulo 2
Problemas parabólicos en dimensión 1 espacial. Coeficientes variables

Resumen
En este capítulo extendemos los métodos estudiados en el capítulo 1 a problemas parabólicos en dimensión 1 con coeficientes variables. Se analizará la convergencia a partir de la consistencia y estabilidad de los esquemas, utilizando diversas técnicas para el análisis de estabilidad según los casos, en particular el método de la energía y el principio del máximo

2.1. Problema con coeficientes variables

Sea una función real $D : [0,1] \to \mathbb{R}$ con propiedades de continuidad y derivabilidad suficientes verificando que existen $\beta > 0$ y $\alpha > 0$ de modo que

$$0 < \alpha \leq D(x) \leq \beta < \infty \quad \forall x \in [0,1] \tag{2.1}$$

y la función

$$f : [0,1] \times [0,T] \to \mathbb{R} \tag{2.2}$$
$$x,t \to f(x,t) \tag{2.3}$$

que supondremos al menos continua.

Consideramos el problema parabólico con coeficientes variables en dimensión 1 espacial siguiente: Hallar la función

$$u : [0,1] \times [0,T] \to \mathbb{R} \tag{2.4}$$
$$x,t \to u(x,t) \tag{2.5}$$

verificando

$$\frac{\partial u}{\partial t} - \frac{\partial u}{\partial x}\big(D(x)\frac{\partial u}{\partial x}\big) = f \quad \forall x \in (0,1), \forall t \in (0,T] \tag{2.6}$$

$$u(0,t) = u(1,t) = 0 \quad t > 0 \tag{2.7}$$

$$u(x,0) = v(x) \quad 0 < x < 1 \tag{2.8}$$

2.2. Aproximación mediante diferencias finitas

Consideremos dos intervalos contiguos genéricos ambos de longitud h

$$[x_{j-1}, x_j] \cup [x_j, x_{j+1}]$$

y los puntos medios de los dos intervalos $x_{j-1/2} = x_j - h/2$ y $x_{j+1/2} = x_j + h/2$ respectivamente. Desarrollando en serie de Taylor $u(x_j,t)$ y $u(x_{j-1},t)$ en un entorno de $x_{j-1/2}$, restando término a término y reordenando obtenemos una aproximación de $(\partial u/\partial x)(x_{j-1/2},t)$:

$$\frac{\partial u}{\partial x}(x_{j-1/2},t) = \frac{u(x_j,t) - u(x_{j-1},t)}{h} - \frac{h^2}{24}\frac{\partial^3 u}{\partial x^3}(x_{j-1/2},t) + \mathcal{O}(h^4) \tag{2.9}$$

análogamente desarrollando en serie de Taylor $u(x_j,t)$ y $u(x_{j+1},t)$ en un entorno de $x_{j+1/2}$ obtenemos

$$\frac{\partial u}{\partial x}(x_{j+1/2},t) = \frac{u(x_{j+1},t) - u(x_j,t)}{h} - \frac{h^2}{24}\frac{\partial^3 u}{\partial x^3}(x_{j+1/2},t) + \mathcal{O}(h^4) \tag{2.10}$$

Aplicamos las aproximaciones anteriores al término

$$\frac{\partial}{\partial x}\big(D(x)\frac{\partial u}{\partial x}\big)$$

Utilizamos diferencias centrales (1.51) para la aproximación de la derivada respecto a x de $D(x)\frac{\partial u}{\partial x}(x_j,t)$

$$\frac{\partial u}{\partial x}\big(D(x)\frac{\partial u}{\partial x}\big)(x_j,t) = \frac{D(x_{j+1/2})\frac{\partial u}{\partial x}(x_{j+1/2},t) - D(x_{j-1/2})\frac{\partial u}{\partial x}(x_{j-1/2},t)}{h} + \mathcal{O}(h^2)$$

Sustituyendo la aproximación de $\frac{\partial u}{\partial x}(x_{j-1/2},t)$ y de $\frac{\partial u}{\partial x}(x_{j+1/2},t)$ dadas por (2.9) y (2.10) obtenemos

$$\frac{\partial u}{\partial x}\big(a(x)\frac{\partial u}{\partial x}\big)(x_j,t) = \frac{D(x_{j+1/2})\frac{u(x_{j+1},t)-u(x_j,t)}{h} - D(x_{j-1/2})\frac{u(x_j,t)-u(x_{j-1},t)}{h} + R}{h} + \mathcal{O}(h^2)$$

Falta evaluar el término R. Tenemos utilizando (2.9) y (2.10)

$$R = D(x_{j-1/2})\frac{h^2}{24}\frac{\partial^3 u}{\partial x^3}(x_{j-1/2},t) - D(x_{j+1/2})\frac{h^2}{24}\frac{\partial^3 u}{\partial x^3}(x_{j+1/2},t) + \mathscr{O}(h^4) \quad (2.11)$$

Desarrollando todos los érminos en un entorno de x_j

$$\begin{aligned}
D(x_{j-1/2}) &= D(x_j) - \frac{h}{2}\frac{dD}{dx}(x_i) + \mathscr{O}(h^2)\\
D(x_{j+1/2}) &= D(x_j) + \frac{h}{2}\frac{dD}{dx}(x_i) + \mathscr{O}(h^2)\\
\frac{\partial^3 u}{\partial x^3}(x_{j-1/2},t) &= \frac{\partial^3 u}{\partial x^3}(x_j,t) - h\frac{\partial^4 u}{\partial x^3}(x_j,t) + \mathscr{O}(h^2)\\
\frac{\partial^3 u}{\partial x^3}(x_{j+1/2},t) &= \frac{\partial^3 u}{\partial x^3}(x_j,t) + h\frac{\partial^4 u}{\partial x^3}(x_j,t) + \mathscr{O}(h^2)
\end{aligned}$$

de donde el término R es de orden $\mathscr{O}(h^3)$ y la aproximación de $\frac{\partial u}{\partial x}\big(D(x)\frac{\partial u}{\partial x}\big)(x_j,t)$ es de orden $\mathscr{O}(h^2)$.

La aproximación anterior

$$\frac{\partial u}{\partial x}\big(D(x)\frac{\partial u}{\partial x}\big)(x_j,t) \approx \frac{D_{j+1/2}\frac{u^n_{j+1}-u^n_j}{h} - D_{j-1/2}\frac{u^n_j-u^n_{j-1}}{h}}{h}$$

se puede expresar como

$$\partial_x (D_{-1/2}\bar{\partial}_x u)^n_j$$

donde interpretamos $(D_{-1/2})_j = D_{j-1/2} = D(x_j - h/2)$. En efecto,

$$\begin{aligned}
\partial_x (D_{-1/2}\bar{\partial}_x u)^n_j &= \frac{(D_{-1/2}\bar{\partial}_x u)^n_{j+1} - (D_{-1/2}\bar{\partial}_x u)^n_j}{h}\\
&= \frac{D_{j+1/2}\bar{\partial}_x u^n_{j+1} - D_{j-1/2}\bar{\partial}_x u^n_j}{h}\\
&= \frac{D_{j+1/2}\frac{u^n_{j+1}-u^n_j}{h} - D_{j-1/2}\frac{u^n_j-u^n_{j-1}}{h}}{h}
\end{aligned}$$

2.2.1. Método general de un paso

El método general de un paso para el problema (2.6)-(2.7)-(2.8) es

$$\partial_t u^n_j - \partial_x (D_{-1/2}\bar{\partial}_x u)^{n+\theta}_j = f^{n+\theta}_j \quad n \geq 0,\ j = 1,...,M \quad (2.12)$$

$$u^n_0 = u^n_{M+1} = 0 \quad n > 0 \quad (2.13)$$

$$u^0_j = v_j \quad j = 1,...,M \quad (2.14)$$

donde $0 \leq \theta \leq 1$

Consistencia

Análogamente a lo visto en la sección (1.2) en el caso con coeficiente D constante para la aproximación de las derivadas con respecto a la variable t y en la subsección anterior para la aproximación de las derivadas con respecto a la variable x tenemos que el error de consistencia $\tau_j^{n+\theta}$ es

$$\tau_j^{n+\theta} = \begin{cases} \mathcal{O}(k+h^2) \text{ si } \theta \neq \dfrac{1}{2} \\ \mathcal{O}(k^2+h^2) \text{ si } \theta = \dfrac{1}{2} \end{cases}$$

Estabilidad

Analizaremos la estabilidad en la norma euclídea, $||.||$ definida por (1.14)

$$||v|| = \left(h\sum_j (v_j)^2\right)^{1/2}$$

Para ello necesitaremos algunas definiciones y propiedades. Para $v = (v_j)_j$, $w = (w_j)_j \in \mathbb{R}^M$ consideramos el producto escalar

$$(v,w) = h\sum_j v_j w_j$$

y las diferencias finitas progresivas y regresivas

$$\partial_x v_j = \frac{v_{j+1}-v_j}{h} \quad \bar{\partial}_x v_j = \frac{v_j - v_{j-1}}{h}$$

Si extendemos v y w con los valores en la frontera $v_0 = w_0 = 0 \quad v_{M+1} = w_{M+1} = 0$ podemos definir

$$(\partial_x v, w) = h\sum_{j=0}^{M} \partial_x v_j w_j = h\sum_{j=0}^{M} \frac{v_{j+1}-v_j}{h} w_j$$

$$(\bar{\partial}_x v, w) = h\sum_{j=1}^{M+1} \bar{\partial}_x v_j w_j = h\sum_{j=1}^{M+1} \frac{v_j - v_{j-1}}{h} w_j$$

Tenemos la siguiente fórmula de suma por partes

Lema 2.1

$$(\partial_x v, w) = -(v, \bar{\partial}_x w) \tag{2.15}$$

Demostración.

$$
\begin{aligned}
(\partial_x v, w) &= h\sum_{j=0}^{M} \partial_x v_j w_j = h\sum_{j=0}^{M} \frac{v_{j+1}-v_j}{h} w_j \\
&= h\sum_{j=0}^{M} \frac{v_{j+1}w_j}{h} - h\sum_{j=0}^{M} \frac{v_j w_j}{h} \\
&= -\big(h\sum_{j=0}^{M} \frac{v_j w_j}{h} - h\sum_{j=0}^{M} \frac{v_{j+1}w_j}{h}\big) \\
&= -\big(h\sum_{j=1}^{M+1} \frac{v_j w_j}{h} - h\sum_{j=1}^{M+1} \frac{v_j w_{j-1}}{h}\big) \\
&= -h\sum_{j=1}^{M+1} v_j \frac{w_j - w_{j-1}}{h} = -(v, \bar{\partial}_x w)
\end{aligned}
$$

■

En lo que sigue necesitaremos las siguientes propiedades:

Lema 2.2 *Con la definición (1.18) se verifica*

$$(\bar{\partial}_t u^n, u^n) = \frac{1}{2}\bar{\partial}_t ||u^n||^2 + \frac{k}{2}||\bar{\partial}_t u^n||^2 \tag{2.16}$$

Demostración.

$$(\bar{\partial}_t u^n, u^n) = (\frac{u^n - u^{n-1}}{k}, u^n) = \frac{1}{k}||u^n||^2 - \frac{1}{k}(u^{n-1}, u^n)$$

Ahora bien,

$$||u^n - u^{n-1}||^2 = ||u^n||^2 - 2(u^n, u^{n-1}) + ||u^{n-1}||^2$$

de donde

$$-\frac{1}{k}(u^{n-1}, u^n) = -\frac{1}{2k}||u^n||^2 - \frac{1}{2k}||u^{n-1}||^2 + \frac{1}{2k}||u^n - u^{n-1}||^2$$

y finalmente

$$
\begin{aligned}
(\bar{\partial}_t u^n, u^n) &= \frac{1}{2k}||u^n||^2 - \frac{1}{2k}||u^{n-1}||^2 + \frac{1}{2k}||u^n - u^{n-1}||^2 \\
&= \frac{1}{2}\bar{\partial}_t ||u^n||^2 + \frac{k}{2}||\bar{\partial}_t u^n||^2
\end{aligned}
$$

■

Lema 2.3 *Con la definición (1.17), y* $u^{n+\theta} = u^n + \theta(u^{n+1} - u^n)$ *para* $0 \le \theta \le 1$ *se verifica*

$$(\partial_t u^n, u^{n+\theta}) = \frac{1}{2}\partial_t ||u^n||^2 + (\theta - \frac{1}{2})k||\partial_t u^n||^2 \tag{2.17}$$

Demostración.

$$
\begin{aligned}
(\partial_t u^n, u^{n+\theta}) &= (\frac{u^{n+1}-u^n}{k}, u^n + \theta(u^{n+1}-u^n)) \\
&= \frac{1}{k}(u^{n+1}-u^n, u^n) + \frac{\theta}{k}(u^{n+1}-u^n, u^{n+1}-u^n) \\
&= \frac{1}{k}(u^{n+1}-u^n, u^n) + \theta k||\partial_t u^n||^2 \\
&= \frac{1}{k}(u^{n+1}, u^n) - \frac{1}{k}||u^n||^2 + \theta k||\partial_t u^n||^2
\end{aligned}
$$

por otra parte,

$$
\frac{1}{k}(u^{n+1}, u^n) = \frac{1}{2k}||u^{n+1}||^2 + \frac{1}{2k}||u^n||^2 - \frac{1}{2k}||u^{n+1}-u^n||^2
$$

de donde sustituyendo $\frac{1}{k}(u^{n+1}, u^n)$ en la expresión anterior obtenemos (2.17). ■

Lema 2.4 *Para todo $v = (v_i)_{i=1}^M \in \mathbb{R}^M$ extendido mediante $v_0 = 0$ y $v_{M+1} = 0$, asociados a un intervalo $[a,b]$, dividido en $M+1$ intervalos de longitud h existe una constante $C = \frac{b-a}{\sqrt{2}}$ tal que*

$$
||v|| \leq C||\bar{\partial}_x v|| \tag{2.18}
$$

donde $||v|| = \left(h\sum_{j=1}^M (v_j)^2\right)^{1/2}$

Demostración. De manera general asumimos que el intervalo de trabajo es $[a,b]$, dividido en $M+1$ intervalos de longitud h

$$
v_j = \sum_{l=1}^{j} v_l - v_{l-1} = h\sum_{l=1}^{j} \frac{v_l - v_{l-1}}{h}
$$

$$
\begin{aligned}
|v_j| &\leq h\sum_{l=1}^{j} |\frac{v_l - v_{l-1}}{h}| \leq h\Big(\sum_{l=1}^{j} (\frac{v_l - v_{l-1}}{h})^2\Big)^{1/2}\Big(\sum_{l=1}^{j} 1^2\Big)^{1/2} \\
&= \Big(h\sum_{l=1}^{j} (\frac{v_l - v_{l-1}}{h})^2\Big)^{1/2}\Big(h\sum_{l=1}^{j} 1^2\Big)^{1/2} \leq (hj)^{1/2}||\bar{\partial}_x v||
\end{aligned}
$$

Elevando al cuadrado

$$
|v_j|^2 \leq (hj)||\bar{\partial}_x v||^2
$$

multiplicando por h y sumando para todo j

$$
\begin{aligned}
||v||^2 = h\sum_{j=1}^{M} |v_j|^2 &\leq \Big(h^2\sum_{j=1}^{M} j\Big)||\bar{\partial}_x v||^2 \\
&= h^2\frac{M(M+1)}{2}||\bar{\partial}_x v||^2 \leq \frac{(b-a)^2}{2}||\bar{\partial}_x v||^2
\end{aligned}
$$

extrayendo la raiz cuadrada

$$||v|| \leq \frac{b-a}{\sqrt{2}}||\bar{\partial}_x v||$$

■

Observación: Observar que la aplicación

$$\begin{aligned} \mathbb{R}^M &\to \mathbb{R} \\ v &\to ||\bar{\partial}_x v|| \end{aligned}$$

es una norma en en el subespacio de vectores de $\mathbb{R}^M$ tales que $v_0 = 0$ o $v_{M+1} = 0$. En efecto, si $||\bar{\partial}_x v|| = 0$, y $v_0 = 0$,

$$\begin{aligned} \frac{v_l - v_{l-1}}{h} = 0 \quad \forall l = 1, ..., M \\ v_l = v_{l-1} \quad \forall l = 1, ..., M \end{aligned}$$

como $v_0 = 0$ resulta $v_l = 0$ para todo $l = 1, ..., M+1$. Las otras propiedades de una norma se deducen de las propiedades de la norma euclídea $||.||$

Lema 2.5 *Para todo $v = (v_i)_{i=1}^M \in \mathbb{R}^M$ extendido mediante $v_0 = 0$ y $v_{M+1} = 0$, asociados a un intervalo $[a,b]$, dividido en $M+1$ intervalos de longitud h existe una constante $C = 2$ tal que se verifica la siguiente desigualdad inversa*

$$||\partial_x v|| \leq \frac{C}{h}||v|| \tag{2.19}$$

Del mismo modo

$$||\bar{\partial}_x v|| \leq \frac{C}{h}||v|| \tag{2.20}$$

Demostración.

$$\begin{aligned} ||\partial_x v||^2 = h\sum_{j=0}^{M}\left(\frac{v_{j+1}-v_j}{h}\right)^2 = \frac{1}{h}\sum_{j=0}^{M}(v_{j+1}-v_j)^2 \\ \leq \frac{1}{h}\sum_{j=0}^{M}(2v_{j+1}^2 + 2v_j^2) = \frac{2}{h}\sum_{j=0}^{M}v_{j+1}^2 + \frac{2}{h}\sum_{j=0}^{M}v_j^2 \\ = \frac{2}{h}\sum_{j=1}^{M}v_j^2 + \frac{2}{h}\sum_{j=1}^{M}v_j^2 \\ = \frac{4}{h}\sum_{j=1}^{M}v_j^2 = \frac{4}{h^2}\left(h\sum_{j=1}^{M}v_j^2\right) = \frac{4}{h^2}||v||^2 \end{aligned}$$

de donde se obtiene (2.19). Análogamente se obtiene (2.20) ■

Demostraremos aquí la estabilidad para el caso de Euler implícito que corresponde al valor $\theta = 1$ en (2.12). Observemos primero que se puede escribir también así:

$$\bar{\partial}_t u_j^n - \partial_x (D_{-1/2} \bar{\partial}_x u)_j^n = f_j^n \quad n > 0,\ j = 1, ..., M \tag{2.21}$$

$$u_0^n = u_{M+1}^n = 0 \quad n > 0 \tag{2.22}$$

$$u_j^0 = v_j \quad j = 1, ..., M \tag{2.23}$$

Teorema 2.1 *La solución de (2.21)-(2.22)-(2.23) verifica*

$$||u^n||^2 \leq ||u^0||^2 + \frac{T_{\text{máx}}}{4\alpha} \operatorname*{máx}_l ||f^l||^2 \tag{2.24}$$

Demostración. Multiplicamos todos los términos de la ecuación (2.21) por hu_j^n y sumamos respecto a j, teniendo en cuenta el lema (2.1)

$$(\bar{\partial}_t u^n, u^n) - \left(\partial_x (D_{-1/2} \bar{\partial}_x u)^n, u^n\right) = (f^n, u^n)$$
$$(\bar{\partial}_t u^n, u^n) + (D_{-1/2} \bar{\partial}_x u^n, \bar{\partial}_x u^n) = (f^n, u^n)$$

Utilizando (2.16) y las desigualdades

$$(D_{-1/2} \bar{\partial}_x u^n, \bar{\partial}_x u^n) \geq \alpha ||\bar{\partial}_x u^n||^2 \tag{2.25}$$

$$(f^n, u^n) \leq ||f^n||.||u^n|| \leq \frac{1}{2\varepsilon} ||f^n||^2 + \frac{\varepsilon}{2} ||u^n||^2$$

con $\varepsilon > 0$, tenemos gracias al lema (2.4)

$$\begin{aligned} \bar{\partial}_t ||u^n||^2 + 2\alpha ||\bar{\partial}_x u^n||^2 &\leq \frac{1}{\varepsilon} ||f^n||^2 + \varepsilon ||u^n||^2 \\ &\leq \frac{1}{\varepsilon} ||f^n||^2 + \frac{\varepsilon}{2} ||\bar{\partial}_x u^n||^2 \end{aligned}$$

tomando $\varepsilon = 4\alpha$ resulta

$$\bar{\partial}_t ||u^n||^2 \leq \frac{1}{4\alpha} ||f^n||^2$$

de donde

$$||u^n||^2 \leq ||u^{n-1}||^2 + \frac{k}{4\alpha} ||f^n||^2$$

y finalmente aplicando recursivamente la estimación anterior

$$\begin{aligned} ||u^n||^2 &\leq ||u^0||^2 + \frac{k}{4\alpha} \sum_{l=1}^{n} ||f^l||^2 \\ &\leq ||u^0||^2 + \frac{T_{\text{máx}}}{4\alpha} \operatorname*{máx}_l ||f^l||^2 \end{aligned}$$

■

En la estimación anterior no hemos utilizado el hecho de que $\alpha > 0$. Un Refinamiento del desarrollo anterior permite mejorar la estimación.

Teorema 2.2 *La solución de (2.21)-(2.22)-(2.23) verifica*

$$||u^n|| \leq \big(\frac{1}{1+2\alpha}\big)^n ||u^0|| + k\sum_{l=1}^{n}\big(\frac{1}{1+2\alpha}\big)^{n-l+1}||f^l||$$

y para los errores

$$||e^n|| \leq \big(\frac{1}{1+2\alpha}\big)^n ||e^0|| + k\sum_{l=1}^{n}\big(\frac{1}{1+2\alpha}\big)^{n-l+1}||\tau^l|| \tag{2.26}$$

donde el error en el paso n es $e^n = (e_j^n)_{j=1}^M$ siendo $e_j^n = u(x_j,t_n) - u_j^n$ y τ^n es el error de consistencia en el paso n. La estimación anterior del error demuestra que los errores en el caso del método de Euler implícito se amortiguan a medida que aumenta el número de pasos.

Demostración. Procedemos como en el teorema anterior, multiplicando todos los términos de la ecuación (2.21) por hu_j^n y sumando respecto a j. Teniendo en cuenta el lema (2.1)

$$||u^n||^2 + k(D_{-1/2}\bar{\partial}_x u^n, \bar{\partial}_x u^n) = (u^{n-1}, u^n) + k(f^n, u^n)$$

por otra parte teniendo en cuenta la desigualdad (2.18) y reordenando términos

$$(D_{-1/2}\bar{\partial}_x u^n, \bar{\partial}_x u^n) \geq \alpha||\bar{\partial}_x u||^2 \geq 2\alpha||u^n||^2$$

resulta,

$$(1+2\alpha)||u^n||^2 \leq ||u^{n-1}||.||u^n|| + k||f^n||.||u^n||$$

$$||u^n|| \leq \frac{1}{1+2\alpha}||u^{n-1}|| + \frac{k}{1+2\alpha}||f^n||$$

y aplicando recursivamente la relación anterior obtenemos el resultado buscado. ■

Ejercicio

Considerar el problema

$$\frac{\partial u}{\partial t} - \frac{\partial}{\partial x}\big(D(x)\frac{\partial u}{\partial x}\big) = 0 \tag{2.27}$$
$$u(0,t) = u(1,t) = 0 \quad t > 0 \tag{2.28}$$
$$u(x,0) = v(x) \quad 0 < x < 1 \tag{2.29}$$

Para todo $w \in L^2(0,1)$ su norma es

$$||w||_{0,(0,1)} = \int_0^1 w^2 dx$$

1. Deducir la igualdad llamada de la energía

$$||u(t)||^2 + 2\int_0^t ||D^{1/2}\frac{\partial u}{\partial x}||^2 dt = ||v||^2 \tag{2.30}$$

donde para todo $w \in L^2(0,1)$ consideramos la norma

$$||w|| = \int_0^1 w^2 dx$$

2. Deducir la correspondiente versión discreta de la igualdad anterior para el esquema de Euler implícito,

$$\bar{\partial}_t u_j^n - \partial_x (D_{-1/2}\bar{\partial}_x u)_j^n = 0 \quad n > 0,\ j = 1,...,M$$
$$u_0^n = u_{M+1}^n = 0 \quad n > 0$$
$$u_j^0 = v_j \quad j = 1,...,M$$

Solución

1. Multiplicamos los términos de la ecuación por u e integramos. Resulta

$$\int_0^1 \frac{\partial u}{\partial t} u\, dx - \int_0^1 \frac{\partial}{\partial x}\big(D(x)\frac{\partial u}{\partial x}\big) u\, dx = 0$$

Teniendo en cuenta $\frac{\partial u}{\partial t} u = \frac{1}{2}\frac{\partial u^2}{\partial t}$ y que la derivada con respecto al tiempo puede salir fuera de la integral, integrando por partes el segundo término

$$\frac{d}{dt}\int_0^1 u^2\, dx + \int_0^1 D(x)\frac{\partial u}{\partial x}\frac{\partial u}{\partial x} dx = 0$$

integrando entre 0 y t obtenemos el resultado

2. Multiplicando por u^n, aplicando el lema (2.2), la suma por partes (lema (2.1)) en el segundo término y sumando para $n = 1,...,N$ obtenemos

$$||u^n||^2 + 2k(D_{-1/2}\bar{\partial}_x u^n, \bar{\partial}_x u^n) = ||u^{n-1}||^2 - k^2 \sum_{n=1}^{N} ||\bar{\partial}_t u^n||^2$$

y aplicando recursivamente la relación anterior

$$||u^n||^2 + 2k\sum_{n=1}^{N}(D_{-1/2}\bar{\partial}_x u^n, \bar{\partial}_x u^n) = ||v||^2 - k^2\sum_{n=1}^{N}||\bar{\partial}_t u^n||^2$$

Ejercicio

Considerar el método general de un paso (2.12)-(2.13)-(2.14).

1. Demostrar que para el análisis de la estabilidad basta analizar el caso homogéneo, es decir, el caso con $f = 0$.
2. Analizar la estabilidad en el caso homogéneo.

Solución

1. El esquema anterior se puede escribir con notación matricial

$$u^{n+1} = \mathbf{T}u^n + k\mathbf{S}f^{n+\theta}$$

$$\mathbf{T} = (\mathbf{I} + \theta\lambda\mathbf{A})^{-1}(\mathbf{I} - (1-\theta)\lambda\mathbf{A})$$

y

$$\mathbf{S} = (\mathbf{I} + \theta\lambda\mathbf{A})^{-1}$$

donde $\lambda = \frac{k}{h^2}$ y

$$\mathbf{A} = \begin{bmatrix} d_1 & c_1 & \dots & \dots & 0 \\ b_2 & d_2 & c_2 & \dots & 0 \\ & \ddots & \ddots & \ddots & \\ & & & & c_{M-1} \\ 0 & 0 & \dots & b_M & d_M \end{bmatrix}$$

con

$$d_j = D_{j+1/2} + D_{j-1/2}$$
$$c_j = -D_{j+1/2}$$
$$b_j = -D_{j-1/2}$$

En el paso n-ésimo tendremos

$$u^n = \mathbf{T}^n u^0 + k\sum_{l=0}^{n-1}\mathbf{T}^{n-1-l}\mathbf{S}f^{l+\theta}$$

de donde si $||\mathbf{T}|| < 1$

$$||u^n|| \leq ||u^0|| + k||\mathbf{S}|| \sum_{l=0}^{n-1} ||f^{l+\theta}|| \tag{2.31}$$

De las propiedades de D resulta que $\mathbf{A}$ es simétrica y definida positiva. Observar además que $||\mathbf{S}|| \leq C$ con la constante C independiente de k y de h. Por lo tanto la estabilidad solo depende de las propiedades de $\mathbf{T}$. Para el problema homogéneo $f = 0$ el esquema se reduce a

$$u^{n+1} = \mathbf{T}u^n$$

y basta pues demostrar que

$$||u^{n+1}|| \leq ||u^n||$$

Veamos además que si $\mathbf{A}$ es simétrica y definida positiva $||S|| < 1$, en efecto,

$$||\mathbf{S}|| = ||(\mathbf{I} + \theta\lambda\mathbf{A})^{-1}|| = \max_{v \neq 0} \frac{||(\mathbf{I} + \theta\lambda\mathbf{A})^{-1}v||}{||v||}$$

$$\frac{||(\mathbf{I} + \theta\lambda\mathbf{A})^{-1}v||^2}{||v||^2} = \frac{||\varphi||^2}{||(\mathbf{I} + \theta\lambda\mathbf{A})\varphi||^2}$$
$$= \frac{||\varphi||^2}{||\varphi||^2 + 2\theta k(A\varphi, \varphi) + \theta^2 k^2 ||\mathbf{A}||^2} \leq 1$$

2. Las ecuaciones (2.12)-(2.13)-(2.14) se escriben para el caso homogéneo

$$\partial_t u_j^n - \partial_x (D_{-1/2}\bar{\partial}_x u)_j^{n+\theta} = 0 \quad n \geq 0,\ j = 1, ..., M$$
$$u_0^n = u_{M+1}^n = 0 \quad n > 0$$
$$u_j^0 = v_j \quad j = 1, ..., M$$

Multiplicando escalarmente por $u^{n+\theta}$, aplicando el lema (2.3) y sumando por partes

$$\frac{1}{2}\partial_t ||u^n||^2 + (\theta - \frac{1}{2})k||\partial_t u^n||^2 + (D_{-1/2}\bar{\partial}_x u^{n+\theta}, \bar{\partial}_x u^{n+\theta}) = 0$$

Caso $\theta \geq 1/2$
El segundo y tercer término del primer miembro son mayores o igual que cero, obteniendo

$$||u^{n+1}|| \leq ||u^n||$$

independientemente del valor de h y k. El método es incondicionalmente estable.
Caso $\theta < 1/2$ El término

$$(D_{-1/2}\bar{\partial}_x u^{n+\theta}, \bar{\partial}_x u^{n+\theta})$$

se puede escribir

$$(D_{-1/2}\bar{\partial}_x u^{n+\theta}, \bar{\partial}_x u^{n+\theta}) = ||D_{-1/2}^{1/2}\bar{\partial}_x u^{n+\theta}||^2$$

Por otra parte utilizando el lema (2.5)

$$\begin{aligned} ||\partial_t u^n||^2 &= ||\partial_x(D_{1/2}\bar{\partial}_x u^{n+\theta})||^2 \\ &\leq \frac{4}{h^2}||D_{1/2}\bar{\partial}_x u^{n+\theta}||^2 \\ &\leq \frac{4}{h^2}\beta||D_{1/2}^{1/2}\bar{\partial}_x u^{n+\theta}||^2 \end{aligned}$$

donde β es la constante definida en (2.1). Finalmente reordenando todos los términos

$$||u^{n+1}||^2 + \big(2-(1-2\theta)\frac{4k\beta}{h^2}\big)||D_{-1/2}\bar{\partial}_x u^{n+\theta}||^2 \leq ||u^n||^2$$

Eligiendo k y h^2 de modo que $(1-2\theta)\frac{4k\beta}{h^2} \leq 2$ tenemos

$$||u^{n+1}||^2 \leq ||u^n||^2$$

La condición de estabilidad es entonces

$$k \leq \frac{h^2}{2\beta(1-2\theta}\big)$$

En la práctica para $\theta < 1/2$ solo se utiliza $\theta = 0$. En ese caso la condición de estabilidad es

$$k \leq \frac{h^2}{2\beta}$$

2.2.2. Estabilidad en la norma de la energía

Consideremos D como en el problema (2.6)-(2.7)-(2.8). y el correspondiente método general de un paso (2.12)-(2.13)-(2.14) definimos la siguiente norma asociada al problema anterior, llamada norma de la energía:

Definición 2.1 *Para* $v^n = (v_1^n, v_2^n, ..., v_M^n) \in \mathbb{R}^M$ *extendido con los valores* $v_0 = v_{M+1} = 0$

$$||v||_E = (D_{-1/2}\bar{\partial}_x v, \bar{\partial}_x v)^{1/2} = ||D_{-1/2}^{1/2}\bar{\partial}_x v|| \tag{2.32}$$

Vamos a estudiar la estabilidad en esta norma.

Teorema 2.3 *La solución de (2.12)-(2.13)-(2.14) verifica*

$$||u^n||_E^2 \leq ||u^0||_E^2 + k\varepsilon \sum_{l=0}^{n-1} ||f^{n+\theta}||^2 \tag{2.33}$$

donde $\varepsilon = 1/2$ *si* $\theta \geq 1/2$ *y* $\varepsilon > 1/2$ *si* $\theta < 1/2$ *y se verifica la condición de estabilidad.*

$$k \leq (1 - \frac{1}{2\varepsilon}) \frac{h^2}{2(1-2\theta)\beta} \tag{2.34}$$

Para ello necesitamos los siguientes resultados previos:

Lema 2.6 *Para* $v^n = (v_1^n, v_2^n, ..., v_M^n) \in \mathbb{R}^M$

$$\bar{\partial}_x \partial_t v^n = \partial_t \bar{\partial}_x v^n$$

Demostración.

$$\bar{\partial}_x \partial_t v_j^n = \bar{\partial}_x \big(\frac{v_j^{n+1} - v_j^n}{k}\big) = \frac{\frac{v_j^{n+1} - v_j^n}{k} - \frac{v_{j-1}^{n+1} - v_{j-1}^n}{k}}{h}$$
$$= \frac{\frac{v_j^{n+1} - v_{j-1}^{n+1}}{h} - \frac{v_j^n - v_{j-1}^n}{h}}{k} = \partial_t \bar{\partial}_x v_j^n$$

■

Lema 2.7 *Con las notaciones anteriores*

$$(D_{-1/2} \bar{\partial}_x u^{n+\theta}, \bar{\partial}_x \partial_t u^n) = \frac{1}{2} \partial_t ||u^n||_E^2 + (\theta - \frac{1}{2}) k ||\partial_t u^n||_E^2$$

Demostración. La demostración es análoga a la demostración del lema (2.3). Gracias al lema (2.6) tenemos en primer lugar

$$(D_{-1/2} \bar{\partial}_x u^{n+\theta}, \bar{\partial}_x \partial_t u^n) = (D_{-1/2} \bar{\partial}_x u^{n+\theta}, \partial_t \bar{\partial}_x u^n)$$

Denotemos $w^n = \bar{\partial}_x u^n$

$$(D_{-1/2}(w^n + \theta(w^{n+1} - w^n)), \partial_t w^n) = (D_{-1/2}(w^n + \theta(w^{n+1} - w^n)), \frac{w^{n+1} - w^n}{k})$$
$$= \frac{1}{k}(D_{-1/2} w^n, w^{n+1} - w^n) + \frac{\theta}{k}(D_{-1/2}(w^{n+1} - w^n), w^{n+1} - w^n)$$
$$= \frac{1}{k}(D_{-1/2} w^n, w^{n+1} - w^n) + \theta k (D_{-1/2} \partial_t w^n, \partial_t w^n)$$
$$= \frac{1}{k}(D_{-1/2} w^n, w^{n+1} - w^n) + \theta k ||D_{-1/2}^{1/2} \partial_t w^n||^2$$

Por otra parte, teniendo en cuenta

$$2(D_{-1/2} w^{n+1}, w^n) = ||D_{-1/2}^{1/2} w^{n+1}||^2 + ||D_{-1/2}^{1/2} w^{n+1}||^2 - ||D_{-1/2}^{1/2}(w^{n+1} - w^n)||^2$$

resulta

$$\begin{aligned}(D_{-1/2}(w^{n+1}-w^n),w^n) &= (D_{-1/2}w^{n+1},w^n) - ||D^{1/2}_{-1/2}w^n||^2\\ &= \frac{1}{2}||D^{1/2}_{-1/2}w^{n+1}||^2 - \frac{1}{2}||D^{1/2}_{-1/2}w^n||^2 - \frac{1}{2}||D^{1/2}_{-1/2}(w^{n+1}-w^n)||^2\\ &= \frac{1}{2}\partial_t||D^{1/2}_{-1/2}w^n||^2 - \frac{k}{2}||D^{1/2}_{-1/2}\partial_t w^n||^2\end{aligned}$$

de donde finalmente

$$(D_{-1/2}(w^n+\theta(w^{n+1}-w^n)),\partial_t w^n) = \frac{1}{2}\partial_t||D^{1/2}_{-1/2}w^n||^2 + (\theta-\frac{1}{2})k||D^{1/2}_{-1/2}\partial_t w^n||^2$$

y sustituyendo w^n por $\bar{\partial}_x u^n$ obtenemos el resulado buscado. ■

Demostración del teorema (2.3).

Multiplicamos escalarmente los términos de la ecuación (2.12) por $\partial_t u^n$, y sumamos por partes

$$||\partial_t u^n||^2 + (D_{-1/2}\bar{\partial}_x u^{n+\theta},\bar{\partial}_x\partial_t u^n) = (f^{n+\theta},\partial_t u^n)$$

Utilizando el resultado del lema (2.7)

$$||\partial_t u^n||^2 + \frac{1}{2}\partial_t||u^n||^2_E + (\theta-\frac{1}{2})k||\partial_t u^n||^2_E \leq \frac{\varepsilon}{2}||f^{n+\theta}||^2 + \frac{1}{2\varepsilon}||\partial_t u^n||^2$$

Distinguiremos ahora los casos $\theta \geq 1/2$ y $\theta < 1/2$.

Caso $\theta \geq 1/2$: Tomando $\varepsilon = 1/2$,

$$||u^{n+1}||^2_E \leq ||u^n||^2_E + \frac{k}{2}||f^{n+\theta}||^2$$

de donde finalmente

$$||u^n||^2_E \leq ||u^0||^2_E + \frac{k}{2}\sum_{l=0}^{n-1}||f^{l+\theta}||^2$$

Caso $\theta < 1/2$: Utilizando el lema (2.5)

$$||\partial_t u^n||_E = ||D^{1/2}_{-1/2}\bar{\partial}_x\partial_t u^n|| \leq \beta^{1/2}||\bar{\partial}_x\partial_t u^n|| \leq \frac{2\beta^{1/2}}{h}||\partial_t u^n||$$

de donde

$$(1-\frac{1}{2\varepsilon})||\partial_t u^n||^2 + \frac{1}{2}\partial_t||u^n||^2_E \leq (\frac{1}{2}-\theta)k\frac{4\beta}{h^2}||\partial_t u^n||^2 + \frac{\varepsilon}{2}||f^{n+\theta}||^2 \qquad (2.35)$$

En este caso la condición de estabilidad será

$$1-\frac{1}{2\varepsilon}-\frac{1}{2}(1-2\theta)k\frac{4\beta}{h^2}\geq 0$$

o bien

$$(1-2\theta)k\frac{2\beta}{h^2}\leq 1-\frac{1}{2\varepsilon}\quad \forall\varepsilon>1/2$$

es decir

$$k<(1-\frac{1}{2\varepsilon})\frac{h^2}{2(1-2\theta)\beta}$$

resultando

$$||u^{n+1}||_E^2\leq ||u^n||_E^2+k\varepsilon||f^{n+\theta}||^2$$

de donde se obtiene (2.33). ■

2.2.3. *Acotación del error en la norma de la energía y en la norma del máximo*

La estabilidad y consistencia proporciona la siguiente estimación del error para el problema (2.12)-(2.13)-(2.14)

Teorema 2.4 *Tenemos la siguiente estimación del error del método (2.12)-(2.13)-(2.14) en la norma de la energía:*

$$||e^n||_E^2\leq ||e^0||_E^2+k\varepsilon\sum_{l=0}^{n-1}||\tau^{l+\theta}||^2 \tag{2.36}$$

con $\varepsilon=1/2$ o $\varepsilon>1/2$ determindos en función del valor de θ según se indica en el teorema (2.3) y siendo $\tau^{l+\theta}$ el error de consistencia en cada paso.

Demostración. La demostración es inmediata a partir de los resultados de estabilidad y consistencia. ■

La estimación del error en la norma de la energía permite obtener una estimación en la norma del máximo utilizando la siguiente propiedad:

Lema 2.8 *Para todo $v=(v_i)_{i=1}^M\in\mathbb{R}^M$ extendido mediante $v_0=0$, asociados a un intervalo $[a,b]$, dividido en $M+1$ intervalos de longitud h, existe una constante $C=\sqrt{b-a}$ tal que*

$$||v||_\infty=\sup_j|v_j|\leq C||\bar{\partial}_x v|| \tag{2.37}$$

Demostración. Pongamos

$$v_j=\sum_{l=1}^{j}v_l-v_{l-1}$$

Para todo $j = 1, ..., M+1$

$$\begin{aligned}
|v_j| &\leq h\sum_{l=1}^{j}\frac{|v_l - v_{l-1}|}{h} \\
&\leq h\Big(\sum_{l=1}^{j}(\frac{v_l - v_{l-1}}{h})^2\Big)^{1/2}\Big(\sum_{l=1}^{j}1^2\Big)^{1/2} \\
&\leq \Big(h\sum_{l=1}^{M+1}(\frac{v_l - v_{l-1}}{h})^2\Big)^{1/2}\Big(h\sum_{l=1}^{j}1^2\Big)^{1/2} \\
&= ||\bar{\partial}_x v||.(jh)^{1/2} \leq \Big((M+1)h\Big)^{1/2}||\bar{\partial}_x v|| = \sqrt{b-a}||\bar{\partial}_x v||
\end{aligned}$$

■

Podemos ahora estimar el error en la norma $||.||_\infty$

Teorema 2.5 *Tenemos la siguiente estimación del error del método (2.12)-(2.13)-(2.14) en la norma del máximo* $||.||_\infty$

$$||e^n||_\infty^2 \leq \frac{1}{\alpha}\Big(||e^0||_E^2 + k\varepsilon\sum_{l=0}^{n-1}||\tau^{l+\theta}||^2\Big)$$

y si $e^0 = 0$

$$||e^n||_\infty^2 \leq \varepsilon\frac{T_{\text{máx}}}{\alpha}\max_{l=0,...,n-1}||\tau^{l+\theta}||^2$$

donde $\varepsilon = 1/2$ *o* $\varepsilon > 1/2$ *determindos en función del valor de* θ *según se indica en el teorema (2.3) y siendo* $\tau^{l+\theta}$ *el error de consistencia en cada paso.*

Demostración. Para una vector $v = (v_j)_{j=1}^M \in \mathbb{R}^M$ extendido con el valor $v_0 = 0$ y con las propiedades (2.1) del coeficiente D tendremos

$$\sqrt{\alpha}||\bar{\partial}_x v|| \leq ||D_{-1/2}^{1/2}\bar{\partial}_x v|| = ||v||_E$$

de donde utilizando la estimación del error (2.36)

$$\begin{aligned}
||e^n||_\infty^2 \leq ||\bar{\partial}_x v||^2 &\leq \frac{1}{\alpha}||e^n||_E^2 \\
&\leq \frac{1}{\alpha}\Big(||e^0||_E^2 + \varepsilon k\sum_{l=1}^{n}||\tau^l||^2\Big)
\end{aligned}$$

Si $e^0 = 0$

$$||e^n||_\infty^2 \leq \varepsilon\frac{T_{\text{máx}}}{\alpha}\max_{l=1,...,n}||\tau^l||^2$$

■

Ejercicio

Considerar la aproximación por diferencias finitas de

$$\frac{\partial u}{\partial t} - \frac{\partial^2 u}{\partial x^2} + \frac{\partial u}{\partial x} = 0 \tag{2.38}$$

$$u(0,t) = u(1,t) = 0 \quad t > 0 \tag{2.39}$$

$$u(x,0) = v(x) \quad 0 < x < 1 \tag{2.40}$$

mediante el esquema numérico

$$\partial_t u_j^n - \partial_x(\bar{\partial}_x u)_j^{n+1} + \bar{\partial}_x u_j^{n+1} = 0 \quad n > 0,\ j = 1,...,M$$
$$u_0^n = u_{M+1}^n = 0 \quad n > 0$$
$$u_j^0 = v_j \quad j = 1,...,M$$

Analizar el orden y la estabilidad del método.

Solución

La aproximación de $\frac{\partial u}{\partial x}(x_j,t_n)$ mediante $\bar{\partial}_x u_j^{n+1}$ es una aproximación de orden 1 en la variable x. Por otra parte el método de Euler implícito es de orden 1 en la variable t. El error de consistencia es pues $\tau_j^n = \mathscr{O}(h) + \mathscr{O}(k)$. Para la estabilidad: Multiplicando escalarmente por u^{n+1} las ecuaciones que definen el esquema propuesto y utilizando (2.17) con $\theta = 1$

$$\frac{1}{2}\partial_t ||u^n||^2 + \frac{1}{2}k||\partial_t u^n||^2 + ||\bar{\partial}_x u^{n+1}||^2$$
$$= -(\bar{\partial}_x u^{n+1}, u^{n+1}) \leq \frac{\varepsilon}{2}||\bar{\partial}_x u^{n+1}||^2 + \frac{1}{2\varepsilon}||u^{n+1}||^2$$

Utilizando el lema (2.4) $||u^{n+1}||^2 \leq \frac{1}{2}||\bar{\partial}_x u^{n+1}||^2$ y teniendo en cuenta que

$$\frac{1}{2}k||\partial_t u^n||^2 \geq 0$$

tendremos

$$\frac{1}{2}\partial_t ||u^n||^2 + ||\bar{\partial}_x u^{n+1}||^2 \leq (\frac{\varepsilon}{2} + \frac{1}{4\varepsilon})||\bar{\partial}_x u^{n+1}||^2$$

El valor de $\varepsilon > 0$ para el que el coeficiente $\frac{\varepsilon}{2} + \frac{1}{4\varepsilon}$ toma el valor mínimo es $\varepsilon = \frac{\sqrt{2}}{2}$ y este valor mínimo es $\frac{\sqrt{2}}{2}$. Finalmente resulta

$$\frac{1}{2}\partial_t||u^n||^2+(1-\frac{\sqrt{2}}{2})||\bar{\partial}_x u^{n+1}||^2\leq 0$$

o bien

$$||u^{n+1}||^2+k(2-\sqrt{2})||\bar{\partial}_x u^{n+1}||^2\leq ||u^n||^2$$

que prueba la estabilidad.

Ejercicio

Considerar la aproximación por diferencias finitas de

$$\frac{\partial u}{\partial t}-\frac{\partial}{\partial x}\big(D_{-1/2}\frac{\partial u}{\partial x}\big)+\frac{\partial u}{\partial x}=0 \tag{2.41}$$

$$u(0,t)=u(1,t)=0 \quad t>0 \tag{2.42}$$

$$u(x,0)=v(x) \quad 0<x<1 \tag{2.43}$$

donde

$$0<\alpha\leq D(x)\leq\beta<\infty$$

mediante el esquema numérico

$$\partial_t u_j^n-\partial_x(D_{-1/2}\bar{\partial}_x u)_j^{n+1/2}+\frac{1}{2}\big(\bar{\partial}_x u_j^{n+1/2}+\partial_x u_j^{n+1/2}\big)=0 \quad n\geq 0,\ j=1,...,M$$

$$u_0^n=u_{M+1}^n=0 \quad n>0$$

$$u_j^0=v_j \quad j=1,...,M$$

Analizar el orden y la estabilidad del método.

Solución

La aproximación del término

$$\frac{\partial u}{\partial x}$$

se puede escribir de la forma

$$\frac{u_{j+1}^{n+1/2}-u_{j-1}^{n+1/2}}{2h}$$

que es una aproximación de orden 2 de la derivada con respecto a x. Por otra parte la aproximación temporal corresponde al método de Crank-Nicolson que es de orden 2. Tenemos pues que el error de consistencia es $\tau_j^n=\mathscr{O}(h^2)+\mathscr{O}(k^2)$.

Estudiemos la estabilidad. Multiplicando escalarmente por $u^{n+1/2}$ las ecuaciones que definen el esquema propuesto

$$(\partial_t u^n, u^{n+1/2}) + (D_{-1/2}\bar{\partial}_x u^{n+1/2}, \bar{\partial}_x u)^{n+1/2}) + \frac{1}{2}(\bar{\partial}_x u^{n+1/2}, u^{n+1/2}) + \frac{1}{2}(\partial_x u^{n+1/2}, u^{n+1/2})$$

teniendo en cuenta el lema (2.3) con $\theta = 1/2$ y que

$$(\partial_x u^{n+1/2}, u^{n+1/2}) = -(u^{n+1/2}, \bar{\partial}_x u^{n+1/2}) = -(\bar{\partial}_x u^{n+1/2}, u^{n+1/2})$$

resulta

$$\frac{1}{2}\partial_t ||u^n||^2 + (D_{-1/2}\bar{\partial}_x u^{n+1/2}, \bar{\partial}_x u^{n+1/2}) = 0$$

de donde finalmente, como $(D_{-1/2}\bar{\partial}_x u^{n+1/2}, \bar{\partial}_x u^{n+1/2}) \geq \alpha ||\bar{\partial}_x u^{n+1/2}||^2 \geq 0$

$$||u^{n+1}||^2 + 2k\alpha ||\bar{\partial}_x u^{n+1/2}||^2 \leq ||u^n||^2$$

Ejercicio

Considerar el problema de contorno siguiente

$$\begin{aligned} -\varepsilon u'' + u' = 0 \quad 0 < x < 1 \\ u(0) = 1; \quad u(1) = 0 \end{aligned}$$

donde $\varepsilon > 0$

1. Calcular la solución exacta.
2. Considerar el esquema en diferencias finitas

$$\begin{aligned} &\frac{\varepsilon}{h^2}(-u_{j-1} + 2u_j - u_{j+1}) + \frac{1}{2h}(u_{j+1} - u_{j-1}) \\ &u_0 = 1; \quad u(1) = 0 \end{aligned}$$

   y calcular la solución
3. Observar el comportamiento de esta solución para valores de $\varepsilon < h/2$ y comparar con la solución exacta.
4. Considerar el esquema descentrado siguiente

$$\begin{aligned} &\frac{\varepsilon}{h^2}(-u_{j-1} + 2u_j - u_{j+1}) + \frac{\alpha}{h}(u_j - u_{j-1}) + \frac{1-\alpha}{2h}(u_{j+1} - u_{j-1}) \\ &u_0 = 1; \quad u(1) = 0 \end{aligned}$$

   con $\alpha \geq 0$. Observar que para $\alpha = \alpha_{crit} > 1 - \frac{2\varepsilon}{h}$ si $\varepsilon \leq \frac{h}{2}$ se suprimen las oscilaciones de la solución numérica.

