Métodos Numéricos de resolución de Ecuaciones en Derivadas Parciales

Enrique Zuazua

11 de febrero de 2007

Departmento de Matemáticas
Universidad Autónoma
28049 Madrid, Spain
enrique.zuazua@uam.es

En este documento recopilamos las notas de los cursos impartidos en la UAM desde el 2001. En un primer bloque presentamos material introductorio a la resolución numérica de Ecuaciones en Derivadas Parciales (EDP) propio de la licenciatura. En un segundo bloque nos centramos esencialmente en ecuaciones de tipo ondas y analizamos la convergencia y propiedades cualitativas de los métodos de diferencias finitas a través de la transformada discreta de Fourier. En el tercer bloque estudiamos los métodos de descomposición como son el método de direccionbes alternadas y el de descomposición de dominios. En el cuarto bloque abordamos los métodos de descenso de gran utilidad en la resolución de los sistemas algebraicos a los que todo método numérico da lugar. En el quinto bloque presentamos los métodos de Galerkin en el marco de los problemas elípticos para después analizar en el bloque sexto los problemas de evolución.

Índice general

Capítulo 1

Introducción al numérico de EDP's

1.1. Introducción y motivación

Estas notas constituyen una breve guía de lo que consideramos puede y debe ser un último capítulo de un curso introductorio al Cálculo Numérico de Ecuaciones Diferenciales. En efecto, tras haber estudiado los elementos básicos del Cálculo Numérico para Ecuaciones Diferenciales Ordinarias (EDO) y los aspectos fundamentales de la aproximación numérica de la ecuación de Laplace es coherente y natural combinar y sintetizar estos conocimientos para introducirse en el mundo de las Ecuaciones en Derivadas Parciales (EDP) de evolución.

La forma en la que las EDP se presentan habitualmente en la modelización de fenómenos de la Ciencia y Tecnología es precisamente la de modelos de evolución en los que se describe la dinámica a lo largo del tiempo de determinada cantidad o variable (también a veces denominada *estado*) que puede representar objetos de lo más diversos que van desde la posición de un satélite en el espacio hasta la dinámica de un átomo, pasando por los índices bursátiles o el grado en que una enfermedad afecta a la población. En otras palabras, los modelos dinámicos o de evolución son los más naturales en la medida que reproducen nuestra propia concepción del mundo: un espacio tri-dimensional que evoluciona y cambia en el tiempo[1].

[1]Si bien la Teoría de la Relatividad establece que es mejor considerar a las cuatro del mismo modo, nuestra percepción $3+1$ está condicionada por razones puramente fisiológicas y culturales.

Cuando el estado o variable de un modelo o sistema de evolución es finito-dimensional, el modelo más natural es un sistema de EDO, cuya dimensión coincide precisamente con el del número de parámetros necesarios para describir dicho estado. Así, por ejemplo, para posicionar una partícula en el espacio necesitamos de tres variables dependientes del tiempo y para describir su dinámica un sistema de tres ecuaciones diferenciales. Pero en muchas ocasiones, como es el caso sistemáticamente en el contexto de la Mecánica de Medios Continuos, la variable de estado es infinito-dimensional. Esto ocurre por ejemplo cuando se pretende describir la deformación de cuerpos elásticos o la temperatura de un cuerpo sólido en los que la deformación o temperatura de cada uno de los puntos de ese medio continuo constituye una variable o incógnita del sistema. Los modelos matemáticos naturales en este caso son las EDP.

En la teoría clásica de EDP éstas se clasifican en tres grandes grupos: *elípticas, parabólicas e hiperbólicas.*

El modelo elíptico por excelencia involucra el *operador de Laplace*

$$\Delta = \sum_{i=1}^{N} \partial^2/\partial x_i^2 \tag{1.1.1}$$

y ha sido objeto de estudio en el capítulo anterior. La variable tiempo está ausente en este modelo. Es por eso que sólo permite describir estados estacionarios o de equilibrio.

Las ecuaciones parabólicas y las hiperbólicas, representadas respectivamente por la *ecuación del calor* y la *de ondas*, son los modelos clásicos en el contexto de las EDP de evolución. Sus características matemáticas son bien distintas. Mientras que la ecuación del calor permite describir fenómenos altamente irreversibles en tiempo en los que la información se propaga a velocidad infinita, la ecuación de ondas es el prototipo de modelo de propagación a velocidad finita y completamente reversible en tiempo.

El operador del calor es

$$\partial_t - \Delta, \tag{1.1.2}$$

de modo que al actuar sobre una función $u = u(x,t)$ que depende de la variable espacio-tiempo $(x,t) \in \mathbf{R}^N \times (0,\infty)$ tiene como resultado

$$[\partial_t - \Delta]\, u = \frac{\partial u}{\partial t} - \sum_{i=1}^{N} \frac{\partial^2 u}{\partial x_i^2}. \tag{1.1.3}$$

Sin embargo, el operador de ondas o de D'Alembert es de la forma

$$\Box = \partial_t^2 - \Delta \tag{1.1.4}$$

y da lugar a

$$\Box u = \left[\partial_t^2 - \Delta\right] u = \frac{\partial^2 u}{\partial t^2} - \Delta u. \tag{1.1.5}$$

La irreversibilidad temporal de (1.1.3) es evidente. Si hacemos el cambio de variable $t \to \widetilde{t} = -t$, el operador (1.1.3) cambia y da lugar al operador del calor retrógrado $\partial_{\widetilde{t}} + \Delta$ mientras que el operador de ondas permanece invariante.