5. Verificar que la solución del esquema descentrado coincide con la solución exacta, $u_j = u(x_j)$, si $r = e^{h/\varepsilon}$, es decir si $\alpha = \alpha_{opt} = \coth(\frac{h}{2\varepsilon}) - \frac{2\varepsilon}{h}$

Indicaciones para la solución

1. La respuesta es
$$u(x) = \frac{e^{1/\varepsilon} - e^{x/\varepsilon}}{e^{1/\varepsilon} - 1} \quad 0 \le x \le 1$$
2. Ensayando una solución de la forma $u_j = a + br^j$ se obtiene
$$a = -\frac{r^{M+1}}{1 - r^{M+1}} \quad b = \frac{1}{1 - r^{M+1}}$$
$$r = \frac{1 + \frac{h}{2\varepsilon}}{1 - \frac{h}{2\varepsilon}}$$
3. La solución oscila si $r < 0$, es decir si $\varepsilon < \frac{h}{2}$
4. La solución es
$$u_j = a + br^j$$
$$r = \frac{1 + (1+\alpha)\frac{h}{2\varepsilon}}{1 - (1-\alpha)\frac{h}{2\varepsilon}}$$

2.3. Problemas con condiciones de contorno de Neuman

Consideramos aquí el problema con condiciones de contorno de Neuman y mixtas. Sea el problema parabólico con coeficientes variables en dimensión 1 con condiciones de contorno de Neuman homogéneas: Hallar la función

$$u : [0,1] \times [0,T] \to \mathbb{R} \tag{2.44}$$
$$x,t \to u(x,t) \tag{2.45}$$

verificando

$$\frac{\partial u}{\partial t} - \frac{\partial}{\partial x}\big(D(x)\frac{\partial u}{\partial x}\big) = f \quad \forall x \in (0,1), \forall t \in (0,T] \tag{2.46}$$
$$(D(x)\frac{\partial u}{\partial x})(0,t) = (D(x)\frac{\partial u}{\partial x})(1,t) = 0 \quad t > 0 \tag{2.47}$$
$$u(x,0) = v(x) \quad 0 < x < 1 \tag{2.48}$$

donde $D : [0,1] - > \mathbb{R}$ con propiedades de continuidad y derivabilidad suficientes verificando que existe β y α tales que

$$0 < \alpha \leq D(x) \leq \beta < \infty \quad \forall x \in [0,1]$$

Vamos a estudiar un método de Euler implícito para este problema. Para poder aproximar bien las condiciones de contorno la malla del intervalo $[0,1]$ no coincidirá con los puntos $x = 0$ y $x = 1$, de modo que los puntos x_j donde aproximamos la solución serán

$$x_j = -\frac{1}{2}h + jh \quad j = 0, ..., j = M+1$$

de este modo podemos elegir una aproximación centrada de las derivadas primeras en los puntos extremos del intervalo y por lo tanto una aproximación de orden 2, es decir $\mathscr{O}(h^2)$, de las ecuaciones en el contorno. El esquema de Euler implícito se

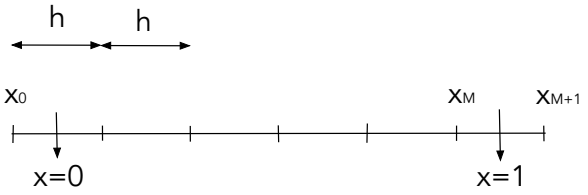


Figura 2.1 Mallado del intervalo

escribe

$$\bar{\partial}_t u_j^n - \partial_x (D_{-1/2} \bar{\partial}_x u)_j^n = f_j^n \quad n \geq 1;\ j = 1, ..., M \tag{2.49}$$

$$D_{1/2} \bar{\partial}_x u_1^n = D_{M+1/2} \bar{\partial}_x u_{M+1}^n = 0 \quad n \geq 1 \tag{2.50}$$

$$u_j^0 = v(-\frac{h}{2} + jh) \quad j = 1, ..., M \tag{2.51}$$

En cada paso tendremos un sistema de $M+2$ ecuaciones con $M+2$ incógnitas. Más adelante veremos que tiene solución única. La aproximación de las condiciones de contorno (2.50) es

$$(\frac{\partial u}{\partial x})(0,t) \approx \frac{u_1^n - u_0^n}{h}$$
$$(\frac{\partial u}{\partial x})(1,t) \approx \frac{u_{M+1}^n - u_M^n}{h}$$

que son aproximaciones centradas en $x = 0$ y $x = 1$ de la derivada con respecto a x. La aproximación es por lo tanto de orden $\mathscr{O}(h^2)$. Por otra parte al ser el método de Euler implícito la aproximación de la la derivada respecto a la variable t es de orden 1. De modo que el error de consistencia será de orden $\tau_j^n = \mathscr{O}(k) + \mathscr{O}(h^2)$. Vamos a continuación a estudiar la estabilidad del método. Para ello necesitamos extender la fórmula de suma por partes al caso en que los valores en los extremos del intervalo no son nulos.

Definición 2.2 *Definimos el producto escalar siguiente*

$$(v,w)_{(r,s)} = h\sum_{j=r}^{s} v_j w_j$$

donde $r = 0,1,2$ *y* $s = M, M+1$ *pueden tomar estos valores según los casos.*

Asimismo denotamos la correspondiente norma mediante $||v||_{(r,s)} = (v,w)^{1/2}_{(r,s)}$

Lema 2.9 *Definimos el producto escalar siguiente*

$$(v,w)_{(r,s)} = h\sum_{j=r}^{s} v_j w_j$$

donde $r = 0,1,2$ *y* $s = M, M+1$ *puede tomar estos valores según los casos.*

Tenemos la siguiente fórmula de suma por partes

$$(\partial_x v, w)_{(r,s)} = h\sum_{j=r}^{s} \frac{v_{j+1}-v_j}{h} w_j = -(v, \bar{\partial}_x w)_{(r+1,s)} + v_{s+1}w_s - v_r w_r \tag{2.52}$$

Demostración.

$$\begin{aligned}
(\partial_x v, w)_{(r,s)} &= h\sum_{j=r}^{s} \frac{v_{j+1}-v_j}{h} w_j = h\sum_{j=r}^{s} \frac{v_{j+1}w_j}{h} - h\sum_{j=r}^{s} \frac{v_j w_j}{h} \\
&= -\big(h\sum_{j=r}^{s} \frac{v_j w_j}{h} - h\sum_{j=r}^{s} \frac{v_{j+1}w_j}{h}\big) \\
&= -\big(h\sum_{j=r}^{s} \frac{v_j w_j}{h} - h\sum_{j=r+1}^{s+1} \frac{v_j w_{j-1}}{h}\big) \\
&= -\big(h\sum_{j=r+1}^{s} \frac{v_j w_j}{h} - h\sum_{j=r+1}^{s} \frac{v_j w_{j-1}}{h} + v_r w_r - v_{s+1}w_s\big) \\
&= -(v, \bar{\partial}_x w)_{(r+1,s)} + v_{s+1}w_s - v_r w_r
\end{aligned}$$

■

Estamos ahora en condiciones de demostrar el siguiente resultado de estabilidad.

Teorema 2.6 *La solución del esquema definido por la ecuaciones (2.49)-(2.50)-(2.51) verifica*

$$||u^n||_{(1,M)} \leq ||u^0||_{(1,M)} + k\sum_{l=1}^{n} ||f^l||_{(1,M)} \tag{2.53}$$

Demostración. Multiplicamos (2.49) escalarmente por u^n con respecto al producto escalar $(.,.)_{(1,M)}$ y sumando por partes el segundo término del primer miembro

$$
\begin{aligned}
&(\bar{\partial}_t u^n, u^n)_{(1,M)} - (\partial_x (D_{-1/2}\bar{\partial}_x u^n, u^n)_{(1,M)} \\
&= (\bar{\partial}_t u^n, u^n)_{(1,M)} + (D_{-1/2}\bar{\partial}_x u^n, \bar{\partial}_x u^n)_{(2,M)} - D_{M+1/2}\bar{\partial}_x u^n_{M+1} u^n_M + D_{1/2}\bar{\partial}_x u^n_1 u^n_1 \\
&= (f^n, u^n)_{(1,M)}
\end{aligned}
$$

Teniendo en cuenta las condiciones de contorno (2.50) tenemos

$$(\bar{\partial}_t u^n, u^n)_{(1,M)} + (D_{-1/2}\bar{\partial} u^n, \bar{\partial} u^n)_{(2,M)} = (f^n, u^n)_{(1,M)}$$

observando que

$$(D_{-1/2}\bar{\partial} u^n, \bar{\partial} u^n)_{(2,M)} \geq 0$$

podemos escribir

$$(\frac{u^n - u^{n-1}}{k}, u^n)_{(1,M)} \leq (f^n, u^n)_{(1,M)}$$

de donde

$$||u^n||^2_{(1,M)} \leq (u^{n-1}, u^n)_{(1,M)} + k(f^n, u^n)_{(1,M)}$$

aplicando la desigualdad de Cauchy-Schwarz y simplificando

$$||u^n||_{(1,M)} \leq ||u^{n-1}||_{(1,M)} + k||f^n||_{(1,M)}$$

aplicando recursivamente la relación anterior obtenemos el resultado buscado. ■

Corolario 2.1 *La solución de (2.49)-(2.50)-(2.51) es única.*

Demostración. En cada paso n tenemos un sistema de $M+2$ ecuaciones con $M+2$ incógnitas . Basta demostrar que si $f^n = 0$ para todo n y $u^0 = 0$ entonces $u^n = 0$ para todo n. En efecto si $f^1 = 0$ y $u^0 = 0$ entonces $||u^1||_{(1,M)} \leq ||u^0||_{(1,M)} = 0$ lo que implica $u^1 = 0$ y aplicamos recursivamente el mismo razonamiento teniendo en cuenta que $f^n = 0$ para $n \geq 2$. La solución es pues idénticamente nula. ■

Ejercicio

Problema con condiciones de contorno mixtas.

$$\frac{\partial u}{\partial t} - \frac{\partial}{\partial x}\big(D(x)\frac{\partial u}{\partial x}\big) = f \quad \forall x \in (0,1), \forall t \in (0,T] \tag{2.54}$$

$$(D(x)\frac{\partial u}{\partial x} - cu)(0,t) = (D(x)\frac{\partial u}{\partial x} + du)(1,t) = 0 \quad t > 0 \tag{2.55}$$

$$u(x,0) = v(x) \quad 0 < x < 1 \tag{2.56}$$

donde $D : [0,1] \to \mathbb{R}$ con propiedades de continuidad y derivabilidad suficientes verificando que existe β y α tales que

$$0 < \alpha \leq D(x) \leq \beta < \infty \quad \forall x \in [0,1]$$

y las constantes c y d verifican $c, d \geq 0$. Analizar el orden y estabilidad del esquema numérico siguiente

$$\bar{\partial}_t u_j^n - \partial_x (D_{-1/2} \bar{\partial}_x u)_j^n = f_j^n \quad j = 1, ..., M;\ n \geq 1 \tag{2.57}$$

$$D_{1/2} \bar{\partial}_x u_1^n - c u_1^n = D_{M+1/2} \bar{\partial}_x u_{M+1}^n + d u_M^n = 0 n \geq 1 \quad n \geq 1 \tag{2.58}$$

$$u_j^0 = v(-\frac{h}{2} + jh) \quad j = 1, ..., M \tag{2.59}$$

Solución

El orden de consistencia con respecto a x es solo $\mathscr{O}(h)$ debido a la aproximación de las condiciones de contorno. Para obtener una aproximación de orden 2 deberiamos aproximar las condiciones de mixtas de la forma

$$D_{1/2} \bar{\partial}_x u_1^n - c \frac{u_1^n + u_0^n}{2} = D_{M+1/2} \bar{\partial}_x u_{M+1}^n + d \frac{u_{M+1}^n + u_M^n}{2} = 0$$

La estabilidad se obtiene del mismo modo que en el caso de las condiciones de Neuman, teniendo en cuenta que $c(u_1^n)^2 \geq 0$ y $d(u_M^n)^2 \geq 0$.

Capítulo 3
Problemas parabólicos en dimensión espacial mayor que 1

Resumen En este capítulo extendemos los métodos estudiados en el capítulo 1 y 2 a problemas parabólicos en dimensión mayor que 1 .

3.1. Formulación del problema de contorno y valor inicial y propiedad de unicidad

Sea Ω un abierto acotado de $\mathbb{R}^d$ y Γ su frontera. Sean $D_{ij} : \Omega \to \mathbb{R}$ con $i, j = 1, ..., d$, funciones con propiedades de continuidad y derivabilidad suficientes verificando las siguientes propiedades:

1. Existe $\gamma < \infty$ tal que

$$0 < D_{ij}(x) \leq \gamma \quad \forall i, j = 1, ..., d \quad \forall x \in \Omega \tag{3.1}$$

2. Elipticidad: Existe $\alpha > 0$ de modo que

$$\sum_{i,j=1}^{d} D_{ij}(x)\xi_i\xi_j \geq \alpha \sum_{i=1}^{d} \xi_i^2 \quad \forall \xi = (\xi_i)_{i=1}^{d} \in \mathbb{R}^d \quad \forall x \in \Omega \tag{3.2}$$

Sea f la función

$$f : \Omega \times [0,T] \to \mathbb{R} \tag{3.3}$$

$$x, t \to f(x,t) \tag{3.4}$$

que supondremos al menos continua.

Consideramos el problema parabólico con coeficientes variables en dimensión d espacial siguiente: Hallar la función

$$u : \Omega \times [0,T] \to \mathbb{R} \tag{3.5}$$

$$x,t \to u(x,t) \tag{3.6}$$

verificando

$$\frac{\partial u}{\partial t} - \sum_{i,j=1}^{d} \frac{\partial}{\partial x}\big(D_{ij}(x)\frac{\partial u}{\partial x}\big) = f \quad \forall x \in \Omega,\ \forall t \in (0,T] \tag{3.7}$$

$$u(x,t) = 0 \quad \text{sobre } \Gamma \quad t > 0 \tag{3.8}$$

$$u(x,0) = v(x) \quad \text{en } \Omega \tag{3.9}$$

La existencia de solución de este problema está fuera del ámbito de este curso. Sin embargo es fácil obtener la unicidad de solución del problema con técnicas elementales. En determinados casos se pueden calcular soluciones a este problema mediante el método de separación de variables, por ejemplo en el caso de un dominio Ω rectangular y $D_{ij} = \delta_{ij}\alpha$ (caso isótropo y coeficientes constantes).

Teorema 3.1 *El problema (3.7)-(3.8)-(3.9) tiene solución única.*

Demostración. Sean u_1 y u_2 dos soluciones. $\bar{u} = u_1 - u_2$ verifica

$$\frac{\partial \bar{u}}{\partial t} - \sum_{i,j=1}^{d} \frac{\partial}{\partial x}\big(D_{ij}(x)\frac{\partial \bar{u}}{\partial x}\big) = 0 \quad \forall x \in \Omega,\ \forall t \in (0,T]$$

$$\bar{u}(x,t) = 0 \quad \text{sobre } \Gamma \quad t > 0$$

$$\bar{u}(x,0) = 0 \quad \text{en } \Omega$$

Multiplicando los términos de la ecuación (3.7) por $\bar{u}$ e integrando por partes el segundo término

$$\frac{1}{2}\frac{d}{dt}\int_\Omega (\bar{u}(x,t))^2\,dx + \sum_{i,j=1}^{d}\int_\Omega D_{ij}\frac{\partial \bar{u}}{\partial x_i}\frac{\partial \bar{u}}{\partial x_j}\,dx = 0$$

$$\frac{1}{2}\frac{d}{dt}\int_\Omega (\bar{u}(x,t))^2\,dx + \alpha\sum_{i,j=1}^{d}\int_\Omega \big(\frac{\partial \bar{u}}{\partial x_i}\big)^2\,dx \le 0$$

$$\frac{1}{2}\frac{d}{dt}\int_\Omega (\bar{u}(x,t))^2\,dx \le 0$$

integrando entre 0 y t

$$\int_\Omega \bar{u}^2(x,t)\,dx \le \int_\Omega \bar{u}^2(x,0)\,dx = 0$$

lo que implica

$$\bar{u}^2(x,t) = 0 \quad \forall x \in \Omega, \quad \forall t > 0$$

■

3.2. Aproximación mediante diferencias finitas

Consideraremos el caso en que $\Omega \in \mathbb{R}^d$ es la unión de rectángulos ($d = 2$) o paralelepípedos ($d = 3$) con interior de intersección vacía. Más precisamente supondremos que Ω se puede recubrir por una malla cuyas celdas son cuadrados ($d = 2$), respectivamente cubos ($d = 3$) de lado de longitud h, como en la figura (3.1). Aso-

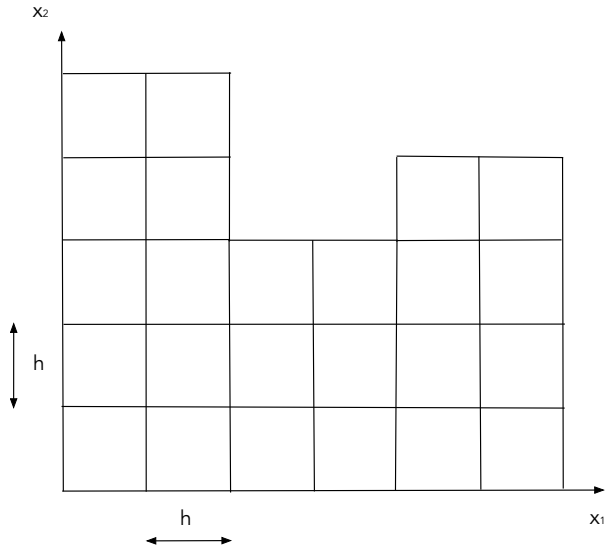


Figura 3.1 Malla de diferencias finitas en 2D

ciados a esta malla, consideramos los puntos nodales (a los que llamamos nodos) de la forma

$$x = \beta h = (\beta_1 h, ..., \beta_d h)$$

donde $\beta = (\beta_1, ..., \beta_d) \in \mathbb{Z}^d$ es un multientero. Designamos mediante Ω_h el conjunto de nodos interiores a Ω y designamos mediante Γ_h el conjunto de nodos sobre Γ. Al conjunto de vecinos de un nodo lo designamos por $x + \delta h$, donde $\delta = (\delta_1, ..., \delta_d) \in \mathbb{Z}^d$ con $|\delta_i| = 1$ $i = 1, ...d$ y que están en $\Omega_h \cup \Gamma_h$

Definición 3.1 *Definimos las siguientes diferencias divididas.*

- *i-ésima diferencia parcial progresiva:*

$$\partial_{x_i} u(x) = \frac{u(x + he_i) - u(x)}{h}$$

- *i-ésima diferencia parcial regresiva:*

$$\bar{\partial}_{x_i} u(x) = \frac{u(x) - u(x - he_i)}{h}$$

donde $e_i = (0, ..., 1, .., 0) \in \mathbb{R}^d$ *donde el* 1 *ocupa el i-ésimo lugar.*

Definición 3.2 *Sea el conjunto de nodos de* $\Omega_h \cup \Gamma_h$ *y el espacio vectorial* $\mathbb{R}^M$ *donde* M *es el cardinal del conjunto de nodos* $\Omega_h \cup \Gamma_h$.

El producto escalar de dos vectores $v = (v(x))_{x \in \Omega_h \cup \Gamma_h}$ *y* $w = (w(x))_{x \in \Omega_h \cup \Gamma_h}$ *en* $\mathbb{R}^M$ *es*

$$(v, w) = h^d \sum_{x = \beta h \in \Omega_h \cup \Gamma_h} v(x) w(x) \tag{3.10}$$

y la norma correspondiente

$$||v|| = (v, v)^{1/2} \tag{3.11}$$

En el análisis de los métodos de diferencias finitas que vamos a estudiar necesitaremos la siguiente fórmula de suma por partes que es la generalización de (2.15).

Lema 3.1 *Para vectores* v *y* w *en* $\mathbb{R}^M$ *que se anulan en* Γ_h *se tiene la siguiente fórmula de suma por partes*

$$(\partial_{x_i} v, w) = -(v, \bar{\partial}_{x_i} w) \tag{3.12}$$

Demostración. La demostración es análoga a la del caso de dimensión 1 dada en el lema (2.1) ■

Para simplificar la descripción de los métodos en diferencias finitas que siguen introduciremos el siguiente operador en diferencias A_h que aproxima el operador en derivadas parciales A definido por

$$Au = -\sum_{i,j=1}^{d} \frac{\partial}{\partial x} \Big(D_{ij}(x) \frac{\partial u}{\partial x} \Big) \tag{3.13}$$

La aproximación en Diferencias Finitas que consideraremos es

$$\mathbf{A}_h v = -\sum_{i,j=1}^{d} \partial_{x_i} \big(D_{ij} \bar{\partial}_{x_j} v \big) \tag{3.14}$$

En particular para vectores $v \in \mathbb{R}^M$ que se anulan sobre Γ_h tenemos la siguiente propiedad

$$\begin{aligned} (\mathbf{A}_h v, v) &= -\sum_{i,j=1}^{d} \big(\partial_{x_i} (D_{ij} \bar{\partial}_{x_j} v), v \big) \\ &= \sum_{i,j=1}^{d} (D_{ij} \bar{\partial}_{x_j} v, \bar{\partial}_{x_i} v) \geq \alpha \sum_{i=1}^{d} ||\bar{\partial}_{x_i} v||^2 \end{aligned} \tag{3.15}$$

Ejemplo: En el caso particular en el que el operador $-A$ es el operador Laplaciano, por ejemplo en dimensión 2

$$A = -\Delta = -\Big(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\Big) \tag{3.16}$$

el operador $\mathbf{A}_h$ está representado, en una malla como la de la figura (3.2), por la matriz

$$\mathbf{A}_h = \frac{1}{h^2}\begin{bmatrix} 4 & -1 & 0 & -1 & 0 & 0 & \dots \\ -1 & 4 & -1 & 0 & -1 & 0 & \dots \\ 0 & -1 & 4 & 0 & 0 & -1 & \dots \\ -1 & 0 & 0 & 4 & -1 & 0 & \dots \\ 0 & -1 & 0 & -1 & 4 & -1 & \dots \\ 0 & 0 & -1 & 0 & -1 & 4 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \end{bmatrix}$$

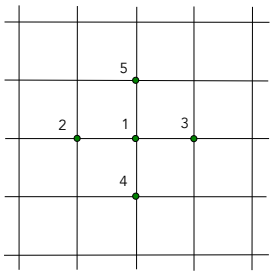


Figura 3.2 Esquema de 5 puntos

3.2.1. Método general de un paso

Para $0 \le \theta \le 1$ el método general de un paso se escribe

$$\partial_t u^n + \theta \mathbf{A}_h u^{n+1} + (1-\theta)\mathbf{A}_h u^n = f^{n+\theta} \quad \text{en } \Omega_h,\ n \ge 0 \tag{3.17}$$

$$u^n = 0 \quad \text{en } \Gamma_h;\ n \ge 0 \tag{3.18}$$

$$u^0(\beta h) = v(\beta h) \quad \text{para } \beta h \in \Omega_h \tag{3.19}$$

y donde $f^{n+\theta}(x) = f(x, t_n + \theta k)$ para cada $x \in \Omega_h$.

En cada paso n dado u^n calculamos u^{n+1} resolviendo el sistema lineal algebraico de ecuaciones siguiente

$$(\mathbf{I} + k\theta\mathbf{A}_h)u^{n+1} = (\mathbf{I} - (1-\theta)k\mathbf{A}_h)u^n + kf^{n+\theta}$$

Como $\mathbf{A}_h$ es definida positiva, $\mathbf{I} + k\theta\mathbf{A}_h$ también lo es y el problema anterior tiene solución única.

Teorema 3.2 *El esquema anterior (3.17)-(3.18)-(3.19) es consistente de orden*

$$\tau = \mathcal{O}(h^2 + k) \quad \text{si } \theta \neq \frac{1}{2}$$
$$\tau = \mathcal{O}(h^2 + k^2) \quad \text{si } \theta = \frac{1}{2}$$

Demostración. La demostración es análoga al caso en dimensión 1. ■

Vamos a estudiar la estabilidad en los distintos casos según el valor de θ. En el siguiente teorema obtenemos la estabilidad en el caso $\theta = 1$.

Teorema 3.3 *Sea $\theta = 1$. En el caso homogéneo, es decir $f = 0$ el esquema (3.17)-(3.18)-(3.19) verifica*

$$||u^n|| < ||u^0|| \quad \forall n > 0$$

El el caso general con $f \neq 0$

$$||u^n|| < ||u^0|| + k \sum_{l=1}^{n} ||f^n|| \quad \forall n > 0$$

El método es por lo tanto incondicionalmente estable.

Demostración. Supongamos primero $f = 0$. Multiplicamos escalarmente los términos de (3.17) por u^{n+1}.

$$(\partial_t u^n, u^{n+1}) + (\mathbf{A}_h u^{n+1}, u^{n+1}) = 0$$

como por una parte, gracias al lema (2.3) con $\theta = 1$

$$(\partial_t u^n, u^{n+1}) = \frac{1}{2}\partial_t ||u^n||^2 + \frac{1}{2}k||\partial_t u^n||^2 \geq \frac{1}{2}\partial_t ||u^n||^2$$

y por otra parte la propiedad (3.15)

$$(\mathbf{A}_h u^{n+1}, u^{n+1}) \geq \alpha ||\bar{\partial}_x u^{n+1}||^2 > 0$$

resulta

$$||u^{n+1}||^2 < ||u^n||^2$$

y finalmente tomando la raiz cuadrada positiva y aplicando recursivamente la desigualdad para $n = 0, 1, \ldots$ obtenemos el resultado buscado.

En el caso $f \neq 0$ podemos escribir el esquema como

$$u^{n+1} = \mathbf{T}u^n + k\mathbf{T}f^{n+1}$$

donde $\mathbf{T} = (\mathbf{I} + k\mathbf{A}_h)^{-1}$. De la demostración de la estabilidad en el caso homogéneo deducimos $||\mathbf{T}|| < 1$. Tomando normas y mayorando en la expresión anterior obte-

nemos

$$||u^{n+1}||^2 < ||u^n||^2 + k||f^{n+1}||$$

y aplicando recursivamente esta desigualdad obtenemos el resultado buscado. ■

Vamos a estudiar el caso general, distinguiendo el caso con $\theta \geq 1/2$ y el caso $\theta = 0$. El caso $0 < \theta < 1/2$ no tiene interés práctico. Para ello necesitaremos la siguiente desigualdad de tipo "desigualdad de Poincaré".

Lema 3.2 *Para todo $v \in \mathbb{R}^M$ donde M es el cardinal del conjunto de nodos $\Omega_h \cup \Gamma_h$ con $v = 0$ sobre Γ_h existe una constante C independiente de v y de h tal que*

$$||v|| \leq C\big(\sum_{j=1}^{d} ||\bar{\partial}_{x_j} v||^2\big)^{1/2} \tag{3.20}$$

Demostración. La demostración es la generalización a dimensión d de la demostración del lema (2.4). Separamos explícitamente la componente d-ésima de las primeras $d-1$-componentes. Denotemos los puntos nodales de la forma

$$\begin{aligned} \tilde{\beta} &\in \mathbb{Z}^{d-1} \\ \beta &= (\tilde{\beta}, \beta_d) \in \mathbb{Z}^d \\ \beta h &= (\tilde{\beta} h, \beta_d h) \end{aligned}$$

En la figura (3.3) se representa en el eje de abcisas las $d-1$ primeras coordenadas y en el eje de ordenadas la d-ésima coordenada. Para $v \in \mathbb{R}^M$

$$v(x) = v(\tilde{\beta} h, \beta_d h) = h \sum_{l=1}^{\beta_d} \frac{v(\tilde{\beta} h, lh) - v(\tilde{\beta} h, (l-1)h)}{h}$$

Para los puntos nodales $x = (\beta h)_\beta$ que no forman parte del mallado extendemos v mediante el valor cero. Tomando valores absolutos y mayorando

$$\begin{aligned} |v(x)| &\leq h \sum_{l=1}^{\beta_d} |\frac{v(\tilde{\beta} h, lh) - v(\tilde{\beta} h, (l-1)h)}{h}| \\ &\leq h\big(\sum_{l=1}^{\beta_d} |\frac{v(\tilde{\beta} h, lh) - v(\tilde{\beta} h, (l-1)h)}{h}|^2\big)^{1/2} \big(\sum_{l=1}^{\beta_d} 1^2\big)^{1/2} \end{aligned}$$

elevando al cuadrado

$$|v(x)|^2 \leq \big(h \sum_{l=1}^{\beta_d} |\bar{\partial}_{x_d} v(\tilde{\beta} h, lh)|^2\big).\big(h\beta_d\big)$$

Multiplicando por h^{d-1} y sumando para todos los $\tilde{\beta} \in \mathbb{Z}^{d-1}$

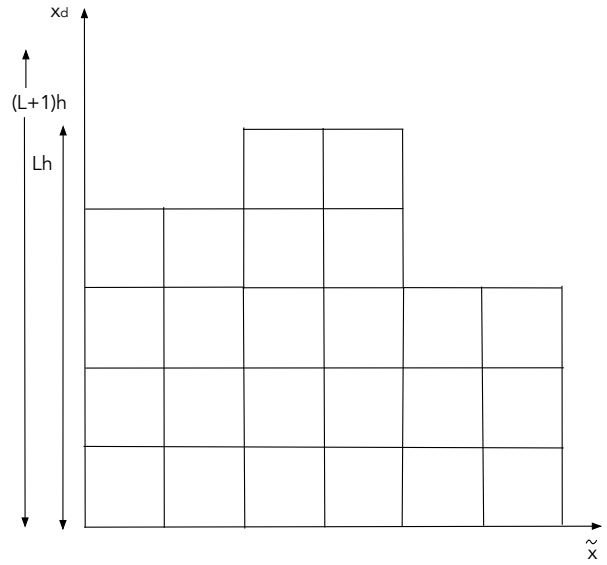


Figura 3.3 Malla en dimension d

$$h^{d-1}\sum_{\tilde{\beta}\in\mathbb{Z}^{d-1}}|v(x)|^2 \le h^d\Big(\sum_{\beta\in\mathbb{Z}^d}|\bar{\partial}_{x_d}v(\beta h)|^2\Big)(h\beta_d) = ||\bar{\partial}_{x_d}v||^2 h\beta_d$$

Multiplicando por h y sumando para $\beta_d = 1,...,L$, teniendo en cuenta que $v(\beta h) = 0$ en los puntos que no pertenecen a Ω_h

$$h^d\sum_{\beta\in\Omega_h}|v(x)|^2 \le ||\bar{\partial}_{x_d}v||^2 h^2\sum_{\beta_d}^{L}\beta_d \le ||\bar{\partial}_{x_d}v||^2 h^2\frac{L(L+1)}{2}$$

llamando $S = (L+1)h$, que es una cota superior de la longitud del dominio Ω en la dirección x_d tenemos

$$||v||^2 \le \frac{S^2}{2}||\bar{\partial}_{x_d}v||^2 \le \frac{S^2}{2}\sum_{i=1}^{d}||\bar{\partial}_{x_i}v||^2$$

$$||v|| \le \frac{S}{\sqrt{2}}\Big(\sum_{i=1}^{d}||\bar{\partial}_{x_i}v||^2\Big)^{1/2}$$

obtenemos el resultado buscado con $C = S/\sqrt{2}$. ■

Observación: Observar que la aplicación

$$\mathbb{R}^M \to \mathbb{R} \tag{3.21}$$

$$v \to \big(\sum_{j=1}^{d} ||\bar{\partial}_{x_i} v||^2\big)^{1/2} \tag{3.22}$$

es una norma en el subespacio $\mathbb{R}^M$ de vectores $v = (v(x))_{x\in\Omega_h\cup\Gamma_h}$ que se anulan sobre Γ_h. En efecto, si $\big(\sum_{j=1}^{d} ||\bar{\partial}_{x_j} v||^2\big)^{1/2} = 0$, para $i = 1, ..., d$

$$v(\beta_1, ..., \beta_i h, ..., \beta_d) - v(\beta_1, ..., (\beta_i - 1)h, ..., \beta_d) = 0 \quad \forall \beta_i = 1, ...L_i + 1$$
$$v(\beta_1, ..., \beta_i h, ..., \beta_d) = v(\beta_1, ..., (\beta_i - 1)h, ..., \beta_d) \quad \forall \beta_i = 1, ...L_i + 1$$

siendo $L_i + 1$ el número de nodos en la dirección i. como $v(..., 0h, ...) = 0$ resulta $v(..., \beta_i h, ...) = 0$ para todo $\beta_i = 1, ..., L_i + 1$. Como esto es cierto para todo i, $v = 0$. Las otras propiedades de norma se deducen de las propiedades de la norma euclídea $||.||$

Teorema 3.4 *La solución de (3.17)-(3.18)-3.19) para $\theta \geq 1/2$ verifica el siguiente resultado de estabilidad*

$$||u^n||^2 \leq ||u^0||^2 + k\frac{C^2}{2\alpha}\sum_{l=0}^{n-1} ||f^{l+\theta}||^2 \tag{3.23}$$

donde C es la constante del lema (3.2)

Demostración. Multiplicamos por $u^{n+\theta}$ los términos de la ecuación (3.17) y aplicando el lema (2.4)

$$\frac{1}{2}\partial_t ||u^n||^2 + (\theta - \frac{1}{2})\frac{k}{2}||\partial_t u^n||^2 + (\mathbf{A}_h u^{n+\theta}, u^{n+\theta}) = (f^{n+\theta}, u^{n+\theta})$$

Minorando el primer miembro, mayorando el segundo miembro aplicando la desigualdad de Cauchy-Schwarz y aplicando el lema (3.2)

$$\frac{1}{2}\partial_t ||u^n||^2 + \alpha\sum_{i=1}^{d} ||\bar{\partial}_{x_i} u^{n+\theta}||^2 \leq \frac{1}{2\varepsilon}||f^{n+\theta}||^2 + \frac{\varepsilon}{2}||u^{n+\theta}||^2$$
$$\leq \frac{1}{2\varepsilon}||f^{n+\theta}||^2 + \frac{C^2\varepsilon}{2}\sum_{i=1}^{d} ||\bar{\partial}_{x_i} u^{n+\theta}||^2$$
$$\frac{1}{2}\partial_t ||u^n||^2 + (\alpha - \frac{C^2\varepsilon}{2})\sum_{i=1}^{d} ||\bar{\partial}_{x_i} u^{n+\theta}||^2 \leq \frac{1}{2\varepsilon}||f^{n+\theta}||^2$$

Elegimos ε de modo que $(\alpha - \frac{C^2\varepsilon}{2}) \geq 0$ por ejemplo $\varepsilon = \frac{2\alpha}{C^2}$ y resulta finalmente

$$||u^{n+1}||^2 \leq ||u^n||^2 + k\frac{C^2}{2\alpha}||f^{n+\theta}||$$

Aplicando recursivamente la estimación anterior para $n = 0, 1, ..., n-1$ obtenemos (3.23). El método es pues incondicionalmente estable. ■

Vamos a estudiar la estabilidad en el caso $\theta = 0$. Para ello necesitaremos la siguiente desigualdad inversa que es una generalización al caso multidimensional del lema (2.5).

Lema 3.3 *Para cada todo $v \in \mathbb{R}^M$ donde donde M es el cardinal del conjunto de nodos $\Omega_h \cup \Gamma_h$ con $v = 0$ sobre Γ_h, existe una constante $C = 2$ independiente de v y de h tal que se verifica la siguiente desigualdad inversa*

$$||\partial_{x_i} v|| \leq \frac{C}{h}||v|| \tag{3.24}$$

Demostración. Podemos suponer sin perder generalidad, ampliando el dominio, que el dominio es un rectángulo (en dimensión 2) o un ortoedro (en dimensión 3) o un hiperparalelepípedo rectángulo en dimensión mayor que 3. Prolongamos el vector v con el valor 0 en los nodos fuera del dominio. En la dirección i-ésima sea L_i el número de intervalos del dominio ampliado. Sea $\beta_i = 0, ..., L_i + 1$.

$$\begin{aligned}
||\partial_{x_i} v||^2 &= h^d \sum_{\beta_i=0}^{L_i} \left(\frac{v(\beta_1, ..., (\beta_i+1)h, ..., \beta_d) - v(\beta_1, ..., \beta_i h, ..., \beta_d)}{h}\right)^2 \\
&= h^{d-2} \sum_{\beta_i=0}^{L_i} (v(\beta_1, ..., (\beta_i+1)h, ..., \beta_d) - v(\beta_1 ..., \beta_i h, ...\beta_d))^2 \\
&\leq h^{d-2} \Big(\sum_{\beta_i=0}^{L_i} 2v^2(\beta_1 ..., (\beta_i+1)h, ...\beta_d) + \sum_{\beta_i=0}^{L_i} 2v^2(\beta_1 ..., \beta_i h, ...\beta_d) \Big) \\
&=\leq h^{d-2} \Big(\sum_{\beta_i=1}^{L_i} 2v^2(\beta_1 ..., \beta_i h, ...\beta_d) + \sum_{\beta_i=1}^{L_i} 2v^2(\beta_1 ..., \beta_i h, ...\beta_d) \Big) \\
&= \frac{4}{h^2} h^d \Big(\sum_{\beta_i=1}^{L_i} v^2(\beta_1 ..., \beta_i h, ...\beta_d) \Big) \\
&\leq \frac{4}{h^2} h^d \Big(\sum_{\beta} v^2(\beta h) \Big)^2
\end{aligned}$$

donde el último sumatorio se extiende a todos los nodos $\beta h \in \Omega_h \cup \Gamma_h$. Extrayendo la raiz cuadrada obtenemos el resultado buscado. ■

En el siguiente teorema obtenemos el resultado de estabilidad para $\theta = 0$.

Teorema 3.5 *El método (3.17)-(3.18)-3.19) para $\theta = 0$ es condicionalmente estable, es decir, el método es estable si se verifica la condición de estabilidad*

$$k \leq \frac{\alpha h^2}{2\gamma d^3} \tag{3.25}$$

siendo $\gamma = \max\{D_{ij}(x);\ x \in \Omega;\ i,j = 1,...,d\}$. *Más precisamente se tiene para la solución* u^n *el siguiente resultado de estabilidad*

$$||u^n|| \leq ||u^0|| + k\sum_{l=0}^{n-1} ||f^l||$$

Demostración. Primeramente observemos que basta estudiar el caso homogéneo, es decir, cuando $f = 0$. En efecto, el esquema numérico se puede escribir

$$u^{n+1} = (\mathbf{I} - k\mathbf{A}_h)u^n + kf^n$$

llamando $\mathbf{T} = (\mathbf{I} - k\mathbf{A}_h)$ y aplicando la fórmula anterior recursivamente

$$u^n = \mathbf{T}^n u^0 + k\sum_{l=0}^{n-1} \mathbf{T}^l f^{n-l-1}$$

Si demostramos que $||\mathbf{T}|| < 1$ siendo aquí la norma $||.||$ la norma matricial inducida por la norma vectorial (3.11) tendremos

$$||u^n|| \leq ||u^0|| + k\sum_{l=0}^{n-1} ||f^{n-l-1}||$$

Para ver que $||\mathbf{T}|| \leq 1$ basta ver que $||u^{n+1}|| \leq ||u^n||$ en el caso $f^n = 0$, que es el caso homogéneo. Estudiemos pues la estabilidad para la ecuación en diferencias finitas

$$\partial_t u^n + \mathbf{A}_h u^n = 0$$

Multiplicando escalarmente por u^n, aplicando el lema (2.3) con $\theta = 0$, minorando el término elíptico y reordenando términos, resulta

$$\frac{1}{2}\partial_t ||u^n||^2 - \frac{k}{2}||\partial_t u^n||^2 + (\mathbf{A}_h u^n, u^n) = 0$$
$$\frac{1}{2}\partial_t ||u^n||^2 + \alpha\sum_{i=1}^{d} ||\bar{\partial}_{x_i} u^n||^2 \leq \frac{k}{2}||\partial_t u^n||^2 = \frac{k}{2}||\mathbf{A}_h u^n||^2$$

y aplicando el lema (3.3)

$$
\begin{aligned}
||\mathbf{A}_h u^n|| &= || \sum_{i,j=1}^{d} \partial_{x_i}(D_{ij}\bar{\partial}_{x_j}u^n)|| \leq \sum_{i,j=1}^{d} ||\partial_{x_i}(D_{ij}\bar{\partial}_{x_j}u^n)|| \\
&\leq \frac{2}{h}\sum_{i,j=1}^{d} ||D_{ij}\bar{\partial}_{x_j}u^n)|| \leq \frac{2\gamma}{h}\sum_{i,j=1}^{d} ||\bar{\partial}_{x_j}u^n)|| \\
&\leq \frac{2\gamma d}{h}\sum_{j=1}^{d} ||\bar{\partial}_{x_j}u^n)|| \leq \frac{2\gamma d}{h}\big(\sum_{j=1}^{d} ||\bar{\partial}_{x_j}u^n)||^2\big)^{1/2}\big(\sum_{j=1}^{d} 1^2\big)^{1/2} \\
&= \frac{2\gamma d\sqrt{d}}{h}\big(\sum_{j=1}^{d} ||\bar{\partial}_{x_j}u^n)||^2\big)^{1/2}
\end{aligned}
$$

Reuniendo los resultados anteriores

$$\frac{1}{2}\partial_t||u^n||^2 + \alpha\sum_{i=1}^{d}||\bar{\partial}_{x_i}u^n||^2 \leq \frac{k}{2}\frac{4\gamma^2 d^3}{h^2}\sum_{j=1}^{d}||\bar{\partial}_{x_j}u^n||^2$$

y agrupando términos

$$\frac{1}{2}\partial_t||u^n||^2 + (\alpha - \frac{k}{h^2}2\gamma^2 d^3)\sum_{j=1}^{d}||\bar{\partial}_{x_j}u^n||^2 \leq 0$$

Si elegimos k y h de modo que

$$(\alpha - \frac{k}{h^2}2\gamma^2 d^3) \geq 0$$

es decir, si se verifica la condición de estabilidad (3.25) tendremos

$$||u^{n+1}|| \leq ||u^n||$$

y por tanto $||\mathbf{T}|| \leq 1$ ■

3.3. Otros métodos: Splitting y direcciones alternadas

Consideraremos el problema modelo siguiente:
En $\Omega \in \mathbb{R}^2$ con Γ la frontera de Ω, hallar u tal que

$$\frac{\partial u}{\partial t} - (\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}) = f \quad \text{en } \Omega,\ t > 0 \tag{3.26}$$

$$u(x,t) = 0 \quad \text{sobre } \Gamma,\ t > 0 \tag{3.27}$$

$$u(x,0 = v(x) \quad \text{en } \Omega \tag{3.28}$$

Si utilizamos diferencias centrales para aproximar el operador

$$-\Delta = -(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2})$$

escribimos

$$\mathbf{A}_h u(x,y) = \mathbf{A}_1 u(x,y) + \mathbf{A}_2 u(x,y)$$

$$\mathbf{A}_1 u(x,y) = -\frac{u(x-h,y) - 2u(x,y) + u(x+h,y)}{h^2}$$
$$\mathbf{A}_2 u(x,y) = -\frac{u(x,y-h) - 2u(x,y) + u(x,y+h)}{h^2}$$

3.3.1. Splitting

La técnica del "Splitting" consiste en descomponer en dos ecuaciones la aproximación temporal

$$\frac{u^{n+1/2} - u^n}{k} + \mathbf{A}_1 u^{n+1/2} = 0 \tag{3.29}$$
$$\frac{u^{n+1} - u^{n+1/2}}{k} + \mathbf{A}_2 u^{n+1} = f^n \tag{3.30}$$

que con notación matricial se escribe

$$(\mathbf{I} + k\mathbf{A}_1)u^{n+1/2} = u^n$$
$$(\mathbf{I} + k\mathbf{A}_2)u^{n+1} = u^{n+1/2} + kf^n$$

eliminando $u^{n+1/2}$

$$(\mathbf{I} + k\mathbf{A}_2)u^{n+1} = (\mathbf{I} + k\mathbf{A}_1)^{-1}u^n + kf^n$$
$$u^{n+1} = (\mathbf{I} + k\mathbf{A}_2)^{-1}(\mathbf{I} + k\mathbf{A}_1)^{-1}u^n + k(\mathbf{I} + k\mathbf{A}_2)^{-1}f^n$$

En cada paso resolvemos sistemas análogos a los del caso 1-dimensional, donde $\mathbf{A}_1$ y $\mathbf{A}_2$ son matrices tridiagonales. Veamos en la práctica como se organizan los cálculos. Consideremos un mallado como el de la figura (3.4), con M puntos nodales en las dirección según el eje x, y L puntos nodales en las dirección según el eje y. El algoritmo de "Splitting" se escribe de la siguiente manera

$$\frac{u_{i,j}^{n+1/2} - u_{i,j}^n}{k} + \frac{-u_{i-1,j}^{n+1/2} + 2u_{i,j}^{n+1/2} - u_{i+1,j}^{n+1/2}}{h^2} = 0 \quad i = 1,...M \; j = 1,...,L$$
$$\frac{u_{i,j}^{n+1} - u_{i,j}^{n+1/2}}{k} + \frac{-u_{i,j-1}^{n+1} + 2u_{i,j}^{n+1} - u_{i,j+1}^{n+1}}{h^2} = f_{i,j}^n \quad i = 1,...M \; j = 1,...,L$$

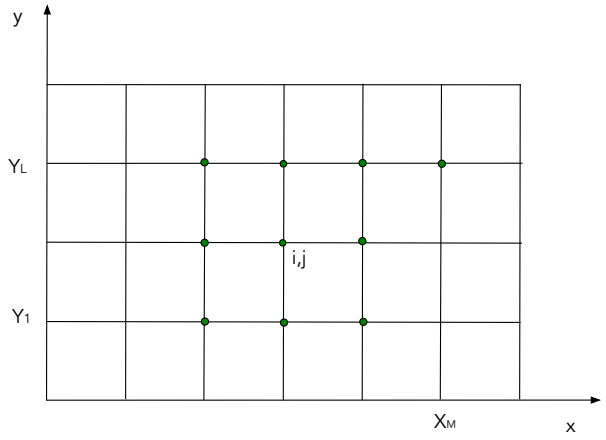


Figura 3.4 Mallado de diferencias finitas en un splitting

El número de incógnitas es $L \times M$. Veamos la complejidad algorítmica:

En el primer paso resolvemos L sistemas tridiagonales de M ecuaciones.

En el segundo paso resolvemos M sistemas tridiagonales de L ecuaciones.

En total resolvemos :

L sistemas tridiagonales de M ecuaciones lo que implica $CL \times M$ operaciones. donde C es una constante independiente de L y de M.

M sistemas tridiagonales de L ecuaciones lo que implica $CM \times L$ operaciones.

En definitiva el número de operaciones totales es $2CL \times M = C'L \times M$ operaciones. Es decir el número de operaciones en cada paso es proporcional al número de incógnitas.

Estudiamos a continuación la consistencia y estabilidad del método (3.29)-(3.30).

Consistencia

Teorema 3.6 *El esquema (3.29)-(3.30) es consistente de orden* 1 *en la variable t y de orden* 2 *en la variable x, es decir que el error de consistencia es*

$$\tau_i^n = \mathscr{O}(k+h^2) \quad \forall i = 1,...,L \times M \; \forall n > 0$$

.

Demostración. Dada una matriz $\mathbf{A}$, si $k||\mathbf{A}|| < 1$ tenemos

$$(\mathbf{I} + k\mathbf{A}^{-1}) = \mathbf{I} - k\mathbf{A} + k^2\mathbf{A}^2...$$

En nuestro caso si elegimos para k suficientemente pequeño, de manera que $k||\mathbf{A}_1|| < 1$ y $k||\mathbf{A}_2|| < 1$ resulta

$$(\mathbf{I}+k\mathbf{A}_1)^{-1}) = \mathbf{I}-k\mathbf{A}_1+k^2\mathbf{A}_1^2...$$
$$(\mathbf{I}+k\mathbf{A}_2)^{-1}) = \mathbf{I}-k\mathbf{A}_2+k^2\mathbf{A}_2^2...$$

de donde

$$\begin{aligned}
u^{n+1} &= (\mathbf{I}+k\mathbf{A}_2^{-1})(\mathbf{I}+k\mathbf{A}_1^{-1})u^n + k(\mathbf{I}+k\mathbf{A}_2^{-1})f^n \\
&= (\mathbf{I}-k\mathbf{A_2}+k^2\mathbf{A}_2^2-...)(\mathbf{I}-k\mathbf{A_1}+k^2\mathbf{A}_1^2-...)u^n \\
&+ k(\mathbf{I}-k\mathbf{A_2}+k^2\mathbf{A}_2^2-...)f^n \\
&= (\mathbf{I}-k(\mathbf{A}_1+\mathbf{A}_2)+\mathscr{O}(k^2))u^n + kf^n + \mathscr{O}(k^2) \\
&= (\mathbf{I}+k\mathbf{A}_h)u^n + kf^n + \mathscr{O}(k^2)
\end{aligned}$$

es decir que el esquema tiene el mismo orden de consistencia que el método de Euler explícito, por tanto el error de consistencia es $\tau_i^n = \mathscr{O}(k+h^2)$. ■

Estabilidad

Estudiemos la estabilidad. Para ello necesitaremos el siguiente resultado previo.

Lema 3.4 *Si* $\mathbf{A}$ *es semidefinida positiva, es decir* $(Av,v) \geq 0$ *para todo v, entonces*

$$||(\mathbf{I}+k\mathbf{A})^{-1}|| \leq 1 \quad \forall k \geq 0$$

Demostración.

$$||(\mathbf{I}+k\mathbf{A})^{-1}|| = \sup_{v\neq 0} \frac{||(\mathbf{I}+k\mathbf{A})^{-1}v||}{||v||}$$

poniendo $\varphi = (\mathbf{I}+k\mathbf{A})^{-1}v$

$$\begin{aligned}
\frac{||(\mathbf{I}+k\mathbf{A})^{-1}v||^2}{||v||^2} &= \frac{||\varphi||^2}{||(\mathbf{I}+k\mathbf{A})\varphi||^2} \\
&= \frac{||\varphi||^2}{||\varphi||^2+k((\mathbf{A}+\mathbf{A}^t)\varphi,\varphi)+k^2||\mathbf{A}\varphi||^2} \leq 1
\end{aligned}$$

■

Teorema 3.7 *El esquema (3.29)-(3.30) es estable y se verifica*

$$||u^n|| \leq ||u^0|| + k\sum_{l=1}^{n} ||f^{l-1}||$$

Demostración. El esquema anterior es de la forma

$$u^{n+1} = \mathbf{T}u^n + k\mathbf{S}f^n$$

donde

$$\mathbf{T} = (\mathbf{I} + k\mathbf{A}_2)^{-1}(\mathbf{I} + k\mathbf{A}_1)^{-1}$$
$$\mathbf{S} = (\mathbf{I} + k\mathbf{A}_2)^{-1}$$

Aplicando recursivamente la relación anterior tendremos

$$u^n = \mathbf{T}^n u^0 + k\sum_{l=1}^{n} \mathbf{T}^{n-l}\mathbf{S}f^{l-1}$$

Bastará ver que $||\mathbf{S}|| \leq 1$ y $||\mathbf{T}|| \leq 1$. En efecto, $\mathbf{S}$ verifica $||\mathbf{S}|| \leq 1$ gracias al lema anterior. Por otra parte

$$||\mathbf{T}|| \leq ||(\mathbf{I} + k\mathbf{A}_2^{-1})||.||(\mathbf{I} + k\mathbf{A}_1)^{-1}|| \leq 1$$

aplicando de nuevo el lema anterior. ■

3.3.2. Método de Splitting basado en el método de Cranc-Nicolson

Veamos un método de "Splitting" construido a partir del método de Cranc-Nicolson en dimensión 1. Supondremos:

- Caso homogéneo, es decir $f = 0$
- $\mathbf{A} = \mathbf{A}_1 + \mathbf{A}_2$ con $\mathbf{A}_1\mathbf{A}_2 = \mathbf{A}_2\mathbf{A}_1$
- $\mathbf{A}_1$ y $\mathbf{A}_2$ definidas positivas.

El esquema es el siguiente:

$$\frac{u^{n+1/2} - u^n}{k} + \mathbf{A}_1\frac{u^{n+1/2} + u^n}{2} = 0 \tag{3.31}$$
$$\frac{u^{n+1} - u^{n+1/2}}{k} + \mathbf{A}_2\frac{u^{n+1} + u^{n+1/2}}{2} = 0 \tag{3.32}$$

Eliminando $u^{n+1/2}$ se puede escribir

$$u^{n+1} = Tu^n$$

con

$$T = (\mathbf{I} + \frac{k}{2}\mathbf{A}_2)^{-1}(\mathbf{I} - \frac{k}{2}\mathbf{A}_2)(\mathbf{I} + \frac{k}{2}\mathbf{A}_1)^{-1}(\mathbf{I} - \frac{k}{2}\mathbf{A}_1)$$

La disposición práctica de los cálculos es la siguiente:

1. Cálculo explícito:

$$u^{n+1/4} = (\mathbf{I} - \frac{k}{2}\mathbf{A}_1)u^n$$

2. Resolución de L sistemas tridiagonales con M incógnitas:

$$(\mathbf{I}+\frac{k}{2}\mathbf{A}_1)u^{n+1/2} = u^{n+1/4}$$

3. Cálculo explícito:

$$u^{n+3/4} = (\mathbf{I}-\frac{k}{2}\mathbf{A}_2)u^{n+1/2}$$

4. Resolución de M sistemas tridiagonales con L incógnitas:

$$(\mathbf{I}+\frac{k}{2}\mathbf{A}_2)u^{n+1} = u^{n+3/4}$$

Estudiemos la consistencia y estabilidad del método con respecto a la variable t. Observemos que el método es aplicable a cualesquiera que sean las matrices $\mathbf{A}_1$ y $\mathbf{A}_2$ con tal de que cumplan las propiedades requeridas de positividad y conmutatividad.

Consistencia

Teorema 3.8 *El esquema (3.31)-(3.32) es consistente de orden* 2 *en la variable t.*

Demostración. Para valores de k tales que $\frac{k}{2}||\mathbf{A}_1|| < 1$ y $\frac{k}{2}||\mathbf{A}_2|| < 1$ tendremos

$$\begin{aligned}\mathbf{T} &= (\mathbf{I}-\frac{k}{2}\mathbf{A}_2+\frac{k^2}{4}\mathbf{A}_2\mathbf{A}_2+...)(\mathbf{I}-\frac{k}{2}\mathbf{A}_2).(\mathbf{I}-\frac{k}{2}\mathbf{A}_1+\frac{k^2}{4}\mathbf{A}_1\mathbf{A}_1+...)(\mathbf{I}-\frac{k}{2}\mathbf{A}_1)\\ &= (\mathbf{I}-\frac{k}{2}\mathbf{A}_2-\frac{k}{2}\mathbf{A}_2+\frac{k^2}{4}\mathbf{A}_2\mathbf{A}_2+\frac{k^2}{4}\mathbf{A}_2\mathbf{A}_2+...)\\ &.(\mathbf{I}-\frac{k}{2}\mathbf{A}_1-\frac{k}{2}\mathbf{A}_1+\frac{k^2}{4}\mathbf{A}_1\mathbf{A}_1+\frac{k^2}{4}\mathbf{A}_1\mathbf{A}_1+...)\end{aligned}$$

Si $\mathbf{A}_1$ y $\mathbf{A}_2$ conmutan

$$\begin{aligned}\mathbf{T} &= (\mathbf{I}-k\mathbf{A}+\frac{k^2}{2}(\mathbf{A}_1^2+2\mathbf{A}_1\mathbf{A}_2+\mathbf{A}_2^2)+...)\\ \mathbf{T} &= (\mathbf{I}-k\mathbf{A}+\frac{k^2}{2}\mathbf{A}^2+...)\end{aligned}$$

El esquema puede representarse de la forma

$$u^{n+1} = u^n - k\mathbf{A}u^n + \frac{k^2}{2}\mathbf{A}^2u^n + ...$$

o bien

$$\frac{u^{n+1}-u^n}{k} + \mathbf{A}(u^n - \frac{1}{2}k\mathbf{A}u^n + ...) = 0$$

Sustituyendo la expresión $k\mathbf{A}u^n$ de la anterior

$$\frac{u^{n+1}-u^n}{k}+\mathbf{A}\big(u^n+\frac{1}{2}(u^{n+1}-u^n)-\frac{k^2}{4}\mathbf{A}^2u^n+...\big)=0$$
$$\frac{u^{n+1}-u^n}{k}+\mathbf{A}\big(\frac{u^{n+1}+u^n}{2}-\frac{k^2}{4}\mathbf{A}^2u^n+...\big)=0$$

Comparando la iteración anterior con la iteración de Cranc-Nicolson concluimos que este esquema es también de orden 2. ■

Estabilidad

Para el estudio de la estabilidad utilizaremos las siguientes propiedades:

Lema 3.5 *Sea* $\mathbf{A}$ *una matriz semidefinida positiva, es decir,* $(\mathbf{A}v,v)\geq 0$ *para todo vector v, y* σ *un número no negativo, entonces*

$$||(\mathbf{I}-\sigma\mathbf{A})(\mathbf{I}+\sigma\mathbf{A})^{-1}||\leq 1$$

Demostración. Pongamos

$$\mathbf{T}=(\mathbf{I}-\sigma\mathbf{A})(\mathbf{I}+\sigma\mathbf{A})^{-1}$$

tenemos

$$||\mathbf{T}||=\sup_{v\neq 0}\frac{||(\mathbf{I}-\sigma\mathbf{A})(\mathbf{I}+\sigma\mathbf{A})^{-1}v||}{||v||}$$

Llamemos $\varphi=(\mathbf{I}+\sigma\mathbf{A})^{-1}v$. Tendremos

$$\frac{||(\mathbf{I}-\sigma\mathbf{A})(\mathbf{I}+\sigma\mathbf{A})^{-1}v||^2}{||v||^2}=\frac{||(\mathbf{I}-\sigma\mathbf{A})\varphi||^2}{||(\mathbf{I}+\sigma\mathbf{A})\varphi||^2}$$
$$=\frac{||\varphi||^2-2\sigma(\mathbf{A}\varphi,\varphi)+\sigma^2||\mathbf{A}\varphi||^2}{||\varphi||^2+2\sigma(\mathbf{A}\varphi,\varphi)+\sigma^2||\mathbf{A}\varphi||^2}\leq 1$$

■

Lema 3.6 *Sea* σ *y* $\mathbf{A}$ *tal que* $||\sigma\mathbf{A}||<1||$ *de modo que* $\mathbf{I}+\sigma\mathbf{A}$ *tiene inversa, entonces*

$$(\mathbf{I}+\sigma\mathbf{A})^{-1}(\mathbf{I}-\sigma\mathbf{A})=(\mathbf{I}-\sigma\mathbf{A})(\mathbf{I}+\sigma\mathbf{A})^{-1}$$

Demostración. Primero veamos que $\mathbf{I}+\sigma\mathbf{A}$ y $\mathbf{I}-\sigma\mathbf{A}$ conmutan.

$$(\mathbf{I}+\sigma\mathbf{A})(\mathbf{I}-\sigma\mathbf{A})=\mathbf{I}-\sigma^2\mathbf{A}^2=(\mathbf{I}-\sigma\mathbf{A})(\mathbf{I}+\sigma\mathbf{A})$$

Finalmente

$$(\mathbf{I}+\sigma\mathbf{A})^{-1}(\mathbf{I}-\sigma\mathbf{A})=(\mathbf{I}+\sigma\mathbf{A})^{-1}(\mathbf{I}-\sigma\mathbf{A})(\mathbf{I}+\sigma\mathbf{A})(\mathbf{I}+\sigma\mathbf{A})^{-1}$$
$$=(\mathbf{I}+\sigma\mathbf{A})^{-1}(\mathbf{I}+\sigma\mathbf{A})(\mathbf{I}-\sigma\mathbf{A})(\mathbf{I}+\sigma\mathbf{A})^{-1}=(\mathbf{I}-\sigma\mathbf{A})(\mathbf{I}+\sigma\mathbf{A})^{-1}$$

■

Podemos ahora demostrar la estabilidad del método.

Teorema 3.9 *El esquema (3.31)-(3.32) es estable.*

Demostración. Basta demostrar que $||\mathbf{T}|| \leq 1$ donde

$$\mathbf{T} = (\mathbf{I} + \frac{k}{2}\mathbf{A}_2)^{-1}(\mathbf{I} - \frac{k}{2}\mathbf{A}_2)(\mathbf{I} + \frac{k}{2}\mathbf{A}_1)^{-1}(\mathbf{I} - \frac{k}{2}\mathbf{A}_1)$$

Tomando normas y aplicando los dos lemas anteriores

$$\begin{aligned} ||\mathbf{T}|| &\leq ||(\mathbf{I} + \frac{k}{2}\mathbf{A}_2)^{-1}(\mathbf{I} - \frac{k}{2}\mathbf{A}_2)||.||(\mathbf{I} + \frac{k}{2}\mathbf{A}_1)^{-1}(\mathbf{I} - \frac{k}{2}\mathbf{A}_1)|| \\ &= ||(\mathbf{I} - \frac{k}{2}\mathbf{A}_2)(\mathbf{I} + \frac{k}{2}\mathbf{A}_2)^{-1}||.||(\mathbf{I} - \frac{k}{2}\mathbf{A}_1)(\mathbf{I} + \frac{k}{2}\mathbf{A}_1)^{-1}|| \leq 1 \end{aligned}$$

■

3.3.3. Métodos de direcciones alternadas

El punto de partida será el sistema semidiscreto siguiente,

$$\frac{du}{dt} + \mathbf{A}_h u = f \quad \text{donde } \mathbf{A}_h = \mathbf{A}_1 + \mathbf{A}_2 \tag{3.33}$$

$$u(0) = u_0 \tag{3.34}$$

donde $\mathbf{A}_1$ y $\mathbf{A}_2$ son las matrices correspondientes a una aproximación de operadores en derivadas parciales. Por ejemplo como en (3.26) que llevará asociado las correspondientes condiciones de contorno. Aquí estudiaremos la aproximación de la (3.33), es decir la semidiscretización con respecto a la variable t, utilizando la descomposición de $\mathbf{A}_h$ como la suma de $\mathbf{A}_1$ y $\mathbf{A}_2$.

El método que se describe a continuación se conoce como el método de Douglas, Peaceman y Rachford. En cada paso de la variable t, que en la mayor parte de las aplicaciones representa el tiempo, dividimos el intervalo $[t_n, t_{n+1}]$ en 2 pasos:

$$\frac{u^{n+1/2} - u^n}{k/2} + \mathbf{A}_1 u^{n+1/2} + \mathbf{A}_2 u^n = f^{n+1/2} \tag{3.35}$$

$$\frac{u^{n+1} - u^{n+1/2}}{k/2} + \mathbf{A}_1 u^{n+1/2} + \mathbf{A}_2 u^{n+1} = f^{n+1/2} \tag{3.36}$$

o bien multiplicando por $1/2$

$$\frac{u^{n+1/2}-u^n}{k}+\frac{1}{2}(\mathbf{A}_1u^{n+1/2}+\mathbf{A}_2u^n)=\frac{1}{2}f^{n+1/2} \tag{3.37}$$

$$\frac{u^{n+1}-u^{n+1/2}}{k}+\frac{1}{2}(\mathbf{A}_1u^{n+1/2}+\mathbf{A}_2u^{n+1})=\frac{1}{2}f^{n+1/2} \tag{3.38}$$

Vamos a estudiar la consistencia y estabilidad del método anterior.

Consistencia

Teorema 3.10 *El método (3.37)-(3.38) es consistente de orden* 2

Demostración. Sumando (3.37) y (3.38) obtenemos

$$\frac{u^{n+1}-u^n}{k}+\mathbf{A}_1u^{n+1/2}+\mathbf{A}_2(\frac{u^{n+1}+u^n}{2})=f^{n+1/2} \tag{3.39}$$

restando (3.37) de (3.38) obtenemos

$$\frac{u^{n+1}-2u^{n+1/2}+u^n}{k}+\mathbf{A}_2(\frac{u^{n+1}-u^n}{2})=0 \tag{3.40}$$

reordenando (3.40)

$$u^{n+1}-2u^{n+1/2}+u^n+k\mathbf{A}_2(\frac{u^{n+1}-u^n}{2})=0$$

$$u^{n+1/2}=\frac{u^{n+1}+u^n}{2}+\frac{k}{2}\mathbf{A}_2(\frac{u^{n+1}-u^n}{2})$$

$$\mathbf{A}_1u^{n+1/2}=\mathbf{A}_1(\frac{u^{n+1}+u^n}{2})+\frac{k}{2}\mathbf{A}_1\mathbf{A}_2(\frac{u^{n+1}-u^n}{2})$$

y sustituyendo en (3.39)

$$\frac{u^{n+1}-u^n}{k}+\mathbf{A}(\frac{u^{n+1}+u^n}{2})+\frac{k^2}{2}\mathbf{A}_1\mathbf{A}_2(\frac{u^{n+1}-u^n}{2k})x=f^{n+1/2}$$

como

$$\frac{u^{n+1}-u^n}{k}=\frac{\partial u}{\partial t}(t_n+\frac{k}{2})+\mathscr{O}(k^2)$$

hemos obtenido el método de Cranc-Nicolson perturbado por un término de orden $\mathscr{O}(k^2)$ ■

Estabilidad

Empezamos observando que un paso del método (3.35)-(3.36) se puede escribir de la forma

$$u^{n+1} = (\mathbf{I}+\frac{k}{2}\mathbf{A}_2)^{-1}(\mathbf{I}-\frac{k}{2}\mathbf{A}_1)(\mathbf{I}+\frac{k}{2}\mathbf{A}_1)^{-1}(\mathbf{I}-\frac{k}{2}\mathbf{A}_2)u^n$$
$$+\frac{k}{2}(\mathbf{I}+\frac{k}{2}\mathbf{A}_2)^{-1}(\mathbf{I}+\frac{k}{2}\mathbf{A}_1)^{-1}f^{n+1/2}+\frac{k}{2}(\mathbf{I}+\frac{k}{2}\mathbf{A}_2)^{-1}f^{n+1/2}$$

Puesto que $||(\mathbf{I}+\frac{k}{2}\mathbf{A}_i)^{-1}|| \leq \frac{1}{1-(k/2)||\mathbf{A}_i||}$ para $i = 1,2$ basta demostrar la estabilidad en el caso homogéneo, es decir, cuando $f = 0$. Si $f = 0$ un paso de este método de direcciones alternadas se escribe $u^{n+1} = \mathbf{T}u^n$ donde

$$\mathbf{T} = (\mathbf{I}+\frac{k}{2}\mathbf{A}_2)^{-1}(\mathbf{I}-\frac{k}{2}\mathbf{A}_1)(\mathbf{I}+\frac{k}{2}\mathbf{A}_1)^{-1}(\mathbf{I}-\frac{k}{2}\mathbf{A}_2)$$

Veremos que $||u^{n+1}||_C \leq ||u^n||_C$ para una determinada norma $||.||_C$.

Teorema 3.11 *Sea $u^{n+1} = \mathbf{T}u^n$ solución de (3.35)-(3.36) con $f = 0$, es decir,*

$$\mathbf{T} = (\mathbf{I}+\frac{k}{2}\mathbf{A}_2)^{-1}(\mathbf{I}-\frac{k}{2}\mathbf{A}_1)(\mathbf{I}+\frac{k}{2}\mathbf{A}_1)^{-1}(\mathbf{I}-\frac{k}{2}\mathbf{A}_2)$$

entonces

$$||u^{n+1}||_C \leq ||u^n||_C$$

donde $||v||_C = (Cv,v)^{1/2}$ para $C = (\mathbf{I}+\frac{k}{2}\mathbf{A}_2^t)(\mathbf{I}+\frac{k}{2}\mathbf{A}_2)$.

Demostración. Hagamos el cambio de variable siguiente

$$w^n = (\mathbf{I}+\frac{k}{2}\mathbf{A}_2)u^n \quad \text{para todo } n$$

tendremos

$$w^{n+1} = (\mathbf{I}-\frac{k}{2}\mathbf{A}_1)(\mathbf{I}+\frac{k}{2}\mathbf{A}_1)^{-1}(\mathbf{I}-\frac{k}{2}\mathbf{A}_2)(\mathbf{I}+\frac{k}{2}\mathbf{A}_2)^{-1}w^n$$

Para la norma $||.||$ tenemos aplicando el lema (3.5)

$$||w^{n+1}|| \leq ||w^n||$$

de donde finalmente

$$||u^{n+1}||_C^2 = (Cu^{n+1},u^{n+1}) = \big((I+\frac{k}{2}\mathbf{A}_2)u^{n+1},(I+\frac{k}{2}\mathbf{A}_2)u^{n+1}\big)$$
$$= (w^{n+1},w^{n+1}) = ||w^{n+1}||^2 \leq ||w^n||^2$$
$$= \big((I+\frac{k}{2}\mathbf{A}_2)u^n,(I+\frac{k}{2}\mathbf{A}_2)u^n\big) = (Cu^n,u^n) = ||u^n||_C^2$$

■

Ejercicios

1. Estudiar el método de factorización aproximada obtenido a partir del método de Euler implícito

$$\frac{u^{n+1}-u^n}{k}+\mathbf{A}u^{n+1}=f^{n+1}$$
$$u(0)=v$$

o bien

$$(\mathbf{I}+k\mathbf{A})u^{n+1}=u^n+kf^{n+1}$$

Supongamos que $\mathbf{A}=\mathbf{A}_1+\mathbf{A}_2$ y aproximemos $(\mathbf{I}+k\mathbf{A})$ de la forma

$$(\mathbf{I}+k\mathbf{A})=((\mathbf{I}+k(\mathbf{A}_1+\mathbf{A}_2))\approx(\mathbf{I}+k\mathbf{A}_1)(\mathbf{I}+k\mathbf{A}_2)=(\mathbf{I}+k\mathbf{A}+k^2\mathbf{A}_1\mathbf{A}_2)$$

El error cometido al sustituir $(\mathbf{I}+k\mathbf{A})$ por $(\mathbf{I}+k\mathbf{A}_1)(\mathbf{I}+k\mathbf{A}_2)$ es del mismo orden que el error de consistencia en el método de Euler. Estudiar el error de consistencia y la estabilidad del método

$$(\mathbf{I}+k\mathbf{A}_1)(\mathbf{I}+k\mathbf{A}_2)u^{n+1}=u^n+kf^{n+1}$$

2. Estudiar el siguiente método de estabilización del método de Euler explícito: el método

$$\frac{u^{n+1}-u^n}{k}+\mathbf{A}_1u^n+\mathbf{A}_2u^n=0$$
$$u(0)=v$$

lo sustituimos por

$$(\mathbf{I}+\frac{k}{2}\mathbf{A}_1)(\mathbf{I}+\frac{k}{2}\mathbf{A}_2)\frac{u^{n+1}-u^n}{k}+\mathbf{A}_1u^n+\mathbf{A}_2u^n=0$$
$$u(0)=v$$

Demostrar que el nuevo método tiene orden 2 y es incondicionalmente estable.

Capítulo 4
Problemas elípticos de segundo orden

Resumen
En este capítulo se estudia el método de Diferencias Finitas para resolver problemas elípticos de segundo orden que aparecen típicamente en problemas estacionarios de difusión. Realizamos el análisis de estabilidad en la norma de la energía y también utilizamos el principio del máximo, que permite obtener resultados de estabilidad en la norma del máximo. Finalmente vemos como el método de direcciones alternadas se puede considerar aquí lo que en definitiva en este contexto es un método iterativo para resolver el sistema de ecuaciones algebraico correspondiente.

4.1. Problema de contorno elíptico de segundo orden

El marco general de este capítulo es:

- Ω será un abierto, acotado y conexo de $\mathbb{R}^d$. Γ la frontera de Ω.
- $D_{ij} : (\bar{\Omega}) \to \mathbb{R} \quad i, j = 1, ..., d$ funciones de clase $C^1(\bar{\Omega})$ verificando la siguiente condición de elipticidad: Existe $\alpha > 0$ tal que

$$\forall \xi \in \mathbb{R}^d \quad \sum_{i,j=1}^{d} D_{ij}(x)\xi_i\xi_j \geq \alpha \sum \xi_i^2 \quad \forall x \in \Omega \tag{4.1}$$

- $f : \Omega \to \mathbb{R}$ una función continua.

El problema de contorno con condiciones de Dirichlet homogéneas es:
Hallar $u : \bar{\Omega} \to \mathbb{R}$ de clase $C^2(\bar{\Omega})$ verificando

$$-\sum_{i,j=1}^{d} \frac{\partial}{\partial x_i}(D_{ij}\frac{\partial u}{\partial x_j}) = f \quad \text{en } \Omega \tag{4.2}$$

$$u = 0 \quad \text{sobre } \Gamma \tag{4.3}$$

La existencia de solución de este problema está fuera del ámbito de este curso. En determinados casos se pueden calcular soluciones a este problema mediante el método de separación de variables, por ejemplo en el caso de un dominio Ω rectangular y $D_{ij} = \delta_{ij}\alpha$ (caso isótropo y coeficientes constantes). Sin embargo es fácil obtener la unicidad de solución del problema con técnicas elementales.

Teorema 4.1 *Si existe una solución u de clase $C^2(\bar{\Omega})$ del problema (4.2)-(4.3), esta es única.*

Demostración. Sean u_1 y u_2 dos soluciones de (4.2)-(4.3). Llamemos $w = u_1 - u_2$ a la diferencia entre las dos. La diferencia de soluciones w verifica

$$-\sum_{i,j=1}^{d} \frac{\partial}{\partial x_i}(D_{ij}\frac{\partial w}{\partial x_j}) = 0 \quad \text{en } \Omega$$
$$w = 0 \quad \text{sobre } \Gamma$$

Multiplicando por w e integrando en Ω con respecto a la medida $dx = dx_1, ..., dx_d$

$$-\sum_{i,j=1}^{d} \int_{\Omega} \left(\frac{\partial}{\partial x_i}(D_{ij}\frac{\partial w}{\partial x_j})\right) w\, dx = 0$$

Integrando por partes

$$\sum_{i,j=1}^{d} \int_{\Omega} D_{ij}\frac{\partial w}{\partial x_i}\frac{\partial w}{\partial x_j}\, dx = 0$$

utilizando la elipticidad (4.1)

$$\alpha \sum_{j=1}^{d} \int_{\Omega} \left(\frac{\partial w}{\partial x_j}\right)^2 dx \leq 0$$

lo que implica

$$\frac{\partial w}{\partial x_j} = 0 \quad \text{en } \Omega \quad j = 1, ..., d$$

por tanto w es una función constante y continua en todo $\bar{\Omega}$, como $w = 0$ sobre Γ implica que $w = 0$ en $\bar{\Omega}$ ■

4.2. Un método de diferencias finitas

Recubrimos Ω con una malla de tamaño característico h y consideramos los conjuntos de puntos, que llamamos nodos, siguientes:

$$\Omega_h = \{x = \beta h = (\beta_1 h, ..., \beta_d h),\ \beta \in \mathbb{Z}^d\ x \in \Omega\}$$
$$\Gamma_h = \{x = \beta h = (\beta_1 h, ..., \beta_d h),\ \beta \in \mathbb{Z}^d\ x \in \Gamma\}$$

Supondremos que Ω tiene las propiedades necesarias para que una descripción como la anterior sea posible, de manera que todos los vecinos de un punto de Ω_h sean o bien nodos de Ω_h o bien de Γ_h. Observemos que estas condiciones son realmente restrictivas. Dominios Ω admisibles en este sentido serían por ejemplo un dominio como el de la figura (3.1).

El método en diferencias se escribe:

$$\mathbf{A}_h u_h(x) = f(x) \quad x \in \Omega_h \tag{4.4}$$
$$u_h(x) = 0 \quad x \in \Gamma_h \tag{4.5}$$

donde

$$\mathbf{A}_h u_h(x) = -\sum_{i,j=1}^{d} \partial_{x_i}\left(D_{ij}\bar{\partial}_{x_j} u_h(x)\right)$$

Primero veamos que el problema (4.4)-(4.5) tiene solución única.

Teorema 4.2 *El problema (4.4)-(4.5) tiene solución única.*

Demostración. Sea M el cardinal del conjunto Ω_h. El problema (4.4)-(4.5) es un sistema de M ecuaciones con M incógnitas, cuya matriz asociada $\mathbf{A}_h$ es definida positiva, en efecto gracias a la propiedad (4.1) y su versión discreta (3.15)

$$(\mathbf{A}_h v, v) \geq \alpha \sum_{j=1}^{d} ||\bar{\partial}_{x_j} v||^2$$

■

Vamos a estudiar la convergencia del método. Para ello observemos que el método es consistente de orden 2 pues según se ha visto en la sección (2.2) y el teorema (3.2) del capítulo anterior tenemos,

$$\mathbf{A}_h u_h(x) = \mathbf{A}u(x) + \tau(x) \quad \forall x \in \Omega_h$$

con $\tau(x) = \mathscr{O}(h^2)$. Solo resta estudiar la estabilidad. Estudiaremos la estabilidad de la solución con respecto a la norma (3.22)

$$v \to \left(\sum_{j=1}^{d} ||\bar{\partial}_{x_i} v||^2\right)^{1/2}$$

Teorema 4.3 *El método en Diferencias Finitas (4.4)-(4.5) verifica*

$$\Big(\sum_{j=1}^{d} ||\bar{\partial}_{x_j} u_h||^2\Big)^{1/2} \leq \frac{C}{\alpha}||f||$$

Demostración. Multiplicando escalarmente por u_h en (4.4)

$$(\mathbf{A}_h u_h, u_h) = (f, u_h)$$
$$\alpha \sum_{j=1}^{d} ||\bar{\partial}_{x_j} u_h||^2 \leq ||f||.||u_h||$$

como gracias al lema (3.2)

$$||u_h|| \leq C\Big(\sum_{i=1}^{d} ||\bar{\partial}_{x_i} u_h||^2\Big)^{1/2}$$

con C una constante independiente de h resulta

$$\alpha \sum_{j=1}^{d} ||\bar{\partial}_{x_j} u_h||^2 \leq C||f||.\Big(\sum_{i=1}^{d} ||\bar{\partial}_{x_i} u_h||^2\Big)^{1/2}$$

simplificando y reordenando obtenemos el resultado buscado. ■

Observación: En el caso que la matriz $\mathbf{A}_h$ es simétrica la aplicación

$$v \to \Big(\sum_{i=1}^{d} ||\bar{\partial}_{x_i} u_h||^2\Big)^{1/2} \tag{4.6}$$

es una norma equivalente a la norma discreta de la energía, definida por

$$v \to (\mathbf{A}_h v, v)^{1/2}$$

Estamos en disposición de demostrar el siguiente teorema de convergencia:

Teorema 4.4 *Sea $e(x) = u(x) - u_h(x)$ la diferencia entre la solución u de (4.2)-(4.3) y la solución u_h de (4.4) y (4.5) en los puntos $x \in \Omega_h$. Tenemos la siguiente estimación del error*

$$\Big(\sum_{i=1}^{d} ||\bar{\partial}_{x_i} e||^2\Big)^{1/2} \leq \frac{C}{\alpha}||\tau|| = \mathcal{O}(h^2)$$

donde τ es el error de consistenacia.

Demostración. Tenemos para todo $x \in \Omega_h$

$$\begin{aligned}\mathbf{A}_h e(x) &= \mathbf{A}_h u(x) - A_h u_h(x) \\ &= Au(x) + \tau(x) - \mathbf{A}_h u_h(x) = f(x) + \tau(x) - f(x) = \tau(x)\end{aligned}$$

es decir

$$\mathbf{A}_h e(x) = \tau(x)$$

y aplicamos el resultado de estabilidad del teorema (4.3) a la anterior ecuación. ■

4.3. Análisis numérico basado en el principio del máximo

En esta sección utilizaremos el principio del máximo para el análisis numérico del método de diferencias finitas. Para simplificar los desarrollos nos limitaremos a la ecuación de Poisson en dimensión 2: Sea Ω un abierto acotado de $\mathbb{R}^2$ de frontera Γ y consideramos el problema

$$-\Delta u = f \quad \text{en } \Omega \tag{4.7}$$
$$u = 0 \quad \text{sobre } \Gamma \tag{4.8}$$

4.3.1. Principio del máximo

Teorema 4.5 *Sea $u \in C^2(\Omega)$ continua en $\bar{\Omega}$ y $f < 0$ en Ω. Entonces u alcanza el máximo V en algún punto de la frontera Γ de Ω.*

Demostración. Razonamos por reducción al absurdo: Suponamos que u alcanza el máximo en un punto interior de $\bar{\Omega}$, es decir, en un punto $x^* \in \Omega$. En este punto tendremos

$$\frac{\partial u}{\partial x}(x^*) = \frac{\partial u}{\partial y}(x^*) = 0$$
$$\frac{\partial^2}{\partial x^2}(x^*) \leq 0 \quad \frac{\partial^2}{\partial y^2}(x^*) \leq 0$$

de modo que en este punto si u es solución de (4.7)

$$-\Delta u(x^*) \geq 0$$

en contradicción con

$$-\Delta u = f < 0$$

■

Deseamos extender esta conclusión al caso $f \leq 0$. En particular se aplicará a la ecuación de Laplace $-\Delta = 0$.

Teorema 4.6 *Supongamos que $f \leq 0$ en Ω verificando la ecuación (4.7) en Ω. Entonces u alcanza el máximo V en algún punto de la frontera Γ de Ω.*

Demostración. Supongamos que $u \leq V$ en Γ demostraremos que $u \leq V$ en Ω. Consideremos la función auxiliar

$$v(x,y) = x^2 + y^2 \geq 0$$

que verifica

$$\Delta v(x,y) = \frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} = 2 + 2 = 4 > 0$$

Para todo $\varepsilon > 0$, la función

$$z(x,y) = u + \varepsilon v = u + \varepsilon(x^2 + y^2)$$

verifica

$$\begin{aligned} &\Delta z = \Delta u + \varepsilon \Delta v = \Delta u + 4\varepsilon > 0 \\ &-\Delta z < 0 \end{aligned}$$

y aplicando el teorema (4.5) z alcanza su máximo sobre Γ, es decir,

$$z(x,y) \leq \max_{(x,y)\in\Gamma} \{z(x,y\} \leq V + \varepsilon R^2\} \quad \forall (x,y) \in \Omega$$

siendo R el radio de un círculo que contiene a $\bar{\Omega}$. Puesto que

$$u = z - \varepsilon(x^2 + y^2) \leq z$$

resulta para todo $(x,y) \in \Omega$

$$u(x,y) \leq z(x,y) \leq V + \varepsilon R^2 \quad \forall \varepsilon > 0$$

pasando al límite cuando $\varepsilon \to 0$ resulta

$$u(x,y) \leq V \quad \forall (x,y) \in \Omega$$

■

Corolario 4.1 *Si u es solución de*

$$\Delta u = 0$$

entonces u alcanza su máximo y su mínimo sobre Γ.

Demostración. Podemos aplicar el principio del máximo a u y a la función $-u$. Encontramos entonces que si

$$V_{\text{mín}} \leq u \leq V_{\text{máx}} \quad \text{sobre } \Gamma$$

entonces

$$V_{\text{mín}} \leq u \leq V_{\text{máx}} \quad \text{en } \Omega$$

■

Corolario 4.2 *La solución del problema*

$$-\Delta u = f \quad \text{en } \Omega \tag{4.9}$$
$$u = u_0 \quad \text{sobre } \Gamma \tag{4.10}$$

es única.

Demostración. Sean u_1 y u_2 dos soluciones de (4.9)-(4.10). $w = u_1 - u_2$ verifica

$$-\Delta w = 0 \quad \text{en } \Omega$$
$$w = 0 \quad \text{sobre } \Gamma$$

lo que implica utilizando el corolario (4.1) $w = 0$ en todo Ω

El siguiente corolario permite establecer la continuidad de la solución con respecto a los datos del problema.

Corolario 4.3 *La solución u de (4.9)-(4.10) verifica*

$$||u||_{\infty,\Omega} \leq ||u_0||_{\infty,\Gamma} + \frac{R^2}{4}||f||_{\infty,\Omega} \tag{4.11}$$

donde para $v \in C(\bar{\Omega})$

$$||v||_{\infty,\Omega} = \max_{(x,y)\in\Omega} |v(x,y)| \tag{4.12}$$
$$||v||_{\infty,\Gamma} = \max_{(x,y)\in\Gamma} |v(x,y)| \tag{4.13}$$

y R es el radio del círculo mínimo que contiene a $\bar{\Omega}$

Demostración. Pongamos $V = ||f||_{\infty,\Omega} = \text{máx}_{(x,y)\in\Omega} |f(x,y)|$. La función

$$v = u + \frac{V}{4}(x^2 + y^2)$$

verifica $v \geq u$, y

$$-\Delta v = -\Delta u - V = f - V \leq 0$$

de donde

$$\begin{aligned} u \leq v &\leq \max_{(x,y)\in\Gamma} \left(u(x,y) + \frac{V}{4}(x^2+y^2)\right) \\ &\leq \max_{(x,y)\in\Gamma} u(x,y) + \frac{V}{4}R^2 \\ &\leq \max_{(x,y)\in\Gamma} u_0(x,y) + \frac{1}{4}R^2 \max_{(x,y)\in\Omega} |f(x,y)| \end{aligned}$$

Aplicando a $-u$ el mismo razonamiento

$$-u \leq \max_{(x,y)\in\Gamma}(-u_0(x,y)) + \frac{1}{4}R^2 \max_{(x,y)\in\Omega}|f(x,y)|$$

es decir,

$$||u||_{\infty,\Omega} \leq ||u_0||_{\infty,\Gamma} + \frac{R^2}{4}||f||_{\infty,\Omega}$$

■

4.3.2. Análisis numérico del Método de Diferencias Finitas utilizando el principio del máximo

Sea como anteriormente los conjunto Ω_h y Γ_h. Para los puntos $P = (x,y) \in \Omega_h$ consideremos el operador en diferencias

$$\Delta_h v(x,y) = \frac{v(x-h,y) - 2v(x,y) + v(x+h,y)}{h^2} + \frac{v(x,y-h) - 2v(x,y) + v(x,y+h)}{h^2}$$

Vamos a estudiar la versión discreta del principio del máximo visto en la subsección anterior.

Teorema 4.7 *Supongamos u definida en los puntos de $\Omega_h \cup \Gamma_h$ y f una función definida en Ω_h tal que $f(x,y) < 0$ para todo $(x,y) \in \Omega_h$. Si u verifica*

$$-\Delta_h u = f \quad \text{en } \Omega_h$$

entonces u alcanza el valor máximo en Γ_h

Demostración. Razonamos por reducción al absurdo. Supongamos que u alcanza su valor máximo V en algún punto $P_0 = (x_0, y_0)$ de Ω_h. Escribamos la ecuación en diferencias correspondiente al punto P_0. Con las notaciones de la figura (4.1) la ecuación en P_0 es

$$-\Delta_h u(P_0) = f(P_0)$$

y llamando

$$\begin{aligned} P_1 &= (x_0 + h, y_0) \\ P_2 &= (x_0 - h, y_0) \\ P_3 &= (x_0, y_0 + h) \\ P_4 &= (x_0, y_0 - h) \end{aligned}$$

tendremos

$$V = u(P_0) = \frac{1}{4}\big(u(P_1)+u(P_2)+u(P_3)+u(P_4)\big) + \frac{h^2}{4}f(P_0)$$
$$\leq \frac{1}{4}\big(V+V+V+V)\big) + \frac{h^2}{4}f(P_0) < V$$

y llegamos a una contradicción. Por tanto u alcanza su máximo en Γ_h ■

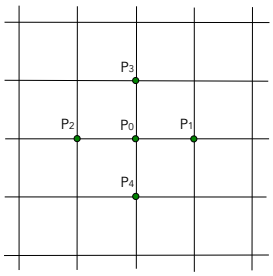


Figura 4.1 Molécula de 5 puntos

Extendemos el resultado anterior al caso en que $f \leq 0$.

Teorema 4.8 *Supongamos u definida en los puntos de $\Omega_h \cup \Gamma_h$ y f una función definida en Ω_h tal que $f(x,y) \leq 0$ para todo $(x,y) \in \Omega_h$. Si u verifica*

$$-\Delta_h u = f \quad \text{en } \Omega_h$$

entonces u alcanza el valor máximo en Γ_h

Demostración. Demostraremos que si $u \leq V$ en Γ_h, entonces $u \leq V$ en Ω_h. Para ello utilizamos, igual que en el caso continuo, la función auxiliar

$$v(x,y) = x^2 + y^2 \geq 0$$

definida en los puntos $(x,y) \in \Omega_h$. Tendremos

$$\Delta_h v(x,y) = \frac{(x-h)^2+y^2-2(x^2+y^2)+(x+h)^2+y^2}{h^2}$$
$$+ \frac{x^2+(y-h)^2-2(x^2+y^2)+x^2+(y+h)^2}{h^2} = 4$$

Consideremos ahora la función, definida para todo $\varepsilon > 0$

$$z(x,y) = u(x,y) + \varepsilon v(x,y) \geq u(x,y)$$

tenemos

$$-\Delta_h z(x,y) = -\Delta_h u(x,y) - \varepsilon \Delta_h v(x,y) = f(x,y) - 4\varepsilon < 0$$

Por el teorema anterior, sabemos que z alcanza el máximo en Γ_h de modo que para todo $(x,y) \in \Omega_h$

$$z(x,y) \leq \max_{(x,y)\in\Gamma_h} \left(u(x,y) + \varepsilon(x^2+y^2)\right) \leq V + \varepsilon R^2$$

siendo R el radio de un círculo que contenga a $\Omega_h \cup \Gamma_h$. Como para todo $(x,y) \in \Omega_h$

$$\begin{aligned} u(x,y) &= z(x,y) - \varepsilon(x^2+y^2) \leq z(x,y) \\ u(x,y) &\leq V + \varepsilon R^2 \quad \forall \varepsilon > 0 \end{aligned}$$

y pasando al límite cuando $\varepsilon \to 0$, $u(x,y) \leq V$ para todo $(x,y) \in \Omega_h$ ■

Análogamente al caso continuo tenemos las consecuencias siguientes:

Corolario 4.4 *Sea u solución de*

$$-\Delta_h u = 0$$

Si

$$V_{\text{mín}} \leq u \leq V_{\text{máx}} \quad \text{en } \Gamma_h$$

entonces

$$V_{\text{mín}} \leq u \leq V_{\text{máx}} \quad \text{en } \Omega_h$$

Demostración. Aplicamos el principio del máximo, teorema (4.8) a u y a $-u$. ■

El principio del máximo nos permite obtener fácilmente la unicidad de solución y al tratarse de un problema en dimensión finita también la existencia de solución del problema discreto.

Corolario 4.5 *Unicidad y existencia de solución del problema*

$$\begin{aligned} -\Delta_h u &= f \quad \text{en } \Omega_h \\ u &= u_0 \quad \text{en } \Gamma_h \end{aligned}$$

Demostración. Sean u_1 y u_2 dos soluciones y sea $w = u_1 - u_2$ su diferencia. w verifica

$$\begin{aligned} -\Delta_h w &= 0 \quad \text{en } \Omega_h \\ w &= 0 \quad \text{en } \Gamma_h \end{aligned}$$

Aplicando el corolario (4.4) anterior resulta $w = 0$ en Ω_h , de ahí la unicidad.

La existencia se deduce de la unicidad al tratarse de un sistema de M ecuaciones con M incógnitas donde M es el cardinal de Ω_h. ■

Del principio del máximo se deduce también la estabilidad, en efecto,

Corolario 4.6 *La solución de*

$$-\Delta_h u = f \quad \text{en } \Omega_h \tag{4.14}$$
$$u = u_0 \quad \text{en } \Gamma_h \tag{4.15}$$

verifica

$$||u||_{\infty,\Omega_h} \leq ||u_0||_{\infty,\Gamma_h} + \frac{1}{4}R^2||f||_{\infty,\Omega_h}$$

donde para v definida en $\Omega_h \cup \Gamma_h$,

$$||v||_{\infty,\Omega_h} = \max_{(x,y)\in\Omega_h} |v(x,y)|$$
$$||v||_{\infty,\Gamma_h} = \max_{(x,y)\in\Gamma_h} |v(x,y)|$$

y R es el radio del círculo mínimo que contenga $\Omega_h \cup \Gamma_h$

Demostración. La demostración es idéntica al la del corolario (4.3) sustituyendo Ω y Γ por Ω_h y Γ_h. ■

La convergencia es ahora una consecuencia de la consistencia (3.2) y de la estabilidad (corolario 4.6).

Teorema 4.9 *Sea u la solución de (4.9)-(4.10) y* u_h *la solución de*

$$-\Delta_h u_h = f \quad \text{en } \Omega_h \tag{4.16}$$
$$u_h = u_0 \quad \text{en } \Gamma_h \tag{4.17}$$

el error $e = u - u_h$ *definido en los puntos de* $\Omega_h \cup \Gamma_h$, *verifica*

$$||e||_{\infty,\Omega_h} \leq \frac{CR^2}{4}h^2$$

siendo C una constante independiente de h

Demostración. Según se ha visto en la sección (2.2) y el teorema (3.2) para una constante C independiente de h

$$\begin{aligned}
&-\Delta_h u = -\Delta u + Ch^2 \quad \text{en } \Omega_h \\
&-\Delta_h e = -\Delta_h u + \Delta_h u_h = -\Delta u + Ch^2 + \Delta_h u_h \\
&= f + Ch^2 - f = Ch^2 \quad \text{en } \Omega_h
\end{aligned}$$

de donde

$$-\Delta_h e = Ch^2 \quad \text{en } \Omega_h$$
$$e = 0 \quad \text{sobre } \Gamma_h$$

y aplicando el resultado de estabilidad del corolario (4.6) obtenemos el resultado buscado. ■

Ejercicio

En este ejercicio se estudia la velocidad de convergencia del método $S.O.R.$ cuando se utiliza para resolver el sistema de ecuaciones correspondiente al problema modelo siguiente: Sea $\Omega = [0,1] \times [0,1]$ y Γ su frontera.

$$\begin{aligned} -\Delta u &= f \quad \text{en} \Omega \\ u &= 0 \quad \text{sobre } \Gamma \end{aligned}$$

mediante el método de diferencias finitas centrales definido por el esquema de 5 puntos como en la figura (4.1) y una malla de $N \times N$ cuadrados, con $h = 1/N$.

La matriz $\mathbf{A}_h$ del sistema correspondiente de ecuaciones es:

$$\mathbf{A}_h = \frac{1}{h^2} \begin{bmatrix} \mathbf{G} & -\mathbf{I} & \dots & \dots & 0 \\ -\mathbf{I} & \mathbf{G} & -\mathbf{I} & \dots & 0 \\ & \ddots & \ddots & \ddots & \\ & & & & -\mathbf{I} \\ 0 & 0 & \dots & -\mathbf{I} & \mathbf{G} \end{bmatrix}$$

donde $\mathbf{G}$ es

$$\mathbf{G} = \begin{bmatrix} 4 & -1 & \dots & \dots & 0 \\ -1 & 4 & -1 & \dots & 0 \\ & \ddots & \ddots & \ddots & \\ & & & & -1 \\ 0 & 0 & \dots & -1 & 4 \end{bmatrix}$$

1. Demostrar que la matriz $\mathbf{A}_h$ de $(N-1)^2$ filas $\times$ $(N-1)^2$ columnas tiene los siguientes valores propios

$$\lambda_{ij} = \frac{4}{h^2}\left(\sin^2\left(\frac{i\pi h}{2}\right) + \sin^2\left(\frac{j\pi h}{2}\right)\right) \quad 1 \le i, j \le N-1$$

no siendo todos diferentes.

Indicación: λ_{ij} es el valor propio correspondiente al vector propio

$$(e^{ij})_{\mu\nu} = 2h \sin(ih\nu\pi)\sin(jh\mu\pi) \quad 1 \le i, j, \nu, \mu \le N-1 \tag{4.18}$$

2. Observar que para $i = j = 1$ se obtiene el valor propio mínimo y para $i = j = N-1$ el correspondiente valor propio máximo, y que son

$$\lambda_{\text{mín}} = \frac{8}{h^2}\sin^2(\frac{\pi h}{2})$$
$$\lambda_{\text{máx}} = \frac{8}{h^2}\cos^2(\frac{\pi h}{2})$$

En particular deducir que $\mathbf{A}_h$ es definida positiva.

3. Demostrar que los vectores $\{e^{ij}\}$ dados por (4.18) son ortonormales para el producto escalar

$$(e^{ij}, e^{kl}) = \sum_{\mu,\nu} e^{ij}_{\nu\mu} e^{kl}_{\nu\mu}$$

Indicación: Observar que $e^{ij}_{\nu\mu} = e^i_\nu e^j_\mu$ donde

$$(e^k)_\nu = \sqrt{2h}\sin(kh\nu\pi)$$

son los vectores propios del correspondiente caso unidimensional.

4. Comprobar que los vectores propios de $\mathbf{A}_h$ son vectores propios de

$$\mathbf{D} = \frac{1}{h^2}\begin{bmatrix} \mathbf{G} & 0 \dots & \dots & 0 \\ 0 & \mathbf{G} & 0 \dots & 0 \\ & \ddots & \ddots & \ddots \\ 0 \dots & 0 & \mathbf{G} & 0 \\ 0 \dots & \dots & 0 & \mathbf{G} \end{bmatrix}$$

y calcular los correspondientes valores propios.
Solución:

$$\frac{2}{h^2}(2-\cos ih\pi)$$

5. Hallar los valores propios de la matriz de iteración de Jacobi

$$\mathbf{J} = \mathbf{I} - \mathbf{D}^{-1}\mathbf{A}_h$$

y hallar su radio espectral $\rho(\mathbf{J})$
Solución:

$$\mu_{ij} = \frac{\cos(j\pi h)}{2-\cos(i\pi h)}$$

$$\rho(\mathbf{J}) = \frac{\cos\pi h}{2-\cos h\pi} = 1 - \pi^2 h^2 + \mathcal{O}(h^4)$$

6. Recordando que el radio espectral de la matriz $\mathcal{L}^*_\omega(\mathbf{G})$ del método de relajación por bloques con parámetro óptimo ω^* viene dado por

$$\rho(\mathcal{L}^*_\omega(\mathbf{G})) = \frac{1-\sqrt{1-\rho^2(\mathbf{J})}}{1+\sqrt{1-\rho^2(\mathbf{J})}}$$

demostrar que

$$\rho(\mathscr{L}_\omega^*(\mathbf{G})) \approx 1 - 2\pi h\sqrt{2} + \mathscr{O}(h^2)$$

4.4. Método de direcciones alternadas para resolver problemas elípticos

El método de direcciones alternadas descrito en la subsección (3.3.3) se puede utilizar para resolver problemas del tipo (4.7)-(4.8) para los cuales la matriz asociada al esquema en diferencias finitas es de la forma $\mathbf{A}_h = \mathbf{A}_1 + \mathbf{A}_2$. En efecto, supongamos que queremos resolver un sistema lineal de ecuaciones

$$\mathbf{A}u = f \tag{4.19}$$

y que $\mathbf{A}$ admite una descomposición de la forma $\mathbf{A} = \mathbf{A}_1 + \mathbf{A}_2$. Sea $p > 0$ un parámetro que elegiremos adecuadamente. Podemos escribir el sistema de ecuaciones (4.19) de cualquiera de las dos formas

$$(\mathbf{A}_1 + p\mathbf{I})u = f + (p\mathbf{I} - \mathbf{A}_2)u$$
$$(\mathbf{A}_2 + p\mathbf{I})u = f + (p\mathbf{I} - \mathbf{A}_1)u$$

Esta descomposición de las ecuaciones sugiere el siguiente método iterativo para resolver el sistema de ecuaciones

$$(\mathbf{A}_1 + p\mathbf{I})u^{n+1/2} = f + (p\mathbf{I} - \mathbf{A}_2)u^n \tag{4.20}$$
$$(\mathbf{A}_2 + p\mathbf{I})u^{n+1} = f + (p\mathbf{I} - \mathbf{A}_1)u^{n+1/2} \tag{4.21}$$

que son las mismas la ecuaciones (3.35) y (3.36) para $p = 2/k$. Aquí n significa el número de la iteración. En definitiva buscamos la solución de (3.33)-(3.34) para $\frac{du}{dt} = 0$ haciendo $t \to \infty$. Si t representa la variable tiempo, estamos buscando la solución estacionaria de (3.33)-(3.34). Estudiemos en primer lugar un resultado general de convergencia. Aquí nos referimos a la convergencia del método iterativo para resolver el sistema algebraico de ecuaciones y no a la convergencia del método de diferencias finitas con respecto a la solución del problema continuo.

Teorema 4.10 *Sea $\mathbf{A} = \mathbf{A}_1 + \mathbf{A}_2$ dos matrices simétricas y definidas positivas, entonces el método de direcciones alternadas definido por (4.20)-(4.21) converge para todo valor de $p > 0$. El método es también convergente aunque una de las matrices sea solo semidefinida positiva.*

Demostración. Llamemos $e^n = u^n - u$ el error en la iteración n. Restando $\mathbf{A}u = f$ de la dos ecuaciones (4.20)-(4.21) obtenemos

$$(\mathbf{A}_1 + p\mathbf{I})e^{n+1/2} = (p\mathbf{I} - \mathbf{A}_2)e^n$$
$$(\mathbf{A}_2 + p\mathbf{I})e^{n+1} = (p\mathbf{I} - \mathbf{A}_1)e^{n+1/2}$$

eliminando de las ecuaciones $e^{n+1/2}$ se obtiene

$$e^{n+1} = \mathbf{G}(p)e^n$$

donde

$$\mathbf{G}(p) = (\mathbf{A_2} + \mathbf{pI})^{-1}(p\mathbf{I} - \mathbf{A}_1)(\mathbf{A_1} + \mathbf{pI})^{-1}(p\mathbf{I} - \mathbf{A}_2)$$

Demostraremos que $\rho(\mathbf{G}(p)) < 1$ lo que implica la convergencia. $\mathbf{G}(p)$ es semejante a

$$\begin{aligned}\mathbf{G}'(p) &= (\mathbf{A_2} + \mathbf{pI})\mathbf{G}(p)(\mathbf{A_2} + \mathbf{pI})^{-1} \\ &= (p\mathbf{I} - \mathbf{A}_1)(\mathbf{A_1} + \mathbf{pI})^{-1}(p\mathbf{I} - \mathbf{A}_2)(\mathbf{A_2} + \mathbf{pI})^{-1}\end{aligned}$$

Tenemos pues que los radios espectrales de $\rho(\mathbf{G}(p))$ y de $\rho(\mathbf{G}'(p))$ son iguales. y para la norma euclídea tendremos,

$$\rho(\mathbf{G}'(p)) \leq ||\mathbf{G}'(p))|| \leq ||(p\mathbf{I} - \mathbf{A}_1)(\mathbf{A_1} + \mathbf{pI})^{-1}||.||(p\mathbf{I} - \mathbf{A}_2)(\mathbf{A_2} + \mathbf{pI})^{-1}||$$

Admitamos provisionalmente el lema (4.1), que dice que si $\mathbf{A}_1$ es simétrica, $(p\mathbf{I} - \mathbf{A}_1)(\mathbf{A_1} + \mathbf{pI})^{-1}$ también es simétrica. Entonces,

$$||(p\mathbf{I} - \mathbf{A}_1)(\mathbf{A_1} + \mathbf{pI})^{-1}|| = \rho((p\mathbf{I} - \mathbf{A}_1)(\mathbf{A_1} + \mathbf{pI})^{-1}) = \max_i \left| \frac{p - \lambda_i(\mathbf{A}_1)}{p + \lambda_i(\mathbf{A}_1)} \right|$$

Para $p > 0$, como todos los valores propios $\lambda_i(\mathbf{A}_1)$ son estrictamente positivos tendremos

$$\max_i \left| \frac{p - \lambda_i(\mathbf{A}_1)}{p + \lambda_i(\mathbf{A}_1)} \right| < 1$$

Análogamente para la matriz $\mathbf{A}_2$

$$\max_i \left| \frac{p - \lambda_i(\mathbf{A}_2)}{p + \lambda_i(\mathbf{A}_2)} \right| < 1$$

y finalmente $\rho(\mathbf{G}'(p)) < 1$. Si una de las dos matrices $\mathbf{A}_1$ o $\mathbf{A}_2$ es únicamente semidefinida positiva, por ejemplo si la matriz $\mathbf{A}_2$ es solo semidefinida positiva

$$\max_i \left| \frac{p - \lambda_i(\mathbf{A}_2)}{p + \lambda_i(\mathbf{A}_2)} \right| = 1$$

pero si $\mathbf{A}_1$ es definida positiva seguiremos teniendo $\rho(\mathbf{G}'(p)) < 1$ ■

Demostremos ahora que $(p\mathbf{I} - \mathbf{A})(\mathbf{A} + p\mathbf{I})^{-1}$ es simétrica.