El operador del calor y de ondas se distinguen también por sus ámbitos de aplicación. Mientras que el primero es habitual en la dinámica de fluidos (a través de una versión más sofisticada, el operador de Stokes) o en fenómenos de difusión (del calor, de contaminantes,...), el operador de ondas y sus variantes intervienen de forma sistemática en elasticidad (frecuentemente a través de sistemas más sofisticados, como el de Lamé, por ejemplo) o en la propagación de ondas acústicas o electromagnéticas (ecuaciones de Maxwell).

La Mecánica de Medios Continuos está repleta también de otras ecuaciones, operadores y modelos, pero en todos ellos, de una u otra manera, encontraremos siempre el operador del calor, de ondas o una variante muy próxima de los mismos.

Frecuentemente los modelos son más sofisticados que una "simple" ecuación aislada. Se trata a menudo de sistemas acoplados de EDP en los que es habitual encontrar tanto componentes parabólicos como hiperbólicos. Es el caso por ejemplo de las ecuaciones de la *termoelasticidad*. En estos casos, si bien un buen conocimiento de los aspectos más relevantes de la ecuación del calor y de ondas aisladamente puede no ser suficiente a causa de las interacciones de los diferentes componentes, sí que resulta indispensable.

Por todo ello es natural e importante entender todos los aspectos matemáticos fundamentales de estas dos piezas clave: la ecuación del calor y la de ondas. Evidentemente esto es también cierto desde el punto de vista del Análisis y del Cálculo Numérico.

Hasta ahora nos hemos referido sólo a las ecuaciones del calor y de ondas en su expresión más sencilla: con coeficientes constantes. Estas ecuaciones, cuando modelizan fenómenos en medios heterogéneos (compuestos por materiales de diversa naturaleza) adoptan formas más complejas y se presentan con coeficientes variables, dependientes de la variable espacial x, de la variable temporal t o de ambas.

Por limitaciones de tiempo nos centraremos esencialmente en el estudio de estas ecuaciones en el caso más sencillo de los coeficientes constantes y lo haremos, sobre todo, en una variable espacial. A pesar de ello, creemos que quien asimile bien los conceptos que aquí expondremos y entienda las técnicas y los resultados principales que presentaremos estará en condiciones de abordar con

éxito situaciones más complejas, incluyendo EDP con coeficientes variables y en varias dimensiones espaciales.

En esta introducción no hemos mencionado para nada otras palabras clave en la modelización de fenómenos complejos como son los términos "no-linealz "no-determinista". La aproximación numérica de modelos de EDP que involucran estos fenómenos queda fuera de los objetivos de este curso pero, nuevamente, se puede asegurar que los elementos que aquí expondremos serán sin duda de gran utilidad, si no indispensables, a la hora de adentrarse en otros modelos más complejos que involucren términos no-lineales y estocásticos.

Habiendo ya motivado la necesidad de proceder al desarrollo de métodos numéricos para la resolución de la ecuación del calor y de la ecuación de ondas, veamos cual es la forma o, más bien, cuáles son las formas más naturales de proceder. Hay al menos tres

a) Discretizamos simultáneamente las variable de espacio y de tiempo. De este modo pasamos directamente de la EDP de evolución a un sistema puramente discreto. Es lo que se denomina una *discretización completa*.

b) Mantenemos la variable temporal continua y discretizamos la variable espacial. En este caso se trata de una *semi-discretización espacial* y el problema se reduce a un sistema de ecuaciones diferenciales de dimensión igual al de nodos espaciales que tenga el mallado utilizado en la discretización espacial. Estos métodos se conocen también como *métodos de líneas.*

c) Mantenemos la variable espacial continua y discretizamos el tiempo. Se trata en este caso de una *semi-discretización temporal*. El sistema se reduce a la resolución iterada, discretamente en tiempo, de ecuaciones de Laplace.

De entre todas estas vías la última es la menos habitual (si bien se trata de un método frecuente a la hora de probar resultados analíticos de existencia de soluciones, idea que inspira, por ejemplo, la teoría de semigrupos no-lineales) y actualmente es objeto de estudio intensivo de cara, en particular, a desarrollar algorítmos paralelizables.

Desde un punto de vista estrictamente computacional sólo la primera es válida y realmente programable en el ordenador. Pero hay varias razones para no descartar la segunda. En primer lugar, cuando realizamos la discretización espacial estamos sustituyendo la dinámica infinito-dimensional de la EDP por una dinámica en dimensión finita. El estudio de la legitimidad de esta sustitución, en sí, es ya un objetivo interesante no sólo desde un punto de vista práctico sino también en el plano conceptual. Por supuesto, en este empeño lo estudiado sobre

la ecuación de Laplace nos resultará de suma utilidad pues la semi-discretización consiste precisamente en discretizar el laplaciano en la variable espacial dejando intacta la variable temporal. Una vez justificada la idoneidad de esta primera discretización espacial, lo cual pasa obviamente por un análisis de la convergencia a medida que el paso del mallado espacial tiende a cero, nos encontramos pues frente a un sistema de EDO. Algunos de los programas comerciales (Matlab, por ejemplo) están equipados de rutinas de resolución de EDO. Esto hace que se puedan obtener con facilidad aproximaciones numéricas y visualizaciones gráficas de las soluciones de dicho sistema de EDO y, por consiguiente de la EDP, lo cual supone sin duda una razón importante para proceder de este modo. Pero no debemos olvidar que la teoría desarrollada en la primera parte de este curso está precisamente orientada a la discretización temporal de sistemas de EDO, con su consiguiente análisis de convergencia. Al final de este doble proceso de aproximación nos encontraremos por tanto con un sistema completamente discreto, igual que si hubiesemos procedido directamente por la primera vía, pero esta vez lo haremos habiendo utilizado las dos teorías de convergencia previamente desarrolladas. Ni que decir tiene que, si realizamos los dos procesos de aproximación (el del laplaciano y el de la EDO) con cuidado, obtendremos no sólo los resultados de convergencia de las soluciones del problema discreto al continuo sino estimaciones del error. Las estimaciones de error junto con el coste computacional del método numérico es lo que al final establece su bondad.

Es por esto que, en cada uno de los ejemplos (ecuación del calor y de ondas) analizaremos las dos primeras vías: discretización completa y semi-discretización espacial. Por supuesto lo que aquí presentaremos no serán más que algunos aspectos, conceptos y resultados básicos y fundamentales. El lector interesado en un análisis más detallado encontrará en los textos de la Bibliografía que incluimos al final de estas notas un excelente material para profundizar en el estudio de este campo, además de diversas y útiles referencias complementarias.

Tanto en la sección dedicada a la ecuación del calor como a la de ondas comenzaremos recordando algunas de sus propiedades analíticas más importantes, para después abordar los aspectos numéricos. No conviene olvidar que la eficacia de un método numérico depende en gran medida de la fidelidad con que consigue reproducir a nivel discreto las propiedades analíticas del modelo continuo.

Antes de concluir esta introducción conviene hacer una observación sobre la notación que usaremos a lo largo de las notas:

1. • El índice j será utilizado para denotar la componente j-ésima de la solución numérica, que será de hecho una aproximación de la solución de

la EDP en el punto nodal $x_j = jh$, siendo $h = \Delta x$ el paso del mallado espacial.

2. • El exponente k se utiliza para denotar la solución numérica en el paso temporal k-ésimo, aproximación de la solución de la EDP en el instante de tiempo $t = k\Delta t$.

3. • El superíndice $\hat{\cdot}$ se utiliza para denotar las componentes de Fourier de las soluciones tanto en el marco continuo como en el discreto.

1.2. La ecuación del calor

Como hemos mencionado anteriormente, la ecuación del calor es el prototipo de ecuación de evolución de tipo parabólico cuyas variantes están presentes de manera sistemática en todos los modelos matemáticos de la difusión y de la Mecánica de Fluidos.

Como hemos dicho antes, la ecuación del calor es un modelo fuertemente irreversible en tiempo en el que la información se propaga a velocidad infinita. Estas propiedades quedarán claramente de manifiesto en la siguiente sección en la que recordamos sus principales propiedades analíticas.

1.2.1. Propiedades básicas de la ecuación del calor

Consideremos en primer lugar el problema de Cauchy

$$\begin{cases} u_t - \Delta u = 0 & \text{en} \quad \mathbf{R}^N \times (0,\infty) \\ u(x,0) = \varphi(x) & \text{en} \quad \mathbf{R}^N. \end{cases} \tag{1.2.1}$$

Se trata de un problema característico en el sentido de Cauchy-Kowaleski (ver F. John [30]). Precisamente por serlo cabe esperar que (1.2.1) esté bien planteado a pesar de que no damos dos datos de Cauchy como es habitual en una ecuación de orden dos, sino sólo una.

La solución fundamental de (1.2.1) se puede calcular explícitamente. Obtenemos así el núcleo de Gauss:

$$G(x,t) = (4\pi t)^{-N/2} \exp\left(- \mid x \mid^2 /4t\right). \tag{1.2.2}$$

No es difícil comprobar que G es efectivamente la solución de (1.2.1) con $\varphi = \delta_0$, la delta de Dirac en $x = 0$[2].

[2]Recordemos que δ_0 es la medida tal que $< \delta_0, \phi >= \phi(0)$ para toda función continua ϕ.

Por consiguiente, para "cualquier" φ,

$$u = G * \varphi \tag{1.2.3}$$

representa la única solución de (1.2.1). (Hemos entrecomillado el cuantificador "cualquier" puesto que se requieren algunas condiciones mínimas sobre φ y la propia solución para que ésta pueda escribirse de manera única como en (1.2.3). Basta por ejemplo con tomar $\varphi \in L^2(\mathbf{R}^N)$ o $\varphi \in L^\infty(\mathbf{R}^N)$ y buscar soluciones u tales que, para cualquier $t > 0$, sean funciones acotadas (véase F. John [30])).

En (1.2.3) $*$ representa la convolución espacial de modo que

$$u(x,t) = (4\pi t)^{-N/2} \int_{\mathbf{R}^N} \exp\left(- \mid x - y \mid^2 /4t\right) \varphi(y) dy. \tag{1.2.4}$$

En esta expresión se observa inmediatamente la velocidad infinita de propagación. En efecto, todos los valores de φ, en cualquier punto y de $\mathbf{R}^n$, intervienen a la hora de calcular u en cualquier punto espacio-temporal (x,t).