Lema 4.1 *Sea* $\mathbf{A}$ *una matriz simétrica y definida positiva de modo que* $\mathbf{A}^{-1}$ *y* $(\mathbf{A} + p\mathbf{I})^{-1}$ *para* $p > 0$ *existen, entonces* $(p\mathbf{I} - \mathbf{A})(\mathbf{A} + p\mathbf{I})^{-1}$ *es simétrica.*

Demostración. Primeramente veamos que $\mathbf{A}$ y $(\mathbf{A} + \mathbf{I})^{-1}$ conmutan. En efecto,

$$\mathbf{A}(\mathbf{A}+\mathbf{I})^{-1} = (\mathbf{A}^{-1})^{-1}(\mathbf{A}+\mathbf{I})^{-1} = \left((\mathbf{A}+\mathbf{I})\mathbf{A}^{-1}\right)^{-1} = (\mathbf{I}+\mathbf{A}^{-1})^{-1}$$
$$(\mathbf{A}+\mathbf{I})^{-1}\mathbf{A} = (\mathbf{I}+\mathbf{A})^{-1}(\mathbf{A}^{-1})^{-1} = \left(\mathbf{A}^{-1}((\mathbf{A}+\mathbf{I})^{-1}\right)^{-1} = (\mathbf{I}+\mathbf{A}^{-1})^{-1}$$

Del mismo modo demostramos que $\mathbf{A}$ y $(\mathbf{A}+p\mathbf{I})^{-1}$ conmutan. Finalmente veamos que $(p\mathbf{I}-\mathbf{A})(\mathbf{A}+\mathbf{pI})^{-1}$ es simétrica. Para ello veamos que coincide con su transpuesta.

$$\begin{aligned}\left((p\mathbf{I}-\mathbf{A})(\mathbf{A}+p\mathbf{I})^{-1}\right)^t &= (\mathbf{A}+p\mathbf{I})^{-t}(p\mathbf{I}-\mathbf{A})^t = p(\mathbf{A}+p\mathbf{I})^{-t} - (\mathbf{A}+p\mathbf{I})^{-1}\mathbf{A} \\ &= p(\mathbf{A}+p\mathbf{I})^{-1} - \mathbf{A}(\mathbf{A}+p\mathbf{I})^{-1} = (p\mathbf{I}-\mathbf{A})(\mathbf{A}+p\mathbf{I})^{-1}\end{aligned}$$

puesto que si $\mathbf{A}$ es simétrica y tambíen lo es $(\mathbf{A}+p\mathbf{I})^{-1}$. ■

Ejercicio

En este ejercicio se estudia la velocidad de convergencia del método de direcciones alternadas cuando se utiliza para resolver el sistema de ecuaciones correspondiente al problema siguiente: Sea $\Omega = [0,1]\times[0,1]$ y Γ su frontera.

$$\begin{aligned}-\Delta u &= f \quad \text{en } \Omega \\ u &= 0 \quad \text{sobre } \Gamma\end{aligned}$$

lo que permite compararlo con la velocidad de convergencia del método $S.O.R.$ con parámetro óptimo. Considerar una malla de $N\times N$ cuadrados, con $h = 1/N$ y sean $\mathbf{A}_1$ y $\mathbf{A}_2$ definidas para cada $(x,y)\in\Omega_h\cup\Gamma_h$ y un vector $v(x,y)_{(x,y)\in\Omega_h\cup\Gamma_h}$ por

$$\begin{aligned}\mathbf{A}_1 v(x,y) &= -\frac{v(x-h,y)-2v(x,y)+v(x+h,y)}{h^2} \quad \text{para } y \text{ fijo} \\ \mathbf{A}_2 v(x,y) &= -\frac{v(x,y-h)-2v(x,y)+v(x,y+h)}{h^2} \quad \text{para } x \text{ fijo}\end{aligned}$$

1. Obtener los valores propios de $\mathbf{A}_1$ y $\mathbf{A}_2$.
   Solución: Los valores propios de $\mathbf{A}_1$ que son iguales a los de $\mathbf{A}_2$ son
   $$\lambda_i = \frac{4}{h^2}\sin^2\left(\frac{ih\pi}{2}\right)$$
2. Calcular el valor propio mínimo $\lambda_{\text{mín}}$ y el valor propio máximo $\lambda_{\text{máx}}$ de $\mathbf{A}_1$ y $\mathbf{A}_2$ matrices asociadas al método de direcciones alternadas.
   Solución:
   $$\begin{aligned}\lambda_{\text{mín}} &= \lambda_1 = \frac{4}{h^2}\sin^2\frac{\pi h}{2} \\ \lambda_{\text{máx}} &= \lambda_{N-1} = \frac{4}{h^2}\cos^2\frac{\pi h}{2}\end{aligned}$$

3. Deducir que el radio espectral de la matriz $\mathbf{G}$ asociada al método de direcciones alternadas es

$$\mathbf{G} = \max_i \left| \frac{p-\lambda_i}{p+\lambda_i} \right|$$

donde λ_i son los valores propios de $\mathbf{A}_1$ que son iguales a los de $\mathbf{A}_2$.
Indicación: En este caso las matrices $\mathbf{A}_1$ y $\mathbf{A}_2$ conmutan y tienen los mismos vectores propios.
4. Hallar el valor óptimo de $p = p_{ópt}$.
Solución: El valor óptimo de p es aquel que minimiza la función

$$\psi(p) = \max\{|\frac{\lambda_{\min}-p}{\lambda_{\min}+p}|, |\frac{\lambda_{\max}-p}{\lambda_{\max}+p}|\}$$

Considerar las gráficas de las funciones

$$p \to \psi(p) = |\frac{\lambda_{\min}-p}{\lambda_{\min}-p}|$$

y

$$p \to \psi(p) = |\frac{\lambda_{\max}-p}{\lambda_{\max}-p}|$$

Obtener el mínimo buscando el punto de interseción de las dos gráficas. El resultado es $p_{ópt} = \sqrt{\lambda_{\min}\lambda_{\max}}$.
5. Deducir que el radio espectral de la matriz $\mathbf{G}$ asociada al método de direcciones alternadas es para $p = p_{ópt}$ es

$$\rho(\mathbf{G}(p_{ópt})) = 1 - 2\pi h + \mathcal{O}(h^2)$$

Capítulo 5
Ecuaciones hiperbólicas

Resumen
En este capítulo se estudia el Método de Diferencias Finitas para resolver problemas hiperbólicos. Se estudiarán brevemente aspectos generales de las ecuaciones hiperbólicas de primer orden lineales. En una subsección estudiamos la resolución numérica mediante el Método de Diferencias Finitas de la ecuación de ondas, que es un ejemplo de problema hiperbólico de segundo orden. Nos limitamos también a analizar un método explícito. La parte principal de este capítulo se dedica a los problemas hiperbólicos de primer orden no lineales. En particular se requiere introducir la noción de solución débil y algunos resultados de existencia y unicidad de estas soluciones. Notablemente el problema de valor inicial asociado a una ecuación hiperbólica no lineal no tiene solución única requiriendo condiciones adicionales para asegurar la unicidad, como es la condición de entropía. Entre todas las soluciones matemáticamente posibles la solución entrópica será la físicamente aceptable. Los métodos numéricos tendrán que adaptarse a esta situación y asegurarse que las soluciones numéricas obtenidas convergen a la solución entrópica.

5.1. Ecuaciones hiperbólicas lineales de primer orden

Nos limitaremos a ecuaciones en dimensión 1 espacial (variable x).

Problema con coeficiente constante

El problema de Cauchy o de valor inicial asociado a una ecuación hiperbólica de primer orden lineal con coeficiente constante es : Sea $a \in \mathbb{R}$, $u_0 : \mathbb{R} \to \mathbb{R}$ una función dada de variable real a valores reales.

$$\frac{\partial u}{\partial t} + a\frac{\partial u}{\partial x} = 0 \quad x \in \mathbb{R},\, t > 0 \tag{5.1}$$

$$u(x,0) = u_0(x) \quad x \in \mathbb{R} \tag{5.2}$$

Vamos a buscar soluciones a este problema. Probamos soluciones de la forma

$$u(x,t) = v(x-at)$$

tendremos

$$\frac{\partial u}{\partial t}(x,t) = v'(x-at).(-a)$$
$$\frac{\partial u}{\partial x}(x,t) = v'(x-at)$$

Si

$$v : \mathbb{R} \to \mathbb{R}$$
$$\xi \to v(\xi)$$

es una función derivable, entonces cualquier función de la forma $(x,t) \to v(x-at)$ es solución de la ecuación diferencial. En efecto,

$$v'(x-at)(-a) + av'(x-at) = 0$$

Impongamos ahora la condición inicial: Para $t = 0$, $v(x) = u(x,0) = u_0(x) \quad \forall x \in \mathbb{R}$. De modo que tenemos $v = u_0$. Es decir finalmente

$$u(x,t) = u_0(x-at)$$

es una solución del problema de Cauchy (5.1)-(5.2), de hecho es la única solución.

Vamos a interpretar gráficamente esta solución. Introducimos la noción de curvas características (recta en este caso) como las soluciones de la ecuación diferencial

$$x : t \to x(t)$$
$$\frac{dx}{dt} = a$$

Integrando, en este caso como a es una constante,

$$x = at + Constante$$

Las curvas características son pues en este caso rectas en el plano $x-t$ de pendiente $1/a$. Observemos que sobre las características la solución $u(x,t)$ de la ecuación diferencial es constante, en efecto, si $t \to x(t)$ es la característica la función $t \to v(t) = u(x(t),t)$ verifica

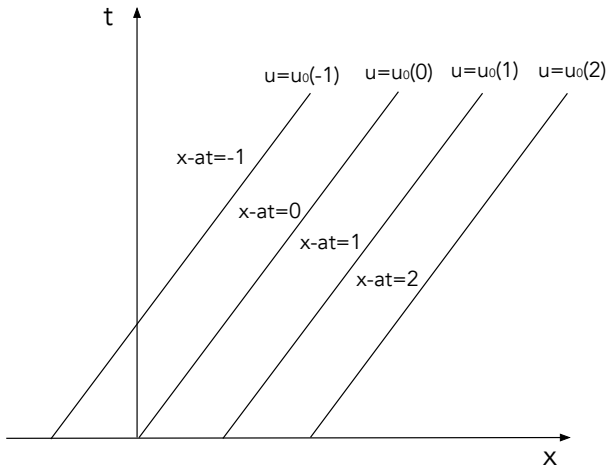


Figura 5.1 Rectas Características

$$\frac{dv}{dt} = \frac{\partial u}{\partial t} + \frac{\partial u}{\partial x}\frac{dx}{dt} = \frac{\partial u}{\partial t} + a\frac{\partial u}{\partial x} = 0$$

por tanto $v(t) = Constante$. Esto es algo que podemos también observar directamente a partir de la solución $u(x,t) = u_0(x-at)$.

Problema con coeficiente variable

Consideremos ahora una ecuación con coeficiente variable. Sea $a : x \to a(x) \in \mathbb{R}$ una función suficientemente regular. El problema de Cauchy se escribe:

$$\frac{\partial u}{\partial t} + a(x)\frac{\partial u}{\partial x} = 0 \quad x \in \mathbb{R},\, t > 0 \tag{5.3}$$

$$u(x,0) = u_0(x) \quad x \in \mathbb{R} \tag{5.4}$$

Introduciendo las curvas características,

$$x : t \to x(t)$$

$$\frac{dx}{dt} = a(x)$$

Sobre estas curvas características la solución es constante, en efecto,

Sobre $t \to v(t) = u(x(t),t)$

$$\frac{dv}{dt} = \frac{\partial u}{\partial t} + \frac{\partial u}{\partial x}\frac{dx}{dt} = \frac{\partial u}{\partial t} + a(x)\frac{\partial u}{\partial x} = 0$$

La solución se puede calcular siguiendo las curvas características.

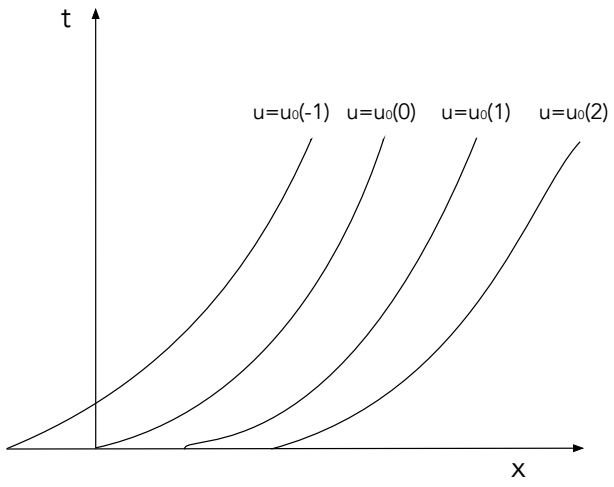


Figura 5.2 Curvas Características

Problema con coeficiente variable en forma conservativa

De manera más general podemos considerar la ecuación con coeficiente variable en forma conservativa, es decir, el coeficiente está bajo el signo de derivación. El problema de Cauchy se escribe:

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x}\big(a(x)u\big) = 0 \quad x \in \mathbb{R},\, t > 0 \tag{5.5}$$

$$u(x,0) = u_0(x) \quad x \in \mathbb{R} \tag{5.6}$$

Derivando la ecuación se puede escribir de la forma ,

$$\frac{\partial u}{\partial x} + a(x)\frac{\partial u}{\partial x} = -a'(x)u$$

Si introducimos la noción de curva característica como las soluciones $t \to v(t) = u(x(t),t)$ de

$$\begin{aligned} \frac{dx}{dt} &= a(x(t)) \\ x(0) &= x_0 \end{aligned}$$

resulta que $t \to v(t) = u(x(t),t)$ verifica

$$\begin{aligned} \frac{dv}{dt} &= \frac{\partial u}{\partial t} + \frac{\partial u}{\partial x}\frac{dx}{dt} \\ &= \frac{\partial u}{\partial t} + a(x(t))\frac{\partial u}{\partial x} \\ &= -a'(x(t)u(x(t),t) \end{aligned}$$

En este caso la solución no es constante sobre las características pero puede determinarse resolviendo el sistema de Ecuaciones Diferenciales Ordinarias siguiente:

$$\begin{aligned} \frac{dx}{dt} &= a(x(t)) \\ x(0) &= x_0 \\ \frac{dv}{dt} &= -a'(x(t))v(t) \\ v(0) &= u_0(x_0) \end{aligned}$$

En los tres casos anteriores se observa que la solución $u(x,t)$ en cualquier punto $(\bar{x},\bar{t})$ depende del valor de la solución inicial u_0 en un único punto $\bar{x}_0$ que será el punto de arranque de la característica que pasa por $(\bar{x},\bar{t})$. Si cambiamos el dato inicial u_0, en otros puntos distintos al punto x_0 pero dejamos fijo $u_0(\bar{x_0})$, la solución en $(\bar{x},\bar{t})$ seguirá siendo la misma. El conjunto

$$D(\bar{x},\bar{t}) = \{\bar{x}_0\}$$

se llama dominio de dependencia del punto $(\bar{x},\bar{t})$.

Discontinuidad de los datos, ley de conservación y solución generalizada

En los cálculos anteriores asumimos la diferenciabilidad de la solución $u(x,t)$. Sin embargo de la observación realizada según la cual la solución a lo largo de las curvas características depende solo del único valor $u_0(x_0)$, resulta claro que la regularidad de u con respecto a la variable x no es imprescindible para la construcción de $u(x,t)$ a partir de $u_0(x)$. Podemos definir una "solución" de la Ecuación en De-

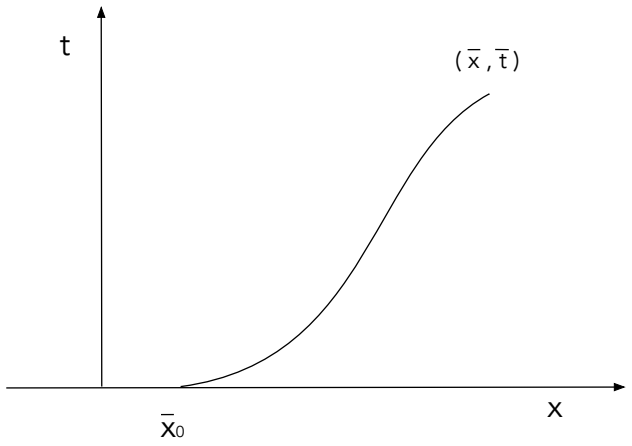


Figura 5.3 Construcción de la solución sobre la característica

rivadas Parciales aunque $u_0(x)$ no sea una función regular. En efecto, si $u_0(x)$ tiene una singularidad en algún punto x_0, por ejemplo una discontinuidad en x_0 o una discontinuidad de su derivada, resulta que $u(x,t)$ tendrá una singularidad del mismo orden a lo largo de la característica que arranca de x_0, pero permanecerá regular a lo largo de las características que arranquen de zonas donde u_0 es regular. Esta es una propiedad fundamental de las ecuaciones hiperbólicas lineales: Las singularidades se propagan a lo largo de las características

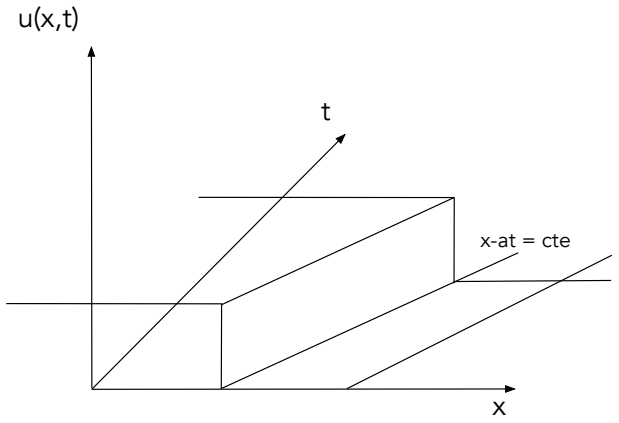


Figura 5.4 Solución Generalizada

Si u_0 no es diferenciable en algún punto, entonces $u(x,t)$ no es una solución de la Ecuación en Derivadas Parciales, en el sentido clásico. Sin embargo, esta función satisface la forma integral de la ley de conservación de la que proviene y que tiene sentido con datos no regulares. Recordemos que la forma integral tiene mayor sentido físico que la propia ecuación diferencial, la cual ha sido deducida a partir de la correspondiente ley de conservación. Tiene pues sentido aceptar esta solución de la ley de conservación como solución generalizada de la Ecuación en Derivadas Parciales. A partir de la Ecuación en Derivadas Parciales podemos recuperar la ley de conservación. En efecto, integrando entre x_1 y x_2 en la ecuación

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x}(a(x)u) = 0 \quad \text{para } x \in \mathbb{R},\ t > 0$$

obtenemos

$$\begin{aligned}
&\int_{x_1}^{x_2} \frac{\partial u}{\partial t}\,dx + \int_{x_1}^{x_2} \frac{\partial}{\partial x}(a(x)u)\,dx = 0\\
&\frac{d}{dt}\int_{x_1}^{x_2} u(x,t)\,dx + a(x_2)u(x_2,t) - a(x_1)u(x_1,t) = 0\\
&\frac{d}{dt}\int_{x_1}^{x_2} u(x,t)\,dx = a(x_1)u(x_1,t) - a(x_2)u(x_2,t)
\end{aligned}$$

También integrando entre t_1 y t_2

$$\int_{x_1}^{x_2} u(x,t_2)\,dx = \int_{x_1}^{x_2} u(x,t_1)\,dx + \int_{t_1}^{t_2} a(x_1)u(x_1,t)\,dt - \int_{t_1}^{t_2} a(x_2)u(x_2,t)\,dt \quad (5.7)$$

que se interpreta como la conservación de una magnitud física de la que $u(x,t)$ es la densidad lineal (cantidad por unidad de longitud) y $a(x)$ es la velocidad de la materia en cuyo seno se halla la magnitud física considerada.

Ejercicio

Sea $a(x) = a = constante$ independiente de x y sea $u_0(x)$ una función integrable. Verificar que la función $u(x,t) = u_0(x-at)$ satisface la forma integral de la ley de conservación (5.7).

Solución

Si llamamos U_0 a una primitiva cualquiera de u_0, tenemos poniendo $y = x - at_2$

$$\begin{aligned}
\int_{x_1}^{x_2} u_0(x-at_2)\,dx &= \int_{y_1=x_1-at_2}^{y_2=x_2-at_2} u_0(y)\,dy\\
&= U_0(x_2-at_2) - U_0(x_1-at_2)
\end{aligned}$$

y también poniendo $y = x - at_1$

$$\int_{x_1}^{x_2} u_0(x-at_1)\,dx = \int_{y_1=x_1-at_1}^{y_2=x_2-at_1} u_0(y)\,dy$$
$$= U_0(x_2-at_1) - U_0(x_1-at_1)$$

y por otra parte poniendo $\tau = x_1 - at$, $d\tau = -a\,dt$

$$\int_{t_1}^{t_2} au_0(x_1-at)\,dt = a\int_{t_1}^{t_2} u_0(x_1-at)\,dt = -\int_{\tau_1=x_1-at_1}^{\tau_2=x_1-at_2} u_0(\tau)\,d\tau$$
$$= -U_0(x_1-at_2) + U_0(x_1-at_1)$$

poniendo $\tau = x_2 - at$, $d\tau = -a\,dt$

$$-\int_{t_1}^{t_2} au_0(x_2-at)\,dt = -a\int_{t_1}^{t_2} u_0(x_2-at)\,dt = \int_{\tau_1=x_2-at_1}^{\tau_2=x_2-at_2} u_0(\tau)\,d\tau$$
$$= U_0(x_2-at_2) - U_0(x_2-at_1)$$

y por tanto verificamos

$$U_0(x_2-at_2) - U_0(x_1-at_2)$$
$$= U_0(x_2-at_1) - U_0(x_1-at_1) - U_0(x_1-at_2) + U_0(x_1-at_1) + U_0(x_2-at_2) - U_0(x_2-at_1)$$

5.2. Métodos numéricos para problemas hiperbólicos lineales

Introducimos un mallado en el plano $x-t$ eligiendo un paso h según la coordenada x que representará habitualmente el espacio y un paso k según la coordenada t que normalmente representará el tiempo. Con las notaciones siguientes

$$x_j = jh \quad j = ...,-1,0,1,2,...$$
$$t_n = nk \quad n = 0,1,2,...$$

También resultará útil definir

$$x_{j+1/2} = x_j + \frac{h}{2} = (j+\frac{1}{2})h$$

El método de diferencias finitas proporcionará aproximaciones u_j^n de la solución $u(x_j,t_n)$ en los nodos de la malla. En el desarrollo y estudio de los diferentes métodos para leyes de conservación es preferible interpretar u_j^n como una aproximación del valor medio de $u(x_j,t_n)$ en el intervalo $[x_{j-1/2},x_{j+1/2}]$, es decir,

$$u_j^n \approx \frac{1}{h}\int_{x_{j-1/2}}^{x_{j+1/2}} u(x,t_n)\,dx$$

Esta interpretación es natural pues si escribimos la ley de conservación en forma

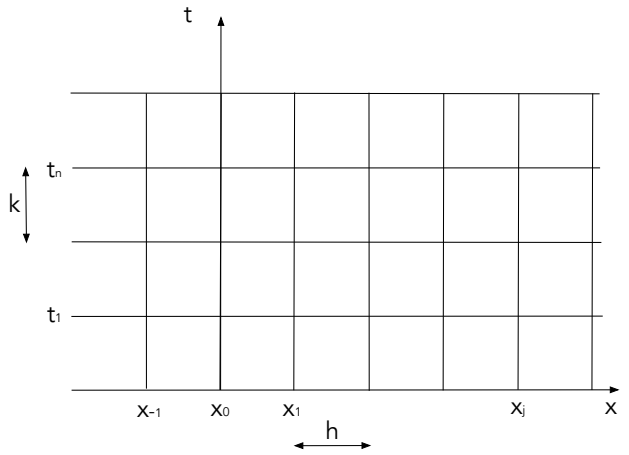


Figura 5.5 Malla de diferencias finitas

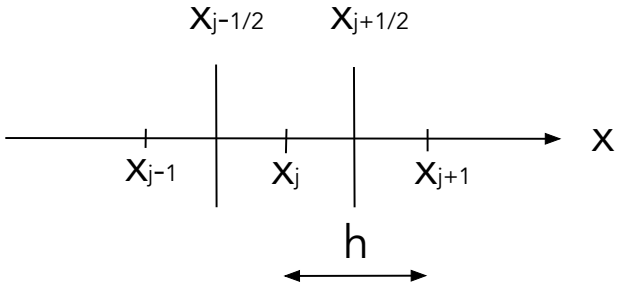


Figura 5.6 Detalle de la malla en x

integral este término es el que aparece. En efecto, integrando entre $[x_{j-1/2}, x_{j+1/2}]$

$$\frac{\partial u}{\partial t} + a\frac{\partial u}{\partial x} = 0$$

resulta

$$\frac{d}{dt}\int_{x_{j-1/2}}^{x_{j+1/2}} u(x,t)\,dx + a[u(x_{j+1/2},t) - u(x_{j-1/2},t)] = 0$$

dividiendo por h

$$\frac{d}{dt}\left(\frac{1}{h}\int_{x_{j-1/2}}^{x_{j+1/2}} u(x,t)\,dx\right) + a\frac{u(x_{j+1/2},t) - u(x_{j-1/2},t)}{h} = 0$$

Como valores iniciales, a partir de $u_0(x)$ definiremos u_j^0 tomando valores puntuales $u_0(x_j)$ o mejor valores medios

$$u_j^0 = \frac{1}{h}\int_{x_{j-1/2}}^{x_{j+1/2}} u_0(x)\,dx$$

Es también conveniente considerar la aproximación definida en todos los puntos del plano $x-t$,

$$u_{h,k}(x,t) = u_j^n \quad \forall (x,t) \in [x_{j-1/2}, x_{j+1/2}] \times [t_n, t_{n+1}]$$

En la práctica tendremos que trabajar en un intervalo acotado $[c,d] \subset \mathbb{R}$ y necesitaremos condiciones de contorno apropiadas en c y/o d. Un caso sencillo se presenta cuando imponemos condiciones de contorno periódicas,

$$u(c,t) = u(d,t) \quad t > 0$$

Este caso es equivalente a un problema de Cauchy con condición inicial periódica. Como la solución permanece periódica solo necesitaremos calcular lo que ocurre en un periodo.

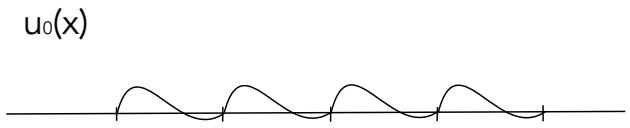


Figura 5.7 Valor Inicial Periódico

5.2.1. Métodos Numéricos para Problemas hiperbólicos lineales de primer orden

Consideraremos primero el problema de Cauchy en dimensión uno espacial (5.1)-(5.2)

$$\begin{aligned} &\frac{\partial u}{\partial t} + a\frac{\partial u}{\partial x} = 0 \quad x \in \mathbb{R},\ t > 0 \\ &u(x,0) = u_0(x) \quad x \in \mathbb{R} \end{aligned}$$

Algunos ejemplos de esquemas numéricos explícitos para este problema son:

- Euler Central

$$u_j^{n+1} = u_j^n - \frac{k}{2h} a(u_{j+1}^n - u_{j-1}^n) \tag{5.8}$$

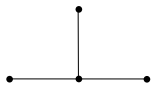


Figura 5.8 Molécula del método de Euler Central

Este método es inestable.

- Euler descentrado a izquierda.

$$u_j^{n+1} = u_j^n - \frac{k}{h} a(u_j^n - u_{j-1}^n) \tag{5.9}$$

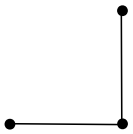


Figura 5.9 Molécula del método de Euler Descentrado a izquierda

- Euler descentrado a derecha.

$$u_j^{n+1} = u_j^n - \frac{k}{h} a(u_{j+1}^n - u_j^n) \tag{5.10}$$

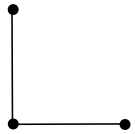


Figura 5.10 Molécula del método de Euler Descentrado a derecha

- Lax-Friedrichs.

$$u_j^{n+1} = \frac{1}{2}(u_{j-1}^n + u_{j+1}^n) - \frac{k}{2h}a(u_{j+1}^n - u_{j-1}^n) \tag{5.11}$$

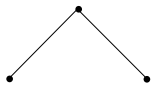


Figura 5.11 Molécula del método de Lax-Friedrichs

- Leapfrog.

$$u_j^{n+1} = u_j^{n-1} - \frac{k}{2h}a(u_{j+1}^n - u_{j-1}^n) \tag{5.12}$$

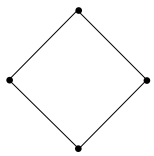


Figura 5.12 Molécula del método Leapfrog

- Lax-Wendroff

$$u_j^{n+1} = u_j^n - \frac{k}{2h}a(u_{j+1}^n - u_{j-1}^n) + \frac{k^2}{2h^2}a^2(u_{j+1}^n - 2u_j^n + u_{j-1}^n) \tag{5.13}$$

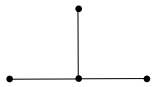


Figura 5.13 Molécula del método de Lax-Wendroff

- Beam-Warming

$$u_j^{n+1} = u_j^n - \frac{k}{2h}a(3u_j^n - 4u_{j-1}^n + u_{j-2}^n) + \frac{k^2}{2h^2}a^2(u_j^n - 2u_{j-1}^n + u_{j-2}^n) \tag{5.14}$$

Figura 5.14 Molécula del método Beam-Warming

Los métodos anteriores salvo el método Leapfrog son métodos de un paso que se pueden escribir de la forma

$$u^{n+1} = H(u^n)$$

donde u^{n+1} representa el vector $(u_j^{n+1})_{j=\ldots,-1,0,1,\ldots}$ de los valores aproximados en el instante t_{n+1}. Más precisamente podemos escribirlos de la forma

$$u^{n+1} = H(u^n;\, j)$$

por ejemplo, para el método de Lax-Friedrichs

$$H(u^n,\, j) = H(u_{j-1}^n,\, u_j^n,\, u_{j+1}^n)$$

que es un método de 3 puntos explícito.

Definición 5.1 *Un método de* $2l+1$ *puntos, de* 1 *paso, explícito es un método que se puede escribir de la forma*

$$u_j^{n+1} = H(u_{j-l}^n, \ldots, u_{j+l}^n) \quad n \geq 0 \quad j \in \mathbb{Z} \tag{5.15}$$

donde $H : \mathbb{R}^{2l+1} \to \mathbb{R}$ *es una función continua de* $2l+1$ *variables.*

O bien utilizando funciones constantes a trozos

$$u(x,t+k) = H(u(x-lh,t),...,u(x+lh,t)) \tag{5.16}$$

Definición 5.2 *Un método numérico de la forma (5.15) es lineal si verifica*

$$H(\alpha u^n + \beta v^n) = \alpha H(u^n) + \beta H(u^n) \tag{5.17}$$

es decir, de la forma

$$u_j^{n+1} = \sum_{i=-l}^{i=l} c_i u_{j+i}^n \quad n \geq 0,\ j \in \mathbb{Z}$$

Los métodos anteriores son todos lineales. Si nos limitamos a un intervalo acotado de $\mathbb{R}$ (completando el problema con adecuadas condiciones de contorno) un método lineal lo podemos expresar de la forma

$$u^{n+1} = Hu^n$$

donde H es una matriz. Por ejemplo en el caso del método de Lax-Friedrichs tenemos

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & & & \dots & & 0 \\ \frac{1}{2}+\frac{ak}{2h} & 0 & \frac{1}{2}-\frac{ak}{2h} & & \dots & & 0 \\ & \ddots & & \ddots & \ddots & & \\ & \dots & \frac{1}{2}+\frac{ak}{2h} & 0 & \frac{1}{2}-\frac{ak}{2h} & \dots & 0 \\ & & & \ddots & & \ddots & \ddots \\ 0 & & \dots & & \frac{1}{2}+\frac{ak}{2h} & 0 & \frac{1}{2}-\frac{ak}{2h} \\ 0 & & \dots & & & 0 & 1 \end{bmatrix}$$

donde hemos tenido en cuenta que en la práctica se trabaja en un intervalo acotado de $\mathbb{R}$ de modo que el índice $j \in [J_{\text{mín}}, J_{\text{máx}}]$. En la matriz anterior la primera fila y la última corresponden a las condiciones de contorno en $x_{J_{\text{mín}}}$ y en $x_{J_{\text{máx}}}$ y que dependerán de cada problema concreto.

Designaremos mediante

$$e_j^n = u(x_j,t_n) - u_j^n \tag{5.18}$$

al error puntual o bien utilizando funciones constantes a trozos

$$E_{k,h}(x,t) = u(x,t) - u_{k,h}(x,t) \tag{5.19}$$

El error global se medirá utilizando diversas normas, p.e., para el error puntual

$$||e^n||_1 = h\sum_j |e_j^n|$$

donde el sumatorio está limitado en la práctica a un número finito de valores de j. Para errores utilizando funciones constantes a trozos

$$||v||_1 = \int_{-\infty}^{\infty} |v(x)|\,dx$$

con la norma $||.||_1$ el error se expresa

$$||E_{k,h}(.,t)||_1 = \int_{-\infty}^{\infty} |E_{k,h}(x,t)|\,dx$$

o bien podremos utilizar otras normas como,

$$||v||_2 = (\int_{-\infty}^{\infty} |v(x)|^2, dx)^{1/2}$$

$$||v||_\infty = \max_{x\in\mathbb{R}} |v(x)|$$

aunque esta última puede dar indicaciones erróneas acerca de la validez de los métodos.

A continuación supondremos siempre que $\lambda = k/h = Constante$.

Definición 5.3 *Un método es de orden p si existe una constante C independiente de k (y por lo tanto también de h) tal que*

$$||E_k(.,t)|| \leq Ck^p$$

Definición 5.4 *Para un esquema de la forma*

$$u(x,t+k) = H(u(x-lh,t),...,u(x+lh,t))$$

el error de consistencia es

$$\tau_{h,k}(x,t) = \frac{1}{k}[u(x,t+k) - H(u(x-lh,t),...,u(x+lh,t))] \tag{5.20}$$

Diremos que el método es consistente si para una determinada norma $||.||$, se tiene

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

cuando $h,k \to 0$ para todo valor de t

Ya que elegimos h y k de manera que $\lambda = h/k$ es constante, bastará poner $||\tau_k(.,t)|| \to 0$ cuando $k \to 0$ para todo t.

Definición 5.5 *Diremos que un método es estable si para cada T existe una constante $C \geq 0$ y un valor $k_0 > 0$ tal que*

$$||H|| \leq C \quad \forall nk \leq T \quad k < k_0 \tag{5.21}$$

más precisamente, existe $\alpha \geq 0$ *tal que y un valor* $k_0 > 0$ *tal que*

$$||H|| \leq 1 + \alpha k \quad \forall nk \leq T \quad k < k_0 \tag{5.22}$$

Si un método verifica (5.22), entonces

$$||H^n|| \leq (1+\alpha k)^n \leq e^{\alpha k n} \leq e^{\alpha T}$$

para todo k y n. Para el análisis de los métodos lineales seguimos el procedimiento utilizado en los capítulos anteriores para los problemas parabólicos y elípticos. Retomemos aquel marco en el contexto de los problemas hiperbólicos. Los problemas hiperbólicos requieren, como veremos, un análisis diferenciado cuando se trata de problemas lineales o de problemas no lineales. Consideramos en esta subsección los problemas hiperbólicos lineales. Precisemos algunos conceptos.

La convergencia es el resultado del siguiente teorema de convergencia, conocido habitualmente como teorema de Lax, que hemos venido utilizando en los capítulos anteriores.

Teorema 5.1 *Un método lineal y consistente es estable si y solo si es convergente.*

Demostración. Demostraremos que si un método lineal y consistente es estable entonces es convergente. Tenemos

$$u^{n+1} = Hu^n + k\tau^n$$

Llamemos $\bar{u}^n = (...u(x-h,t_n), u(x,t_n), u(x+h,t_n), ...u(x+jh,t_n), ...)$ al vector correspondiente a la solución exacta en el instante t_n, y $e^n = \bar{u}^n - u^n$ al error en el paso n. Tendremos si el método es consistente con error de consistencia τ^n en el paso n

$$e^{n+1} = He^n + k\tau^n$$

aplicando recursivamente la anterior relación para $n = 1, ...n-1$

$$e^n = H^n e^0 + k \sum_{l=0}^{n-1} H^{n-l-1} \tau^l$$

tomando normas y mayorando

$$\begin{aligned} ||e^n|| &\leq ||H^n||.||e^0|| + k \sum_{l=0}^{n-1} ||H^{n-l-1}||.||\tau^l|| \\ &\leq e^{\alpha T}||e^0|| + Te^{\alpha T} \max_l ||\tau^l|| \to e^{\alpha T}||e^0|| \quad \text{cuando } k \to 0 \end{aligned}$$

Finalmente si $||e^0|| \to 0$ cuando $h = k/\lambda \to 0$, tendremos

$$||e^n|| \to 0 \quad \text{cuando } k \to 0$$

■

Vamos a estudiar la convergencia de alguno de los métodos anteriores.

Análisis Numérico del método de Lax-Friedrichs para problemas lineales

Consistencia

Suponemos $\lambda = k/h = Constante$. El esquema de Lax-Friedrichs (5.11) se puede escribir de la forma

$$\frac{1}{k}\Big(u_j^{n+1} - \frac{1}{2}(u_{j-1}^n + u_{j+1}^n)\Big) + \frac{1}{2h}a(u_{j+1}^n - u_{j-1}^n) = 0$$

reemplazando por la solución exacta

$$\frac{1}{k}\Big(u(x,t+k) - \frac{1}{2}\big(u(x-h,t)+u(x+h,t)\big)\Big) + \frac{1}{2h}a(u(x+h,t)-u(x-h,t)) = \tau_k(x,t)$$

Desarrollando cada término en serie de Taylor en un entorno de $u(x,t)$

$$\begin{aligned}
&\frac{1}{k}\Big(u + k\frac{\partial u}{\partial t} + \frac{k^2}{2}\frac{\partial^2 u}{\partial t^2} + \mathscr{O}(k^3) \\
&- \frac{1}{2}(u - h\frac{\partial u}{\partial x} + \frac{h^2}{2}\frac{\partial^2 u}{\partial x^2} - \frac{h^3}{6}\frac{\partial^3 u}{\partial x^3} + \mathscr{O}(h^4) + u + h\frac{\partial u}{\partial x} + \frac{h^2}{2}\frac{\partial^2 u}{\partial x^2} + \frac{h^3}{6}\frac{\partial^3 u}{\partial x^3} + \mathscr{O}(h^4))\Big) \\
&+ \frac{1}{2h}a\Big(u + h\frac{\partial u}{\partial x} + \frac{h^2}{2}\frac{\partial^2 u}{\partial x^2} + \frac{h^3}{6}\frac{\partial^3 u}{\partial x^3} + \mathscr{O}(h^4) - u + h\frac{\partial u}{\partial x} - \frac{h^2}{2}\frac{\partial^2 u}{\partial x^2} + \frac{h^3}{6}\frac{\partial^3 u}{\partial x^3} + \mathscr{O}(h^4)\Big) \\
&= \frac{\partial u}{\partial t} + \frac{k}{2}\frac{\partial^2 u}{\partial t^2} + \mathscr{O}(k^2) - \frac{h^2}{2k}\frac{\partial^2 u}{\partial x^2} + \mathscr{O}(h^4) + a\frac{\partial u}{\partial x} + \mathscr{O}(h^2) \\
&= \frac{\partial u}{\partial t} + a\frac{\partial u}{\partial x} + \frac{k}{2}(\frac{\partial^2 u}{\partial t^2} - \frac{h^2}{k^2}\frac{\partial^2 u}{\partial x^2}) + \mathscr{O}(k^2)
\end{aligned}$$

donde hemos tenido en cuenta que $\lambda = k/h$ es constante. En consecuencia el error de consistencia es

$$\tau_k(x,t) = \frac{k}{2}(\frac{\partial^2 u}{\partial t^2} - \lambda^2\frac{\partial^2 u}{\partial x^2}) + \mathscr{O}(k^2) \to 0 \quad \text{cuando } k \to 0$$

Estabilidad

Una condición necesaria y suficiente de estabilidad para el método de Lax-Friedrichs es

$$|\frac{ak}{h}| \leq 1 \tag{5.23}$$

Esta condición se conoce como la condición C.F.L. (Courant-Friedrichs-Lewy).

Demostremos la suficiencia. Utilizaremos la norma $||.||_1$. Tomando normas en (5.11) y mayorando

$$||u^{n+1}||_1 = h\sum_j |u_j^{n+1}| = h|\sum_j \frac{1}{2}(1-\frac{ak}{h})u_{j+1}^n + \frac{1}{2}(1+\frac{ak}{h})u_{j-1}^n|$$
$$\leq \frac{h}{2}\Big(\sum_j |(1-\frac{ak}{h})u_{j+1}^n| + \sum_j |(1+\frac{ak}{h})u_{j-1}^n|\Big)$$

Si $|\frac{ak}{h}| \leq 1$, $-1 \leq \frac{ak}{h} \leq 1$, entonces

$$1+\frac{ak}{h} \geq 0$$
$$1-\frac{ak}{h} \geq 0$$

y los términos entre paréntesis pueden salir del valor absoluto y del sumatorio quedando

$$||u^{n+1}||_1 \leq \frac{h}{2}\Big((1-\frac{ak}{h})\sum_j |u_{j+1}^n| + (1+\frac{ak}{h})\sum_j |u_{j-1}^n|\Big)$$
$$= \frac{1}{2}\Big((1-\frac{ak}{h})||u^n||_1 + (1+\frac{ak}{h})||u^n||_1\Big) = ||u^n||_1$$

Convergencia

Si se cumple la condición de estabilidad (5.23) el método de Lax-Friedrichs es convergente de orden $p=1$. En efecto, el error $e_j^n = u(x_j,t_n) - u_j^n$ verifica

$$e_j^{n+1} = \frac{1}{2}(e_{j-1}^n + e_{j+1}^n) - \frac{k}{2h}a(e_{j+1}^n - e_{j-1}^n) + k\tau_j^n$$

donde hemos puesto $\tau_j^n = \tau_k(x_j,t_n)$. El anterior resultado de estabilidad proporciona para el error la estimación

$$||e^{n+1}||_1 \leq ||e^n||_1 + k||\tau^n||_1$$

Aplicando recursivamente esta expresión, y para n tal que $kn \leq T$

$$||e^n||_1 \leq ||e^0||_1 + k\sum_{l=0}^{n-1} ||\tau^l||_1 \leq |e^0||_1 + T \max_{l=0,\ldots,n-1} ||\tau^l||_1$$

Si el error inicial $||e^0||_1 \to 0$ cuando $h = k/\lambda \to 0$

$$||e^n||_1 \leq ||e^0||_1 + T \max_{l=0,\ldots,n-1} ||\tau^l||_1 \to 0 \quad \text{cuando } k \to 0$$

Ejemplo

Aplicamos el método de Lax-Friedrichs al problema (5.1)-(5.2)

$$\begin{aligned} &\frac{\partial u}{\partial t} + a\frac{\partial u}{\partial x} = 0 \quad x \in \mathbb{R},\, t > 0 \\ &u(x,0) = u_0(x) \quad x \in \mathbb{R} \end{aligned}$$

con u_0 dado por

$$u_0(x) = \begin{cases} 1 & \text{si} \quad x < 0 \\ 0 & \text{si} \quad x > 0 \end{cases}$$

Los datos son $a = 1$, $h = 0.01$ y $k = 0.001$. De modo que $\lambda = 0.1$. Se ha resuelto el problema en $x \in [-2.5, 2.5]$ de modo que los valores en los extremos no se ven afectados por la solución para $t \in [0, 1]$. La solución para distintos tiempos se representa en la figura (5.15)

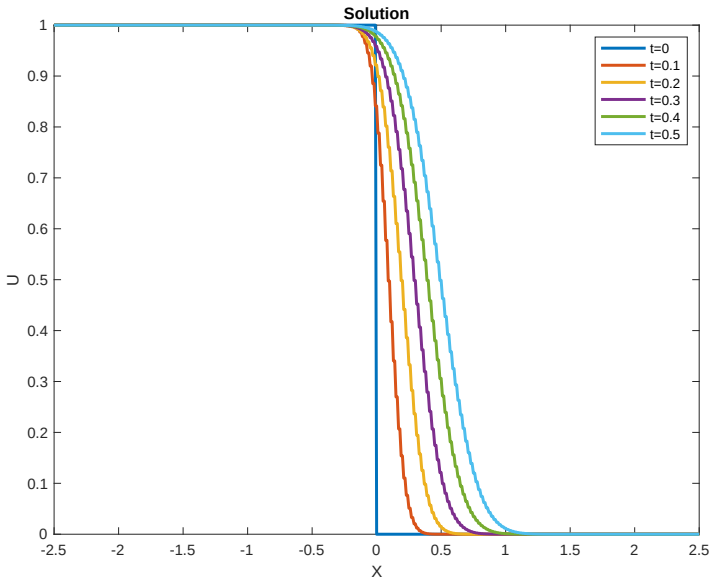


Figura 5.15 Solución numérica con el método de Lax-Friedrichs en distintos tiempos

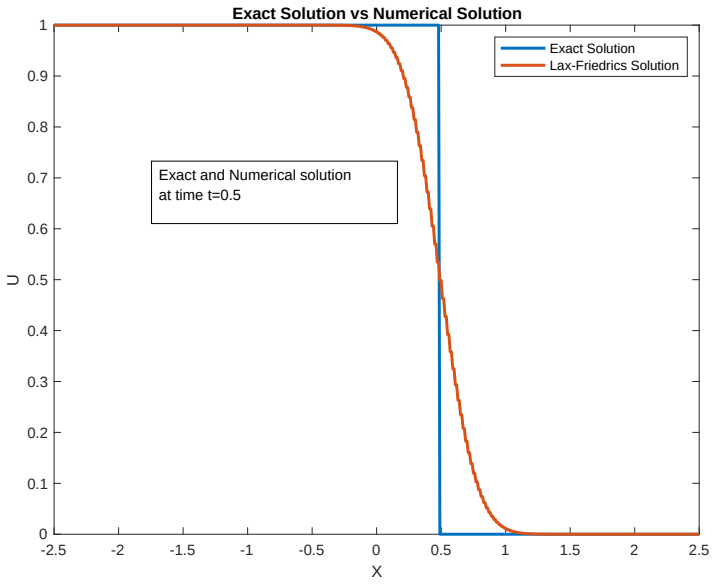


Figura 5.16 Solución exacta vs solución obtenida con el método de Lax-Friedrichs

En la figura (5.16) comparamos la solución exacta $u(.,0.5)$ en el instante t=0.5 con la solución numérica obtenida mediante el método de Lax-Friedrichs u^{500}.

Vamos a estudiar más detalladamente el error de consistencia en el método de Lax-Friedrichs. Derivando con respecto al tiempo la ecuación (5.1)

$$\frac{\partial^2 u}{\partial t^2} = -a\frac{\partial^2 u}{\partial t \partial x} = -a\frac{\partial}{\partial x}(\frac{\partial u}{\partial t}) = a^2\frac{\partial^2 u}{\partial x^2}$$

Podemos escribir el error de consistencia como

$$\begin{aligned}\tau_k(x,t) &= \frac{k}{2}(\frac{\partial^2 u}{\partial t^2} - \frac{h^2}{k^2}\frac{\partial^2 u}{\partial x^2}) + \mathcal{O}(k^2)\\ &= \frac{k}{2}(a^2 - \frac{h^2}{k^2})\frac{\partial^2 u}{\partial x^2} + \mathcal{O}(k^2)\\ &= \frac{h^2}{2k}(\frac{k^2a^2}{h^2} - 1)\frac{\partial^2 u}{\partial x^2} + \mathcal{O}(k^2)\end{aligned}$$

Si consideramos la ecuación modificada

$$\frac{\partial u}{\partial t} + a\frac{\partial u}{\partial x} - \frac{h^2}{2k}(1 - \frac{k^2a^2}{h^2})\frac{\partial^2 u}{\partial x^2} = 0$$

podemos observar que el esqema de Lax-Friedrichs es una aproximación de ésta con un error de consistencia de orden $p = 2$. De hecho la solución de

$$\frac{\partial u}{\partial t} + a\frac{\partial u}{\partial x} - D\frac{\partial^2 u}{\partial x^2} = 0$$

con

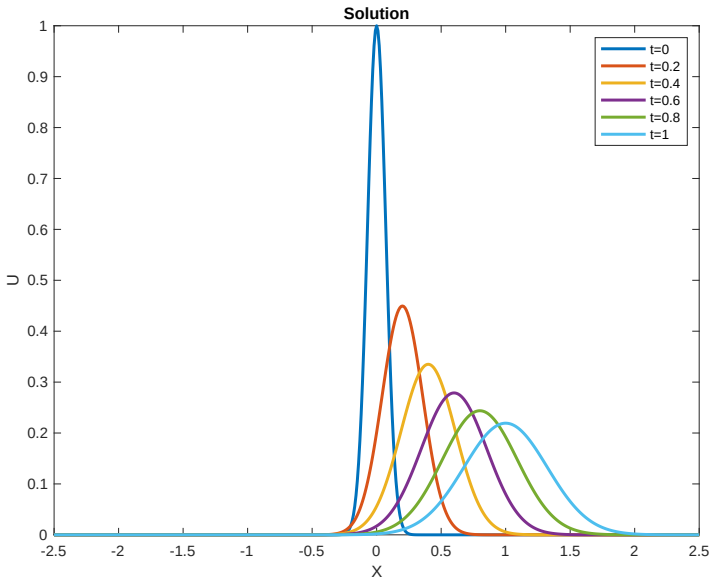
$$D = \frac{h^2}{2k}(1 - \frac{k^2a^2}{h^2})$$

está bien definida si $D > 0$. Precisamente ésta es la condición de estabilidad.

Ejercicio

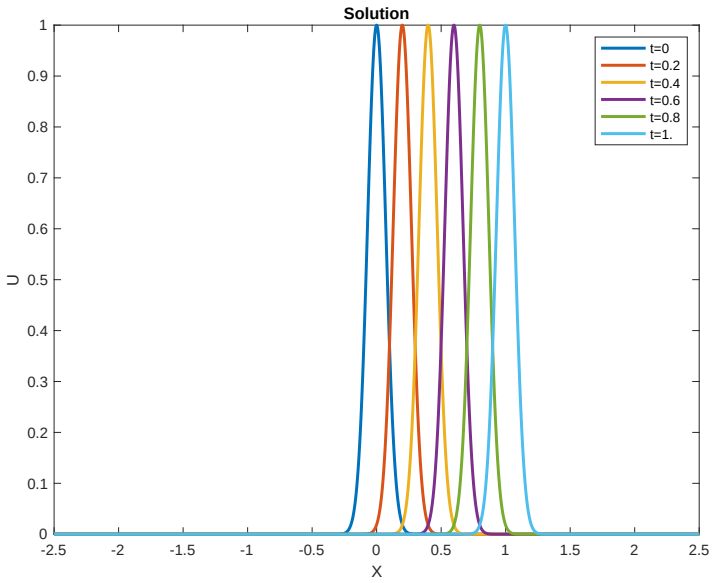
1. Aplicar el método de Lax-Friedrichs al problema (5.1)-(5.2) con la condición inicial u_0 dada por
$$u_0(x) = e^{(-x^2/0.01)}$$
y parámetros del problema dados por $a = 1$, $h = 0.01$ y $k = 0.001$. De modo que $\lambda = 0.1$.
2. Con la misma solución inicial que en el apartdao anterior pero con los parámetros parámetros del problema dados por $a = 1$, $h = 0.001$ y $k = 0.001$. De modo que $\lambda = 1$.
3. Interpretar los resultados en los dos casos anteriores. ¿Porqué, salvo errores de redondeo, se obtiene la solución exacta en el caso $\lambda = 1$?

Resultados numéricos



1.

Figura 5.17 Solución numérica con el método de Lax-Friedrichs en distintos tiempos, $\lambda = 0.1$



2.

Figura 5.18 Solución numérica con el método de Lax-Friedrichs en distintos tiempos, $\lambda = 1$

Análisis Numérico del método de Lax-Wendroff para problemas lineales

Consistencia

Vamos a probar que el método de Lax-Wendroff es consistente de orden 2 con la ecuación

$$\frac{\partial u}{\partial t}+a\frac{\partial u}{\partial x}$$

y es consistente de orden 3 con la ecuación

$$\frac{\partial u}{\partial t}+a\frac{\partial u}{\partial x}+\frac{ah^2}{6}(1-\frac{a^2k^2}{h^2})\frac{\partial^3 u}{\partial x^3}$$

El método de Lax-Wendroff está basado en el desarrollo de Taylor

$$u(x,t+k)=u(x,t)+k\frac{\partial u}{\partial t}(x,t)+\frac{1}{2}k^2\frac{\partial^2 u}{\partial t^2}+...$$

y en la observación según la cual de

$$\frac{\partial u}{\partial t}(x,t)=-a\frac{\partial u}{\partial x}(x,t)$$

se deduce

$$\frac{\partial^2 u}{\partial t^2}=-a\frac{\partial^2 u}{\partial t\partial x}=-a\frac{\partial}{\partial x}(\frac{\partial u}{\partial t})=a^2\frac{\partial^2 u}{\partial x^2} \tag{5.24}$$

El desarrollo de Taylor anterior se transforma en

$$u(x,t+k)=u(x,t)-ka\frac{\partial u}{\partial x}(x,t)+\frac{1}{2}k^2a^2\frac{\partial^2 u}{\partial x^2}+...$$

El esquema de Lax-Wendroff se obtiene al retener los tres primeros términos y utilizar diferencias centrales para aproximar las derivadas respecto a x.