En (1.2.4) es también fácil comprobar el enorme efecto regularizante de la ecuación del calor. En efecto, basta que $\varphi \in L^1(\mathbf{R}^N)$ o que $\varphi \in L^\infty(\mathbf{R}^N)$ para que la solución $u(\cdot,t)$[3] en cada instante $t > 0$ sea una función de $C^\infty(\mathbf{R}^N)$. Este efecto regularizante implica también la irreversibilidad temporal.[4]

De la fórmula (1.2.4) se deducen otras propiedades de la solución de la ecuación del calor:

- *Principio del máximo*: Si $\varphi \geq 0$ entonces $u \geq 0$ y en realidad $u > 0$ en $\mathbf{R}^N \times (0,\infty)$ salvo que $\varphi \equiv 0$.

- *Conservación de la masa*:

$$\int_{\mathbf{R}^N} u(x,t)dx = \int_{\mathbf{R}^N} \varphi(x)dx, \; \forall t > 0. \tag{1.2.5}$$

- *Decaimiento*:

$$\parallel u(t) \parallel_{L^\infty(\mathbf{R}^N)} \leq Ct^{-N/2} \parallel \varphi \parallel_{L^1(\mathbf{R}^N)}, \; \forall t > 0. \tag{1.2.6}$$

Todas ellas admiten claras interpretaciones físicas y obedecen, efectivamente, al comportamiento habitual en un proceso de difusión.

[3] Interpretamos la función $u = u(x,t)$ como una función del tiempo t que, a cada instante t, tiene como imagen una función de x que varía en el tiempo.

[4] En efecto, si la ecuación del calor estuviese bien puesta en el sentido retrógrado del tiempo, como la solución es regular para $t > 0$, volviendo hacia atrás en el tiempo, obtendríamos en el instante inicial $t = 0$ una función $C^\infty(\mathbf{R}^N)$. De este modo acabaríamos probando que toda función de $L^1(\mathbf{R}^N)$ o $L^\infty(\mathbf{R}^N)$ está en $C^\infty(\mathbf{R}^N)$, cosa falsa evidentemente.

Consideramos ahora el problema de la difusión del calor en un dominio acotado Ω de $\mathbf{R}^N$. En esta ocasión, con el objeto de que el sistema de ecuaciones sea completo tenemos también que imponer condiciones de contorno que determinen la interacción del medio Ω con el medio circundante. Desde un punto de vista matemático las condiciones más simples son las de Dirichlet. Obtenemos así el sistema

$$\begin{cases} u_t - \Delta u = 0 & \text{en} \quad \Omega \times (0,\infty) \\ u = 0 & \text{en} \quad \partial\Omega \times (0,\infty) \\ u(x,0) = \varphi(x) & \text{en} \quad \Omega. \end{cases} \tag{1.2.7}$$

Las condiciones de contorno $u = 0$ en $\partial\Omega$ indican que las paredes del dominio Ω se mantienen a temperatura constante $u = 0$. En la práctica, frecuentemente, se utilizan otras condiciones de contorno no tanto sobre la variable u que en la ecuación del calor representa la temperatura, sino sobre el flujo de calor a través de la frontera. Así, por ejemplo, en el caso en que queramos representar que el dominio Ω está completamente aislado de su entorno impondremos condiciones de flujo nulo, i.e.

$$\frac{\partial u}{\partial n} = 0 \text{ en } \partial\Omega \times (0,\infty).$$

Aquí $\partial/\partial n$ denota el operador derivada normal y n es el vector normal exterior unitario a $\partial\Omega$ que varía en función de la geometría del dominio al variar el punto $x \in \partial\Omega$. Se trata de una derivada direccional, de modo que

$$\frac{\partial}{\partial n} = \nabla \cdot n,$$

donde ∇ denota el operador gradiente $\nabla = \left(\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_N}\right)$ y $\cdot$ el producto escalar euclideo en $\mathbf{R}^N$.

Pero, con el objeto de simplificar y no hacer demasiado larga la presentación, en estas notas nos limitaremos a considerar las condiciones de contorno de Dirichlet como en (1.2.7).

En este caso la solución no es tan fácil de obtener explícitamente como lo fue para el problema de Cauchy en $\mathbf{R}^N$. Son diversos los métodos disponibles para su resolución: Galerkin, semigrupos, series de Fourier, ... El lector interesado en el estudio de estos métodos puede consultar el texto de L. Evans [7].

Nosotros nos centraremos en el problema de una sola dimensión espacial. Consideraremos por lo tanto el sistema

$$\begin{cases} u_t - u_{xx} = 0, & 0 < x < \pi, \quad t > 0 \\ u(0,t) = u(\pi,t) = 0, & t > 0 \\ u(x,0) = \varphi(x), & 0 < x < \pi. \end{cases} \tag{1.2.8}$$

En este caso la solución puede obtenerse fácilmente mediante el desarrollo en series de Fourier. En efecto, las funciones trigonométricas

$$w_l(x) = \sqrt{\frac{2}{\pi}}\,\mathrm{sen}(lx),\ l \geq 1 \tag{1.2.9}$$

constituyen una base ortonormal de $L^2(0,\pi)$ (v́ease Lema 2.1).

Por lo tanto, para cualquier función $\varphi \in L^2(0,\pi)$ la solución u de (1.2.8) se puede escribir en la forma

$$u(x,t) = \sum_{l=1}^{\infty} \hat{\varphi}_l e^{-l^2 t} w_l(x) \tag{1.2.10}$$

donde $\{\hat{\varphi}_l\}_{l\geq 1}$ son los coeficientes de Fourier de la función φ, i.e.

$$\hat{\varphi}_l = \int_0^{\pi} \varphi(x) w_l(x) dx. \tag{1.2.11}$$

Esta expresión de la solución en series de Fourier nos resultará de gran utilidad a la hora de abordar la aproximación numérica de la solución. En realidad, las propias sumas parciales de la serie proporcionan ya una manera sistemática de aproximar la solución. Así, para cada $M \in \mathbf{N}$ podemos introducir

$$u_M(x,t) = \sum_{j=1}^{M} \hat{\varphi}_j e^{-j^2 t} w_l(x), \tag{1.2.12}$$

y es entonces fácil comprobar que

$$\| u(t) - u_M(t) \|_{L^2(0,\pi)} \leq e^{-(M+1)^2 t/2} \| \varphi \|_{L^2(0,\pi)},\ \forall t \geq 0, \tag{1.2.13}$$

lo cual indica, efectivamente, que la aproximación de u mediante u_M mejora a medida que $M \to \infty$.

En vista de la aparente simplicidad de este método de aproximación cabe entonces preguntarse: >Para qué necesitamos otros métodos?

Las razones son diversas, pero hay una particularmente importante. Si bien en este caso la obtención de las funciones de base $\{w_l\}_{l\geq 1}$ (que son, en realidad, autofunciones del operador de Laplace involucrado en la ecuación del calor con condiciones de contorno de Dirichlet) es muy simple por encontrarnos en una dimensión espacial, en varias dimensiones espaciales el problema es mucho más complejo, pues pasa por calcular las autofunciones del problema:

$$\begin{cases} -\Delta w = \lambda w & \text{en} \quad \Omega \\ w = 0 & \text{en} \quad \partial\Omega. \end{cases} \tag{1.2.14}$$

Antes que nada conviene señalar que las autofunciones w_l de (1.2.9) se obtienen precisamente al resolver el análogo uni-dimensional de (1.2.14). En este caso el problema de autovalores es un sencillo problema de Sturm-Liouville que se escribe en la forma

$$\begin{cases} -w'' = \lambda w, & 0 < x < \pi \\ w(0) = w(\pi) = 0. \end{cases} \tag{1.2.15}$$

Los autovalores son en este caso

$$\lambda_l = l^2,\ l \geq 1 \tag{1.2.16}$$

y las autofunciones correspondientes, una vez normalizadas en $L^2(0,\pi)$, las funciones trigonométricas (1.2.9).

Si bien la teoría espectral garantiza la existencia de una sucesión de autofunciones que constituyen una base ortogonal de $L^2(\Omega)$ ([2]), su forma depende de la geometría del dominio Ω y, por supuesto, su cálculo explícito es imposible salvo para dominios muy particulares ([30]). Por lo tanto, en varias dimensiones espaciales, la utilización de estas autofunciones exige previamente el desarrollo de métodos numéricos para su aproximación, tan elaborados (o más) como los que vamos a necesitar para aproximar la propia ecuación del calor directamente.

Este hecho, junto con otro igualmente importante como es que para muchas ecuaciones (no-lineales, coeficientes dependientes del espacio-tiempo, etc.) la resolución mediante series de Fourier no es posible, aconsejan que desarrollemos métodos numéricos que permitan abordar sistemáticamente la ecuación del calor y sus variantes, sin pasar por la Teoría Espectral.

Sí que conviene sin embargo utilizar este formalismo de Fourier para entender las aproximaciones que los diferentes esquemas proporcionan a la ecuación del calor y el modo en que afectan a las diferentes componentes de las soluciones en función de la frecuencia de su oscilación espacial.

Volvamos entonces a la ecuación (1.2.8) y a su solución (1.2.10).

En la expresión (1.2.10) se observa un comportamiento de u distinto al del problema de Cauchy en $\mathbf{R}^N$.

En efecto, en este caso es fácil comprobar que la solución decae exponencialmente cuando $t \to \infty$:

$$\| u(t) \|^2_{L^2(0,\pi)} = \sum_{j=1}^{\infty} | \hat{\varphi}_j |^2 \, e^{-2j^2 t} \leq e^{-2t} \sum_{j=1}^{\infty} | \hat{\varphi}_j |^2 = e^{-2t} \| \varphi \|^2_{L^2(0\pi)} \, . \tag{1.2.17}$$

Esta propiedad de decaimiento puede también obtenerse directamente de la ecuación (1.2.8) mediante el *método de la energía*, sin hacer uso del desarrollo

en serie de Fourier de la solución. En efecto, multiplicando en (1.2.8) por u e integrando por partes se obtiene que

$$0 = \int_0^\pi (u_t - u_{xx})\, u dx = \frac{1}{2}\frac{d}{dt}\int_0^\pi u^2 dx + \int_0^\pi u_x^2 dx,$$

o, lo que es lo mismo,

$$\frac{1}{2}\frac{d}{dt}\int_0^\pi u^2 dx = -\int_0^\pi u_x^2 dx. \tag{1.2.18}$$

Utilizamos ahora la desigualdad de Poincaré ([2])

$$\int_0^\pi u_x^2 dx \geq \int_0^\pi u^2 dx,\ \forall u \in H_0^1(0,\pi) \tag{1.2.19}$$

que, combinada con la identidad (1.2.18), proporciona la desigualdad

$$\frac{d}{dt}\int_0^\pi u^2 dx \leq -2\int_0^\pi u^2 dx. \tag{1.2.20}$$

Integrando esta desigualdad (1.2.20) obtenemos exactamente la tasa exponencial de decaimiento de la solución que predijimos en (1.2.17).