El error de consistencia es:

$$\begin{aligned}\tau(x,t)=&\frac{u(x,t+k)-u(x,t)}{k}+\frac{a}{2h}\big(u(x+h,t)-u(x-h,t)\big)\\&-\frac{1}{2}\frac{a^2k}{h^2}\big(u(x+h,t)-2u(x,t)+u(x-h,t)\big)\end{aligned} \tag{5.25}$$

El desarrollo de Taylor de estos términos nos proporciona los restos

$$\frac{u(x,t+k)-u(x,t)}{k} = \frac{\partial u}{\partial t}(x,t) + \frac{k}{2}\frac{\partial^2 u}{\partial t^2}(x,t) + \frac{k^2}{6}\frac{\partial^3 u}{\partial t^3}(x,t) + \mathcal{O}(k^3)$$
$$\frac{u(x+h,t)-u(u(x-h,t)}{2h} = \frac{\partial u}{\partial x}(x,t) + \frac{h^2}{6}\frac{\partial^3 u}{\partial x^3}(x,t) + \mathcal{O}(h^4)$$
$$\frac{u(x+h,t)-2u(x,t)+u(u(x-h,t)}{h^2} = \frac{\partial^2 u}{\partial x^2}(x,t) + \frac{h^2}{12}\frac{\partial^4 u}{\partial x^4}(x,t) + \mathcal{O}(h^4)$$

Sustituyendo en (5.25)

$$\begin{aligned}\tau(x,t) = {} & \frac{\partial u}{\partial t} + a\frac{\partial u}{\partial x} + \frac{k}{2}\frac{\partial^2 u}{\partial t^2}(x,t) + \frac{k^2}{6}\frac{\partial^3 u}{\partial t^3}(x,t) \\ & + \frac{ah^2}{6}\frac{\partial^3 u}{\partial x^3}(x,t) - \frac{a^2 k}{2}\frac{\partial^2 u}{\partial x^2}(x,t) - \frac{a^2 kh^2}{24}\frac{\partial^4 u}{\partial x^4}(x,t) \\ & + \mathcal{O}(k^3) + \mathcal{O}(h^4) + \mathcal{O}(h^4)\end{aligned}$$

teniendo en cuenta (5.24) y que la solución exacta verifica la ecuación

$$\tau(x,t) = \frac{k^2}{6}\frac{\partial^3 u}{\partial t^3}(x,t) + \frac{ah^2}{6}\frac{\partial^3 u}{\partial x^3}(x,t) + \mathcal{O}(kh^2)$$

Derivando una vez más con respecto al tiempo en (5.24)

$$\frac{\partial^3 u}{\partial t^3} = a^2\frac{\partial^2}{\partial x^2}(\frac{\partial u}{\partial t}) = -a^3\frac{\partial^3 u}{\partial x^3}$$

de modo que podemos expresar el error de consistencia como

$$\tau(x,t) = \frac{ah^2}{6}(1-\frac{a^2k^2}{h^2})\frac{\partial^3 u}{\partial x^3} + \mathcal{O}(k^3)$$

Por lo tanto el esquema de Lax-Wendroff es consistente de orden 2 con la ecuación

$$\frac{\partial u}{\partial t} + a\frac{\partial u}{\partial x}$$

y es consistente de orden 3 con la ecuación

$$\frac{\partial u}{\partial t} + a\frac{\partial u}{\partial x} + \frac{ah^2}{6}(1-\frac{a^2k^2}{h^2})\frac{\partial^3 u}{\partial x^3}$$

Las ecuaciones del tipo

$$\frac{\partial u}{\partial t} + a\frac{\partial u}{\partial x} = \mu\frac{\partial^3 u}{\partial x^3}$$

se llaman dispersivas y la solución presenta oscilaciones.

Estabilidad

Utilizaremos el método de Von-Neumann. Resultará cómodo trabajar con funciones constantes a trozos en $L^2(\mathbb{R})$ y analizar la estabilidad en la norma

$$||v||_2 = (\int_{-\infty}^{\infty} |v(x)|^2\, dx)^{1/2}$$

Con funciones constantes a trozos un paso del esquema de Lax-Wendroff se escribe poniendo $\lambda = \frac{ka}{h}$

$$u^{n+1}(x) = u^n(x) - \frac{\lambda}{2}(u^n(x+h) - u^n(x-h)) + \frac{\lambda^2}{2}(u^n(x+h) - 2u^n(x) + u^n(x-h))$$

La condición de estabilidad es

$$||u^n||_2 \leq C||u^0||_2 \quad \forall n > 0$$

Ahora la transformada de Fourier en $L^2(\mathbb{R})$ definida por

$$\hat{v}(\xi) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-ix\xi} v(x)\, dx$$

es una isometría de $L^2(\mathbb{R})$ en $L^2(\hat{\mathbb{R}})$, es decir $||\hat{v}||_2 = ||v||_2$. Aplicado la transformada de Fourier a los términos del esquema de Lax-Wendroff resulta

$$\hat{u}^{n+1}(\xi) = G(\xi)\hat{u}^n(\xi)$$

donde $G(\xi)$ es

$$\begin{aligned} G(\xi) &= 1 - \frac{\lambda}{2}(e^{ih\xi} - e^{-ih\xi}) - \frac{\lambda^2}{2}(2 - \left(e^{ih\xi} + e^{-ih\xi}\right)) \\ &= 1 - \lambda^2(1 - \cos(h\xi) - i\lambda \sin(h\xi) \end{aligned}$$

$G(\xi)$ es el llamado factor de amplificación y la condición de estabilidad conocida como condición de Von-Neumann es

$$|G(\xi)| \leq 1 \quad \forall \xi \in \mathbb{R} \tag{5.26}$$

En el caso del esquema de Lax-Wendroff (5.13) tendremos

Teorema 5.2 *El esquema de Lax-Wendroff es $L^2(\mathbb{R})$-estable si y solo si se verifica la condición*

$$\lambda = \frac{ka}{h} < 1$$

Demostración. En efecto, calculemos $|G(\xi)|^2$.

$$\begin{aligned}
|G(\xi)|^2 &= \Big(1-\lambda^2(1-\cos(h\xi)\Big)^2+\lambda^2\sin^2(h\xi)\\
&= 1-2\lambda^2(1-\cos(h\xi))+\lambda^4\big(1-\cos(h\xi)\big)^2+\lambda^2(1-\cos^2(h\xi))\\
&= 1-2\lambda^2+2\lambda^2\cos(h\xi)+\lambda^4(1-\cos(h\xi))^2+\lambda^2-\lambda^2\cos^2(h\xi)\\
&= 1-\lambda^2(1-2\cos(h\xi)+\cos^2(h\xi))+\lambda^4(1-\cos(h\xi))^2\\
&= 1-\lambda^2(1-\cos(h\xi))^2+\lambda^4(1-\cos(h\xi))^2\\
&= 1-\lambda^2(1-\lambda^2)(1-\cos(h\xi))^2\geq 0
\end{aligned}$$

Finalmente la condición de estabilidad (5.26) se traduce en

$$1-\lambda^2\geq 0 \quad \Leftrightarrow \quad \lambda=\frac{ka}{h}\leq 1$$

que es la condición C.F.L..

Hasta aquí se ha probado que la condición (5.26) es suficiente para la estabilidad. Para ver que la condición (5.26) es necesaria para la estabilidad basta ver que si no se cumple la condición C.F.L., es decir si $\lambda=\frac{ka}{h}>1$ entonces $||u^n||_2\to\infty$ para algún valor inicial u^0, o lo que es lo mismo, $||\hat{u}^n||_2\to\infty$ para algún valor inicial $\hat{u}^0$. Supongamos pues $\lambda=\frac{ka}{h}>1$,

$$\hat{u}^n=G^n(\xi)\hat{u}^0$$

y tomando módulos

$$|\hat{u}^n|=|G(\xi)|^n|\hat{u}^0|$$

Para $h\xi\neq 2m\pi,\ m=0,1,2,...,\ |G(\xi)|^2=1-\lambda^2(1-\lambda^2)(1-\cos(h\xi))^2>1$ si $\lambda>1$. Eligiendo $\hat{u}^0$ de modo que su soporte no contenga a $2m\pi,\ m=0,1,2,...$ tendremos

$$|G(\xi)|=(|1-\lambda^2(1-\lambda^2)(1-\cos(h\xi))^2|)^{1/2}=\eta>1$$

y finalmente en los puntos ξ del soporte de $\hat{u}^0$

$$|\hat{u}^n(\xi)|=|G(\xi)|^n|\hat{u}^0(\xi)|=\eta^n\to\infty \quad \text{cuando } n\to\infty$$

de donde tomando la norma en $L^2(\hat{\mathbb{R}})$

$$||\hat{u}^n||_2\to\infty \quad \text{cuando } n\to\infty$$

y por tanto

$$||u^n||_2\to\infty \quad \text{cuando } n\to\infty$$

■

Convergencia

La convergencia es consecuencia inmediata del teorema de Lax.

Ejercicio

Aplicar el método de Lax-Wendroff para resolver el problema (5.1)-(5.2) con u_0 dado por

$$u_0(x) = \begin{cases} 1 & \text{si} \quad x < 0 \\ 0 & \text{si} \quad x > 0 \end{cases}$$

Solución

Los datos son $a = 1$, $h = 0.01$ y $k = 0.001$. De modo que $\lambda = 0.1$. Se ha resuelto el problema en $x \in [-2.5, 2.5]$ de modo que los valores en los extremos no se ven afectados por la solución para $t \in [0, 1]$. La solución para distintos tiempos se representa en la figura (5.19)

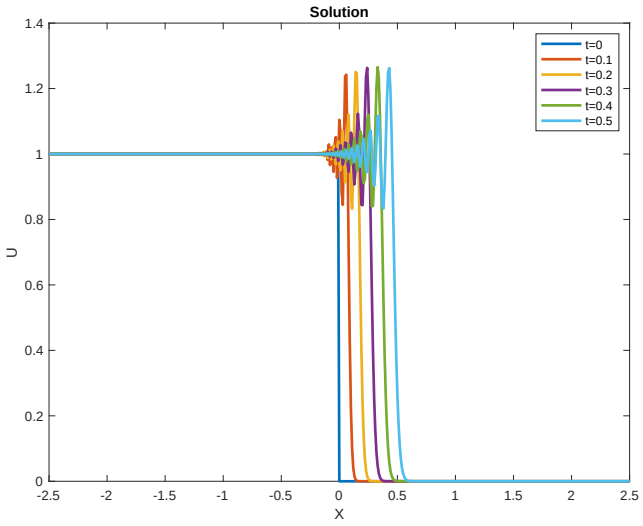


Figura 5.19 Solución numérica con el método de Lax-Wendroff en distintos tiempos

En la figura (5.20) comparamos la solución exacta $u(.,0.5)$ en el instante t=0.5 con la solución numérica obtenida mediante el método de Lax-Wendroff u^{500}.

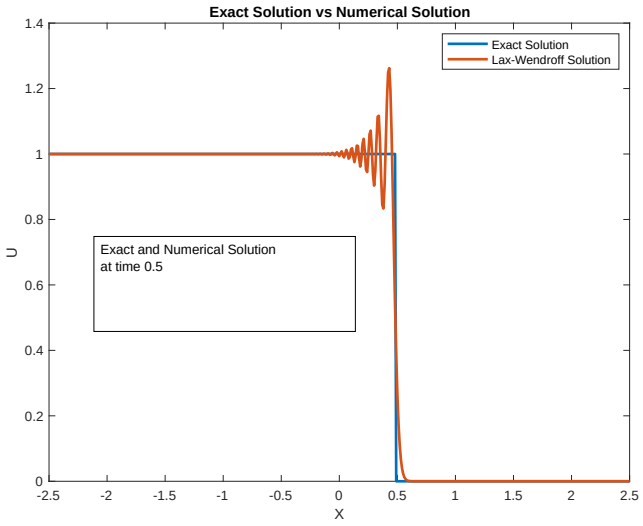


Figura 5.20 Solución exacta vs solución obtenida con el método de Lax-Wendroff

Ejercicio

1. Aplicar el método de Lax-Wendroff al problema (5.1)-(5.2) con la condición inicial u_0 dada por

$$u_0(x) = e^{(-x^2/0.01)}$$

y parámetros del problema dados por $a = 1$, $h = 0.01$ y $k = 0.001$. De modo que $\lambda = 0.1$.

2. Con la misma solución inicial que en el apartado anterior pero con los parámetros parámetros del problema dados por $a = 1$, $h = 0.001$ y $k = 0.001$. De modo que $\lambda = 1$.
3. Interpretar los resultados en los dos caso anteriores.

Ejercicio

1. Demostrar que el método de Euler central es inestable.
   Indicación
   Analizar la estabilidad en $L^2(\mathbb{R})$ mediante el método de Fourier y el criterio de Von-Neumann.
2. Analizar la consistencia y estabilidad de los métodos de Euler descentrados.

5.2.2. Métodos Numéricos para Problemas hiperbólicos lineales de segundo orden

El ejemplo más inmediato de un problema hiperbólico de segundo orden es la ecuación de una cuerda vibrante. Si suponemos esta cuerda sujeta por sus extremos, como por ejemplo la cuerda de un violín el problema que los describe es

$$\frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2} \tag{5.27}$$

$$u(0,t) = u(1,t) = 0 \quad \text{para } t \in [0,T] \tag{5.28}$$

$$u(x,0) = f(x) \quad \text{en } x \in (0,1) \tag{5.29}$$

$$\frac{\partial u}{\partial t}(x,0) = g(x) \quad \text{en } x \in (0,1) \tag{5.30}$$

Hemos supuesto sin perder generalidad que la longitud de la cuerda es $L = 1$. Para una cuerda de longitud finita es preciso dar las condiciones en los extremos de la cuerda (condiciones de contorno). Puesto que la ecuación es de segundo orden en la variable t (tiempo) se necesitan dos condiciones iniciales para tener determinado la solución, por lo que suponemos conocido el valor de $u(.,0)$ (desplazamiento inicial) y el de $\frac{\partial u}{\partial t}(.,0)$ (velocidad inicial). La ecuación, las condiciones de contorno y las condiciones iniciales permiten asegurar la existencia y unicidad de solución del problema. c es un parámetro físico, relacionado con las propiedades físicas de la cuerda, concretamente $c = \frac{\sigma}{\rho}$ donde σ es la tensión de la cuerda y ρ la densidad longitudinal de la misma.

Vamos a construir un Método de Diferencias Finitas para resolver (5.27)-(5.28)-(5.29)-(5.30). Consideramos como en la sección (1.2) en el plano $x-t$ un rectángulo $[0,1]\times[0,T]$ y una malla de diferencias finitas como la de la figura (1.1) con los parámetros h y k correspondientes representando el tamaño de la malla en las direcciones x y t respectivamente. Con las mismas notaciones de la sección (1.2), u_j^n representará una aproximación del valor exacto de la solución u del problema (1.10)-(1.11)-1.12) en este punto, es decir $u_j^n \approx u(x_j,t_n)$ con $x_j = jh$ y $t_n = nk$. Para cada valor de n tenemos definido un vector $u^n = (u_1^n, u_2^n, ..., u_M^n) \in \mathbb{R}^M$ en el espacio euclídeo M-dimensional. Un Método de Diferencias Finitas para resolver el problema anterior es

$$\bar{\partial}_t(\partial_t)u_j^n = c^2\bar{\partial}_x(\partial_x)u_j^n \quad j = 1,...,M;\ n \geq 0 \tag{5.31}$$

$$u_0^n = u_{M+1}^n = 0 \quad n > 0 \tag{5.32}$$

$$u_j^0 = f(x_j) \quad j = 1,...,M \tag{5.33}$$

$$\frac{u_j^1 - u_j^{-1}}{2k} = g(x_j) \quad j = 1,...,M \tag{5.34}$$

La ecuación (5.31) desarrollada se escribe

$$\frac{u_j^{n+1} - 2u_j^n + u_j^{n-1}}{k^2} = c^2\frac{u_{j+1}^n - 2u_j^n + u_{j-1}^n}{h^2} \tag{5.35}$$

La ecuación (5.34) necesita una aclaración. En principio la solución exacta no está definida para $t < 0$ de modo que debemos aclarar como determinar el valor de u_j^{-1}. Para aproximar la condición inicial (5.34) extendemos la solución para valores $t < 0$. Al imponer las condiciones iniciales mediante (5.33) y (5.34) obtenemos un sistema de dos ecuaciones con tres incógnitas u_j^{-1}, u_j^0 y u_j^1. La ecuación en diferencias finitas (5.35) proporciona una tercera ecuación

$$u_j^1 = 2u_j^0 - u_j^{-1} + \frac{c^2k^2}{h^2}(u_{j+1}^0 - 2u_j^0 + u_{j-1}^0) \tag{5.36}$$

El término u_j^{-1} se puede eliminar de (5.34) y (5.36) y podemos obtener u_j^1 ya que conocemos u_j^0. El valor de u_j^1 es

$$u_j^1 = u_j^0 + kg(x_j) + \frac{c^2k^2}{2h^2}(u_{j+1}^0 - 2u_j^0 + u_{j-1}^0)$$

De esta manera la aproximación (5.34) de la condición inicial (5.30) es consistente de orden $\mathcal{O}(k^2)$ como se indica a continuación.

Consistencia

Sustituyendo en los términos de la expresión (5.35) la solución exacta de (5.27)-(5.28)-(5.29)-(5.30) y utilizando el desarrollo de Taylor vemos que el error de consistencia es del orden $\mathscr{O}(h^2)+\mathscr{O}(k^2)$. Del mismo modo sustituyendo la solución exacta en (5.34) tenemos también que el orden de consistencia con respecto a la ecuación (5.30) es también de orden $\mathscr{O}(k^2)$.

Estabilidad I

Una vez obtenido u_j^{-1} y u_j^0 la solución u_j^n para $n=1,2,...$ se obtiene aplicando recursivamente

$$u_j^{n+1}=2u_j^n-u_j^{n-1}-\mu^2(-u_{j+1}^n+2u_j^n-u_{j-1}^n)$$

que podemos expresar matricialmente de la forma

$$u^{n+1}=2u^n-u^{n-1}-\mu^2\mathbf{A}u^n \tag{5.37}$$

donde $\mu^2=\frac{c^2k^2}{h^2}$ y $\mathbf{A}$ es la matriz tridiagonal

$$\mathbf{A}=\begin{bmatrix} 2 & -1 & \dots & \dots & 0 \\ -1 & 2 & \dots & \dots & 0 \\ & \ddots & \ddots & \ddots & \\ & & & & -1 \\ 0 & 0 & \dots & -1 & 2 \end{bmatrix}$$

Sean $(\beta_p,\ w_p)_p \quad p=1,...,M$ los correspondiente pares de valores propios y vectores propios de $\mathbf{A}$. Expresando u^{n+1}, u^n y u^{n-1} en función de la base ortogonal de vectores propios tenemos

$$\begin{aligned} u^{n+1}&=\sum_p \rho_p^{(n+1)}w_p \\ u^n&=\sum_p \rho_p^{(n)}w_p \\ u^{n-1}&=\sum_p \rho_p^{(n-1)}w_p \\ \mathbf{A}u^n&=\sum_p \rho_p^{(n)}\beta_p w_p \end{aligned}$$

la expresión (5.37) se escribe

$$\sum_p \rho_p^{(n+1)} w_p = 2\sum_p \rho_p^{(n)} w_p - \sum_p \rho_p^{(n-1)} w_p - \mu^2 \sum_p \beta_p \rho_p^{(n)} w_p$$

y teniendo en cuenta que la ortogonalidad de la base de vectores propios resulta para cada valor de p

$$\rho_p^{(n+1)} = 2\rho_p^{(n)} - \rho_p^{(n-1)} - \mu^2 \beta_p \rho_p^{(n)} \tag{5.38}$$

La ecuación (5.38) es una ecuación en diferencias homogénea para cada p. Buscaremos soluciones del tipo $\rho_p^{(n)} = r^n$. Entonces sustituyendo en (5.38) y multiplicando por r

$$r^2 - (2 - \mu^2 \beta_p) r + 1 = 0$$

Las soluciones son

$$r_\pm = \frac{2 - \mu^2 \beta_p \pm \sqrt{(2 - \mu^2 \beta_p)^2 - 4}}{2}$$

Llamando $\sigma_p = -\mu^2 \beta_p$

$$r_\pm = \frac{2 + \sigma_p \pm \sqrt{(2 + \sigma_p)^2 - 4}}{2} \tag{5.39}$$

La solución general de la ecuación en diferencias es una combinación lineal de r_+ y r_-.

Soluciones complejas conjugadas

Si $-2 < 2 + \sigma_p < 2$ las dos soluciones son complejas conjugadas y por tanto del mismo módulo cuyo valor es la unidad, en efecto

$$\frac{(2 + \sigma_p)^2}{4} + \frac{4 - (2 + \sigma_p)^2}{4} = 1$$

y tenemos estabilidad. La condición de estabilidad se traduce en

$$-2 < 2 - \mu^2 \beta_p < 2 \quad \forall p = 1, ..., M$$

es decir

$$0 < \mu^2 \beta_p < 4 \quad \forall p = 1, ..., M$$

La primera desigualdad se cumple siempre pues $\beta_p > 0$. Teniendo en cuenta que $\beta_p = 2(1 - \cos(hp\pi))$, el mayor valor de β_p corresponde a β_M para el cual tenemos $\beta_M < 4$, de modo que la condición de estabilidad es

$$\mu^2 \leq 1$$

es decir

$$\frac{ck}{h} \leq 1$$

que es la condición C.F.L.

Soluciones reales simples Si $2 < 2+\sigma_p < -2$, las raíces son reales con producto igual a 1. De modo que una de ellas será mayor que 1 en valor absoluto dando lugar a un comportamiento inestable.

Solución real doble

Este caso no se puede dar. En efecto, si $\sigma_p + 2 = 2$, debería ocurrir que $-\mu^2\beta_p = 0$ pero esto no sucede ya que $\beta_p > 0 \quad \forall p$.

Ejercicio

Calcular la solución de (5.31)-(5.32)-(5.33)-(5.34).

Solución

La solución general de la ecuación en diferencias para dos raices complejas conjugadas del polinomio característico son (el módulo hemos visto que es igual 1)

$$\rho_p^{(n)} = A_p \cos(n\vartheta_p) + B_p \sin(n\vartheta_p)$$

donde $\vartheta_p = \arctan \frac{\sqrt{(2+\sigma_p)^2-4}}{2+\sigma_p}$. La solución en el paso n será

$$u^n = \sum_p \rho_p^{(n)} w_p = \sum_p \left(A_p \cos(n\vartheta_p) + B_p \sin(n\vartheta_p)\right) w_p$$

Para calcular A_p y B_p utilizamos las condiciones de contorno, es decir los valores de u^0 y u^1. Tendremos por una parte para $n = 0$

$$u^0 = (u_j^0)_{j=1}^M = (f(x_j))_{j=1}^M = \sum_p A_p \, w_p$$

Multiplicando escalarmente por w_q

$$(u^0, w_q) = \sum_p A_p (w_p, w_q) = \frac{1}{2} A_q$$

de donde

$$A_q = 2(u^0, w_q) = 2h\sum_j u_j^0 w_{q,j} = 2h\sum_j u_j^0 \sin(jh\pi q)$$
$$= 2h\sum_j f(x_j)\sin(jh\pi q)$$

Por otra parte para calcular B_p utilizamos u^1

$$u^1 = (u_j^1)_{j=1}^M$$

$$(u^1, w_q) = \sum_p \big((A_p\cos\vartheta_p + B_p\sin\vartheta_p)w_p, w_q\big)$$
$$= \frac{1}{2}A_q\cos\vartheta_q + \frac{1}{2}B_q\sin\vartheta_q$$

Sustituyendo el valor de A_q

$$B_q = 2\frac{(u^1 - u^0\cos\vartheta_q, w_q)}{\sin\vartheta_q} = \frac{2h}{\sin\vartheta_q}\sum_j (u_j^1 - u_j^0\cos\vartheta_q)\sin(jh\pi q)$$
$$= \frac{2h}{\sin\vartheta_q}\sum_j \Big((f(x_j)(1-\cos\vartheta_q) + kg(x_j) + \frac{\mu^2}{2}\big(f(x_{j+1}) - 2f(x_j) + f(x_{j-1})\big)\Big)\sin(jh\pi q)$$

Estabilidad II y Convergencia

Con el objetivo de demostrar la convergencia del método necesitamos realizar un análisis de la estabilidad del problema no homogéneo. Para ello será útil reformular el problema (5.27)-(5.28)-(5.29)-(5.30) como un sistema de primer orden en la variable t de la siguiente forma

$$\frac{\partial u}{\partial t} = z \tag{5.40}$$
$$\frac{\partial z}{\partial t} = c^2\frac{\partial^2 u}{\partial x^2} \tag{5.41}$$
$$u(0,t) = u(1,t) = 0 \quad \text{para } t \in [0,T] \tag{5.42}$$
$$u(x,0) = f(x) \quad \text{en } x \in (0,1) \tag{5.43}$$
$$z(x,0) = g(x) \quad \text{en } x \in (0,1) \tag{5.44}$$

Mediante el desarrollo de Taylor, teniendo en cuenta (5.27)

$$\frac{u(x_j,t_n+k)-u(x_j,t_n)}{k} = \frac{\partial u}{\partial t}(x_j,t_n) + \frac{c^2k}{2}\big(\frac{\partial^2 u}{\partial x^2}(x_j,t_n)\big) + \mathscr{O}(k^2)$$
$$= \frac{\partial u}{\partial t}(x_j,t_n) + \frac{c^2k}{2}\frac{u(x_j-h,t_n)-2u(x_j,t_n)+u(x_j+h,t_n)}{h^2} + \mathscr{O}(k^2) + \mathscr{O}(kh^2)$$

Por otra parte desarrollando en serie Taylor $z(x,t_n+k)$ en un entorno de (x,t_n) y $z(x,t_n)$ en un entorno de $z(x,t_n+k)$ y restando se obtiene

$$\frac{z(x_j,t_n+k)-z(x_j,t_n)}{k} = \frac{1}{2}\Big(\frac{\partial z}{\partial t}(x_j,t_n+k) + \frac{\partial z}{\partial t}(x_j,t_n)\Big)$$
$$+ \frac{k}{4}\Big(\frac{\partial^2 z}{\partial t^2}(x_j,t_n) - \frac{\partial^2 z}{\partial t^2}(x_j,t_n+k)\Big) + \mathscr{O}(k^2)$$
$$= \frac{1}{2}\Big(\frac{\partial z}{\partial t}(x_j,t_n+k) + \frac{\partial z}{\partial t}(x_j,t_n)\Big) + \mathscr{O}(k^2)$$
$$= \frac{c^2}{2}\Big(\frac{u(x_j-h,t_n+k)-2u(x_j,t_n+k)+u(x_j+h,t_n+k)}{h^2}$$
$$+ \frac{u(x_j-h,t_n)-2u(x_j,t_n)+u(x_j+h,t_n)}{h^2}\Big) + \mathscr{O}(k^2) + \mathscr{O}(h^2)$$

En consecuencia el esquema numérico se escribe

$$\frac{u_j^{n+1}-u_j^n}{k} = z_j^n + \frac{c^2k}{2}\frac{u_{j-1}^n-2u_j^n+u_{j+1}^n}{h^2} \tag{5.45}$$
$$\frac{z_j^{n+1}-z_j^{n+1}}{k} = \frac{c^2}{2}\Big(\frac{u_{j-1}^n-2u_j^n+u_{j+1}^n}{h^2} + \frac{u_{j-1}^{n+1}-2u_j^{n+1}+u_{j+1}^{n+1}}{h^2}\Big) \tag{5.46}$$
$$u_0^n = u_{M+1}^n = 0 \tag{5.47}$$
$$u_j^0 = f(x_j) \tag{5.48}$$
$$z_j^0 = g(x_j) \tag{5.49}$$
$$\tag{5.50}$$

Observando que $u_j^{n+1}-2u_j^n+u_j^{n-1} = (u_j^{n+1}-u_j^n)-(u_j^n-u_j^{n-1})$ y utilizando (5.46) obtenemos el esquema (5.31)-(5.32)-(5.33)-(5.34).

Poniendo $u^n = (u_j^n)_j,\ j=1,...,M$ y $z^n = (z_j^n)_j,\ j=1,...,M$ el esquema anterior se escribe con notación matricial y reordenando términos

$$u^{n+1} = u^n - \frac{k^2}{2}\mathbf{A}_h u^n + kz^n \tag{5.51}$$
$$z^{n+1} = z^n - \frac{k}{2}(\mathbf{A}_h u^{n+1} + \mathbf{A}_h u^n) \tag{5.52}$$
$$u^0 = (f(x_j))_{j=1,...,M} \tag{5.53}$$
$$z^0 = (g(x_j))_{j=1,...,M} \tag{5.54}$$

donde aquí $\mathbf{A}_h$ es la matriz

$$\mathbf{A}_h = \frac{c^2}{h^2}\mathbf{A} = \frac{c^2}{h^2}\begin{bmatrix} 2 & -1 & \dots & \dots & 0 \\ -1 & 2 & \dots & \dots & 0 \\ & \ddots & \ddots & \ddots & \\ & & & & -1 \\ 0 & 0 & \dots & -1 & 2 \end{bmatrix}$$

Consideremos ahora los errores $e_j^n = u(x_j,t_n) - u_j^n$ y $d_j^n = z(x_j,t_n) - z_j^n$. Para $e^n = (e_j^n)_j,\ j = 1,...,M$ y $d^n = (d_j^n)_j,\ j = 1,...,M$ tendremos

$$e^{n+1} = e^n - \frac{k^2}{2}\mathbf{A}_h e^n + kd^n + k^2\varepsilon^n$$
$$d^{n+1} = d^n - \frac{k^2}{2}(\mathbf{A}_h e^{n+1} + \mathbf{A}_h e^n) + k\eta^n$$

donde ε_j^n y η_j^n son términos de orden $\mathscr{O}(k^2+h^2)$ que son errores de consistencia. Utilizando los valores propios y los vectores propios de $\mathbf{A}_h$, $\beta_{h,p} = \frac{c^2}{h^2}\beta_p = 2\frac{c^2}{h^2}\big(1 - \cos(2\pi p)\big)$ y w_p (que son los mismos vectores propios de $\mathbf{A}$) podemos expresar los vectores anteriores en función de sus componentes en la base de los vectores propios w_p. Para $p = 1,...,M$ llamemos a estas componentes

$$e_p^n = (e^n, w_p),\ d_p^n = (d^n, w_p),\ \varepsilon_p^n = (\varepsilon^n, w_p),\ \eta_p^n = (\eta^n, w_p)$$

$$e_p^{n+1} = e_p^n - \frac{k^2}{2}\beta_{h,p}e_p^n + kd_p^n + k^2\varepsilon_p^n$$
$$d_p^{n+1} = d_p^n - \frac{k}{2}(\beta_{h,p}e^{n+1} + \beta_{h,p}e^n) + k\eta_p^n$$

sustituyendo en la seguna ecuación el valor de e_p^{n+1}

$$e_p^{n+1} = (1 - \frac{k^2}{2}\beta_{h,p})e_p^n + kd_p^n + k^2\varepsilon_p^n$$
$$d_p^{n+1} = -k\beta_{h,p}(1 - \frac{k^2\beta_{h,p}}{4})e_p^n + (1 - \frac{k^2\beta_{h,p}}{2})d_p^n - \frac{k^2\beta_{h,p}}{2}\varepsilon_p^n + k\eta_p^n$$

Poniendo $\theta_p = \sqrt{\beta_p}k$

$$\begin{bmatrix} e_p^{n+1} \\ \frac{d_p^{n+1}}{\sqrt{\beta_{h,p}}} \end{bmatrix} = \begin{bmatrix} (1 - \frac{\theta_p^2}{2}) & \theta_p \\ -\theta_p(1 - \frac{\theta_p^2}{4}) & (1 - \frac{\theta_p^2}{2}) \end{bmatrix} \begin{bmatrix} e_p^n \\ \frac{d_p^n}{\sqrt{\beta_{h,p}}} \end{bmatrix} + k \begin{bmatrix} 1 & 0 \\ -\frac{\theta_p}{2} & 1 \end{bmatrix} \begin{bmatrix} k\varepsilon_p^n \\ \frac{\eta_p^n}{\sqrt{\beta_{h,p}}} \end{bmatrix}$$

Utilizando la siguiente notación vectorial

$$X_p^n = \begin{bmatrix} e_p^n \\ \dfrac{d_p^n}{\sqrt{\beta_{h,p}}} \end{bmatrix} \quad \mathbf{B}_p(\theta_p) = \begin{bmatrix} (1-\dfrac{\theta_p^2}{2}) & \theta_p \\ -\theta_p(1-\dfrac{\theta_p^2}{4}) & (1-\dfrac{\theta_p^2}{2}) \end{bmatrix}$$

$$\mathbf{D}_p(\theta_p) = \begin{bmatrix} 1 & 0 \\ -\dfrac{\theta_p}{2} & 1 \end{bmatrix} \quad E_p^n = \begin{bmatrix} k\varepsilon_p^n \\ \dfrac{\eta_p^n}{\sqrt{\beta_{h,p}}} \end{bmatrix}$$

la relación de recurrencia se escribe

$$X_p^{n+1} = \mathbf{B}(\theta_p)X_p^n + k\mathbf{D}(\theta_p)E_p^n \quad p = 1,...,M$$

y aplicando recursivamente la expresión anterior

$$X_p^n = \mathbf{B}^n(\theta_p)X_p^0 + k\sum_{l=0}^{n-1}\mathbf{B}^{n-l-1}(\theta_p)\mathbf{D}(\theta_p)E_p^l \quad p = 1,...,M$$

de donde

$$|X_p^n| \leq ||\mathbf{B}^n(\theta_p)||.|X_p^0| + k\sum_{l=0}^{n-1}||\mathbf{B}^{n-l-1}(\theta_p)||.||\mathbf{D}(\theta_p)||.|E_p^l| \quad p = 1,...,M$$

Aquí $|.|$ es la norma euclídea en $\mathbb{R}^2$ y $||.||$ la correspondiente norma matricial euclídea. Tomando ahora la norma euclídea en $\mathbb{R}^M$

$$\begin{aligned}(\sum_{p=0}^{M}|X_p^n|^2)^{1/2} &\leq \sup_{1\leq p\leq M}||\mathbf{B}^n(\theta_p)||(\sum_{p=0}^{M}|X_p^0|^2)^{1/2} \\ &+ k\sum_{l=0}^{n-1}\Big(\sup_{1\leq p\leq M}(||\mathbf{B}^{n-l+1}(\theta_p)||.||\mathbf{D}(\theta_p)||)(\sum_{p=0}^{M}|E_p^l|^2)^{1/2}\Big)\end{aligned} \tag{5.55}$$

Por otra parte tenemos,

$$|e_p^n| \leq |X_p^n| = \big(|e_p^n|^2 + \frac{|d_p^n|^2}{\beta_{h,p}}\big)^{1/2}$$

y también existe un constante $c_1 > 0$ independiente de h tal que

$$\beta_{h,p} \geq \beta_{h,1} \geq c^2\lambda_1 > 0$$

donde λ_1 es el mínimo valor propio del operador $-\frac{d^2}{dx^2}$. Más precisamente en el caso de la matriz $\mathbf{A}_h$ tenemos

$$\beta_{h,1} = \frac{2c^2}{h^2}(1-\cos(\pi h)) = \frac{2c^2}{h^2}\big(\frac{(\pi h)^2}{2} - \frac{(\pi h)^4}{4!} + \mathscr{O}(h^6)\big) > c^2\pi^2 > 0$$

para $0 < h < 1$. De donde existe una constante c_1 independiente de h tal que

$$|X_p^n| \leq c_1(|e_p^n|^2 + |d_p^n|^2)^{1/2}$$

y finalmente para la norma euclídea

$$||e^n|| = \big(h\sum_{p=0}^{M}|e_p^n|^2\big)^{1/2} \leq \frac{1}{2}\big(\sum_{p=0}^{M}|X_p^n|^2\big)^{1/2} \leq c_1(||e^n|| + ||d^n||)$$

donde hemos utilizado que para la norma euclídea de un vector v, si $(v_p)_p$ son las componentes en la base de vectores propios $\{w_p\}_p$

$$||v||^2 = (\sum_p v_p w_p, \sum_q v_q w_q) = (\sum_q (v_q)^2)(w_q, w_q) = \frac{1}{2}\sum_q (v_q)^2$$

Del mismo modo tenemos también

$$\frac{1}{2}\big(\sum_{p=0}^{M}|E_p^n|^2\big)^{1/2} \leq c_1(k||\varepsilon^n|| + ||\eta^n||)$$

Obtenemos finalmente utilizando (5.55)

$$\begin{aligned} ||e^n|| \leq c_1\Big(\big(\sup_{1\leq p\leq M}||\mathbf{B}^n(\theta_p)||\big)\big(||e^0|| + ||d^0||\big) \\ + k\sum_{l=0}^{n-1}\Big(\sup_{1\leq p\leq M}\big|||\mathbf{B}^{n-l+1}(\theta_p)||.||\mathbf{D}(\theta_p)||\big)\Big)\big(k||\varepsilon^l|| + ||\eta^l||\big)\Big) \end{aligned} \tag{5.56}$$

Definición 5.6 *Con las notaciones anteriores diremos que el método (5.51)-(5.52)-(5.53)-(5.54) es estable si existe una constante C independiente de p tal que*

$$||\mathbf{B}(\theta_p)|| \leq C \tag{5.57}$$

Vamos a estimar $||\mathbf{B}(\theta_p)||$ y $||\mathbf{D}(\theta_p)||$ donde la norma es la norma matricial subordinada a la norma euclídea en $\mathbb{R}2$. Para asegurar la estabilidad necesitamos tener $||\mathbf{B}(\theta_p)|| \leq C$ con la constante C independiente de p. Empezamos estudiando el radio espectral $\rho(\mathbf{B}(\theta_p))$ de $\mathbf{B}(\theta_p)$. Como

$$\rho^n(\mathbf{B}(\theta_p)) = \rho(\mathbf{B}^n(\theta_p)) \leq ||\mathbf{B}^n(\theta_p))||$$

una condición necesaria de estabilidad es $\rho(\mathbf{B}(\theta_p)) < 1$. Calculemos las raíces $\mu_\pm$ del polinomio

$$det[\mathbf{B}(\theta_p) - \mu I)] = 0$$

es decir las soluciones de

$$\mu^2 - (2-\alpha)\mu + 1 = 0$$

donde hemos puesto $\alpha = \theta_p^2$. Las soluciones son

$$\mu_\pm = \frac{2-\alpha \pm \sqrt{\alpha(\alpha-4)}}{2}$$

Si $\alpha < 4$ tenemos dos soluciones complejas conjugadas de producto igual a 1. De modo que

$$\rho(\mathbf{B}(\theta_p)) = 1$$

Si $\alpha = 4$ tenemos una raiz doble $\mu_+ = \mu_- = -1$. Si $\alpha \geq 4$ entonces al menos una de las raíces tiene módulo mayor que 1 y el radio espectral $\rho(\mathbf{B}(\theta_p)) > 1$ y el método no es estable.

Una condición necesaria de estabilidad es entonces $\alpha \leq 4$, es decir

$$\alpha = \theta_p^2 = k^2\beta_{h,p} = \frac{c^2k^2}{h^2}\beta_p \leq 4 \quad \forall p = 1,...,M$$

Con $\beta_M = 2(1-\cos(hM\pi) \leq 4$ obtenemos la condición necesaria de estabilidad

$$\frac{ck}{h} \leq 1$$

Para demostrar que la condición de estabilidad anterior es suficiente buscaremos una matriz $\mathbf{G}(\theta_p)$ tal que

$$\mathbf{G}(\theta_p)\mathbf{B}(\theta_p)\mathbf{G}^{-1}(\theta_p)$$

sea una matriz normal. Recordemos que una matriz normal es aquella que conmuta con su adjunta (traspuesta en el caso de matrices reales). Recordemos también para una matriz normal la norma euclídea de la matriz coincide con su radio espectral.

Poniendo

$$\begin{aligned}\mathbf{B}(\theta_p) &= \mathbf{G}^{-1}(\theta_p)\mathbf{G}(\theta_p)\mathbf{B}(\theta_p)\mathbf{G}^{-1}(\theta_p)\mathbf{G}(\theta_p)\\ \mathbf{B}^n(\theta_p) &= \mathbf{G}^{-1}(\theta_p)\big(\mathbf{G}(\theta_p)\mathbf{B}(\theta_p)\mathbf{G}^{-1}(\theta_p)\big)^n\mathbf{G}(\theta_p)\end{aligned}$$

resulta

$$\begin{aligned}||\mathbf{B}^n(\theta_p)|| &\leq ||\mathbf{G}^{-1}(\theta_p)||.||(\mathbf{G}(\theta_p)\mathbf{B}(\theta_p)\mathbf{G}^{-1}(\theta_p))^n||.||\mathbf{G}(\theta_p)||\\ &= ||\mathbf{G}^{-1}(\theta_p)||.||\mathbf{G}(\theta_p)||\rho^n(\mathbf{G}(\theta_p)\mathbf{B}(\theta_p)\mathbf{G}^{-1}(\theta_p))\\ &= ||\mathbf{G}^{-1}(\theta_p)||.||\mathbf{G}(\theta_p)||\rho(\mathbf{B}^n(\theta_p))\end{aligned}$$

Lema 5.1 *Una matriz real* 2×2, $\mathbf{A} = (a_{i,j})_{1\leq i,j\leq 2}$ *es normal si y solo si o bien la matriz es simétrica o bien se verifica* $a_{11} = a_{22}$, $a_{12} = -a_{21}$

Demostración. A es normal si y solo si $A^tA = AA^t$. Identificando términos

$$a_{21}^2 = a_{12}^2$$
$$(a_{12} - a_{21})(a_{11} - a_{22}) = 0$$

de donde $a_{12} = \pm a_{21}$. Si $a_{12} = a_{21}$ entonces la matriz es simétrica. Si $a_{12} = -a_{21}$ entonces $a_{11} = a_{22}$. ■

Para encontrar $\mathbf{G}(\theta_p)$ podemos buscar una matriz de la forma

$$\mathbf{G}(\theta_p) = \begin{bmatrix} 1 & 0 \\ s(\theta_p) & t(\theta_p) \end{bmatrix}$$

Tenemos

$$\mathbf{G}^{-1}(\theta_p) = \begin{bmatrix} 1 & 0 \\ -\frac{s(\theta_p)}{t(\theta_p)} & \frac{1}{t(\theta_p)} \end{bmatrix}$$

Calculando $\mathbf{G}\mathbf{B}\mathbf{G}^{-1}$ e imponiendo las condiciones del lema se obtiene

$$\mathbf{G}(\theta_p) = \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{\sqrt{1-\frac{\theta_p^2}{4}}} \end{bmatrix}$$

y

$$\mathbf{G}^{-1}(\theta_p) = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{1-\frac{\theta_p^2}{4}} \end{bmatrix}$$

Por otra parte como tenemos para una matriz que la norma euclídea está mayorada por la norma de Frobenius

$$||\mathbf{G}||^2 \leq 1 + \frac{1}{1-\frac{\theta_p^2}{4}} = \frac{8-\theta_p^2}{4-\theta_p^2}$$
$$||\mathbf{G}^{-1}||^2 \leq 1 + 1 - \frac{\theta_p^2}{4} = \frac{8-\theta_p^2}{4}$$

Finalmente como $\rho(\mathbf{B}(\theta_p)) \leq 1$ resulta

$$||\mathbf{B}^n(\theta_p)|| \leq ||\mathbf{G}^{-1}(\theta_p)||.||\mathbf{G}(\theta_p)|| \leq \frac{8-\theta_p^2}{2\sqrt{4-\theta_p^2}}$$

La condición suficiente de estabilidad será $0 < \theta_p^2 < 4$. Si esta se cumple, existe ξ tal que $0 < \xi < 1$ y $\theta_p^2 = 4(1-\xi)$. Entonces

$$||\mathbf{B}^n(\theta_p)|| \leq \frac{1+\xi}{\sqrt{\xi}} = C(\xi) \tag{5.58}$$

Eligiendo el valor más desfavorable para p que es $p = M$, el valor de ξ es $\xi = 1 - \frac{c^2k^2}{2h^2}(1-\cos(M\pi h))$ y en función de $\mu = ck/h \leq 1$ tendremos $1-\mu^2 < \xi < 1$.

Para la convergencia necesitamos estimar $||\mathbf{D}(\theta_p)||$. Tendremos

$$||\mathbf{D}(\theta_p)||^2 \leq 1+1+\frac{\theta_p^2}{4} = 2+\frac{\theta_p^2}{4} \leq 3 \tag{5.59}$$

$$||\mathbf{D}(\theta_p)|| \leq \sqrt{3} \tag{5.60}$$

donde de nuevo hemos tenido en cuenta la condición $\theta_p^2 < 4$.

Estamos en condiciones de enunciar el siguiente teorema de convergencia:

Teorema 5.3 *Con la condición de estabilidad*

$$\frac{ck}{h} \leq 1$$

existe una constante $C > 0$ independiente de h y de k tal que el error del método (5.51)-(5.52)-(5.53)-(5.54) verifica

$$||e^n|| \leq C\Big(||e^0|| + ||d^0|| + k\sum_{l=0}^{n-1}(k||\varepsilon^l|| + ||\eta^l||)\Big) \tag{5.61}$$

y en consecuencia el método es convergente de orden $\mathscr{O}(k^2)+\mathscr{O}(h^2)$.

Demostración. La estimación (5.61) se deduce de forma inmediata de (5.56) y de las estimaciones (5.58) y (5.60). La convergencia y el orden de convergencia se deduce de (5.61) y del error de consistencia poniendo

$$\begin{aligned} k\sum_{l=0}^{n-1}(k||\varepsilon^l|| + ||\eta^l||) &\leq k(k\sup_l||\varepsilon^l|| + \sup_l||\eta^l||)\sum_{l=0}^{n-1}1 \\ &\leq T(k\sup_{l=0}^{n-1}||\varepsilon^l|| + \sup_{l=0}^{n-1}||\eta^l||) = \mathscr{O}(k^2)+\mathscr{O}(h^2) \end{aligned}$$

siendo $T = nk$ ■

Ejercicio

1. Aplicar el método (5.31)-(5.32)-(5.33)-(5.34) en el intervalo $x \in [-1,1]$ y para un tiempo entre $t \in [0,1]$ con las condiciones de contorno ,

$$u(-1,t) = u(1,t) = 0 \quad t \in [0,1]$$

y la condiciones iniciales

$$f(x) = e^{(-x^2/0.01)} \quad x \in [-1,1]$$

y $g(x) = 0 \quad x \in [-1,1]$. y los parámetros $c^2 = 1$, $h = 0.001$ y $k = 0.001$. De modo que $\mu = 1$.

2. Con la misma solución inicial que en el apartado anterior pero con los parámetros $c^2 = 5$, $h = 0.001$ y $k = 0.001$. De modo que $\mu = 0.5$.
3. Interpretar los resultados en los dos casos anteriores. Particularmente observar que las ondas al rebotar se invierten en la frontera.

Resultados numéricos

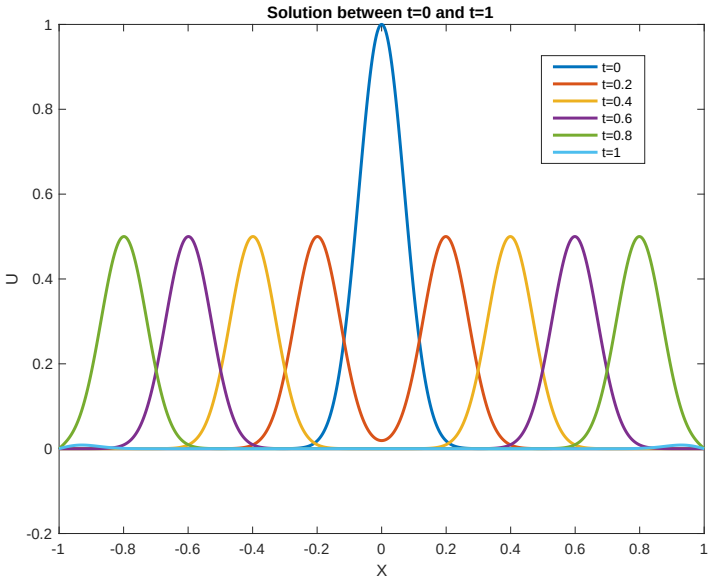


Figura 5.21 Solución numérica en $t \in [0, 1]$

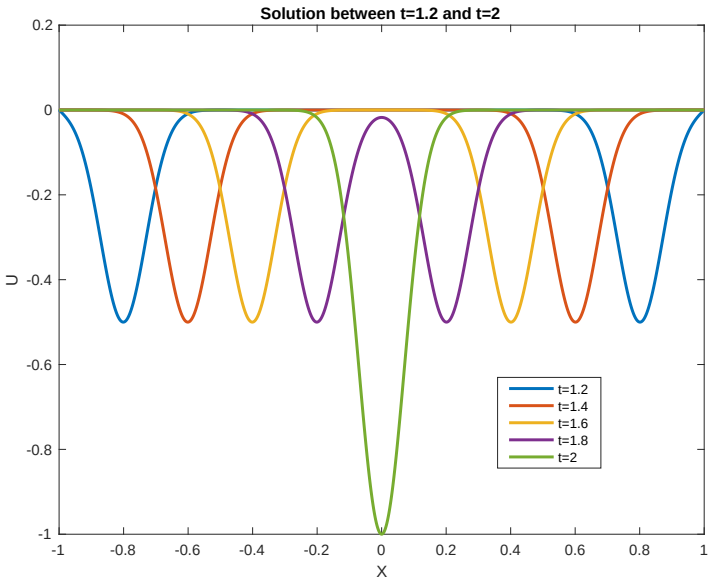


Figura 5.22 Solución numérica en $t \in [1.2, 2]$ para $c^2 = 1,\ \mu = 1$

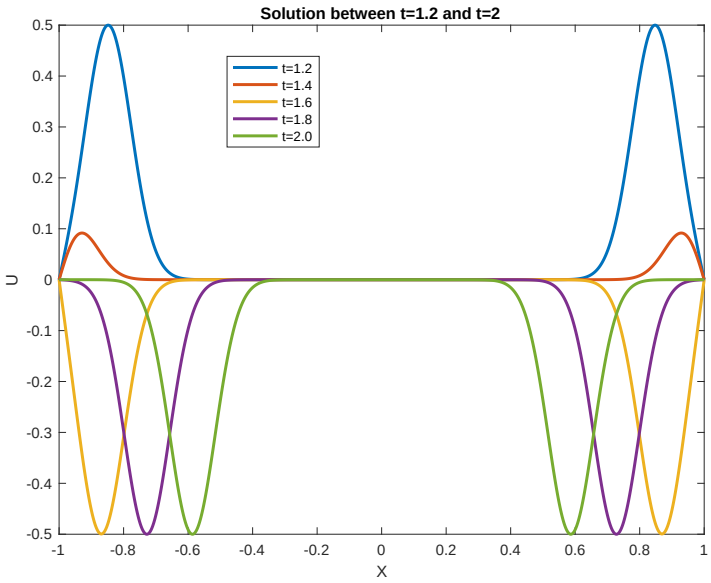


Figura 5.23 Solución numérica en $t \in [1.2, 2]$ para $c^2 = 0.5,\ \mu = \sqrt{0.5}$

Ejercicio

En este ejercicio se estudiará como ejemplo de ecuación hiperbólica de segundo orden en dimensión 2 la ecuación de ondas. Sea Ω un conjunto abierto y acotado de R^2 y Γ su frontera. En este ejercicio se supondrá que $\Omega = [0,1] \times [0,1]$.

$$\frac{\partial^2 u}{\partial t^2} = c^2 (\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}) \quad x \in \Omega \quad t \in (0,T) \tag{5.62}$$

$$u(x,t) = u(x,t) = 0 \quad x \in \Gamma \quad \text{para } t \in [0,T] \tag{5.63}$$

$$u(x,0) = f(x) \quad x \in \Omega \tag{5.64}$$

$$\frac{\partial u}{\partial t}(x,0) = g(x) \quad x \in \Omega \tag{5.65}$$

Con las mismas notaciones que en la sección (3.2) considerar el esquema en diferencias finitas siguiente

$$\bar{\partial}_t(\partial_t) u^n = c^2 \mathbf{A}_h u^n \quad \text{en } \Omega_h,\ n \geq 0 \tag{5.66}$$

$$u^n = 0 \quad \text{en } \Gamma_h,\ n \geq 0 \tag{5.67}$$

$$u^0(x) = f(x) \quad \forall x \in \Omega_h \tag{5.68}$$

$$\frac{u^1(x) - u^{-1}(x)}{2k} = g(x) \quad \forall x \in \Omega_h \tag{5.69}$$

$$\tag{5.70}$$

Analizar la consistencia, estabilidad y convergencia el método anterior.

Indicación

Seguir los pasos del caso unidimensional, utilizando la descomposición espectral del operador en diferencias finitas $\mathbf{A}_h$

5.3. Ecuaciones hiperbólicas no lineales

5.3.1. Introducción

Consideramos en esta sección la ecuación no lineal escalar y el problema de valor inicial asociado

$$\frac{\partial u}{\partial t} + \frac{\partial f(u)}{\partial x} = 0 \quad x \in \mathbb{R}, \quad t > 0 \tag{5.71}$$

$$u(x,0) = u_0(x) \quad x \in \mathbb{R} \tag{5.72}$$

donde f es una función no lineal de u.

En general, aunque la solución inicial $u_0(x)$ sea una función regular, el problema (5.71)-(5.72) puede no tener soluciones clásicas $u \in C^1(Q_T) \cup C^0(\bar{Q}_T)$ donde $Q_T = \mathbb{R} \times (0,T)$ para todo $T > 0$. En efecto, las discontinuidades pueden aparecer para un tiempo T finito. La ecuación (5.71) la podemos escribir de la forma:

$$\frac{\partial u}{\partial t} + a(u)\frac{\partial u}{\partial x} = 0 \quad x \in \mathbb{R}, \quad t > 0$$

donde $a(u) = f'(u)$.

Consideremos las curvas características

$$\begin{aligned} &x : t \to x(t) \\ &\frac{dx}{dt} = a(u(x,t)) \end{aligned}$$

Si el valor de la solución $u(x,t)$ sobre las características es $t \to v(t) = u(x(t),t)$ resulta,

$$\begin{aligned} \frac{dv}{dt} &= \frac{\partial u}{\partial t} + \frac{\partial u}{\partial x}\frac{dx}{dt} \\ &= \frac{\partial u}{\partial t} + a(u)\frac{\partial u}{\partial x} = 0 \end{aligned}$$

es decir, la solución es constante sobre las características. Volviendo a la ecuación de las características

$$\frac{dx}{dt} = a(u(x,t))$$

si buscamos la característica que pasa por el punto $(x_i, 0)$, es decir, $x(0) = x_i$, tendremos que resolver el problema de valor inicial

$$\begin{aligned} &\frac{dx}{dt} = a(u(x,t)) = a(u(x_0)) \\ &x(0) = x_i \end{aligned}$$

puesto que u es constante sobre la característica. Por tanto las características son rectas. La característica que pasa por el punto $(x_i, 0)$ tendrá por ecuación

$$x = a(u_0(x_i))t + x_i$$

La pendiente de la recta es

$$m = \frac{1}{a(u_0(x_i))}$$

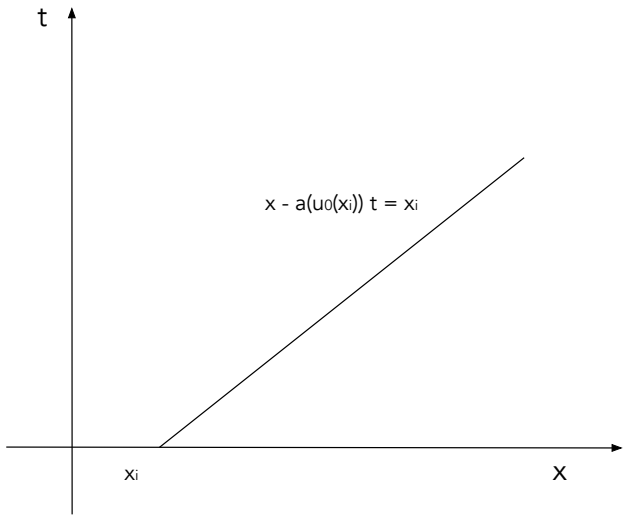


Figura 5.24 Característica de una hiperbólica no lineal

Supongamos que a_0, u_0, x_1, x_2 sean tales que $x_1 < x_2$ y $a(u_0(x_1)) > a(u_0(x_2))$, por tanto

$$m_1 = \frac{1}{a(u_0(x_1))} < \frac{1}{a(u_0(x_2))} = m_2$$

A lo largo de la característica nacida del punto $(x_i, 0)$ la solución será $u_0(x_i)$. Por tanto en el punto de corte P de las dos caracterśticas la solución no puede ser continua. Este hecho es por otra parte independiente de la regularidad de las funciones u_0 y f. Consideremos el caso de la ecuación de Burgers que corresponde a $f(u) = u^2/2$

$$\frac{\partial u}{\partial t} + \frac{1}{2}\frac{\partial u^2}{\partial x} = 0 \quad x \in \mathbb{R}, \quad t > 0 \tag{5.73}$$

que podemos escribir

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = 0 \quad x \in \mathbb{R}, \quad t > 0 \tag{5.74}$$

Supongamos que la condición inicial

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

es tal que $x \to u_0(x)$ es decreciente, por tanto si $x_1 < x_2$ entonces $u_0(x_1) > u_0(x_2)$ y

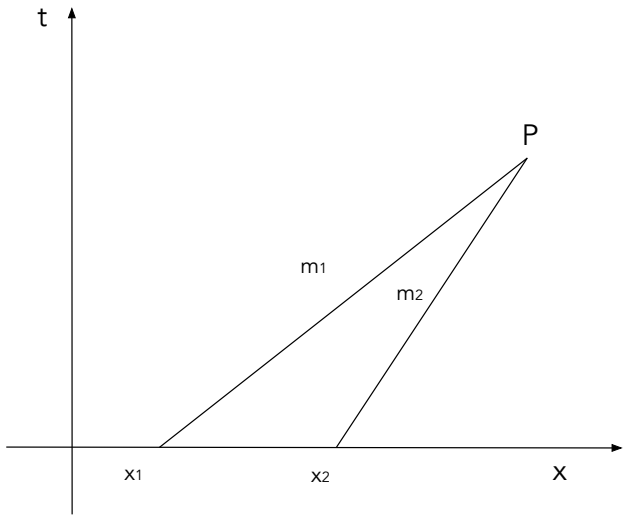


Figura 5.25 Generación de una discontinuidad

$$m_1 = \frac{1}{u_0(x_1)} < \frac{1}{u_0(x_2)} = m_2$$

y aparecerá una discontinuidad en tiempo finito.

Ejercicio

Demostrar que la solución de

$$\frac{\partial u}{\partial t} + \frac{1}{2}\frac{\partial u^2}{\partial x} = 0 \quad x \in \mathbb{R}, \quad t > 0$$
$$u(x,0) = u_0(x)$$

donde $u_0'(x)$ es negativa en algún punto, entonces la solución presentará una discontinuidad en el instante

$$T_b = \frac{-1}{\text{mín}_{x\in\mathbb{R}}\, u_0'(x)}$$

Solución

Se trata de ver en qué instante $\partial u/\partial x$ se hace ∞ sobre la característica. Derivando la ecuación (5.74) respecto a x

$$\frac{\partial}{\partial t}(\frac{\partial u}{\partial x})+(\frac{\partial u}{\partial x})^2+u\frac{\partial^2 u}{\partial x^2}=0 \tag{5.75}$$

Llamemos

$$q(t)=\frac{\partial u}{\partial x}(x(t),t))$$

e introduciendo la derivada total

$$\begin{aligned}\frac{dq}{dt}&=\frac{\partial}{\partial t}(\frac{\partial u}{\partial x})+\frac{\partial^2 u}{\partial x^2}\frac{dx}{dt}\\&=\frac{\partial}{\partial t}(\frac{\partial u}{\partial x})+u\frac{\partial^2 u}{\partial x^2}\end{aligned}$$

sustituyendo en (5.75)

$$\begin{aligned}&\frac{dq}{dt}+q^2=0\\&q(0)=\frac{\partial u}{\partial x}(x(0),0)\end{aligned}$$

Integrando

$$\begin{aligned}&\frac{dq}{q^2}=-dt\\&-\frac{1}{q}+C=-t\end{aligned}$$

donde C es la constante de integración cuyo valor es $C=1/q(0)$ de donde

$$q(t)=\frac{q(0)}{1+tq(0)}$$

Si $q(0)$ es negativo, es decir,

$$\frac{\partial u}{\partial x}(x(0),0)=u_0'(x(0))<0$$

para $t_m=-1/q(0)$, $q(t)$ se hace infinito. Es decir el primer instante en el que aparecerá la discontinuidad será

$$t_b=\min t_m=\frac{-1}{\min_{x\in\mathbb{R}} u_0'(x)}$$

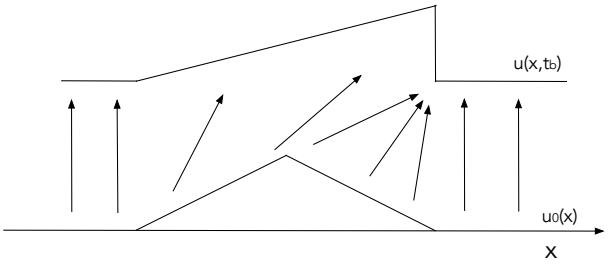


Figura 5.26 Ejemplo de generación de una discontinuidad en la ecuación de Burgers

5.3.2. Soluciones débiles de una ley de conservación

Según hemos visto aunque la solución inicial u_0 sea una función regular, el problema (5.71)-(5.72) no admite siempre una solución clásica $u \in C^1(Q_T) \cap C(\bar{Q}_T)$, con $Q_T = \mathbb{R} \times (0,T)$ para todo tiempo $T > 0$. En efecto, hemos visto que pueden aparecer discontinuidades al cabo de un tiempo finito $t_{\text{mín}} < \infty$.

Conviene pues introducir la noción de solución débil o generalizada del problema (5.71)-(5.72). Una idea es formular directamente la ley de conservación en su forma integral de la que proviene (5.71). Otra idea consiste en considerar la solución (5.71)-(5.72) como el límite de soluciones de un problema con viscosidad cuando esta viscosidad tiene como límite cero. Una tercera posibilidad es la siguiente:

Observemos que si u es solución clásica de (5.71)-(5.72) y φ es una función de clase C^1 y soporte compacto en $\mathbb{R} \times [0,\infty)$, es decir, $\varphi \in C_0^1(\mathbb{R} \times [0,\infty))$ tendremos multiplicando ambos miembros de la ecuación (5.71) por φ e integrando por partes

$$\begin{aligned}&\int_0^\infty \int_{-\infty}^\infty (\frac{\partial u}{\partial t} + \frac{\partial f(u)}{\partial x})\varphi \, dx\, dt \\ &= -\int_0^\infty \int_{-\infty}^\infty (u\frac{\partial \varphi}{\partial t} + f(u)\frac{\partial \varphi}{\partial x})\, dx\, dt - \int_{-\infty}^\infty u(x,0)\varphi(x,0)\, dx = 0\end{aligned}$$

así que la solución u de (5.71)-(5.72) verifica

$$\int_0^\infty \int_{-\infty}^\infty (u\frac{\partial \varphi}{\partial t} + f(u)\frac{\partial \varphi}{\partial x})\, dx\, dt + \int_{-\infty}^\infty u(x,0)\varphi(x,0)\, dx = 0 \tag{5.76}$$

para todo $\varphi \in C_0^1(\mathbb{R} \times [0,\infty))$

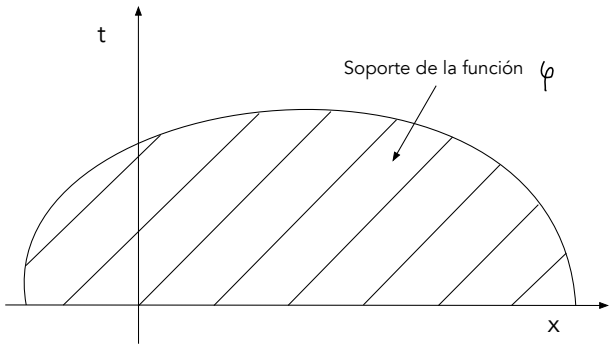


Figura 5.27 Soporte de la función φ

Observemos que (5.76) tiene sentido siempre que $u \in L^\infty_{loc}(\mathbb{R}\times[0,\infty))$. Daremos la siguiente definición:

Definición 5.7 *Diremos que $u \in L^\infty_{loc}(\mathbb{R}\times[0,\infty))$ es una solución débil del problema (5.71)-(5.72) si $u(x,t) \in \Omega$ c.t.p. y verifica (5.76) para toda función $\varphi \in C^1_0(\mathbb{R}\times[0,\infty))$.*

En la definición anterior Ω es el conjunto de estados. En el caso $d = 1$, $\Omega = \mathbb{R}$ o $\Omega = \mathbb{R}^+$.

Por construcción, una solución clásica de (5.71)-(5.72) es también una solución débil. Recíprocamente, si u es solución de (5.76) y verifica

$$u \in C^1(\mathbb{R}\times(0,\infty)) \cap C^0(\mathbb{R}\times[0,\infty))$$

es de hecho una solución clásica.

Si en (5.76) tomamos $\varphi \in C^\infty_0(\mathbb{R}\times(0,\infty))$, resulta

$$\int_0^\infty \int_{-\infty}^\infty (u\frac{\partial \varphi}{\partial t} + f(u)\frac{\partial \varphi}{\partial x})\,dx\,dt = 0 \quad \forall \varphi \in C^\infty_0(\mathbb{R}\times(0,\infty)) \tag{5.77}$$

Si u verifica (5.77) decimos que u verifica la ecuación en derivadas parciales

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

en el sentido de las distribuciones.

5.3.3. Soluciones de clase C^1 "a trozos"

Consideremos ahora las posibles soluciones de (5.76) que son de clase C^1, excepto en determinadas líneas de discontinuidad $x = \psi(t)$ del plano $x-t$. Sea pues

Γ una línea de discontinuidad de clase C^1, en la cual u admite una discontinuidad de primera especie situada en el semi-plano $\mathbb{R}\times(0,\infty)$

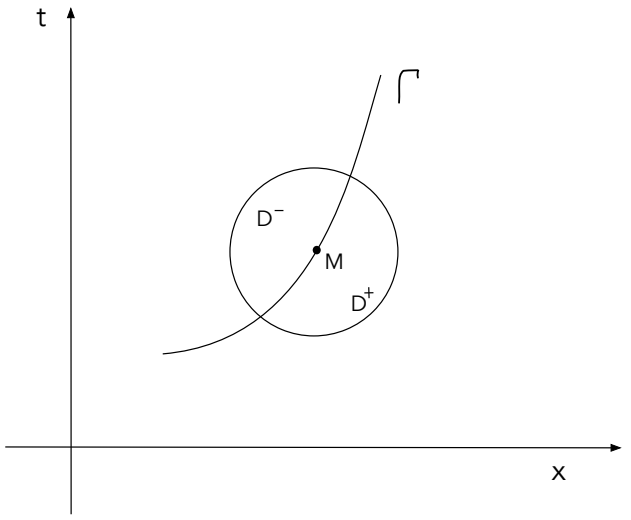


Figura 5.28 Líneas de discontinuidad

Sea M un punto de Γ y D un disco de centro M contenido en $\mathbb{R}\times(0,\infty)$. Designemos D^+ (resp. D^-) la región del disco D situada a la derecha (resp. a la izquierda) de la curva Γ y llamemos:

$u^+ =$ límite por la derecha de u sobre Γ

$u^- =$ límite por la izquierda de u sobre Γ

Sea $\varphi\in C_0^\infty(D)$ (función de clase C^∞ y soporte compacto en D). Si u es solución débil de (5.71)-(5.72), u verifica (5.76) y por lo tanto

$$\begin{aligned}0&=\int\int_D(u\frac{\partial\varphi}{\partial t}+f(u)\frac{\partial\varphi}{\partial x})\,dxdt\\&=\int\int_{D^+}(u\frac{\partial\varphi}{\partial t}+f(u)\frac{\partial\varphi}{\partial x})\,dxdt\\&+\int\int_{D^-}(u\frac{\partial\varphi}{\partial t}+f(u)\frac{\partial\varphi}{\partial x})\,dxdt\end{aligned}$$

como las funciones u y φ son regulares en D^+ se obtiene por aplicación de la fórmula de Green,

$$\begin{aligned}&\int\int_{D^+}(u\frac{\partial\varphi}{\partial t}+f(u)\frac{\partial\varphi}{\partial x})\,dx\,dt\\ =&-\int\int_{D^+}(\frac{\partial u}{\partial t}+\frac{\partial f(u)}{\partial x})\varphi\,dx\,dt\\ &+\int_{\Gamma\cap D^+}(u^+n_t^+ + f(u^+)n_x^+)\varphi\,d\gamma\end{aligned}$$

donde $\mathbf{n}^+=(n_x^+,n_t^+)$ es la normal a Γ dirigida hacia el exterior de D^+.

Análogamente se obtiene

$$\begin{aligned}&\int\int_{D^-}(u\frac{\partial\varphi}{\partial t}+f(u)\frac{\partial\varphi}{\partial x})\,dx\,dt\\ =&-\int\int_{D^-}(\frac{\partial u}{\partial t}+\frac{\partial f(u)}{\partial x})\varphi\,dx\,dt\\ &+\int_{\Gamma\cap D^-}(u^-n_t^- + f(u^-)n_x^-)\varphi\,d\gamma\end{aligned}$$

donde $\mathbf{n}^-=-\mathbf{n}^+$ es la normal a Γ dirigida hacia el exterior de D^-.

Deducimos que, para toda función $\varphi\in C_0^\infty(D)$ tenemos

$$\begin{aligned}0=&-\int\int_{D^+}(\frac{\partial u}{\partial t}+\frac{\partial f(u)}{\partial x})\varphi\,dx\,dt\\ &-\int\int_{D^-}(\frac{\partial u}{\partial t}+\frac{\partial f(u)}{\partial x})\varphi\,dx\,dt\\ &+\int_{\Gamma\cap D}\big((u^+-u^-)n_t^+ + (f(u^+)-f(u^-)n_x^+)\big)\varphi\,d\gamma\end{aligned}$$

en particular tomando φ con soporte, sop $\varphi\subset D^+$ (resp. sop $\varphi\subset D^-$ obtenemos

$$\frac{\partial u}{\partial t}+\frac{\partial f(u)}{\partial x}=0\quad\text{en } D^+(\text{resp. } D^-)$$

y donde las derivadas en la anterior expresión son derivadas en el sentido clásico. Finalmente sustituyendo en la expresión anterior,

$$\int_{\Gamma\cap D}\big((u^+-u^-)n_t^+ + (f(u^+)-f(u^-)n_x^+\big)\varphi\,d\gamma=0 \tag{5.78}$$

para toda función $\varphi\in C_0^\infty(D)$.

Vamos a interpretar esta condición:

Sea

$$\begin{aligned}&\mathbb{R}\to\mathbb{R}^2\\ &t\to(x(t),t)\end{aligned}$$

la ecuación de la curva Γ en el plano $x-t$. El campo tangente a la curva Γ viene dado por $\mathbf{t}=(\frac{dx}{dt},1)^t$. El vector normal unitario $\mathbf{n}$ será

$$\mathbf{n}^+ = \mathbf{n} = (n_x, n_t)^t = \frac{(-1, \frac{dx}{dt})^t}{\sqrt{1+(\frac{dx}{dt})^2}}$$

de manera que introduciendo los saltos, $[u] = u^+ - u^-$ y $[f(u)] = f(u^+) - f(u^-)$ podemos escribir (5.78)

$$\int_{\Gamma \cap D} ([u]n_t + [f(u)]n_x)\varphi \, d\gamma = 0 \quad \forall \varphi \in C_0^\infty(D)$$

es decir

$$[u]n_t + [f(u)]n_x = 0 \quad \text{sobre } \Gamma$$

y sustitutendo los valores de n_x y n_t

$$[u]\frac{dx}{dt} - [f(u)] = 0 \quad \text{sobre } \Gamma$$

Introduciendo la velocidad de propagación de la discontinuidad, $s = \frac{dx}{dt}$ obtenemos la siguiente condición para toda solución u sobre las líneas de discontinuidad:

$$s[u] = [f(u)] \tag{5.79}$$

llamada condición de Rankine-Hugoniot, que es una condición necesaria que debe satisfacer toda solución débil de (5.71)-(5.72) sobre las discontinuidades.

Tenemos también un recíproco de la conclusión anterior, de hecho podemos enunciar el siguiente teorema:

Teorema 5.4 *Una función $u = u(x,t)$, C^1 "a trozos" verifica (5.77), es decir verifica la Ecuación en Derivadas Parciales (5.71) en el sentido de las distribuciones si y solo si las dos condiciones siguientes se satisfacen:*

1. *u es solución clásica de (5.71) en los dominios de regularidad de u.*
2. *u verifica la condición de Rankine-Hugoniot (5.79) sobre las líneas de discontinuidad.*

Demostración. Ya se ha demostrado que las condiciones (1) y (2) del enunciado del teorema son necesarias. Recíprocamente, si u de clase C^1 "a trozos" verifica (1) y (2) demostraremos que u es solución de (5.77).

Sea $\varphi \in C_0^\infty(\mathbb{R} \times (0,\infty)$.

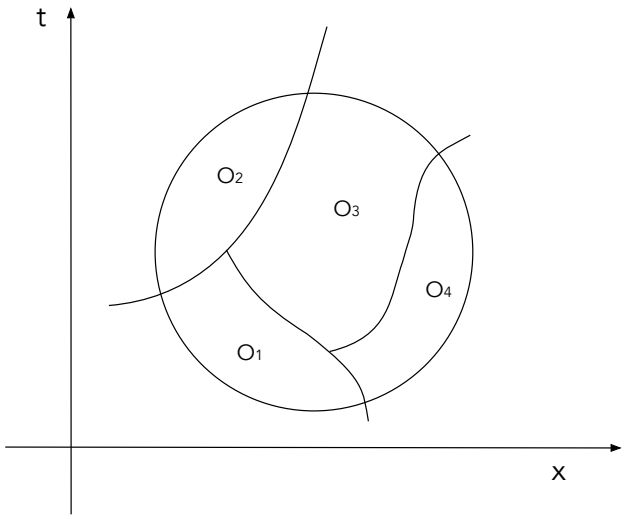


Figura 5.29 Dominio y líneas de discontinuidad

Descomponemos el soporte de φ en partes $\mathscr{O}_i$ cuyas fronteras son las líneas de discontinuidad de u. Tendremos,

$$\begin{aligned}&\int_0^\infty \int_{-\infty}^\infty (u\frac{\partial \varphi}{\partial t} + f(u)\frac{\partial \varphi}{\partial x})\,dx\,dt\\ &= \sum_i \int\int_{\mathscr{O}_i} (u\frac{\partial \varphi}{\partial t} + f(u)\frac{\partial \varphi}{\partial x})\,dx\,dt\end{aligned}$$

utilizando la fórmula de Green

$$\begin{aligned}&\int_0^\infty \int_{-\infty}^\infty (u\frac{\partial \varphi}{\partial t} + f(u)\frac{\partial \varphi}{\partial x})\,dx\,dt\\ &= -\sum_i \int\int_{\mathscr{O}_i} (\frac{\partial u}{\partial t} + \frac{\partial f(u)}{\partial x})\varphi\,dx\,dt\\ &\quad + \sum_i \int_{\partial \mathscr{O}_i} (u^i n_t^i + f(u^i) n_x^i)\varphi\,d\gamma\end{aligned}$$

donde u^i es el límite de u sobre $\partial\mathscr{O}_i$ para valores interiores de $\mathscr{O}_i$ y $\mathbf{n}^i = (n_x^i, n_t^i)^t$ es la normal exterior sobre $\partial\mathscr{O}_i$ de $\mathscr{O}_i$. Debido a la condición (1) se anulan las integrales en $\mathscr{O}_i$ y debido a la condición (2) se anulan las integrales sobre las fronteras de $\mathscr{O}_i$. ■

Ejemplos

Veamos algunos ejemplos y las correspondientes soluciones débiles de la ecuación de Burgers para distintas condiciones iniciales.

Ejemplo 1

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = 0 \quad x \in \mathbb{R},\, t > 0$$

con la condición inicial

$$u(x,0) = u_0(x) = \begin{cases} 1 & \text{si} \quad x \leq 0 \\ 1-x & \text{si} \quad 0 \leq x \leq 1 \\ 0 & \text{si} \quad 1 \geq x \end{cases}$$

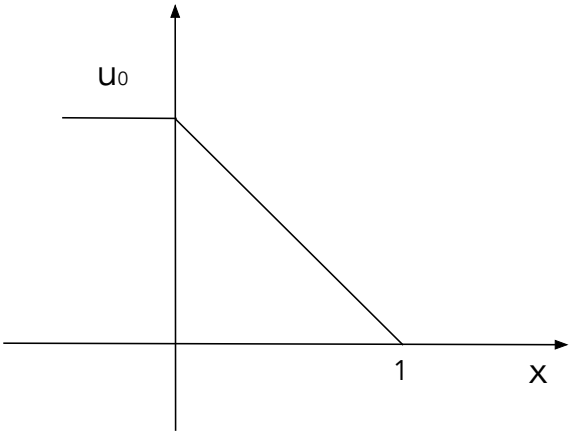


Figura 5.30 Condición inicial

Como u_0 es decreciente en una parte de $\mathbb{R}$, la solución se hará discontinua en un tiempo finito, concretamente en

$$t_M = \frac{-1}{\text{mín}_{x\in\mathbb{R}}\, u_0'(x)} = \frac{-1}{-1} = 1$$

En una discontinuidad se deberá cumplir la condición de Rankine-Hugoniot. En este caso $f(u) = \frac{1}{2}u^2$ y la condición de Rankine-Hugoniot es

$$s(u^+ - u^-) = \frac{1}{2}((u^+)^2 - (u^-)^2)$$

Para construir una solución utilizaremos el método de las características:

La característica que pasa por $(x_0, 0)$ será la solución de

$$\begin{aligned} \frac{dx}{dt} &= u_0(x_0) \\ x(0) &= x_0 \end{aligned}$$

donde

$$u_0(x_0) = \begin{cases} 1 & \text{si} \quad x_0 \leq 0 \\ 1 - x_0 & \text{si} \quad 0 \leq x_0 \leq 1 \\ 0 & \text{si} \quad 1 \geq x_0 \end{cases}$$

Las características son las rectas

$$x = x_0 + u(x_0)t = \begin{cases} x(t) = x_0 + t & \text{si} \quad x_0 \leq 0 \\ x(t) = x_0 + (1 - x_0)t & \text{si} \quad 0 \leq x_0 \leq 1 \\ x(t) = x_0 & \text{si} \quad 1 \geq x_0 \end{cases}$$

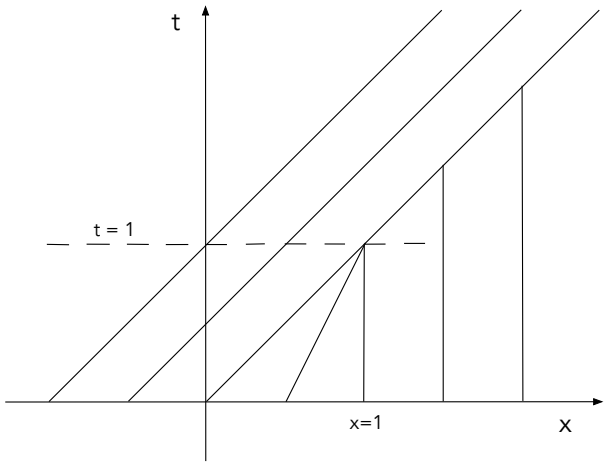


Figura 5.31 Rectas Características

Para $t < 1$, las características no se cortan y tenemos una solución continua.

Para calcular la solución $u(x,t)$ en un punto genérico (x,t), buscaremos el pie de la característica que pasa por este punto (x,t), es decir despejamos x_0 de la ecuación de la característica:

$$x = x(t) = \begin{cases} x = x_0 + t \quad \Rightarrow x_0 = x - t \quad \text{si} \quad x - t \leq 0 \Rightarrow x \leq t \\ x = x_0 + (1 - x_0)t \quad \Rightarrow x_0 = \dfrac{x-t}{1-t} \quad \text{si} \quad 0 \leq \dfrac{x-t}{1-t} \Rightarrow t \leq x \leq 1 \\ x = x_0 \quad \Rightarrow x_0 = x \quad \text{si} \quad 1 \geq x_0 \Rightarrow 1 \leq x \end{cases}$$

de donde la solución $u(x,t)$ vendrá dada por

$$u(x,t) = u_0(x_0) = \begin{cases} u_0(x-t) \quad \text{si} \quad x \leq t \\ u_0\left(\dfrac{x-t}{1-t}\right) \quad \text{si} \quad t \leq x \leq 1 \\ u_0(x) \quad \text{si} \quad 1 \geq x \end{cases}$$

es decir,

$$u(x,t) = u_0(x_0) = \begin{cases} 1 \quad \text{si} \quad x \leq t \\ 1 - \dfrac{x-t}{1-t} = \dfrac{1-x}{1-t} \quad \text{si} \quad t \leq x \leq 1 \\ 0 \quad \text{si} \quad 1 \geq x \end{cases}$$

Si representamos la solución para distintos tiempos t

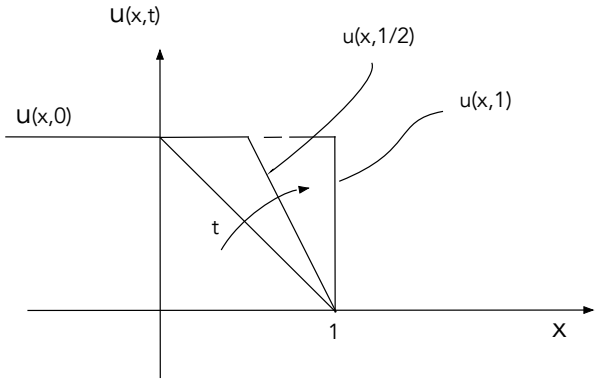


Figura 5.32 Solución para distintos valores de t

Por ejemplo, para $t = 1/2$

$$u(x,\frac{1}{2}) = \begin{cases} 1 & \text{si} \quad x \leq \dfrac{1}{2} \\ 2(1-x) & \text{si} \quad \dfrac{1}{2} \leq x \leq 1 \\ 0 & \text{si} \quad 1 \geq x \end{cases}$$

Para $t = 1$, aparece una discontinuidad. La condición de Rankine-Hugoniot determina como se propagará esta discontinuidad. La velocidad s de propagación de la discontinuidad vendrá dada por

$$s = \frac{1}{2}\frac{(u^+)^2 - (u^-)^2}{u^+ - u^-} = \frac{1}{2}(u^+ + u^-) = \frac{1}{2}$$

La línea de discontinuidad será

$$\frac{dx}{dt} = s = \frac{1}{2} \Rightarrow x = \frac{1}{2}t + C$$

como para $t = 1, x = 1$, resulta $1 = \frac{1}{2} + C$, de donde la constante $C = \frac{1}{2}$ y finalmente tendremos

$$x = \frac{1}{2}t + \frac{1}{2}$$

o bien

$$t = 2x - 1$$

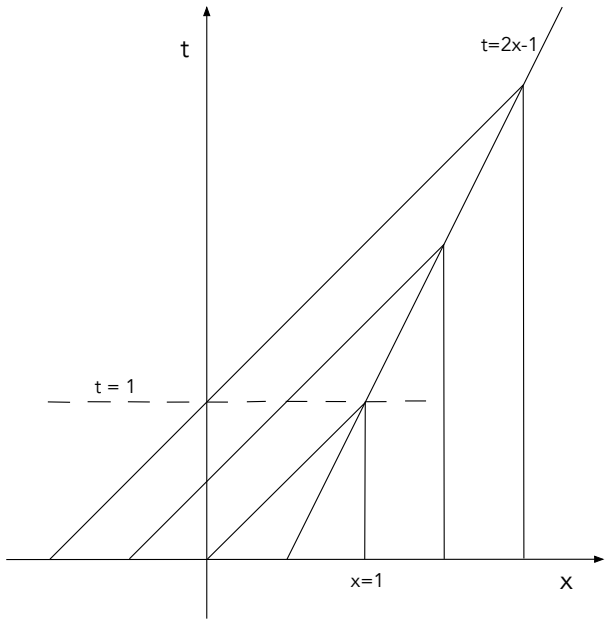


Figura 5.33 Ejemplo de línea de discontinuidad en la ecuación de Burgers

Ejemplo 2

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = 0 \quad x \in \mathbb{R},\, t > 0$$

con la condición inicial

$$u(x,0) = u_0(x) = \begin{cases} a & \text{si} \quad x < 0 \\ b & \text{si} \quad x > 0 \end{cases}$$

Veremos que en ciertos casos se pueden construir varias soluciones débiles del problema anterior. Es decir, la solución puede no ser única.

Caso $a > b$

En este caso las rectas características se cortan de manera que la solución es necesariamente discontinua.

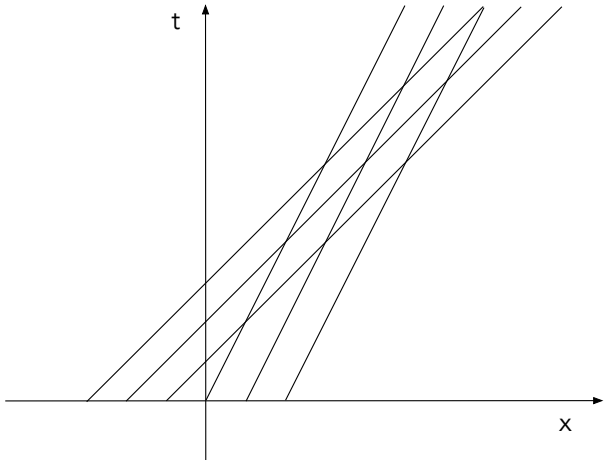


Figura 5.34 Características en el ejemplo 2. Caso $a > b$

Veamos que la solución

$$u(x,t) = \begin{cases} a & \text{si} \quad x < \dfrac{a+b}{2}t \\ b & \text{si} \quad x > \dfrac{a+b}{2}t \end{cases}$$

es una solución débil del problema considerado. La condición de Rankine-Hugoniot se satisface con

$$s = \frac{1}{2}(u^+ + u^-) = \frac{a+b}{2}$$

La ecuación de la línea de discontinuidad es

$$\frac{dx}{dt} = s$$
$$x(0) = 0$$

de donde

$$x = \frac{a+b}{2}t$$

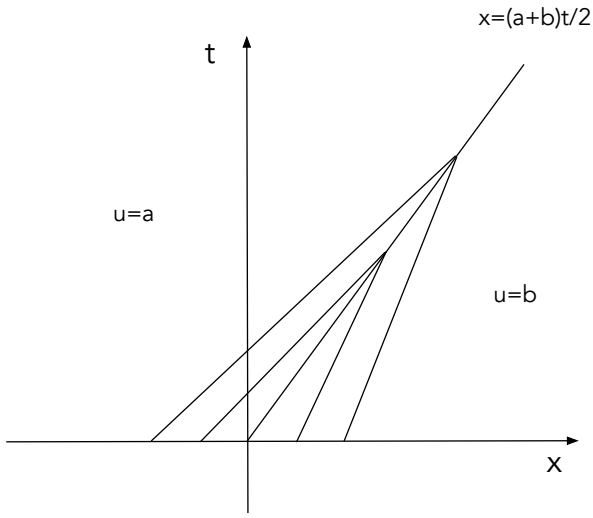


Figura 5.35 Características y línea de discontinuidad en el ejemplo 2

Evidentemente, fuera de la línea de discontinuidad la solución $u = constante$ es solución clásica.

Caso $a < b$

En este caso las características no se cortan. La solución anterior (discontinua) sigue siendo solución débil

$$u(x,t) = \begin{cases} a & \text{si} \quad x < \dfrac{a+b}{2}t \\ b & \text{si} \quad x > \dfrac{a+b}{2}t \end{cases}$$

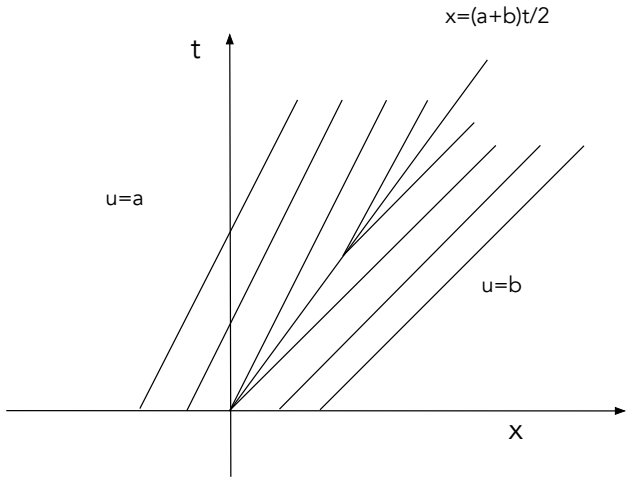


Figura 5.36 Características y línea de discontinuidad en el ejemplo 2. Caso $a < b$

Sin embargo como las características no se cortan podemos tratar de buscar una solución continua. Primero observemos que toda función de la forma $u(x,t) = v(\frac{x}{t})$ es solución clásica en las zonas de regularidad, en efecto,

$$\frac{\partial u}{\partial t} = v'(\frac{x}{t}).(-\frac{x}{t^2})$$
$$\frac{\partial u}{\partial x} = v'(\frac{x}{t}).\frac{1}{t}$$
$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = v'(\frac{x}{t}).(-\frac{x}{t^2} + \frac{1}{t}v(\frac{x}{t})) = 0$$

de donde

$$v'(\frac{x}{t}) = 0 \quad \Rightarrow \quad v(\frac{x}{t}) = constante$$
$$-\frac{x}{t^2} + \frac{1}{t}v(\frac{x}{t}) = 0 \quad \Rightarrow \quad v(\frac{x}{t}) = \frac{x}{t}$$

Teniendo en cuenta la condición inicial, resulta

$$u(x,t) = \begin{cases} a & \text{si} \quad x \leq at \\ \dfrac{x}{t} & \text{si} \quad at \leq x \leq bt \\ b & \text{si} \quad bt \leq x \end{cases}$$

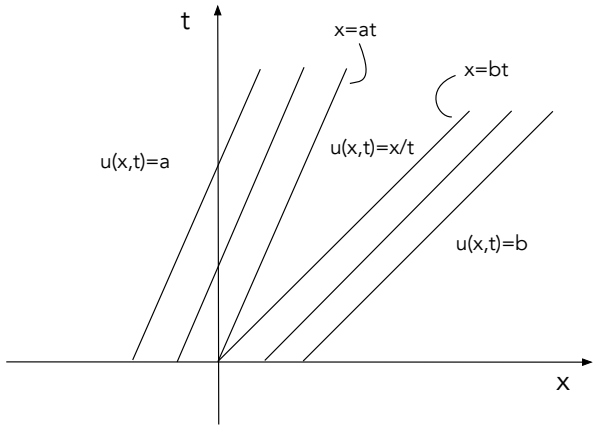


Figura 5.37 Solución continua en el ejemplo 2. Caso $a < b$

Esta solución es continua, salvo en $t = 0$. La discontinuidad desaparece para $t > 0$.

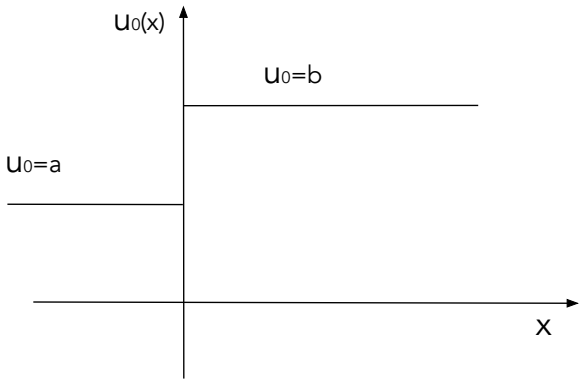


Figura 5.38 Solución inicial en el ejemplo 2.

En el instante $t = 1$ la solución viene dada por

$$u(x,t) = \begin{cases} a & \text{si} \quad x \leq a \\ x & \text{si} \quad a \leq x \leq b \\ b & \text{si} \quad b \leq x \end{cases}$$

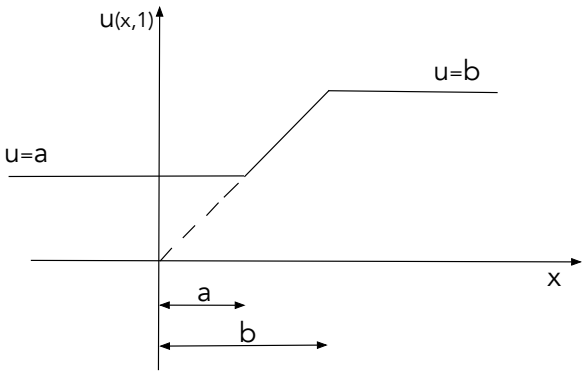


Figura 5.39 Solución en el instante $t = 1$ en el ejemplo 2.

5.3.4. Noción de Entropía

En el ejemplo anterior hemos visto que podemos encontrar varias soluciones débiles para el mismo problema de valor inicial. Es decir que no podemos asegurar la unicidad de solución. Este hecho nos está diciendo que nuestro modelo es incompleto. Utilizando el lenguaje de la física, para que el problema tenga solución única necesitamos añadir alguna ecuación o condición más que nos permita elegir entre todas las soluciones matemáticamente posibles aquella que elige la física del problema. La noción de Entropía permite resolver esta cuestión. A continuación trataremos de responder a la siguiente pregunta: Dada una solución clásica de (5.71) nos preguntamos si esta solución satisface una ley de conservación adicional de la forma

$$\frac{\partial U(u)}{\partial t} + \frac{\partial F(u)}{\partial x} = 0 \tag{5.80}$$

donde U y F son funciones regulares de $\Omega \subset \mathbb{R}^p \to \mathbb{R}$ donde Ω es el conjunto donde está u. Aquí $p = 1$ pues estamos en el caso de una ley de conservación escalar. Para un sistema de leyes de conservación $p > 1$. Veamos que este es el caso si U y F verifican

$$F'(u) = U'(u).f'(u) \tag{5.81}$$

En efecto, si u es solución de (5.71) y si F y U verifican la relación (5.81) entonces

$$\frac{\partial U}{\partial t} = U'(u)\frac{\partial u}{\partial t}$$
$$\frac{\partial F}{\partial x} = F'(u)\frac{\partial u}{\partial x} = U'(u)f'(u)\frac{\partial u}{\partial x}$$

y se verifica (5.80). De forma más rigurosa la siguiente definición define la noción de Entropía en sentido matemático:

Definición 5.8 *Sea $\Omega \subset \mathbb{R}^p$ un conjunto convexo. Una función convexa*

$$U : \Omega \to \mathbb{R}$$

se llama Entropía para un sistema de leyes de conservación (5.71) si existe una función

$$F : \Omega \to \mathbb{R}$$

llamada Flujo de Entropía tal que la relación

$$F'(u) = U'(u).f'(u)$$

tiene lugar.

Notemos que esta noción se aplica a un sistema de leyes de conservación y no solo a una ley de conservación escalar que es la que estamos considerando en estos apuntes. En el caso $p = 1$ toda función convexa U es una Entropía. Basta tomar para F una primitiva de $U'f'$. Una familia (de un parámetro) particular de entropías es

$$U(u) = |u-k| \quad k \in \mathbb{R} \tag{5.82}$$

La familia de flujos de Entropía correspondiente es

$$F(u) = sgn(u-k)(f(u)-f(k)) \tag{5.83}$$

En efecto,

$$F'(u) = sgn(u-k)f'(u)$$
$$U'(u) = sgn(u-k)$$

entonces

$$F'(u) = U'(u).f'(u)$$

Comentario

Si consideramos ahora soluciones débiles y efectuamos a partir de (5.80) cálculos análogos a los efectuados con (5.71) encontraremos sobre las discontinuidades una condición de la forma

$$s[U(u)] = [F(u)]$$

Normalmente esta condición resulta ser incompatible con la condición de Rankine-Hugoniot. En general pues, si u no es una función regular, para una Entropía U y Flujo de Entropía F no se cumple (5.80) en el sentido de las distribuciones. La pregunta es ¿Cuales son las soluciones físicamente admisibles? Admitiremos que estas

son las llamadas soluciones viscosas, es decir, aquellas que son el límite cuando $\varepsilon \to 0$ en un sentido a precisar de las soluciones de

$$\frac{\partial u_\varepsilon}{\partial t} + \frac{\partial f(u_\varepsilon)}{\partial x} - \varepsilon \frac{\partial^2 u_\varepsilon}{\partial x^2} = 0 \tag{5.84}$$

Queremos ahora relacionar las llamadas soluciones viscosas con alguna condición de Entropía. El siguiente teorema responde a esta cuestión.

Teorema 5.5 *Supongamos que (5.71) admite una Entropía U con Flujo de Entropía F. Sea $(u_\varepsilon)_\varepsilon$ una sucesión de funciones regulares solución de (5.84) tales que*

$$\begin{aligned} &1) \quad ||u_\varepsilon||_{L^\infty(\mathbb{R}\times[0,\infty))} \leq C \quad \text{independiente de } \varepsilon \\ &2) \quad u_\varepsilon \to u \quad \text{cuando } \varepsilon \to 0 \quad \text{c.t.p. en } \mathbb{R}\times[0,\infty) \end{aligned}$$

Entonces u es solución de (5.71) en el sentido de las distribuciones y satisface la condición de Entropía siguiente

$$\frac{\partial U(u)}{\partial t} + \frac{\partial F(u)}{\partial x} \leq 0 \tag{5.85}$$

en el sentido de las distribuciones sobre $\mathbb{R}\times[0,\infty)$, es decir, para toda función $\varphi \in C_0^\infty(\mathbb{R}\times(0,\infty))$, tal que $\varphi \geq 0$

$$\int_0^\infty \int_{-\infty}^\infty (U(u)\frac{\partial \varphi}{\partial t} + F(u)\frac{\partial \varphi}{\partial x})\, dx\, dt \geq 0 \tag{5.86}$$

Demostración. En primer lugar veamos que u satisface (5.71) en el sentido de las distribuciones. Multiplicando la ecuación (5.84) por $\varphi \in C_0^\infty(\mathbb{R}\times(0,\infty))$ e integrando por partes tenemos

$$\int_0^\infty \int_{-\infty}^\infty (u_\varepsilon \frac{\partial \varphi}{\partial t} + f(u_\varepsilon)\frac{\partial \varphi}{\partial x} - \varepsilon u_\varepsilon \frac{\partial^2 \varphi}{\partial x^2})\, dx\, dt = 0 \tag{5.87}$$

Aplicando el teorema de la convergencia dominada, como $||u_\varepsilon||_{L^\infty} \leq C$ tendremos para toda función $\varphi \in C_0^\infty(\mathbb{R}\times(0,\infty))$ y llamando K al soporte de φ

$$\begin{aligned} &\int_K u_\varepsilon \varphi\, dx\, dt \to \int_K u\varphi\, dx\, dt \quad \text{cuando } \varepsilon \to 0 \\ &\int_K u_\varepsilon \frac{\partial \varphi}{\partial t}\, dx\, dt \to \int_K u \frac{\partial \varphi}{\partial t}\, dx\, dt \qquad \text{cuando } \varepsilon \to 0 \\ &\int_K \varepsilon u_\varepsilon \frac{\partial^2 \varphi}{\partial x^2}\, dx\, dt \to 0 \quad \text{cuando } \varepsilon \to 0 \\ &\int_K f(u_\varepsilon)\frac{\partial \varphi}{\partial x}\, dx\, dt \to \int_K f(u)\frac{\partial \varphi}{\partial x} \quad \text{cuando } \varepsilon \to 0 \end{aligned}$$

En la ecuación (5.87) haciendo $\varepsilon \to 0$, obtenemos para toda función $\varphi \in C_0^\infty(\mathbb{R}\times(0,\infty))$

$$\int_0^\infty \int_{-\infty}^\infty (u\frac{\partial \varphi}{\partial t} + f(u)\frac{\partial \varphi}{\partial x})\, dx\, dt = 0$$

Es decir, u verifica (5.71) en el sentido de las distribuciones.

Pasemos ahora a la Entropía. Sea U una función de Entropía de clase C^2, es decir que verifica

$$F'(u) = U'(u).f'(u)$$

Multiplicando (5.84) por $U'(u_\varepsilon)$

$$U'(u_\varepsilon)\frac{\partial u_\varepsilon}{\partial t} + U'(u_\varepsilon)\frac{\partial f(u_\varepsilon)}{\partial x} - \varepsilon U'(u_\varepsilon)\frac{\partial^2 u_\varepsilon}{\partial x^2} = 0$$
$$\frac{\partial U(u_\varepsilon)}{\partial t} + \frac{\partial F(u_\varepsilon)}{\partial x} = \varepsilon U'(u_\varepsilon)\frac{\partial^2 u_\varepsilon}{\partial x^2}$$

Ahora bien

$$\frac{\partial^2 U(u_\varepsilon)}{\partial x^2} = \frac{\partial}{\partial x}\big(\frac{\partial U(u_\varepsilon)}{\partial x}\big) = \frac{\partial}{\partial x}\big(U'(u_\varepsilon)\frac{\partial u_\varepsilon}{\partial x}\big)$$
$$= U''(u_\varepsilon)\big(\frac{\partial u_\varepsilon}{\partial x}\big)^2 + U'(u_\varepsilon)\frac{\partial^2 u_\varepsilon}{\partial x^2}$$

de donde

$$\varepsilon U'(u_\varepsilon)\frac{\partial^2 u_\varepsilon}{\partial x^2} = \varepsilon\frac{\partial^2 U(u_\varepsilon)}{\partial x^2} - \varepsilon U''(u_\varepsilon)\big(\frac{\partial u_\varepsilon}{\partial x}\big)^2$$

Como U es convexa, $U'' \geq 0$ y por lo tanto

$$\frac{\partial U(u_\varepsilon)}{\partial t} + \frac{\partial F(u_\varepsilon)}{\partial x} = \varepsilon U'(u_\varepsilon)\frac{\partial^2 u_\varepsilon}{\partial x^2}$$
$$\leq \varepsilon\frac{\partial^2 U(u_\varepsilon)}{\partial x^2}$$

Sea $\varphi \in C_0^\infty(\mathbb{R}\times(0,\infty))$ con $\varphi \geq 0$. Multiplicando la anterior expresión por φ e integrando por partes

$$-\int_0^\infty \int_{-\infty}^\infty \big(U(u_\varepsilon)\frac{\partial \varphi}{\partial t} + F(u_\varepsilon)\frac{\partial \varphi}{\partial x}\big)\, dx\, dt \leq \varepsilon \int_0^\infty \int_{-\infty}^\infty U(u_\varepsilon)\frac{\partial^2 \varphi}{\partial x^2}\, dx\, dt$$
$$\int_0^\infty \int_{-\infty}^\infty \big(U(u_\varepsilon)\frac{\partial \varphi}{\partial t} + F(u_\varepsilon)\frac{\partial \varphi}{\partial x}\big)\, dx\, dt \geq -\varepsilon \int_0^\infty \int_{-\infty}^\infty U(u_\varepsilon)\frac{\partial^2 \varphi}{\partial x^2}\, dx\, dt$$

pasando al límite cuando $\varepsilon \to 0$

$$\int_0^\infty \int_{-\infty}^\infty \big(U(u)\frac{\partial \varphi}{\partial t} + F(u)\frac{\partial \varphi}{\partial x}\big)\, dx\, dt \geq 0$$

para toda función $\varphi \in C_0^\infty(\mathbb{R}\times(0,\infty))$ con $\varphi \geq 0$. ■

Recopilamos los anteriores resultados y los completamos en el siguiente teorema:

Teorema 5.6 *Si u es una solución débil de (5.71) de clase C^1 "a trozos", u satisface la condición de entropía*

$$\int_0^\infty \int_{-\infty}^\infty \big(U(u)\frac{\partial \varphi}{\partial t} + F(u)\frac{\partial \varphi}{\partial x}\big)\, dx\, dt \geq 0$$

para toda función $\varphi \in C_0^\infty(\mathbb{R}\times(0,\infty))$ con $\varphi \geq 0$, si y solo si

1. *u es solución clásica en los dominios de regularidad de u.*
2. *u satisface en las líneas de discontinuidad la condición de Rankine-Hugoniot y la condición de entropía*
$$s[U(u)] \geq [F(u)] \tag{5.88}$$
donde s es la velocidad de propagación de la discontinuidad.

Demostración. La demostración es análoga a la del teorema (5.4) utilizando la relación (5.86) en lugar de la forma débil de la ley de conservación (5.77). ■

Condición de Oleinik

Vamos ahora a expresar la condición de entropía anterior de una forma más fácil de utilizar. Es la llamada condición de Oleinik.