Observación 1.2.1 La *desigualdad de Poincaré* (ver [B]) garantiza que

$$\int_0^\pi |a'(x)|^2\, dx \geq \int_0^\pi |a(x)|^2\, dx,\ \forall a \in H_0^1(0,\pi). \tag{1.2.21}$$

La mejor constante de la desigualdad (1.2.21) viene caracterizada por el siguiente principio de minimalidad que involucra el cociente de Rayleigh:

$$\lambda_1 = \min_{a \in H_0^1(0,\pi)} \frac{\displaystyle\int_0^\pi |a'(x)|^2\, dx}{\displaystyle\int_0^\pi a^2(x) dx}. \tag{1.2.22}$$

En este caso $\lambda_1 = 1$ puesto que se trata del primer autovalor λ_1 del operador $-d^2/dx^2$ en $H_0^1(0,\pi)$ que posee una sucesión de autovalores (1.2.16).

■

1.2.2. Semi-discretización espacial: El método de Fourier

Esta sección está dedicada a estudiar las semi-discretizaciones espaciales de la ecuación del calor $1-d$ (unidimensional) (1.2.8).

Lo haremos en el caso más sencillo en el que el operador de Laplace espacial se aproxima mediante el esquema clásico y sencillo de tres puntos. Analizaremos la

convergencia del método tanto mediante series de Fourier como por estimaciones de energía.

Si bien los resultados de esta sección se refieren a un problema muy sencillo como es (1.2.8), en el transcurso de la misma desarrollaremos una metodología susceptible de ser adaptada a situaciones más complejas. Esto es así, muy en particular en lo referente al método de la energía, de fácil aplicación a otras condiciones de contorno, coeficientes variables, ecuaciones no-lineales, etc.

Consideramos por tanto un paso $h > 0$ del mallado espacial. Para simplificar la presentación suponemos que $h = \pi/(M+1)$ con $M \in \mathbf{N}$, de modo que la partición que definen los nodos

$$x_j = jh,\ j = 0, \ldots, M+1 \tag{1.2.23}$$

descompone el intervalo $[0, \pi]$ en $M+1$ subintervalos de longitud h : $I_j = [x_j,\ x_{j+1}]$, $j = 0, \ldots,\ M$. Obsérvese que el primer y el último nodo corresponden a los extremos del intervalo, i.e. $x_0 = 0$, $x_{M+1} = \pi$.

Utilizando la clásica aproximación de tres puntos para el operador d^2/dx^2 (que, como vimos, es de orden dos) obtenemos de manera natural la siguiente semi-discretización de (1.2.8):

$$\begin{cases} u_j' + \frac{2u_j - u_{j+1} - u_{j-1}}{h^2} = 0, & j = 1, \ldots, M, \quad t > 0 \\ u_0 = u_{M+1} = 0, & t > 0 \\ u_j(0) = \varphi_j, & j = 1, \ldots, M. \end{cases} \tag{1.2.24}$$

Este sistema constituye un conjunto de M ecuaciones diferenciales de orden uno, lineales, acopladas de tres en tres con M incógnitas. En vista de que, por las condiciones de contorno, $u_0 \equiv u_{M+1} \equiv 0$, las genuinas incógnitas del problema son las M funciones $u_j(t)$, $j = 1, \ldots,\ M$.

Cada una de las funciones $u_j(t)$ proporciona una aproximación de la solución $u(\cdot, t)$ en el punto $x = x_j$. A medida que el paso h de la discretización tiende a cero tenemos más y más puntos en el mallado. Cabe por tanto esperar que obtengamos progresivamente estimaciones más finas de la solución. Sin embargo, tal y como veremos, no basta con que una aproximación parezca coherente para poder garantizar su convergencia. El objetivo principal de este capítulo es precisamente desarrollar una teoría que nos permita discernir si un método numérico es convergente o no.

Queda sin embargo por determinar una buena elección de los datos iniciales φ_j, $j = 1, ..., M$. Las posibilidades son diversas y algunas de ellas serán analizadas a lo largo de estas notas. Cuando la función φ del dato inicial de la ecuación del calor es continua, lo más sencillo es tomar sus valores en los nodos como

dato inicial del problema semi-discreto, i.e. $\varphi_j = \varphi(x_j)$. Cuando φ no es continua sino meramente integrable podemos también hacer medias del dato inicial $\varphi = \varphi(x)$ en intervalos en torno a los nodos. También es posible elegir los datos iniciales del sistema semi-discreto truncando la serie de Fourier del dato inicial de la ecuación del calor.

Las posibilidades son diversas pero, si un método es convergente, ha de serlo para cualquier elección razonable de los datos iniciales. Esto dependerá esencialmente del esquema elegido para aproximar la ecuación y las condiciones de contorno.