Teorema 5.7 *La condición de entropía*

$$s[U(u)] \geq [F(u)]$$

con $s = [f(u)]/[u]$ puede escribirse de una de las dos formas siguientes y equivalentes

$$\frac{f(u^+)-f(k)}{u^+-k} \leq s \quad \forall k \quad \text{entre } u^- \text{ y } u^+$$
$$\frac{f(u^-)-f(k)}{u^--k} \geq s \quad \forall k \quad \text{entre } u^- \text{ y } u^+$$

Demostración. Consideremos la función de Entropía y Flujo de Entropía siguiente

$$U(u) = |u-k| \quad F(u) = sgn(u-k)(f(u)-f(k)) \quad \forall k$$

La condición de entropía se escribe

$$s(|u^+-k|-|u^--k|) \geq sgn(u^+-k)(f(u^+)-f(k)) - sgn(u^--k)(f(u^-)-f(k)) \quad \forall k$$

Supongamos $u^+ > u^-$, primero observemos que para $k \geq u^+ > u^-$ junto con $k \leq u^- < u^+$ obtenemos la condición de Rankine-Hugoniot. En efecto, para $k \geq u^+ > u^-$

$$|u^+ - k| = k - u^+$$
$$|u^- - k| = k - u^-$$

de donde

$$|u^+ - k| - |u^- - k| = k - u^+ - k + u^- = u^- - u^+$$

y también

$$sgn(u^+ - k) = -1$$
$$sgn(u^- - k) = -1$$

de donde

$$s(u^- - u^+) \geq -(f(u^+) - f(k)) + (f(u^-) - f(k)) = f(u^-) - f(u^+)$$

Por otra parte para $k \leq u^- < u^+$

$$|u^+ - k| = u^+ - k$$
$$|u^- - k| = u^- - k$$

$$sgn(u^+ - k) = 1$$
$$sgn(u^- - k) = 1$$

$$s(u^+ - u^-) \geq f(u^+) - f(u^-)$$

que junto con la anterior proporciona

$$s[u] = [f(u)]$$

que es la relación de Rankine-Hugoniot.

Caso $u^- < k < u^+$

$$|u^+ - k| = u^+ - k$$
$$|u^- - k| = k - u^-$$

$$sgn(u^+ - k) = 1$$
$$sgn(u^- - k) = -1$$

La condición de entropía se escribe

$$s(|u^+-k|-|u^--k|) \geq sgn(u^+-k)(f(u^+)-f(k)) - sgn(u^--k)(f(u^-)-f(k))$$
$$s(u^++u^--2k) \geq (f(u^+)-f(k))+(f(u^-)-f(k)) = f(u^+)+f(u^-)-2f(k)$$

Sumando esta última con la condición de Rankine-Hugoniot

$$s(u^+-u^-) = f(u^+)-f(u^-)$$

obtenemos

$$2s(u^+-k) \geq 2((f(u^+)-f(k))$$

es decir

$$s \geq \frac{f(u^+)-f(k)}{u^+-k}$$

y restándole la condición de Rankine-Hugoniot obtenemos

$$2s(u^--k) \geq 2((f(u^-)-f(k))$$

es decir (al dividir por un número negativo cambia el sentido de la desigualdad)

$$s \leq \frac{f(u^-)-f(k)}{u^--k}$$

Caso $u^+ < k < u^-$

$$|u^+-k| = k-u^+$$
$$|u^--k| = u^--k$$

$$sgn(u^+-k) = -1$$
$$sgn(u^--k) = 1$$

La condición de entropía se escribe

$$s(2k-u^+-u^-) \geq 2f(k)-f(u^+)-f(u^-)$$

Sumando a esta expresión la condición de Rankine-Hugoniot

$$2s(k-u^-) \geq 2f(k)-2f(u^-)$$

de donde

$$s \leq \frac{f(u^-)-f(k)}{u^--k}$$

y restándole la condición de Rankine-Hugoniot

$$s \geq \frac{f(u^+) - f(k)}{u^+ - k}$$

■

Vamos ahora a considerar un caso particular.

Caso en que f es estrictamente convexa

En este caso la condición de entropía se escribe

$$u^+ < u^-$$

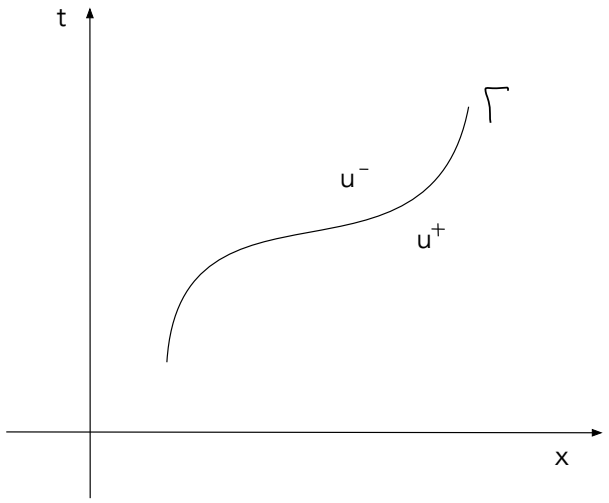


Figura 5.40 Condición de Oleinik en el caso en que f es estrictamente convexa

En efecto, consideremos por el contrario que $u^- < u^+$

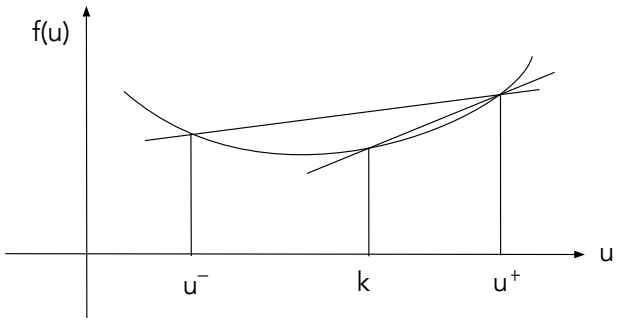


Figura 5.41 Caso no entrópico

Para todo k entre u^- y u^+ tendremos

$$\frac{f(u^+)-f(u^-)}{u^+-u^-} = s \leq \frac{f(u^+)-f(k)}{u^+-k}$$

que es el caso excluido. Por el contrario si $u^- > u^+$

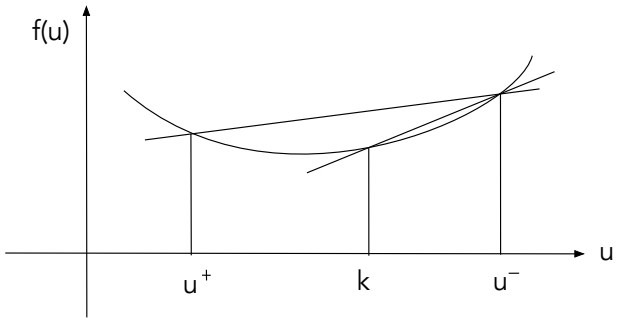


Figura 5.42 Caso entrópico

$$\frac{f(u^+)-f(u^-)}{u^+-u^-} = s \leq \frac{f(u^-)-f(k)}{u^--k}$$

que es el caso admisible.

5.3.5. *Resolución del problema de Riemann*

La idea básica de partida de muchos métodos numéricos para resolver problemas hiperbólicos no lineales de primer orden es la solución del problema de Riemann. El problema de Riemann asociado a la ecuación (5.71) es el siguiente:

$$\frac{\partial u}{\partial t} + \frac{\partial f(u)}{\partial x} = 0 \quad x \in \mathbb{R}, \quad t > 0 \tag{5.89}$$

$$u(x,t) = u_0(x) = \begin{cases} a & \text{si} \quad x < 0 \\ b & \text{si} \quad x > 0 \end{cases} \tag{5.90}$$

Nos limitaremos al caso en que f es estrictamente convexa.

1. Si $a = b$, la solución $u(x,t) = a = b$ es evidentemente una solución entrópica.
2. Si $a > b$, la discontinuidad es admisible y se propagará según una línea de discontinuidad determinada por la condición de Rankine-Hugoniot.

$$s = \frac{f(u^+) - f(u^-)}{u^+ - u^-} = \frac{f(b) - f(a)}{b - a}$$

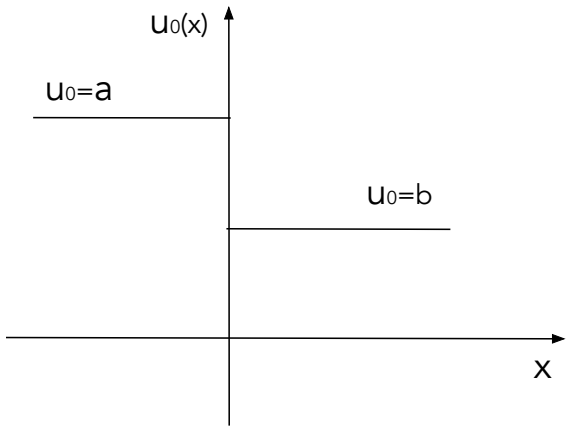


Figura 5.43 Condición inicial del problema de Riemann en el caso $a > b$

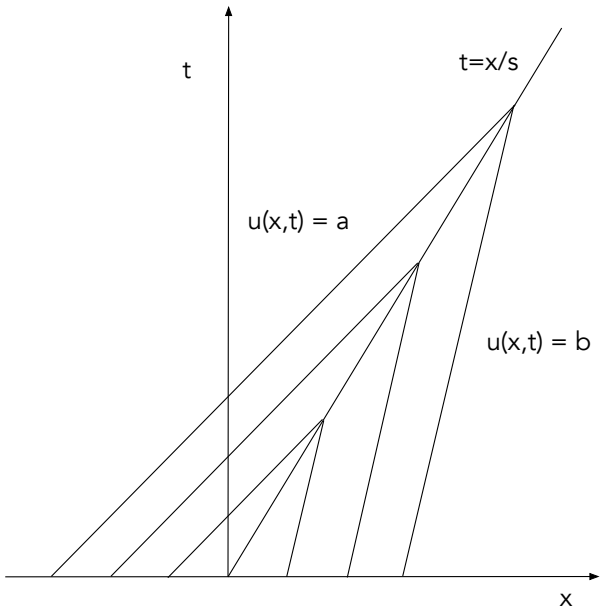


Figura 5.44 Solución del problema de Riemann en el caso $a > b$

3. Si $a < b$, la discontinuidad no es admisible y no se propagará. Probamos una solución de la forma $u(x,t) = v\left(\frac{x}{t}\right)$. Esta es una solución clásica en las zonas de regularidad, en efecto,

$$\frac{\partial u}{\partial t} = -\frac{x}{t^2} v'\left(\frac{x}{t}\right)$$
$$\frac{\partial f(u)}{\partial x} = f'\left(v\left(\frac{x}{t}\right)\right)\frac{1}{t} v'\left(\frac{x}{t}\right).$$
$$\frac{\partial u}{\partial t} + \frac{\partial f(u)}{\partial x} = -\frac{x}{t^2} v'\left(\frac{x}{t}\right) + f'\left(v\left(\frac{x}{t}\right)\right)\frac{1}{t} v'\left(\frac{x}{t}\right) = 0$$
$$\frac{1}{t} v'\left(\frac{x}{t}\right)\left(-\frac{x}{t} + f'\left(v\left(\frac{x}{t}\right)\right)\right) = 0$$

de donde

$$v'\left(\frac{x}{t}\right) = 0 \quad \Rightarrow \quad v\left(\frac{x}{t}\right) = constante$$
$$f'\left(v\left(\frac{x}{t}\right)\right) = \frac{x}{t} \quad \Rightarrow v\left(\frac{x}{t}\right) = (f')^{-1}\left(\frac{x}{t}\right)$$

f es estrictamente convexa lo que implica que f es creciente y la ecuación $f'\left(v(\xi)\right) = \xi$ tiene solución única y $u\left(\frac{x}{t}\right) = (f')^{-1}\left(\frac{x}{t}\right)$ para $\xi \in [f'(a), f'(b)]$. Teniendo en cuenta la condición inicial, resulta

$$u(x,t) = w_R(\frac{x}{t}, a, b)$$

donde, poniendo $\xi = \frac{x}{t}$

$$w_R(\xi, a, b) = \begin{cases} a & \text{si} \quad \xi \leq f'(a) \\ (f')^{-1}(\xi) & \text{si} \quad f'(a) \leq \xi \leq f'(b) \\ b & \text{si} \quad f'(b) \leq \xi \end{cases}$$

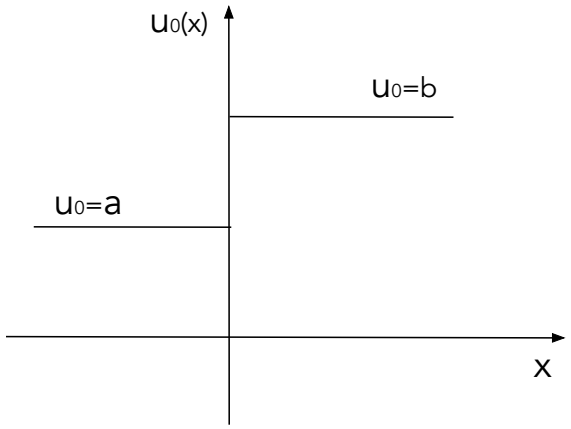


Figura 5.45 Condición inicial del problema de Riemann en el caso $a < b$

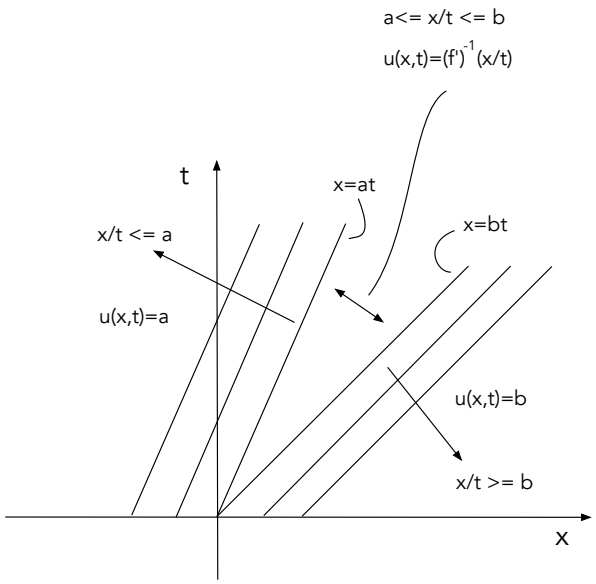


Figura 5.46 Solución continua en el problema de Riemann. Caso $a < b$

Esta solución es continua , salvo en $t = 0$. La discontinuidad desaparece para $t > 0$.

5.3.6. Resultados de existencia y unicidad

Terminaremos esta sección dando los resultados de existencia y unicidad de solución débil del problema (5.71)-(5.72) que son imprescindibles para entender los métodos numéricos que se abordarán en la sección siguiente.

Empezamos por introducir el espacio $BV(\mathbb{R})$ de funciones de variación acotada. Recordemos que una medida μ sobre $\mathbb{R}$ es una forma lineal continua sobre $C_0(\mathbb{R})$, espacio de funciones continuas de soporte compacto. Dada una función $v \in L^1_{loc}(\mathbb{R})$ se define la variación total de v mediante

$$TV(v) = \sup\{\int_{-\infty}^{\infty} v \frac{d\varphi}{dx},\ \varphi \in C_0^1(\mathbb{R}),\ ||\varphi||_{L^\infty(\mathbb{R})} \leq 1\} \tag{5.91}$$

en general tenemos $TV(v) = \infty$. Tenemos entonces la siguiente definición

$$BV(\mathbb{R}) = \{v \in L^1_{loc}(\mathbb{R});\ TV(g) < \infty\} \tag{5.92}$$

Si una función v pertenece al espacio de Sobolev $W^{1,1}(\mathbb{R})$ tenemos

$$TV(v) = \int_{-\infty}^{\infty} |\frac{dv}{dx}| \, dx < \infty$$

y por tanto $W^{1,1}(\mathbb{R}) \subset BV(\mathbb{R})$.

Si $v \in BV(\mathbb{R})$ entonces $\frac{dv}{dx}$ es una medida de Radon y $TV(v)$ es la masa total que se designa habitualmente como

$$\int_{-\infty}^{\infty} |\frac{dv}{dx}|$$

Tenemos el siguiente resultado de existencia de solución:

Si $u_0 \in L^\infty(\mathbb{R}) \cap L^2(\mathbb{R}) \cap BV(\mathbb{R})$, el problema (5.71)-(5.72) admite al menos una solución entrópica $u \in C(0,\infty; L^1_{loc}(\mathbb{R}))$ tal que $u(.,t) \in BV(\mathbb{R})$ para todo t. Además

$$||u(.,t)||_{L^\infty(\mathbb{R})} \leq ||u_0||_{L^\infty(\mathbb{R})} \tag{5.93}$$

$$TV(u(.,t)) \leq TV(u_0) \tag{5.94}$$

Finalmente se puede demostrar el siguiente resultado debido a Kružkov que implica la unicidad:

Consideremos la familia de Entropías-Flujo de Entropías (5.82)-(5.83) verificando la condición de entropía (5.86). Sean

$$u, v \in L^\infty(\mathbb{R} \times (0,\infty) \cap C(0,\infty;\, L^1_{loc}(\mathbb{R})$$

dos soluciones débiles entrópicas de (5.71) correspondientes a las condiciones iniciales u_0 y v_0 respectivamente. Si

$$M = \sup_{|\xi| \leq \text{máx}\{||u_0||_{L^\infty}, ||v_0||_{L^\infty}\}} |f'(\xi)|$$

Tenemos para todo $a > 0$ y para todo $t > 0$

$$\int_{-a}^{a} |u(x,t) - v(x,t)| \, dx \leq \int_{-a-Mt}^{a+Mt} |u_0(x) - v_0(x)| \, dx$$

5.4. Métodos Numéricos para Problemas hiperbólicos no lineales

Como en el caso lineal introducimos un mallado en el plano $x-t$ eligiendo un paso h según la coordenada x que representará habitualmente el espacio y un paso k según la coordenada t que normalmente representará el tiempo. Con las mismas notaciones

$$x_j = jh \quad j = ..., -1, 0, 1, 2, ...$$
$$t_n = nk \quad n = 0, 1, 2, ...$$

y también resultará útil definir

$$x_{j+1/2} = x_j + \frac{h}{2} = (j + \frac{1}{2})h$$

El método de diferencias finitas proporcionará aproximaciones u_j^n de la solución $u(x_j,t_n)$ en los nodos de la malla. Del mismo modo que en el caso lineal en el desarrollo y estudio de los diferentes métodos para leyes de conservación es preferible interpretar u_j^n como una aproximación del valor medio de $u(x_j,t_n)$ en el intervalo $[x_{j-1/2},x_{j+1/2}]$, es decir,

$$u_j^n \approx \frac{1}{h}\int_{x_{j-1/2}}^{x_{j+1/2}} u(x,t_n)\,dx$$

Esta interpretación es natural pues si escribimos la ley de conservación en forma integral este término es el que aparece. En efecto, integrando entre $[x_{j-1/2},x_{j+1/2}]$

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

resulta

$$\frac{d}{dt}\int_{x_{j-1/2}}^{x_{j+1/2}} u(x,t)\,dx + f(u(x_{j+1/2},t)) - f(u(x_{j-1/2},t)) = 0$$

Como valores iniciales, a partir de $u_0(x)$ definiremos u_j^0 tomando valores puntuales $u_0(x_j)$ o mejor valores medios

$$u_j^0 = \frac{1}{h}\int_{x_{j-1/2}}^{x_{j+1/2}} u_0(x)\,dx$$

Es también conveniente considerar la aproximación definida en todos los puntos del plano $x-t$,

$$u_{h,k}(x,t) = u_j^n \quad \forall (x,t) \in [x_{j-1/2},x_{j+1/2}] \times [t_n,t_{n+1}]$$

En la práctica tendremos que trabajar en un intervalo acotado de $\mathbb{R}$. En la figura se representa el caso de un problema de Riemann con condición inicial periódica con discontinuidad no admisible.

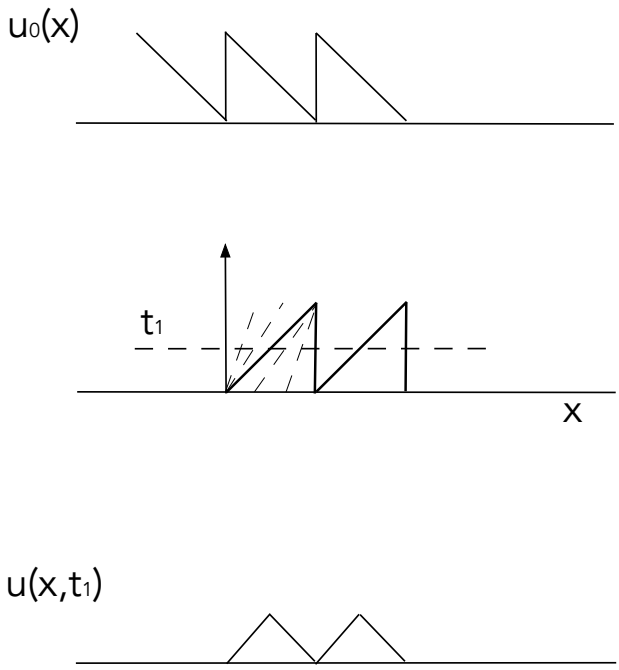


Figura 5.47 Valor Inicial Periódico con discontinuidad no admisible: $u^- < u^+$

5.4.1. Introducción: Definiciones y resultados generales

Los problemas hiperbólicos no lineales tienen, como se ha visto en la sección (5.3) propiedades cualitativas muy diferentes a los problemas lineales. En consecuencia los métodos numéricos para estos problemas no lineales se deben adaptar a esta circustancia. Un método puede presentar inestabilidad no lineal aunque sea estable en su versión lineal. También un método puede converger a una solución que no es una solución débil del problema original o converger hacia una solución no entrópica. Veamos un ejemplo:

Si consideramos la ecuación de Burgers

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = 0$$

Un método bastante natural sería

$$u_j^{n+1} = u_j^n - \frac{k}{h}u_j^n(u_j^n - u_{j-1}^n)$$

Sin embargo este método que es apropiado para el cálculo de soluciones clásicas, no converge en general hacia soluciones débiles discontinuas. En efecto, consideremos la condición inicial

$$u_0(x) = \begin{cases} 1 & \text{si} \quad x < 0 \\ 0 & \text{si} \quad x > 0 \end{cases}$$

La condición inicial para el problema discreto será

$$u_j^0 = \begin{cases} 1 & \text{si} \quad j < 0 \\ 0 & \text{si} \quad j \geq 0 \end{cases}$$

Calculemos $u^1 = (u_j^1)_{j \in \mathbb{Z}}$

$$\begin{aligned} u_{-1}^1 &= u_{-1}^0 - \frac{k}{h} u_{-1}^0 (u_{-1}^0 - u_{-2}^0) = 1 - \frac{k}{h} 1(1-1) = 1 \\ u_0^1 &= u_0^0 - \frac{k}{h} u_0^0 (u_0^0 - u_{-1}^0) = 0 - \frac{k}{h} 0(0-1) = 0 \\ u_1^1 &= u_1^0 - \frac{k}{h} u_1^0 (u_1^0 - u_0^0) = 0 - \frac{k}{h} 0(0-0) = 0 \end{aligned}$$

es decir

$$u_j^1 = u_j^0 \quad \forall j$$

$$u_j^n = u_j^0 \quad \forall j, \forall n$$

Cuando $h \to 0$, la función converge claramente hacia la función $u(x,t) = u_0(x)$ que no es la solución débil del problema.

Con otros datos iniciales podemos obtener resultados que pueden parecer a primera vista plausibles y que sin embargo son incorrectos. Por ejemplo con

$$u_0(x) = \begin{cases} 1.2 & \text{si} \quad x < 0 \\ 0.4 & \text{si} \quad x > 0 \end{cases}$$

obtenemos una solución que viaja a una velocidad inadecuada.

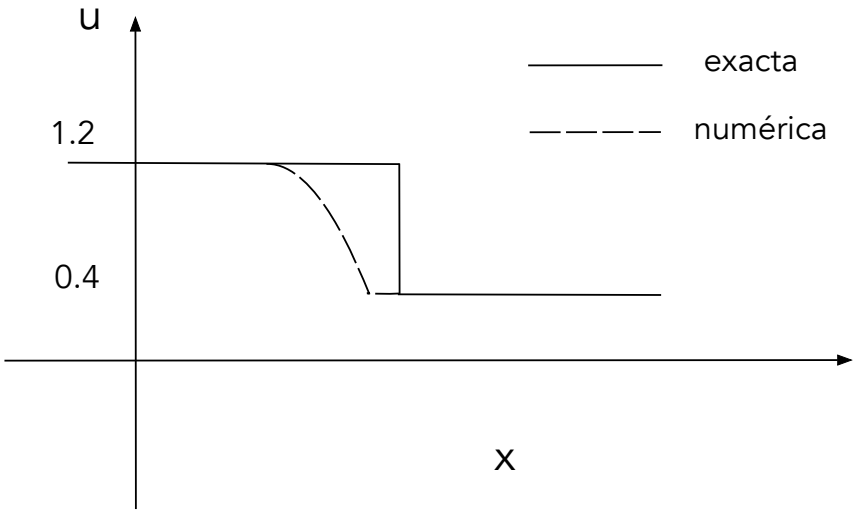


Figura 5.48 Solución numérica con velocidad no correcta

Con el fin de evitar estos efectos indeseados nos limitaremos a los llamados métodos conservativos que responden de forma natural a las leyes de conservación de la que provienen las ecuaciones. Empezaremos definiendo lo que entendemos por método conservativo.

Definición 5.9 *Dado un método de* $2l+1$ *puntos*

$$u_j^{n+1} = H(u_{j-l}^n, ..., u_{j+l}^n)$$

diremos que se puede poner en forma conservativa, o más brevemente, que es un método conservativo si existe una función continua de $2l$ *variables* $g : \mathbb{R}^{2l} \to \mathbb{R}$, *llamada función de flujo numérico tal que*

$$H(v_{j-l}, ..., v_{j+l}) = v_0 - \lambda \left(g(v_{-l+1}, ..., v_l) - g(v_{-l}, ..., v_{l-1}) \right) \tag{5.95}$$

es decir, el esquema numérico correspondiente se expresa de la forma

$$u_j^{n+1} = u_j^n - \lambda \left(g(u_{j-l+1}^n, ..., u_{j+l}^n) - g(u_{j-l}^n, ..., u_{j+l-1}^n) \right)$$

o con notación abreviada

$$u_j^{n+1} = u_j^n - \lambda \left(g_{j+1/2}^n - g_{j-1/2}^n \right) \tag{5.96}$$

En el caso de un esquema de tres puntos tenemos

$$u_j^{n+1} = u_j^n - \lambda \Big(g(u_j^n, u_{j+1}^n) - (g(u_{j-1}^n, u_j^n) \Big)$$

La forma conservativa es muy natural si consideramos u_j^n como una aproximación del valor medio de $u(x_j,t_n)$ en la celda correspondiente. Si $u(x,t)$ satisface la ley de conservación

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

una forma integral de esta ley es

$$\int_{x_{j-1/2}}^{x_{j+1/2}} u(x,t_{n+1})\,dx = \int_{x_{j-1/2}}^{x_{j+1/2}} u(x,t_n)\,dx$$
$$- \Big(\int_{t_n}^{t_{n+1}} f(u(x_{j+1/2},t))\,dt - \int_{t_n}^{t_{n+1}} f(u(x_{j-1/2},t))\,dt \Big)$$

Dividiendo por h y llamando

$$\bar{u}(x_j,t_n) = \frac{1}{h} \int_{x_{j-1/2}}^{x_{j+1/2}} u(x,t_n)\,dx$$

$$\bar{u}(x_j,t_{n+1}) = \bar{u}(x_j,t_n) - \frac{1}{h} \Big(\int_{t_n}^{t_{n+1}} f(u(x_{j+1/2},t))\,dt - \int_{t_n}^{t_{n+1}} f(u(x_{j-1/2},t))\,dt \Big)$$

La función flujo numérico juega pues el papel del flujo a través de $x_{j+1/2}$ (resp. $x_{j-1/2}$) en el intervalo $[t_n,t_{n+1}]$.

Veamos un primer ejemplo de método conservativo: El esquema de Lax-Friedrichs.

$$u_j^{n+1} = \frac{u_{j+1}^n + u_{j-1}^n}{2} - \frac{\lambda}{2} \Big(f(u_{j+1}^n) - f(u_{j-1}^n) \Big)$$

que se puede escribir de la forma

$$u_j^{n+1} = u_j^n - \frac{\lambda}{2} \Big(f(u_{j+1}^n) - f(u_{j-1}^n) \Big) + \frac{1}{2} (u_{j+1}^n - 2u_j^n + u_{j-1}^n)$$

y que corresponde a la función de flujo numérico

$$g^{LF}(u,v) = \frac{1}{2} \Big(f(u) + f(v) \Big) - \frac{1}{2\lambda} (v-u)$$

En efecto,

$$\begin{aligned} u_j^{n+1} &= u_j^n - \lambda\left(g^{LF}(u_j^n, u_{j+1}^n) - g^{LF}(u_{j-1}^n, u_j^n)\right) \\ &= u_j^n - \frac{\lambda}{2}\left(f(u_j^n) + f(u_{j+1}^n) - f(u_{j-1}^n) - f(u_j^n)\right) \\ &+ \frac{1}{2}(u_{j+1}^n - u_j^n - u_j^n + u_{j-1}^n) \\ &= u_j^n - \frac{\lambda}{2}\left(f(u_{j+1}^n) - f(u_{j-1}^n)\right) + \frac{1}{2}(u_{j+1}^n - 2u_j^n + u_{j-1}^n) \end{aligned}$$

En el marco de los métodos numéricos conservativos la noción de consistencia toma una forma muy sencilla. En efecto, tenemos la siguiente propiedad:

Propiedad 5.1 *Un método conservativo es consistente con la ley de conservación y es al menos de orden* 1*, si se verifica*

$$g(v,...,v) = f(v) \quad \forall v \tag{5.97}$$

Demostración. Pongamos

$$G(x,t) = g(u(x-lh,t),...,u((x+(l-1)h,t)$$

$$\frac{u(x,t+k)-u(x,t)}{k} + \frac{G(x+h,t)-G(x,t)}{h} = \frac{\partial u}{\partial t}(x,t) + \mathscr{O}(k) + \frac{\partial G}{\partial x}(x,t) + \mathscr{O}(h)$$
$$= \frac{\partial u}{\partial t}(x,t) + \frac{\partial G}{\partial x}(x,t) + \mathscr{O}(k)$$

donde tenemos en cuenta que $\lambda = k/h = constante$. Consideremos ahora la relación (5.97) y derivemos en el punto u. Para ello observemos que $f = g \circ V$ donde

$$g : \mathbb{R}^{2l} \to \mathbb{R}$$

y

$$\begin{aligned} V : \mathbb{R} &\to \mathbb{R}^{2l} \\ v &\to (v,...,v)^t \end{aligned}$$

Tenemos por la regla de la cadena

$$\begin{aligned} f'(u) &= g'(V(u)) \circ V'(u) \\ &= (\partial_1 g,...,\partial_{2l} g)(u,...,u).\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \\ &= \sum_{i=1}^{2l} \partial_i g(u,...,u) \end{aligned}$$

de donde

$$\frac{\partial G}{\partial x}(x,t) = \sum_{i=1}^{2l} \partial_i g(u,...,u)\frac{\partial u}{\partial x}(x,t) + \mathscr{O}(k)$$
$$= f'(u)\frac{\partial u}{\partial x} + \mathscr{O}(k) = \frac{\partial f(u)}{\partial x} + \mathscr{O}(k)$$

y finalmente

$$\frac{u(x,t+k)-u(x,t)}{k} + \frac{G(x+h,t)-G(x,t)}{h}$$
$$= \frac{\partial u}{\partial t}(x,t) + \frac{\partial G}{\partial x}(x,t) + \mathscr{O}(k)$$
$$= \frac{\partial u}{\partial t}(x,t) + \frac{\partial f(u)}{\partial x}(x,t) + \mathscr{O}(k) = \mathscr{O}(k)$$

■

De forma inmediata obtenemos

Corolario 5.1 *El esquema de Lax-Friedrichs es consistente de orden* 1

Demostración.

$$g^{LF}(u,v) = \frac{1}{2}\Big(f(u)+f(v)\Big) - \frac{1}{2\lambda}(v-u)$$

y por lo tanto

$$g^{LF}(u,v) = f(u)$$

■

El siguiente teorema nos proporciona una caracterizacón de los métodos de orden 2.

Teorema 5.8 *Sea un método numérico para resolver (5.71)-(5.72) de la forma (5.16) consistente y que puede ponerse en forma conservativa (5.96) y en el que H es de clase C^3. Entonces para una solución u de (5.71) suficientemente regular, y con $\lambda = k/h = constante$, el error de consistencia se puede expresar de la siguiente forma*

$$\frac{1}{k}\Big(u(x,t+k) - H\big(u(x-lh,t),...,u(x+lh,t)\big)\Big)$$
$$= -k\frac{\partial}{\partial x}\Big(\beta(u,\lambda)\frac{\partial u}{\partial x}(x,t)\Big) + \mathscr{O}(k^2) \tag{5.98}$$

donde

$$\beta(u,\lambda) = \frac{\sum_{j=-l}^{j=l} j^2 \partial_j H(u,...,u)}{2\lambda^2} - \frac{(f'(u))^2}{2}$$

Para simplificar la demostración desglosaremos la misma mediante algunos lemas previos.

Lema 5.2 *Sea* $g : \mathbb{R}^{2l} \to \mathbb{R}$ *tal que* $g(u,...,u) = f(u)$. *Entonces*

$$f'(u) = \sum_j \partial_j g(u,...,u) \tag{5.99}$$

Demostración. Ver la demostración de la propiedad (5.1). ■

Lema 5.3 *Sea* $H : \mathbb{R}^{2l+1} \to \mathbb{R}$ *y* $g : \mathbb{R}^{2l} \to \mathbb{R}$ *con*

$$H(v_{-l},...,v_l) = v_0 - \lambda\left(g(v_{-l+1},...,v_l) - g(v_{-l},...,v_{l-1})\right)$$

la derivada parcial j*-ésima de* H *viene dada por*

$$\partial_j H = \delta_j^0 - \lambda\left(\partial_{j-1} g - \partial_j g\right)$$

Demostración. Para la demostración resultará más cómodo trabajar con índices j tales que $1 \leq j \leq 2l+1$. Introducimos las aplicaciones auxiliares

$$\hat{T} : \mathbb{R}^{2l+1} \to \mathbb{R}^{2l}$$
$$(v_1,...,v_{2l+1})^t \to (v_2,...,v_{2l+1})^t$$

$$\check{T} : \mathbb{R}^{2l+1} \to \mathbb{R}^{2l}$$
$$(v_1,...,v_{2l+1})^t \to (v_1,...,v_{2l})^t$$

$$E : \mathbb{R}^{2l+1} \to \mathbb{R}$$
$$(v_1,...,v_{2l+1})^t \to v_{l+1}$$

Podemos escribir

$$H(v_1,...,v_{2l+1}) = E(v_1,...,v_{2l+1}) - \lambda\left(g(\hat{T}(v_1...,v_{2l+1}) - g(\check{T}(v_1,...,v_{2l+1})\right)$$

Aplicando la regla de la cadena, la matriz jacobiana de H es

$$H' = E' - \lambda(g'.\hat{T}' - g'.\check{T}')$$

Tenemos por una parte

$$E' = (0,...,1,...,0) \quad \text{donde el 1 ocupa el lugar } l+1$$

por otra parte

$$g'.\hat{T}' = (\partial_1 g, ..., \partial_{2l} g).\begin{pmatrix} 0 & 1 & 0 \dots & 0 \\ 0 & 0 & 1 \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & 0 \dots & 1 \end{pmatrix} = (0, \partial_1 g, ..., \partial_{2l} g)$$

y también

$$g'.\check{T}' = (\partial_1 g, ..., \partial_{2l} g).\begin{pmatrix} 1 & 0 & 0 \dots & 0 \\ 0 & 1 & 0 \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 \dots & 1 & 0 \end{pmatrix} = (\partial_1 g, ..., \partial_{2l} g, 0)$$

y como

$$\partial_j H = H'.\begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix}$$

donde el 1 ocupa la j-ésima fila. Finalmente obtenemos

$$\partial_j H = \Big((0,..,1,...,0) - \lambda(0 - \partial_1 g, \partial_1 g - \partial_2 g, ..., \partial_{j-1} g - \partial_j g, ..., \partial_{2l} g - 0).\Big).\begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} \quad \text{(lugar } j\text{)}$$

De modo que podemos escribir para $j = 1, ..., 2l+1$

$$\partial_j H = \delta_j^{l+1} - \lambda(\partial_{j-1} g - \partial_j g)$$

donde adoptamos las extensiones $\partial_0 g = 0$ y $\partial_{2l+1} g = 0$

■

Demostración del teorema (5.8)

Desarrollaremos en serie de Taylor, en un entorno de $u(x,t)$ por una parte – $u(x,t+k)$ y por otra $H(u(x-lh,t), ..., u(x+lh,t))$. Para $u(x,t+k)$ tenemos, escribiendo eventualmente para simplificar la escritura u en lugar de $u(x,t)$ y lo mismo para otros términos evaluados en el punto (x,t)

$$u(x,t+k) = u + k\frac{\partial u}{\partial t} + \frac{k^2}{2}\frac{\partial^2 u}{\partial t^2} + \mathcal{O}(k^3)$$

como u es solución de

$$\frac{\partial u}{\partial t} = -\frac{\partial f(u)}{\partial x} = -f'(u)\frac{\partial u}{\partial x}$$

y también derivando

$$\begin{aligned}\frac{\partial^2 u}{\partial t^2} &= -\frac{\partial}{\partial t}\Big(\frac{\partial f(u)}{\partial x}\Big)\\ &= -\frac{\partial}{\partial x}\Big(\frac{\partial f(u)}{\partial t}\Big) = \frac{\partial}{\partial x}\Big((f'(u))^2\frac{\partial u}{\partial x}\Big)\end{aligned}$$

de manera que

$$u(x,t+k) = u - kf'(u)\frac{\partial u}{\partial x} + \frac{k^2}{2}\frac{\partial}{\partial x}\Big((f'(u))^2\frac{\partial u}{\partial x}\Big) + \mathscr{O}(k^3) \tag{5.100}$$

Pasemos al desarrollo de Taylor de H. Como el método es conservativo

$$H(u,...,u) = u$$

y desarrolando en un entorno de $(u,...,u)$

$$H(u(x-lh,t),...,u(x+lh,t)) = u + \sum_{j=-l}^{l}\partial_j H(u,...,u)\big(u(x+jh,t)-u(x,t)\big)$$

$$+\frac{1}{2}\sum_{i,j=-l}^{l}\partial^2_{i,j}H(u,...,u)((u(x+ih,t)-u(x,t))(u(x+jh,t)-u(x,t)) + \mathscr{O}(h^3)$$

De nuevo, mediante el desarrollo de Taylor

$$u(x+jh,t)-u(x,t) = jh\frac{\partial u}{\partial x} + \frac{(jh)^2}{2}\frac{\partial^2 u}{\partial x^2} + \mathscr{O}(h^3)$$

de donde

$$\big(u(x+ih,t)-u(x,t)\big)\big(u(x+jh,t)-u(x,t)\big) = ijh^2(\frac{\partial u}{\partial x})^2 + \mathscr{O}(h^3)$$

Por lo tanto

$$H(u(x-lh,t),...,u(x+lh,t)) = u + h\sum_{j=-l}^{l} j\partial_j H(u,...,u)\frac{\partial u}{\partial x}$$

$$+\frac{h^2}{2}\Big(\sum_{j=-l}^{l} j^2\partial_j H(u,...,u)\frac{\partial^2 u}{\partial x^2} + \sum_{i,j=-l}^{l} ij\partial^2_{i,j}H(u,...,u)(\frac{\partial u}{\partial x})^2\Big) + \mathscr{O}(h^3)$$

Vamos a evaluar cada uno de los términos del este desarrollo.

- Primero teniendo en cuenta que el método es conservativo y el lema (5.3), renumerando el índice j entre $-l$ y l

$$\partial_j H = \delta^0_j - \lambda(\partial_{j-1}g - \partial_j g)$$

De donde deducimos

$$\sum_{j=-l}^{l} j\partial_j H(u,...,u) = -\lambda \sum_{j=-l}^{l} j\big(\partial_{j-1} g(u,...,u) - \partial_j g(u,...,u)\big)$$

Observando

$$\sum_{j=-l}^{l} j\partial_{j-1} g(u,...,u) = \sum_{j=-l}^{l} (j+1)\partial_j g(u,...,u)$$

resulta

$$\begin{aligned}\sum_{j=-l}^{l} j\partial_j H(u,...,u)(u,...,u) &= -\lambda \sum_{j=-l}^{l} ((j+1)-j)\partial_j g(u,...,u) \\ &= -\lambda f'(u) \end{aligned} \tag{5.101}$$

- Evaluamos ahora los términos en derivadas segundas de H. Derivando de nuevo

$$\partial^2_{i,j} H = -\lambda(\partial^2_{i-1,j-1} g - \partial^2_{i,j} g)$$

$$\begin{aligned}&\sum_{i,j=-l}^{l} (i-j)^2 \partial^2_{i,j} H(u,...,u) \\ &= -\lambda\Big(\sum_{i,j=-l}^{l} (i-j)^2 \partial^2_{i-1,j-1} g(u,...,u) - \sum_{i,j=-l}^{l} (i-j)^2 \partial^2_{i,j} g(u,...,u)\Big)\end{aligned}$$

cambiando en el primer sumando los índices i, j por $i+1, j+1$ como

$$(i+1)-(j+1) = i-j$$

$$\sum_{i,j=-l}^{l} (i-j)^2 \partial^2_{i,j} H(u,...,u) = 0 \tag{5.102}$$

- Vamos a transformar los términos en h^2. Tenemos,

$$\sum_{j=-l}^{l} j^2 \partial_j H(u,...,u)\frac{\partial^2 u}{\partial x^2} + \sum_{i,j=-l}^{l} ij\partial^2_{i,j} H(u,...,u)(\frac{\partial u}{\partial x})^2$$
$$= \frac{\partial}{\partial x}\Big(\sum_{j=-l}^{l} j^2 \partial H_j(u,...,u)\frac{\partial u}{\partial x}\Big) - \Big(\sum_{j=-l}^{l} j^2 \partial^2_{i,j} H(u,...,u)(\frac{\partial u}{\partial x})^2\Big)$$
$$+ \sum_{i,j=-l}^{l} ij\partial^2_{i,j} H(u,...,u)(\frac{\partial u}{\partial x})^2$$
$$= \frac{\partial}{\partial x}\Big(\sum_{j=-l}^{l} j^2 \partial_j H(u,...,u)\frac{\partial u}{\partial x}\Big) + \sum_{i,j=-l}^{l} (ij-j^2)\partial^2_{i,j} H(u,...,u)(\frac{\partial u}{\partial x})^2$$

Por la desigualdad de Schwarz de las derivadas cruzadas

$$\sum_{i,j=-l}^{l} i^2 \partial^2_{i,j} H(u,...,u) = \sum_{i,j=-l}^{l} j^2 \partial^2_{i,j} H(u,...,u)$$

y podemos escribir

$$\sum_{i,j=-l}^{l} (ij-j^2)\partial^2_{i,j} H(u,...,u)$$
$$= \frac{1}{2}\Big(\sum_{i,j=-l}^{l} 2ij\partial^2_{i,j} H(u,...,u) - \sum_{i,j=-l}^{l} j^2 \partial^2_{i,j} H(u,...,u) - \sum_{i,j=-l}^{l} i^2 \partial^2_{i,j} H(u,...,u)\Big)$$
$$= -\frac{1}{2}\sum_{i,j=-l}^{l} (i-j)^2)\partial^2_{i,j} H(u,...,u) = 0$$

donde hemos tenido en cuenta (5.102).

Finalmente el desarrollo de H se puede escribir

$$H(u(x-lh,t),...,u(x+lh,t))$$
$$= u - \lambda h f'(u)\frac{\partial u}{\partial x} + \frac{h^2}{2}\frac{\partial}{\partial x}\Big(\sum_{j=-l}^{l} j^2 \partial_j H(u,...,u)\frac{\partial u}{\partial x}\Big) + \mathscr{O}(h^3) \qquad (5.103)$$

Restando (5.103) de (5.100) y teniendo en cuenta que $h = k/\lambda$

$$u(x,t+k) - H(u(x-lh,t),...,u(x+lh,t))$$
$$= \frac{k^2}{2}\frac{\partial}{\partial x}\Big((f'(u))^2 - \sum_{j=-l}^{l} \frac{j^2}{\lambda^2}\frac{\partial H}{\partial x}(u,...,u)\Big)\frac{\partial u}{\partial x} + \mathscr{O}(k^3)$$

dividiendo por k y reordenando términos obtenemos (5.98)

■

De forma inmediata obtenemos,

Corolario 5.2 *Si* $\beta(u,\lambda) \neq 0$ *el esquema*

$$u(x,t+k) = H(u(x-lh,t),...,u(x+lh,t)$$

es de orden 1 *con respecto a la ecuación (5.71) y de orden* 2 *para la ecuación*

$$\frac{\partial u}{\partial t} + \frac{\partial f(u)}{\partial x} - \lambda h \frac{\partial}{\partial x}\Big(\beta(u,\lambda)\frac{\partial u}{\partial x}\Big) = 0 \tag{5.104}$$

En el marco de los problemas no lineales la noción de estabilidad hay que adaptarla. Las apropiadas condiciones de estabilidad junto con la consistencia permitirán demostrar la convergencia. Un primer resultado de convergencia es el siguiente teorema de Lax-Wendroff.

Teorema 5.9 *Consideremos sucesiones* $(k)_k$ *y* $(h)_h$ *de parámetros* $k,h \to 0$.

1. $u_k(x,t)$ *la solución aproximada correspondiente para cada k (resp. h,* $\lambda = h/k$*) generada por el método conservativo (5.96).*
2. *El método verifica* $g(v,...,v) = f(v) \quad \forall v$*, es decir el método es consistente.*
3. *Supongamos que* u_k *converge hacia u cuando* $k \to 0$ *en el sentido de la topología* $L^1_{loc}(\mathbb{R}\times[0,\infty))$*, es decir, para todo* $\bar{\Omega} = [a,b]\times[0,T]$ *en el plano* $x-t$ *tenemos*

$$||u_k - u||_{1,\bar{\Omega}} = \int_0^T \int_a^b |u_k(x,t) - u(x,t)|\, dx\, dt \to 0 \quad \text{cuando } k \to 0 \tag{5.105}$$

y también

$$u_k(x,t) \to u(x,t) \quad \text{c.t.p.} \tag{5.106}$$

4. *Supondremos además que se verifica la condición de estabilidad siguiente: Existe una constante* $C > 0$ *tal que*

$$||u_k||_{L^\infty(\mathbb{R}\times(0,\infty)} \leq C \tag{5.107}$$

Entonces u es solución débil del problema (5.71), es decir, verifica (5.76) que es

$$\int_0^\infty \int_{-\infty}^\infty \Big(u\frac{\partial \varphi}{\partial t} + f(u)\frac{\partial \varphi}{\partial x}\Big)\, dx\, dt = \int_{-\infty}^\infty u(x,0)\varphi(x,0)\, dx$$

para todo $\varphi \in C_0^1(\mathbb{R}\times[0,\text{ínf}))$

Demostración. Sea $\varphi \in C_0^1(\mathbb{R}\times[0,\text{ínf}))$ y multipliquemos los términos de la expresión (5.96) por $\varphi(x_j,t_n) = \varphi_j^n$

$$u_j^{n+1}\varphi_j^n = u_j^n\varphi_j^n - \lambda\Big(g_{j+1/2} - g_{j-1/2})\Big)\varphi_j^n$$

sumando obtenemos

$$\sum_{n=0}^{\infty}\sum_{j=-\infty}^{\infty}(u_j^{n+1}-u_j^n)\varphi_j^n+\lambda\sum_{n=0}^{\infty}\sum_{j=-\infty}^{\infty}\Big(g_{j+l/2}-g_{j-l/2})\Big)\varphi_j^n=0$$

y realizando las siguientes sumas por partes

$$\sum_{n\geq 0}(u_j^{n+1}-u_j^n)\varphi_j^n=-\sum_{n\geq 1}u_j^n(\varphi_j^n-\varphi_j^{n-1})-u_j^0\varphi_j^0$$

y también

$$\sum_{j\in\mathbb{Z}}(g_{j+1/2}^n-g_{j-1/2}^n)\varphi_j^n=-\sum_{j\in\mathbb{Z}}g_{j+1/2}^n(\varphi_{j+1}^n-\varphi_j^n)$$

podemos escribir

$$h\sum_{n\geq 1}\sum_{j\in\mathbb{Z}}u_j^n(\varphi_j^n-\varphi_j^{n-1})+k\sum_{n\geq 0}\sum_{j\in\mathbb{Z}}g_{j+1/2}^n(\varphi_{j+1}^n-\varphi_j^n)+h\sum_{j\in\mathbb{Z}}u_j^0\varphi_j^0=0$$

Utilizando las extensiones a todo el plano $x-t$ mediante funciones constantes a trozos

$$\begin{aligned}
u_k(x,t)&=u_j^n \quad x_{j-1/2}\leq x<x_{j+1/2} \quad t_n\leq t<t_{n+1}\\
\varphi_k(x,t)&=\varphi_j^n \quad x_{j-1/2}\leq x<x_{j+1/2} \quad t_n\leq t<t_{n+1}\\
g_k(x,t)&=g_{j+1/2}^n \quad x_j\leq x<x_{j+1} \quad t_n\leq t<t_{n+1}
\end{aligned}$$

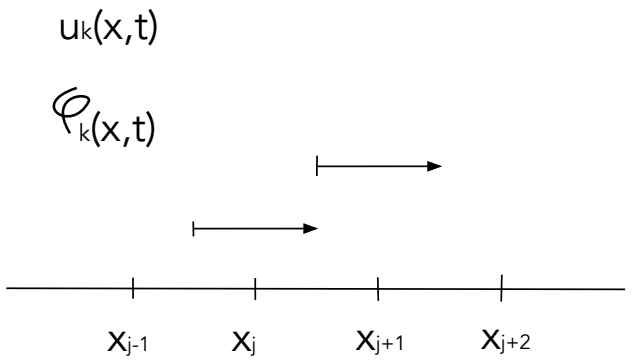


Figura 5.49 Extensión a funciones constantes a trozos

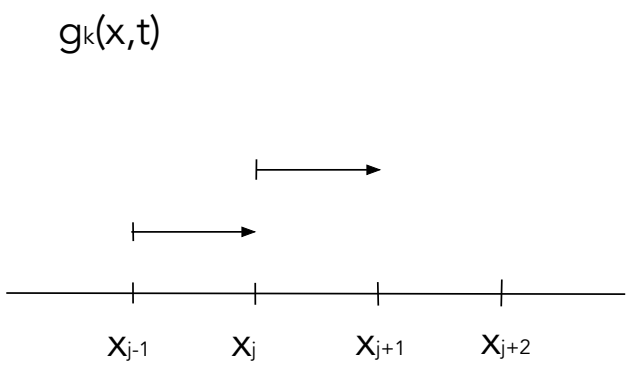


Figura 5.50 Extensión a funciones constantes a trozos

Los anteriores sumatorios se interpretan como integrales observando

$$h\sum_{j\in\mathbb{Z}} u_j^0\varphi_j^0 = \int_{-\infty}^{\infty} u_k(x,0)\varphi_k(x,0)\,dx$$

y análogamente para los otros sumatorios dobles

$$\int_0^\infty\int_{-\infty}^{\infty} u_k(x,t)\frac{\varphi_k(x,t)-\varphi_k(x,t-k)}{k}\,dx\,dt+$$
$$\int_0^\infty\int_{-\infty}^{\infty} g_k(x,t)\frac{\varphi_k(x+h/2,t)-\varphi(x-h/2,t)}{h}\,dx\,dt+\int_{-\infty}^{\infty} u_k(x,0)\varphi_k(x,0)\,dx=0$$

Pasemos al límite cuando $k\to 0$ (resp. $h\to 0$), tendremos

$$\varphi_k\to\varphi$$
$$\frac{\varphi_k(x,t+k)-\varphi_k(x,t)}{k}\to\frac{\partial\varphi}{\partial t}$$
$$\frac{\varphi_k(x+h/2,t)-\varphi_k(x-h/2,t)}{h}\to\frac{\partial\varphi}{\partial x}$$

uniformemente. Por otra parte

$$u_k\to u \quad \text{en } L^1_{loc}(\mathbb{R}\times[0,\infty)) \text{ y c.t.p.}$$

por lo tanto

$$\int_0^\infty\int_{-\infty}^{\infty} u_k(x,t)\frac{\varphi_k(x,t)-\varphi_k(x,t-k)}{k}\,dx\,dt\to\int_0^\infty\int_{-\infty}^{\infty} u\frac{\partial\varphi}{\partial t}\,dx\,dt$$
$$\int_{-\infty}^{\infty} u_k(x,0)\varphi_k(x,0)\,dx\to\int_{-\infty}^{\infty} u_0(x)\varphi(x,0)\,dx$$

Finalmente estudiemos el paso al límite en el término

$$\int_0^\infty \int_{-\infty}^\infty g_k(x,t) \frac{\varphi_k(x+h/2,t) - \varphi(x-h/2,t)}{h} \, dx\,dt$$

Tenemos $g_k(x,t) = g(u_k(x-(l-1)k,t),...,u_k(x+lk,t))$. Vamos a demostrar que $g_k(x,t) \to f(u(x,t))$ cuando $k \to 0$. Análogamente por la propiedades (5.105) y (5.106) y la propiedad de consistencia, como g es continua se puede demostrar que

$$g_k(x,t) \to f(u) \quad \text{c.t.p.}$$

y gracias a (5.107) y el teorema de convergencia dominada de Lebesgue

$$g_k(x,t) \to f(u) \quad \text{en } L^1_{loc}(\mathbb{R} \times [0,\infty))$$

y pasando al límite y obtenemos

$$\int_0^\infty \int_{-\infty}^\infty g_k(x,t) \frac{\varphi_k(x+h/2,t) - \varphi(x-h/2,t)}{h} \, dx\,dt \to \int_0^\infty \int_{-\infty}^\infty f(u) \frac{\partial \varphi}{\partial x} \, dx\,dt$$

■

El teorema de Lax-Wendroff nos indica en qué condiciones la solución dada por un esquema numérico converge a una solución débil del problema de partida. En la demostración ha sido necesario suponer que el esquema numérico sea conservativo para poder pasar al límite. En el caso no conservativo, no es posible en general demostrar que el límite u es una solución débil del problema.

Queda todavía por estudiar si la solución límite débil es entrópica. Supongamos pues (U,F) es la pareja de entropía U asociada al flujo de entropía F. Sabemos que la solución entrópica u satisface (5.85) lo que justifica la siguiente definición.

Definición 5.10 *Diremos que un método numérico en diferencias finitas es consistente con la condición de entropía (5.85) si existe una función G continua de* $\mathbb{R}^{2l} \to \mathbb{R}$ *llamada flujo de entropía numérico, verificando*

$$G(v,v,...,v) = F(v)$$

para la cual se tiene la desigualdad

$$U_j^{n+1} \leq U_j^n - \lambda (G_{j+1/2}^n - G_{j-1/2}^n)$$

donde

$$U_j^n = U(v_j^n)$$
$$G_{j+1/2}^n = G(v_{j-l+1}^n,...,v_{j+l}^n)$$

Diremos entonces que el método numérico es entrópico si es consistente con toda condición de entropía.

Estamos en condiciones de precisar el teorema (5.9).

Teorema 5.10 *Supongamos que se verifican las hipótesis del teorema (5.9) y consideremos un método de diferencias finitas consistente con la condición de entropía (5.85). Entonces el límite u es una solución débil del problema (5.71)-(5.72) que verifica la condición de entropía (5.85). Si el método de diferencias finitas es entrópico entonces u es la solución entrópica de (5.71)-(5.72)*

Demostración. La demostración es análoga a la del teorema (5.9). ■

En el caso de ecuaciones no lineales no disponemos de un método general para estudiar la estabilidad de los métodos numéricos correspondientes. La estabilidad de los métodos lineales aplicados a la ecuación linealizada no es suficiente para determinar la estabilidad global del esquema numérico. Una idea general a la hora de construir métodos numéricos para problemas no lineales es buscar métodos cuyas soluciones conserven en la medida de lo posible algunas propiedades de la solución exacta. En particular en los problemas hiperbólicos no lineales queremos que la solución numérica obtenida converja hacia una solución débil del problema y además converja hacia la solución físicamente admisible, es decir la solución entrópica. Buscaremos pues criterios que nos aseguren que el esquema numérico en cuestión converge.

5.4.2. Esquemas Monótonos

Las soluciones débiles del problema (5.71) entrópicas verifican la siguiente propiedad: Si $u_1(x,t)$ es la solución correspondiente al valor inicial $u_1(x,0) = v_1(x)$ y $u_2(x,y)$ la solución correspondiente al valor inicial $u_2(x,0) = v_2(x)$ de manera que

$$v_1(x) \geq v_2(x)$$

en casi todos los puntos $x \in \mathbb{R}$ entonces

$$u_1(x,t) \geq u_2(x,t)$$

en casi totos los puntos $x \in \mathbb{R}$. Esta propiedad da lugar a la siguiente definición de esquema monótono:

Definición 5.11 *Diremos que un esquema numérico de la forma*

$$u_j^{n+1} = H(u_{j-l}^n, ..., u_{j+l}^n)$$

es monótono si la función H es una función creciente de sus argumentos, es decir,
Para todo $j \in \mathbb{Z}$

$$v_j \geq w_j \Rightarrow H(v_{j-l}^n, ..., v_{j+l}^n) \geq H(w_{j-l}^n, ..., w_{j+l}^n)$$

de modo que tendremos

$$v_j^0 \geq w_j^0 \Rightarrow v_j^n \geq w_j^n \quad \forall n \geq 0$$

Veremos a continuación algunos ejemplos de esquemas monótonos.

Método de Lax-Friedrichs

Vamos a estudiar en que condiciones el método de Lax-Friedrichs es monótono. El flujo numérico del método de Lax-Friedrichs es

$$g^{LF}(u,v) = \frac{1}{2}\big(f(u)+f(v)\big) - \frac{1}{2\lambda}(v-u)$$

y el método es

$$\begin{aligned} H(v_{-1},v_0,v_1) &= v_0 - \lambda\left(g^{LF}(v_0,v_1) - g^{LF}(v_{-1},v_0)\right) \\ &= \frac{v_1+v_{-1}}{2} - \frac{\lambda}{2}\big(f(v_1)-f(v_{-1})\big) \end{aligned}$$

Para que el método sea monótono es necesario y suficiente que

$$\partial_l H \geq 0 \quad l = -1,0,-1$$

Tendremos pues,

$$\begin{aligned} \partial_{-1}H &= \frac{1}{2} + \frac{\lambda}{2}f'(v_{-1}) \\ \partial_0 H &= 0 \\ \partial_1 H &= \frac{1}{2} - \frac{\lambda}{2}f'(v_1) \end{aligned}$$

Por lo tanto el esquema de Lax-Friedrichs será monótono si

$$\begin{aligned} \frac{1}{2} - \frac{\lambda}{2}f'(v_{j+1}) \geq 0 \quad \forall j \in \mathbb{Z} \\ \frac{1}{2} + \frac{\lambda}{2}f'(v_{j-1}) \geq 0 \quad \forall j \in \mathbb{Z} \end{aligned}$$

es decir si

$$-1 \leq \lambda f'(v_j) \leq 1 \Leftrightarrow \lambda \max_j |f'(v_j^n)| \leq 1$$

que es la condición de Courant-Friedrichs-Lewy (C.F.L.)

Método de Godunov

El método de Godunov se basa en la resolución de problemas locales de Riemann. Recordemos que la única solución entrópica del problema de Riemann (5.89)-(5.90) es:

1. Si $a = b$, la solución $u(x,t) = a = b$ es evidentemente una solución entrópica.
2. Si $a > b$, la discontinuidad es admisible y se propagará según una línea de discontinuidad determinada por la condición de Rankine-Hugoniot.

$$s = \frac{f(u^+) - f(u^-)}{u^+ - u^-} = \frac{f(b) - f(a)}{b - a}$$

3. Si $a < b$, la discontinuidad no es admisible y no se propagará. Es una solución clásica en las zonas de regularidad, teniendo en cuenta la condición inicial, resulta

$$u(x,t) = w_R(\frac{x}{t}, a, b)$$

donde, poniendo $\xi = \frac{x}{t}$

$$w_R(\xi, a, b) = \begin{cases} a & \text{si} \quad \xi \leq f'(a) \\ (f')^{-1}(\xi) & \text{si} \quad f'(a) \leq \xi \leq f'(b) \\ b & \text{si} \quad f'(b) \leq \xi \end{cases}$$

Esta solución es continua, salvo en $t = 0$. La discontinuidad desaparece para $t > 0$.

Supongamos ahora que tenemos una aproximación $u^n = (u_j^n)_{j \in \mathbb{Z}}$ de la solución en el tiempo t_n. El esquema de Godunov consiste en construir u^{n+1} como sigue

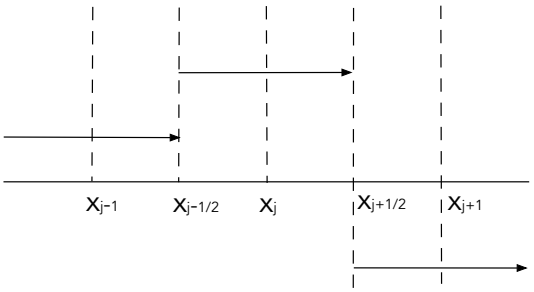


Figura 5.51 Construcción del Método de Godunov. Problemas de Riemann locales

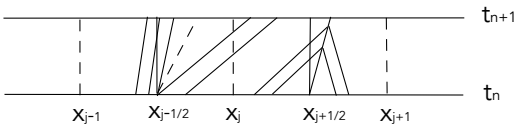


Figura 5.52 Construcción del Método de Godunov. Solución de los problemas de Riemann locales

- Primer paso:
  Se resuelve exactamente el problema

$$\frac{\partial w}{\partial t} + \frac{\partial f(w)}{\partial x} = 0 \quad x \in \mathbb{R} \quad t \in (t_n, t_{n+1}]$$
$$w(x,t_n) = v(x,t)$$

  donde

$$v(x,t_n) = u_j^n \quad x \in (x_{j-1/2}, x_{j+1/2}) \quad j \in \mathbb{Z}$$

  La función v pertenece a $L^\infty(\mathbb{R})$ y este problema tiene una única solución entrópica que podemos calcular explícitamente al menos para valores del paso k suficientemente pequeño. En consecuencia consideramos los problemas de Riemann locales centrados en los puntos $x_{j+1/2}$ para $j \in \mathbb{Z}$

$$\frac{\partial u}{\partial t} + \frac{\partial f(u)}{\partial x} = 0 \quad x \in \mathbb{R} \quad t \in (0,k]$$

$$u(x,0) = \begin{cases} u_j^n & \text{si} \quad x < x_{j+1/2} \\ u_{j+1}^n & \text{si} \quad x > x_{j+1/2} \end{cases}$$

  Dos problemas de Riemann vecinos no interaccionan con tal que la condición

$$\lambda \, \text{máx}\{|f'(v)|;\, v \in [u_{j-1}, u_j]\} \leq \frac{1}{2} \quad j \in \mathbb{Z}$$

  se cumpla. De hecho esta condición asegura que la onda que parte del punto $x_{j-1/2}$ no alcanzará las líneas $x = x_{j-1}$ y $x = x_j$ antes del tiempo k.

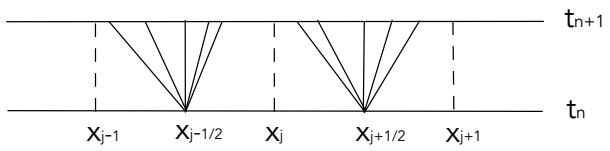


Figura 5.53 Problemas de Riemann locales

Las solución será una superposición de los problemas de Riemann locales.

$$u(x,t_{n+1}) = w_R((x-x_{j-1/2})/k, u_j^n, u_{j+1}^n) \quad x \in (x_j, x_{j+1}] \quad j \in \mathbb{Z}$$

- Segundo paso:
  u_j^{n+1} se define tomando el valor medio de la solución anterior en el intervalo $(x_{j-1/2}, x_{j+1/2})$

$$u_j^{n+1} = \frac{1}{h}\int_{x_{j-1/2}}^{x_{j+1/2}} w_R(x,t_{n+1})\, dx$$

es decir, suponiendo que la condición *C.F.L.* se cumple obtenemos

$$u_j^{n+1} = \frac{1}{h}\Big(\int_0^{h/2} w_R(x/k; u_j^n, u_{j+1}^n)\, dx + \int_{-h/2}^{0} w_R(x/k; u_{j-1}^n, u_j^n)\, dx\Big)$$

Para calcular el valor medio u_j^{n+1} resulta práctico recurrir a la forma integral de la ley de conservación. Integrando en el rectángulo $[x_{j-1/2}, x_{j+1/2}] \times [t_n, t_{n+1}]$

$$\begin{aligned}
&\int_{t_n}^{t_n}\int_{x_{j-1/2}}^{x_{j+1/2}} \Big(\frac{\partial u}{\partial t} + \frac{\partial f(u)}{\partial x}\Big)\, dx dt \\
&= \int_{x_{j-1/2}}^{x_{j+1/2}} (u(x,k) - u(x,0))\, dx + \int_0^k (f(u(x_{j+1/2}-0,t) - f(u(x_{j-1/2}+0,t))\, dt \\
&= h(u_j^{n+1} - u_j^n) - k\Big(f(w_R(0-; u_j^n, u_{j+1}^n) - f(w_R(0+; u_{j-1}^n, u_j^n)\Big) = 0
\end{aligned}$$

El esquema de Godunov se puede escribir

$$u_j^{n+1} = u_j^n - \lambda\Big(f(w_R(0-; u_j^n, u_{j+1}^n) - f(w_R(0+; u_{j-1}^n, u_j^n)\Big)$$

El esquema de Godunov es pues conservativo. Además

$$f(w_R(0-;a,b)) = f(w_R(0+;a,b)) = f(w_R(0;a,b))$$

en efecto, la función $\xi \to f(w_R(\xi;a,b))$ es continua en $\xi = 0$ pues la velocidad de propagación de la discontinuidad $s = dx/dt = \xi$ es igual a 0 en $\xi = 0$ y por la condición de Rankine-Hugoniot $[f(w_R(0)] = 0$

$$f(w_R(0-;a,b)) = f(w_R(0+;a,b))$$

de modo que el flujo numérico correspondiente al método de Godunov es

$$g^G(u,v) = f(w_R(0;u,v))$$

El cálculo anterior es correcto siempre que las ondas que parten de $x_{j-1/2}$ no interseccionan con de $x = x_{j+1/2}$ y recíprocamente. Por lo tanto el esquema de Godunov es válido con la condición

$$\lambda |f'(u_j^n)| \leq 1$$

que es la condición *C.F.L.* Con esta condición el esquema de Godunov es monótono pues es el resultado de dos pasos, cada uno de ellos monótono.

Cuando f es lineal, es decir $f'(v) = a = constante$ o más generalmente si f es estrictamente convexa, en las regiones en que f es monótona, el esquema de Godunov tiene una expresión sencilla. Tenemos,

$$w_R(0;u,v) = \begin{cases} u & \text{si} \quad f' > 0 \\ v & \text{si} \quad f' < 0 \end{cases}$$

de modo que

$$g^G(u,v) = \begin{cases} f(u) & \text{si} \quad f' > 0 \\ f(v) & \text{si} \quad f' < 0 \end{cases}$$

que es el método conocido como método "upwind".

Método de Engquist-Osher

El método de Godunov necesita resolver exactamente los problemas de Riemann locales en cada punto $x_{j+1/2}$ de la malla. Esto puede ser relativamente costoso en las aplicaciones. Vamos a introducir un método próximo al esquema de Godunov pero que no necesita resolver explícitamente el problema de Riemann. El método de Engquist-Osher se escribe

$$u_j^{n+1} = u_j^n - \frac{\lambda}{2}\big(f(u_{j+1}^n) - f(u_{j-1}^n)\big) + \frac{\lambda}{2}\Big(\int_{u_j^n}^{u_{j+1}^n} |f'(\xi)|\, d\xi - \int_{u_{j-1}^n}^{u_j^n} |f'(\xi)|\, d\xi\Big) \tag{5.108}$$

que corresponde a la función de flujo numérico

$$g^{EO}(u,v) = \frac{1}{2}\big(f(u) + f(v)\big) - \frac{1}{2}\int_u^v |f'(\xi)|\, d\xi$$

El método de Engquist-Osher es monótono si se verifica la condición *C.F.L.*. En efecto, si

$$-1 \leq \lambda f'(v_j) \leq 1 \Leftrightarrow \lambda \max_j |f'(v_j^n)| \leq 1$$

tenemos

$$\begin{aligned}
\partial_{-1} H(v_{j-1}^n, v_j^n, v_{j+1}^n) &= \frac{\lambda}{2}\left(f'(v_{j-1}) + |f'(v_{j-1}^n)|\right) \geq 0 \\
\partial_0 H(v_{j-1}^n, v_j^n, v_{j+1}^n) &= 1 - \lambda |f'(v_j^n| \geq 0 \\
\partial_1 H(v_{j-1}^n, v_j^n, v_{j+1}^n) &= \frac{\lambda}{2}\left(|f'(v_{j+1}^{n)}| - f'(u_{j+1}^n)\right) \geq 0
\end{aligned}$$

Consideremos el caso particular importante en el que la función f es estrictamente convexa. Si introducimos el punto $\bar{u} \in \mathbb{R}$ en el que $f'(\bar{u}) = 0$ pongamos

$$\begin{aligned}
f_+(v) &= f(\max(v, \bar{u})) \\
f_-(v) &= f(\min(v, \bar{u}))
\end{aligned}$$

de manera que

$$\begin{aligned}
f(v) &= f_+(v) + f_-(v) - f(\bar{u}) \\
|f'(v)| &= f'_+(v) - f'_-(v)
\end{aligned}$$

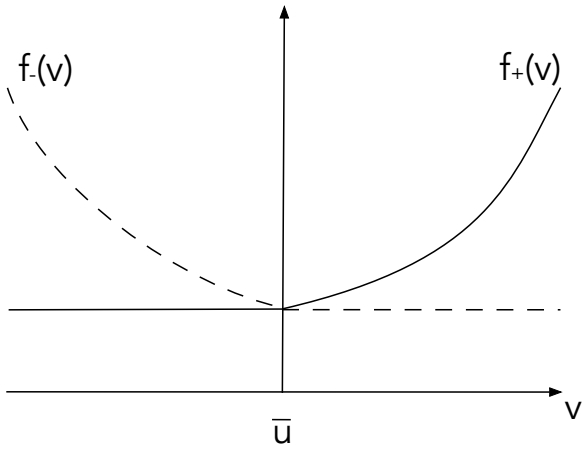


Figura 5.54 Gráficas de $f+$ y de $f-$

Tendremos pues,

$$\begin{aligned} g^{EO}(u,v) &= \frac{1}{2}\big(f(u)+f(v)\big) - \frac{1}{2}\Big(\big(f_+(v)-f_+(u)\big) - \big(f_-(v)-f_-(u)\big)\Big) \\ &= \frac{1}{2}\big(f(u)+f_+(u)-f_-(u)+f(v)-f_+(v)+f_-(v)\big) \\ &= f_+(u)+f_-(v)-f(\bar{u}) \end{aligned}$$

Como el flujo numérico g está definido salvo una constante, podemos tomar en el caso estrictamente convexo

$$g^{EO}(u,v) = f_+(u)+f_-(v)$$

lo que da el esquema

$$v_j^{n+1} = v_j^n - \lambda\Big(\big(f_-(v_{j+1}^n) - f_-(v_j^n)\big) + \big(\big(f_+(v_{j+1}^n) - f_+(v_j^n)\big)\Big)$$

Si consideramos por ejemplo la ecuación de Burgers, es decir $f(u) = \frac{1}{2}u^2$ y ($\bar{u} = 0$), tendremos

$$\begin{aligned} f_+(v) &= \frac{1}{2}v_+^2 \\ f_-(v) &= \frac{1}{2}v_-^2 \end{aligned}$$

de manera que

$$g^{EO}(u,v) = \begin{cases} \dfrac{1}{2}u^2 & \text{si} \quad u,v \geq 0 \\ \dfrac{1}{2}v^2 & \text{si} \quad u,v \leq 0 \\ \dfrac{1}{2}(u^2+v^2) & \text{si} \quad u \geq 0\ v \leq 0 \end{cases}$$

Ejercicio

Aplicar el método de Engquist-Osher para resolver el problema (5.71) donde $f(u) = \frac{1}{2}u^2$ y con la condición inicial u_0 dada por

$$u_0(x) = \begin{cases} 1 & \text{si} \quad x < -1.25 \\ 0.5 & \text{si} \quad -1.25 \leq x < 0 \\ 1 & \text{si} \quad 0 \leq x \end{cases}$$

1. Comparar los resultados para distintos valores de $\lambda = \frac{k}{h}$ y explicar los resultados.
2. Comparar los resultados con los que se obtienen utilizando el método de Lax-Friedrichs.

Solución

Con los datos son $h = 0.01$ y $k = 0.001$. De modo que $\lambda = 0.1$. La solución para distintos tiempos se representa en la figura

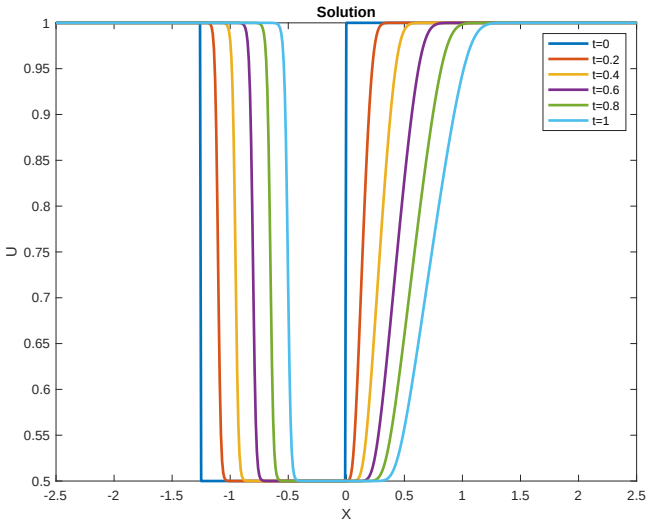


Figura 5.55 Solución numérica con el método de Engquist-Osher en distintos tiempos

La monotonía es una propiedad muy fuerte y los esquemas monótonos tienen una importante limitación. En particular los métodos conservativos monótonos son a lo sumo de orden1. Para ver esto necesitamos el teorema (5.8) de caracterización de los esquemas de orden 2.

Vamos a demostraresta limitación de los esquemas monótonos.

Teorema 5.11 *Sea un método numérico para resolver (5.71)-(5.72) de la forma (5.16) consistente y que puede ponerse en forma conservativa y en el que H es de clase C^3. Si el esquema es monótono es como máximo de orden 1.*

Demostración. Con las notaciones del teorema (5.8) demostraremos que si un esquema numérico en diferencias finitas es monótono tenemos que $\beta(u,\lambda) \geq 0$ y $\beta(u,\lambda) = 0$ solo en el caso trivial. Si el esquema es monótono $\partial_j H \geq 0$, y podemos escribir teniendo en cuenta (5.101)

$$-\lambda f'(u) = \sum_{j=-l}^{l} j\partial_j H(u,...,u) = \sum_{j=-l}^{l} j\sqrt{\partial_j H(u,...,u)}\sqrt{\partial_j H(u,...,u)}$$

y utilizando la desigualdad de Cauchy-Schwarz

$$(\lambda f'(u))^2 \leq \sum_{j=-l}^{l} j^2\partial_j H(u,...,u) \sum_{j=-l}^{l} \partial_j H(u,...,u)$$

Ahora si derivamos la relación respecto a u

$$H(u,...,u) = u$$

obtenemos

$$\sum_{j=-l}^{l} \partial_j H(u,...,u) = 1$$

de modo que

$$(\lambda f'(u))^2 \leq \sum_{j=-l}^{l} j^2\partial_j H(u,...,u) \tag{5.109}$$

por lo que

$$\beta(u,\lambda) = \frac{\sum_{j=-l}^{j=l} j^2\partial_j H(u,...,u)}{2\lambda^2} - \frac{(f'(u))^2}{2} \geq 0$$

Por otra parte tenemos que $\beta(u,\lambda) = 0$ si y solo la desigualdad (5.109) es una igualdad y esto esto ocurre si y solo en la desigualdad de Cauchy-Schwarz los vectores son proporcionales, es decir en nuestro caso, si para todo j

$$j^2\partial_j H(u,...,u)$$

es proporcional a

$$\partial_j H(u,...,u)$$

para todo j y para todo u, es decir, si y solamente si

$$\partial_j H(u,...,u) = 0$$

salvo a lo sumo para un $j = j_0$. Se deduce que

$$1 = \sum_{j=-l}^{l} \partial_j H(u,...,u) = \partial_{j_0} H(u,...,u)$$

y

$$-\lambda f'(u) = \sum_{j=-l}^{l} j\partial_j H(u,...,u) = j_0 \partial_{j_0} H(u,...,u) = j_0$$

Estamos pues en un caso lineal, $f'(u) = a =$ constante con $\lambda = -j_0/a$, $j_0 = -ak/h$ y

$$H(u_{-l},...,u_l) = u_{j_0}$$

el esquema es simplemente

$$u_j^{n+1} = u_{j+j_0}^n$$

Este esquema numérico es exacto, en efecto el error de consistencia es igual a 0 pues si $u(x,t)$ es la solución exacta en (x,t)

$$u(x,t+k) - H\big(u(x-lh,t),...,u(x+lh,t)\big) = u(x,t+k) - u(x+j_0h,t)$$
$$= u_0(x-a(t+k)) - u_0(x+j_0h-at) = 0$$

■

Para terminar este apartado, observemos que si un esquema es monótono, la ecuación (5.104) es una ecuación parabólica. El término de perturbación

$$-\lambda h \frac{\partial}{\partial x}\Big(\beta(u,\lambda)\frac{\partial u}{\partial x}\Big)$$

aparece como un término de viscosidad. Es por lo tanto de esperar que la solución obtenida mediante un método monótono tenga como límite la solución entrópica.

5.4.3. Esquemas de Variación Total Decreciente (T.V.D.)

La monotonía es una propiedad muy fuerte. Se pueden considerar métodos con propiedades no tan restrictivas y que son suficientes para asegurar la convergencia

de los métodos hacia una solución entrópica. Una clase importante son los métodos de Variación Total Decreciente (en inglés "Total Variation Diminishing"). Por otra parte la aplicación del teorema de convergencia (5.9) necesita la verificación de la consistencia y las propiedades de estabilidad suficientes que permitan además asegurar la convergencia de la sucesión generada por el método. Las propiedades de estabilidad mediante el apropiado razonamiento de compacidad permitirá obtener subsucesiones convergentes hacia la solución buscada. Empezaremos considerando diferentes normas asociadas a una sucesión $v = (v_j)_{j\in\mathbb{Z}}$.

1. Norma L^1

$$||v||_{L^1} = h\sum_{j\in\mathbb{Z}} |v_j|$$

2. Norma L^∞

$$||v||_{L^\infty} = \max_{j\in\mathbb{Z}} |v_j|$$

3. Variación Total de $v = (v_j)_{j\in\mathbb{Z}}$

$$TV(v) = \sum_{j\in\mathbb{Z}} |v_{j+1} - v_j|$$

A continuación cuando escribamos $H(v)$ donde $v = (v_j)_{j\in\mathbb{Z}}$, entendemos que es la sucesión $(H(v)_j)_{j\in\mathbb{Z}} = \big(H(v_{j-l},...,v_{j+l})\big)_{j\in\mathbb{Z}}$

Definición 5.12 *Diremos que que un esquema en diferencias finitas (5.15)*

$$u_j^{n+1} = H(u_{j-l},...,u_{j+l})$$

es de Variación Total Decreciente (T.V.D.) si verifica

$$TV(H(v)) \leq TV(v) \quad \forall v = (v_j)_{j\in\mathbb{Z}} \tag{5.110}$$

Según la definición son métodos de variación total no creciente. Lamentablemente, también los métodos T.V.D. conservativos de 3 puntos son a lo sumo de orden1, aunque a diferencia de los métodos monótonos se pueden construir métodos T.V.D. de orden 2 de 5 puntos. La utilidad de esta noción es que a partir de ella se pueden deducir propiedades de estabilidad requeridas en los teoremas de convergencia.

Definición 5.13 *Diremos que un método en diferencias finitas (5.15) es L^∞ estable si existe una constante $C > 0$ independiente de n y de k tal que la sucesión u^n generada por el método verifica*

$$||u^n||_{L^\infty} \leq C \quad \forall n \geq 0 \tag{5.111}$$

Vamos a ver que relación existe entre esquemas monótonos, esquemas T.V.D. y estabilidad L^∞. Empezaremos con el lema de Crandall-Tartar

Lema 5.4 *(Crandall-Tartar). Sea (Ω,μ) un espacio de medidad y C un subconjunto de $L^1(\Omega)$ tal que*

- $$f,g \in C \Rightarrow \text{máx}(f,g) \in C \tag{5.112}$$
- *Sea T una aplicación $T : C \to L^1(\Omega)$ que satisface*

$$\int_\Omega T(f)\,d\mu = \int_\Omega f\,d\mu \quad \forall f \in C \tag{5.113}$$

Entonces las tres propiedades siguientes son equivalentes

1. *T preserva el orden, es decir,*
   $f,g \in C$ con $f \le g$ c.t.p. $\Rightarrow$ $T(f) \le T(g)$ c.t.p.
2. $$\int_\Omega (T(f)-T(g))_+\,d\mu \le \int_\Omega (f-g)_+\,d\mu \quad \forall f,g \in C$$
3. $$\int_\Omega |T(f)-T(g)|\,d\mu \leqslant \int_\Omega |f-g|\,d\mu \quad \forall f,g \in C$$

Demostración. ▪ $1 \Rightarrow 2$. Como $\text{máx}(f,g) \in C$ y T preserva el orden

$$T(\text{máx}(f,g)) \ge T(g)$$

y igualmente

$$T(\text{máx}(f,g)) \ge T(f)$$

restando $T(g)$ de ambos lados

$$\begin{aligned} T(\text{máx}(f,g)) - T(g) &\ge 0 \\ T(\text{máx}(f,g)) - T(g) &\ge T(f) - T(g) \end{aligned}$$

y por lo tanto

$$T(\text{máx}(f,g)) - T(g) \ge \text{máx}(T(f)-T(g),0)$$

es decir

$$T(\text{máx}(f,g)) - T(g) \ge (T(f)-T(g))_+$$

Utilizando (5.112)

$$\begin{aligned} \int_\Omega (T(f)-T(g))_+\,d\mu &\le \int_\Omega T(\text{máx}(f,g)) - T(g)\,d\mu \\ &= \int_\Omega (\text{máx}(f,g)) - g)\,d\mu \end{aligned}$$

y finalmente teniendo en cuenta que

$$\text{máx}(f,g) = g + (f-g)_+$$

$$\int_\Omega (T(f)-T(g))_+ \, d\mu \leq \int_\Omega (f-g)_+ \, d\mu$$

- $2 \Rightarrow 3$ Utilizando la identidad

$$|f-g| = (f-g)_+ + (g-f)_+$$

tenemos

$$\begin{aligned}\int_\Omega |T(f)-T(g)| \, d\mu &= \int_\Omega (T(f)-T(g))_+ \, d\mu + \int_\Omega (T(g)-T(f))_+ \, d\mu \\ &\leq \int_\Omega (f-g)_+ \, d\mu + \int_\Omega (g-f)_+ \, d\mu = \int_\Omega |f-g| \, d\mu\end{aligned}$$

- $3 \Rightarrow 1$. En la práctica demostraremos $3 \Rightarrow 2 \Rightarrow 1$. Aplicando la identidad

$$(f-g)_+ = \big(|f-g| + (f-g)\big)/2$$

y la propiedad (5.113)

$$\begin{aligned}\int_\Omega (T(f)-T(g))_+ \, d\mu &= \frac{1}{2}\int_\Omega \big(|T(f)-T(g)| + T(f) - T(g)\big) \, d\mu \\ &\leq \frac{1}{2}\int_\Omega \big(|f-g| + f - g\big) \, d\mu = \int_\Omega (f-g)_+ \, d\mu\end{aligned}$$

Si suponemos ahora que $f \leq g$ c.t.p., tendremos $(f-g)_+ = 0$ de modo que

$$\int_\Omega (T(f)-T(g))_+ \, d\mu \leq 0$$

lo que implica $T(f) \leq T(g)$ c.t.p.

■

Veremos ahora que la monotonía implica la estabilidad L^∞, la estabilidad L^1, la propiedad T.V.D. así como que la aplicación H es una contracción en L^1.

Teorema 5.12 *Supongamos que el método (5.15) es conservativo y monótono. Entonces es L^∞ estable, H es una contracción en L^1 y es T.V.D. Más precisamente,*

1. *Estabilidad L^∞:*

$$||u^n||_{L^\infty} \leq ||u^0||_{L^\infty} \tag{5.114}$$

2. *Estabilidad L^1:*

$$||u^n||_{L^1} \leq ||u^0||_{L^1} \tag{5.115}$$

3. Para cada par de sucesiones $(u_j)_{j\in\mathbb{Z}}$, $(v_j)_{j\in\mathbb{Z}}$, *tenemos*

$$||H(u)-H(v)||_{L^1} \leq ||u-v||_{L^1} \tag{5.116}$$

4. Estabilidad de la Variación Total:

$$TV(u^n) \leq TV(u^0) \tag{5.117}$$

Demostración. Empezamos verificando que (propiedad del tipo máximo)

$$\min_{j-l\leq i\leq j+l} u_i \leq (H(u))_j \leq \max_{j-l\leq i\leq j+l} u_i \tag{5.118}$$

Esto implicara la estabilidad L^∞. En efecto, sea w la sucesión constante

$$w_j = \max_{i\in\mathbb{Z}} u_i = c \quad \forall j \in \mathbb{Z}$$

Como el esquema es conservativo

$$(H(w))_j = c$$

y como el método es monótono

$$H(u) \leq H(w)$$

Como $H(u)$ solo depende de $2l+1$ variables obtenemos la segunda desigualdad de (5.118). Análogamente, si elegimos w la sucesión

$$w_j = \min_{i\in\mathbb{Z}} u_i = c \quad \forall j \in \mathbb{Z}$$

y como el método es monótono

$$H(w) \leq H(u)$$

que es la primera desigualdad de (5.118). (5.118) implica

$$|H(u)_j| \leq \max_{j-l\leq i\leq j+l} |u_i|$$

y por lo tanto

$$||H(u)||_{L^\infty} \leq ||u||_{L^\infty}$$

de donde recursivamente obtenemos (5.114).

Para demostrar que H es una contracción en L^1 utilizamos el lema (5.4). Tendremos teniendo en cuenta (5.118)

$$|H(u)_j| \leq \max_{j-l\leq i\leq j+l} |u_i| \leq \sum_{i=j-l}^{j+l} |u_i|$$

$$||H(u)||_{L^1} = h\sum_{j\in\mathbb{Z}} |H(u)_j| \leq (2l+1)h\sum_{j\in\mathbb{Z}} |u_j| = (2l+1)||u||_{L^1}$$

De modo que H es una aplicación de L^1 en L^1 y además como para un método conservativo $\sum_{j\in\mathbb{Z}} H(u_{j-l},...,u_{j+l}) = \sum_{j\in\mathbb{Z}} u_j$, es decir H preserva la integral. Como el esquema es monótono, $u \leq v \Rightarrow H(u) \leq H(v)$ y gracias al lema (5.4) esta propiedad es equivalente a (5.116).

Ahora como $H(0) = 0$ deducimos primeramente

$$||H(u)||_{L^1} \leq ||u||_{L^1}$$

y obtenemos (5.115). Por otra parte podemos escribir

$$TV(H(u)) = \sum_{j\in\mathbb{Z}} |H(u)_{j+1} - H(u)_j| = \frac{1}{h}||H(w) - H(u)||_{L^1}$$

donde $w = (w_j)_{j\in\mathbb{Z}}$, con $w_j = u_{j+1}$. Finalmente se obtiene gracias a (5.116)

$$TV(H(u)) \leq \frac{1}{h}||w-u||_{L^1} = \sum_{j\in\mathbb{Z}} |u_{j+1} - u_j| = TV(u)$$

y tenemos (5.117). ■

Hemos visto que un esquema monótono es T.V.D. La recíproca no es en general cierto, sin embargo podemos demostrar que los esquemas T.V.D. preservan la monotonía, lo que en la práctica quiere decir que no producen nuevos extremos relativos. Vamos a precisar esto:

Definición 5.14 *Diremos que el método (5.15) preserva la monotonía si para toda sucesión $(v_j)_j$ monótona la sucesión $(H(v)_j)_j$ es monótona.*

Vamos a verificar que todo esquema T.V.D. preserva la monotonía. En efecto, sea $(v_j)_j$ una sucesión monótona tal que $TV(v) < \infty$. Como el esquema es de $2l+1$ puntos basta considerar una sucesión

$$v_j \begin{cases} v_i & j < J- \\ \text{monótona} & J_- \leq j \leq J_+ \\ v_d & j > J+ \end{cases}$$

de manera que

$$TV(v) = |v_d - v_i|$$

Razonando por reducción al absurdo supongamos que $w = H(v)$ no sea una sucesión monótona, de manera que exista un mínimo local v_m y un máximo local v_M.

tendremos

$$TV(w) \geq |v_d - v_i| + |v_M - v_m| \geq TV(v)$$

lo que es una contradicción.

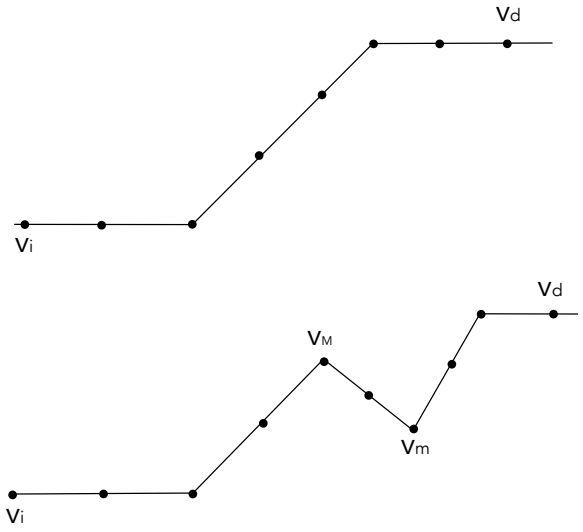


Figura 5.56 Demostración de que un método TVD preserva la monotonía

Tenemos pues las implicaciones siguientes

Esquema monótono $\Rightarrow$ Esquema T.V.D. $\Rightarrow$ Esquema que preserva la monotonía.

Los esquemas T.V.D. son pues menos restrictivos que los esquemas monótonos. Vamos a ver que con alguna hipótesis adicional los esquemas T.V.D. permiten obtener las propiedades de estabilidad suficientes para asegurar la convergencia de los mismos. El siguiente teorema proporciona para los métodos T.V.D. el resultado análogo al del teorema (5.12) para un método que sea monótono.

Teorema 5.13 *Supongamos que el esquema (5.15) es conservativo y que la función de flujo numérico asociada g es localmente lipschitziana. Entonces si el esquema es T.V.D. y L^∞- estable, existe una constante $C > 0$ que depende únicamente de la constante de estabilidad L^∞ tal que*

1.

$$||v(.,t)||_{L^1} \leq ||v(.,0)||_{L^1} + C\,T\,TV(v(.,0)) \quad 0 \leq t \leq T \tag{5.119}$$

2.

$$TV(v(.,t) \leq TV(v(.,0) \tag{5.120}$$

3.

$$||v(.,t) - v(.,s)||_{L^1} \leq C(|t-s|+k)TV(v(.,0)) \tag{5.121}$$

Demostración. Primero observemos que (5.120) no es más que la propiedad del método T.V.D., pues

$$TV(H(u)) \leq TV(u) \Rightarrow TV(H^n(u)) \leq TV(u)$$

Verifiquemos la estimación (5.119). Pongamos

$$||u||_{L^\infty} = \max_{j\in\mathbb{Z}} |v_j|$$
$$||u||_{L^1} = h\sum_{j\in\mathbb{Z}} |v_j|$$

Podemos escribir

$$H(u)_j - u_j = -\lambda(g_{j+1/2} - g_{j-1/2})$$

donde recordemos

$$g_{j+1/2} = g(u_{j-l+1}, ..., u_{j+l})$$

y

$$g_{j-1/2} = g(u_{j-l}, ..., u_{j+l-1})$$

Puesto que g es localmente lipschitziana, tenemos

$$|H(u)_j - u_j| \leq \lambda C_1(|u_{j-l+1} - u_{j-l}| + ... + |u_{j+l} - u_{j+l-1}|)$$

donde $C_1 = C_1(g, ||u||_{L^\infty})$. Así pues

$$||H(u) - u||_{L^1} \leq 2l\,C_1\,k\,TV(v)$$

Si suponemos que $||H^l(u)||_{L^\infty} \leq C_2$ para todo l, se obtiene

$$||H^{l+1}(u) - H^l(u)||_{L^1} \leq 2l\,C_2\,k\,TV(H^l(u)) \leq C_3\,k\,TV(u)$$

de donde

$$||H^n(u) - u||_{L^1} \leq \sum_{l=0}^{n-1} ||H^{l+1}(u) - H^l(u)||_{L^1} \leq C_3\,n\,k\,TV(u)$$

es decir,

$$||H^n(u)||_{L^1} \leq ||u||_{L^1} + C_3\, T\, TV(u) \quad 0 \leq t_n \leq T$$

y obtenemos (5.119). Para demostrar (5.121), observemos que si $t_m \leq s \leq t_{m+1}$ y $t_n \leq t \leq t_{n+1}$ tenemos

$$|t_n - t_m| \leq |t-s| + k$$

en efecto, podemos poner, con $0 \leq \alpha < 1$ y $0 \leq \beta < 1$

$$s = t_m + \alpha k \quad t = t_n + \beta k$$

de donde

$$|t_n - t_m| = |t - \beta k - s + \alpha k| = |t - s + (\alpha - \beta)k| \leq |t-s| + |\alpha - \beta| k \leq |t-s| + k$$

también pues v es constante a trozos

$$v(.,t) - v(.,s) = v(.,t_n) - v(.,t_m)$$

Si por ejemplo, $n > m$,

$$H^n(v) - H^m(u) = \sum_{l=m}^{n-1} \left(H^{l+1}(u) - H^l(u)\right)$$

de donde

$$\begin{aligned} ||H^n(u) - H^m(u)||_{L^1} &\leq C_3(n-m)kTV(u) = C_3(t_n - t_m)TV(u) \\ &\leq C_3(|t-s|+k)TV(u) \end{aligned}$$

■

Veamos la estabilidad L^∞ de los esquemas T.V.D. Lo demostraremos para u^0 de soporte compacto. Está claro que u^n es también de soporte compacto. Para una sucesión de soporte compacto tenemos

$$u_i = \sum_{j=i}^{\infty} u_j - u_{j+1}$$

de donde

$$||u||_{L^\infty} \leq TV(u)$$

de manera que si el esquema es T.V.D.

$$||u(.,t)||_{L^\infty} \leq TV(u(.,0)$$

Se puede demostrar que para un esquema T.V.D. con $u^0 \in BV(\mathbb{R})$ se tiene

$$||u||_{L^\infty} \leq \liminf_{j \to -\infty} |u_j^0| + TV(u^0)$$

Como consecuencia de los anteriores resultados obtenemos el siguiente resultado de convergencia para esquema T.V.D. que completa el resultado general del teorema(5.9).

Teorema 5.14 *Sea u una solución débil del problema (5.71)-(5.72). Sea (5.15) un método de diferencias finitas T.V.D. y L^∞ estable que satisface las hipótesis del teorema (5.13). Designamos mediante u_k la función definida en todo $(0,T)\times\mathbb{R}$ por*

$$u_k(x,t) = u_j^n \quad \forall\,(x,t)\in[x_{j-1/2},x_{j+1/2}]\times[t_n,t_{n+1}]$$

y donde $\lambda = k/h =$ constante

Entonces existe una sucesión de valores k, con $k\to 0$ tal que la correspondiente sucesión de soluciones $(u_k)_k$ verifica

$$u_k \to u \quad \text{en } L^\infty\big(0,T;L^1_{loc}(\mathbb{R})\big)$$

para todo $T>0$. Si además el método es entrópico la sucesión $(u_k)_k$ es única y converge hacia la única solución entrópica.

Demostración. Sea $T>0$. La inclusión canónica de $BV(\mathbb{R})\cap L^1(\mathbb{R})$ es compacta. De las estimaciones obtenidas en el teorema (5.13) y de la demostración del teorema de Ascoli se deduce que se puede extraer una subsucesión de $(u_k)_k$ que designaremos también mediante $(u_k)_k$ tal que

$$u_k \to u \quad \text{en } L^\infty\big(0,T;L^1_{loc}(\mathbb{R})\big)$$

y podemos suponer también que

$$u_k \to u \quad c.t.p.$$

La estabilidad L^∞ permite aplicar el teorema (5.9) de donde se deduce que u es solución débil del problema (5.71)-(5.72).

Si el método es además entrópico se deduce del teorema (5.10) que el límite u es la solución entrópica. Finalmente $(u_k)_k$ admite un único punto de acumulación y es toda la sucesión la que converge hacia u en $L^\infty\big(0,T;L^1_{loc}(\mathbb{R})\big)$ ■

5.4.4. Esquemas Entrópicos

Hemos visto en el teorema (5.12) que los esquemas monótonos son T.V.D. Vamos a verificar que también son entrópicos.

Teorema 5.15 *Un esquema monótono y consistente es entrópico.*

Demostración. Hemos visto que basta considerar entropías de la forma

$$U(u) = |u-r|, \quad F(u) = sgn(u-r)\big(f(u)-f(r)\big)$$

Se trata ahora de encontrar una función flujo de entropía numérica tal que

$$G(u,u,...,u) = sgn(u-r)\big(f(u)-f(r)\big)$$

y

$$|u_j^{n+1}-r| - |u_j^n - r| + \lambda\,(G_{j+1/2} - G_{j-1/2}) \leq 0$$

Pongamos,

$$a \vee b = \text{máx}(a,b), \quad a \wedge b = \text{mín}(a,b)$$

y definimos

$$G(u_{l+1},...,u_l) = g(u_{-l+1} \vee r,...,u_l \vee r) - g(u_{-l+1} \wedge r,...,u_l \wedge r)$$

y veamos que este flujo numérico verifica las condiciones requeridas. Calculamos

$$\begin{aligned} G_{j+1/2} - G_{j-1/2} = {} & g(u_{j-l+1} \vee r,...,u_{j+l} \vee r) - g(u_{j-l} \vee r,...,u_{j+l-1} \vee r) \\ & - g(u_{j-l+1} \wedge r,...,u_{j+l} \wedge r) + g(u_{j-l} \wedge r,...,u_{j+l-1} \wedge r) \end{aligned}$$

De la definición de método conservativo (5.95) obtenemos

$$\begin{aligned} -\lambda\,(G_{j+1/2} - G_{j-1/2}) = {} & H(u_{j-l} \vee r,...,u_{j+l} \vee r) - u_j \vee r \\ & - H(u_{j-l} \wedge r,...,u_{j+l} \wedge r) + u_j \wedge r \end{aligned}$$

y observando que

$$u_j \vee r - u_j \wedge r = |u_j - r|$$

obtenemos la igualdad

$$|u_j - r| - \lambda\,(G_{j+1/2} - G_{j-1/2}) = H(u_{j-l} \vee r,...,u_{j+l} \vee r) - H(u_{j-l} \wedge r,...,u_{j+l} \wedge r) \tag{5.122}$$

Ahora, para un método monótono y teniendo en cuenta la consistencia (de modo que $H(r,...,r) = r$)

$$H(u_{j-l} \vee r,...,u_{j+l} \vee r) \geq H(u_{j-l},...,u_{j+l}) \vee H(r,...,r) \geq u_j^{n+1} \vee r$$

y análogamente

$$H(u_{j-l} \wedge r,...,u_{j+l} \wedge r) \leq H(u_{j-l},...,u_{j+l}) \wedge H(r,...,r) \leq u_j^{n+1} \wedge r$$

que junto con (5.122) nos proporciona

$$|u_j^{n+1} - r| - |u_j^n| + \lambda\,(G_{j+1/2} - G_{j-1/2}) \leq 0$$

Finalmente la consistencia se verifica fácilmente, en efecto se tiene

$$G(u,...,u) = g(u \vee r,...,u \vee r) - g(u \wedge r,...,u \wedge r) = f(u \vee r) - f(u \wedge r)$$
$$= sgn(u-r)\big(f(u) - f(r)\big)$$

■

5.4.5. Comentarios adicionales

En esta sección nos hemos limitado a métodos de diferencias finitas explícitos aplicados a leyes de conservación escalares y en dimensión 1. Hay muchas referencias en la bibliografía donde se tratan métodos semidiscretos, métodos implícitos o métodos con mallados adaptativos, así como la extensión a problemas en dimensión mayor que 1. Estas notas deben pues considerarse simplemente como el punto de partida básico que proporcionrá los conocimientos mínimos necesarios para abordar el problema general de los métodos numéricos para problemas hiperbólicos no lineales en toda su extensión. Se ha hecho así pues, aún en el caso de una ley de conservación escalar y en dimensión 1, aparecen ya la problemática ligada a los problemas hiperbólicos como es la falta de regularidad de las soluciones, y de ahí la necesidad de introducir el concepto de solución débil, y en el caso no lineal, la falta de unicidad, siendo necesario completar la ley de conservación con alguna condición adicional, como la condición de entropía que permite al modelo matemático elegir la solución físicamente admisible. Estas condiciones se deben reflejar de alguna manera plausible en el método numérico.

Aunque se ha tomado como punto de partida un esquema numérico bastante general, hemos limitado el análisis numérico a métodos monótonos y métodos T.V.D. En particular un método monótono es T.V.D. La propiedad T.V.D. es algo menos restrictiva que la monotonía pero permite obtener los resultados de estabilidad suficientes para poder asegurar la convergencia de la solución numérica. Se ha visto que un método monótono tiene la limitación de ser a lo sumo de orden 1. Asimismo un método T.V.D. de 3 puntos también tiene esta limitación. Para construir métodos T.V.D. de orden superior a 1 es preciso acudir a métodos de al menos 5 puntos. También se ha introducido la noción de método entrópico que asegura que la solución numérica converge hacia la solución entrópica. En particular se ha visto que un método monótono es entrópico.

Una manera de construir métodos de segundo orden es, siguiendo la idea del método de Godunov, utilizar una aproximación lineal a trozos en lugar de una aproximación constante a trozos. Estos métodos se sustentan en la forma integral de la ley de conservación e involucra la solución exacta o aproximada de problemas de Riemann locales. Otras extensiones son el método P.P.M. (“piecewise parabolic method”) y E.N.O. (“Essentially non-oscilatory method”) que son esquemas de tipo Godunov de alto orden. Una característica común de los métodos T.V.D. es un procedimiento de limitadores de flujo para conservar la propiedad T.V.D. Ello implica que la precisión de segundo orden en las regiones de regularidad suficiente se obvia en las zonas donde no existe esta regularidad.

Referencias

1. Thomée, V., Finite difference methods for linear parabolic equations, Handbook of Numerical Analysis Volume I, Pages 5-196. Ciarlet, P.G. , Lions J.L (Editors), Ed. North-Holland, (1990)
2. Marchuk, V., Splitting and alternating direction methods, Handbook of Numerical Analysis Volume I, Pages 197-462. Ciarlet, P.G. , Lions J.L (Editors), Ed. North-Holland, (1990)
3. Godlewski, E., Raviart, P.A.: Hyperbolic systems of conservations laws. Ed. Elllipses (1984)
4. Ames W.F.: Numerical methods for Partial Diffferential Equations. Ed. Academic Press, (1975)
5. Godunov S.K.: Ecuaciones de la Fśica Matemática. Ed. Mir (1984)