Frecuentemente utilizaremos una notación vectorial para simplificar las expresiones. Introducimos por tanto el vector columna $\vec{u} = \vec{u}(t)$ que representa a la incógnita del sistema (1.2.24):

$$\vec{u}(t) = \begin{pmatrix} u_1(t) \\ \vdots \\ u_M(t) \end{pmatrix}, \tag{1.2.25}$$

y la matriz tridiagonal:

$$A_h = \frac{1}{h^2} \begin{pmatrix} 2 & -1 & 0 & 0 \\ -1 & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & -1 \\ 0 & 0 & -1 & 2 \end{pmatrix}. \tag{1.2.26}$$

Así, el sistema (1.2.24) se escribe

$$\begin{cases} \vec{u}\,'(t) + A_h \vec{u}(t) = 0, \quad t > 0 \\ \vec{u}(0) = \vec{\varphi}. \end{cases} \tag{1.2.27}$$

Obviamente la solución $\vec{u}$ de (1.2.27) también depende de h de modo que sería más legítimo denotarla mediante el subíndice h: $\vec{u}_h$. Pero, para simplificar la notación, escribiremos simplemente $\vec{u}$, salvo que este hecho pueda conducir a confusión.

En la sección anterior vimos que que la ecuación del calor continua verifica el *principio del máximo* que garantiza que si el dato inicial es no-negativo, la solución lo es para todo x y todo t. Esto es cierto para el problema de Cauchy en todo el espacio pero también lo es para el problema de Dirichlet con condiciones de contono nulas. El sistema (1.2.24) refleja también esta propiedad de naturaleza física. En efecto, supongamos que el dato inicial de (1.2.24) es positivo, i.e. $\varphi_j > 0$, $j = 1, ..., M$. Entonces, $u_j(t) > 0$ para todo $j = 1, ..., M$

y todo $t > 0$. Con el objeto de probar esta afirmación argumentamos del modo siguiente. En primer lugar observamos que, por continuidad, existe $\tau > 0$ tal que $u_j(t) > 0$ para todo j y todo $0 \leq t \leq \tau$. Sea τ^* el primer instante de tiempo en el que la solución se anula en alguna de sus componentes que denotaremos mediante el índice j^*. Tenemos entonces:

- $u_j(t) > 0$, $\forall j = 1, ..., M$, $\forall 0 \leq t < \tau^*$.
- $u_{j^*}(\tau^*) = 0$.
- $u'_{j^*}(\tau^*) \leq 0$.
- $u_j(\tau^*) \geq 0$, $j = 1, ..., M$.

Haciendose uso de estas propiedades escribimos la ecuación de (1.2.24) correspondiente al índice j^* en el instante $t = \tau^*$. Obtenemos $u_{j^*+1}(\tau^*) + u_{j^*-1}(\tau^*) \leq 0$, lo cual implica, en virtud de las propiedades anteriores, que $u_{j^*+1}(\tau^*) = u_{j^*-1}(\tau^*) = 0$. Iterando este argumento se obtiene que $u_j(\tau^*) = 0$ para todo $j = 1, ..., M$. Por unicidad de las soluciones de la ecuación diferencial (1.2.24) esto implica que $\vec{u} \equiv 0$, lo cual está en contradicción con el hecho de que el dato inicial sea positivo. Esto prueba que el principio del máximo se verifica también para el sistema semi-discreto (1.2.24).

En el dato inicial de u_j hemos tomado el valor exacto del dato inicial φ de la ecuación del calor en el punto x_j. Esto exige que el dato φ sea continuo. Pero existen otras elecciones del dato inicial de la ecuación semi-discreta como decíamos anteriormente. En particular, la elección puede hacerse a través de los coeficientes de Fourier de φ.

El punto de vista de Fourier no sólo es sumamente útil a la hora de entender la teoría analítica de las EDP sino también su Análisis Numérico. En efecto, la solución de (1.2.27) puede escribirse en serie de Fourier en la base de autovectores de la matriz A_h. En este caso será simplemente una suma finita de M términos pues se trata efectivamente de un problema finito-dimensional de dimensión M. Para ello es preciso introducir el espectro de la matriz A_h.

Consideremos por tanto el problema de autovalores:

$$A_h \vec{W} = \lambda \vec{W}. \tag{1.2.28}$$

Los autovalores y autovectores solución de (1.2.28) pueden calcularse de forma explícita:

$$\lambda_l(h) = \frac{4}{h^2} \operatorname{sen}^2 \left(l \frac{h}{2} \right), \ \ell = 1, ..., M. \tag{1.2.29}$$

Los autovectores correspondientes son

$$\vec{W}_l(h) = \sqrt{\frac{2}{\pi}} \begin{pmatrix} \text{sen}(lx_1) \\ \vdots \\ \sin(lx_M) \end{pmatrix}, \, l = 1, ..., M. \tag{1.2.30}$$

El lector interesado puede encontrar una prueba de este hecho en [14], Lema 10.5.

A partir de ahora las componentes del vector $\vec{W}_l(h)$ serán denotadas mediante $(W_{l,j})_{j=1,...,M}$.

En (1.2.29)-(1.2.30) se observan varias analogías con los autovalores y autofunciones del operador de Laplace expresados en (1.2.9) y (1.2.16). En efecto, para cada índice $l \geq 1$ fijo tenemos

$$\lambda_l \to l^2, \text{ cuando } h \to 0, \tag{1.2.31}$$

lo cual refleja que los autovalores del problema discreto (1.2.28) aproximan a los del continuo (1.2.15) a medida que $h \to 0$, que es a su vez consecuencia de la convergencia del esquema de tres puntos para la aproximación del Laplaciano probada en el capítulo anterior. Por otra parte, los autovectores $\vec{W}_l(h)$ del problema discreto (1.2.28) no son más que una restricción a los puntos del mallado de las autofunciones (1.2.9) del problema continuo. Esto explica por tanto la proximidad de ambos espectros. Es oportuno indicar también que los autovectores $\vec{W}_l(h)$ dependen de h de dos maneras: Por su número de componentes $M = \pi/h - 1$ y por el valor de cada una de ellas.

Conviene sin embargo señalar que, en general, cuando se abordan problemas más sofisticados (en varias dimensiones espaciales, por ejemplo) no es frecuente que se dé la coincidencia exacta entre los autovectores del problema discreto y las autofunciones del continuo sino simplemente la convergencia a medida que $h \to 0$, si bien, para que ésto sea cierto, es indispensable que el esquema numérico elegido para la aproximación de la ecuación de Laplace sea convergente.

Por último es interesante también observar que de la expresión explícita de los autovalores $\lambda_l(h)$ del problema discreto se deduce que

$$\lambda_l(h) \leq \lambda_l = l^2. \tag{1.2.32}$$

Las soluciones del problema discreto, como hemos visto, son vectores columna de $\mathbf{R}^M$ y en el caso del problema de evolución, funciones regulares del tiempo t a valores en $\mathbf{R}^M$. En $\mathbf{R}^M$ consideramos la norma euclidea escalada

$$\| \vec{e} \|_h = \left[h \sum_{j=1}^{M} | e_j |^2 \right]^{1/2}, \, \forall \vec{e} = (e_1, ..., e_M)^t, \tag{1.2.33}$$

y su producto escalar asociado

$$\langle \vec{e},\, \vec{f} \rangle_h = h \sum_{j=1}^{M} e_j f_j. \tag{1.2.34}$$

El factor de escala introducido en la norma y producto escalar ($\sqrt{h}$ y h respectivamente) es importante pues garantiza que, cuando $h \to 0$, estas normas y producto escalar aproximan a las correspondientes de $L^2(0,\pi)$:

$$\| e \|_{L^2(0,\pi)} = \left(\int_0^\pi e^2(x)dx \right)^{1/2}, \tag{1.2.35}$$

$$\langle e, f \rangle_{L^2(0,\pi)} = \int_0^\pi e(x)f(x)dx. \tag{1.2.36}$$

En efecto, en vista de que $h = \pi/(M+1)$, se observa inmediatamente que (1.2.33) y (1.2.34) son versiones discretas de las integrales (1.2.35) y (1.2.36), semejantes a las sumas de Riemann.

Es por tanto natural, abusando un poco del lenguaje, referirse a este producto euclideo escalado, como el producto en L^2.

Es fácil comprobar que los autovectores $\vec{W}_l(h)$ de (1.2.30) son ortonormales en el producto escalar (1.2.34).

En efecto, tenemos el siguiente resultado:

Lemma 1.2.1 *Para cada h de la forma $h = \pi/(M+1)$, con $M \in \mathbf{N}$ y para cada $l \in \mathbf{N}$, $1 \le l \le M$, se tiene*

$$h \sum_{j=1}^{M} sen^2(ljh) = \pi/2. \tag{1.2.37}$$

Asismismo, si $l, l' \in \mathbf{N}$ con $1 \le l, l' \le M$, $l \ne l'$,

$$h \sum_{j=1}^{M} \operatorname{sen}(ljh) \operatorname{sen}(l'jh) = 0. \tag{1.2.38}$$

Por consiguiente,

$$||\vec{W}_l(h)||_h = 1, \quad l = 1, ..., M; \quad < \vec{W}_l(h), \vec{W}_k(h) >_h = \delta_{lk}, \quad l, k = 1, ..., M, \tag{1.2.39}$$

y

$$< A_h \vec{W}_l(h), \vec{W}_l(h) >_h = h \sum_{j=0}^{M} \frac{|W_{l,j+1} - W_{l,j}|^2}{h^2} = \lambda_l(h), \quad l = 1, ..., M. \tag{1.2.40}$$

Demostración del Lema 1.2.1. Para todo par l, l' con $1 \leq l, l' \leq M$ y $l \neq l'$ se tiene:

$$h\sum_{j=1}^{M}\sin(ljh)\sin(l'jh) = \frac{h}{2}\sum_{j=1}^{M}[\cos(j(l'-l)h) - \cos(j(l+l')h)]$$

$$= \frac{h}{2}Re\left(\sum_{j=0}^{M}e^{i(l'-l)jh} - \sum_{j=0}^{M}e^{i(l+l')jh}\right),$$

donde Re denota la parte real de un número complejo. Aplicando la fórmula de la suma para una serie geométrica tenemos

$$\begin{aligned}\textstyle\sum_{j=0}^{M}e^{i(l'\pm l)jh} &= \frac{e^{i(l'\pm l)\pi}-1}{e^{i(l'\pm l)h}-1} = \frac{(-1)^{(l'\pm l)}-1}{e^{i(l'\pm l)h}-1}\\ &= \frac{(-1)^{(l'\pm l)}-1}{\cos(l'\pm l)h + i\sin(l'\pm l)h - 1}.\end{aligned}$$

Aplicando las fórmulas trigonométricas

$$\cos(x) = 1 - 2\sin^2\frac{x}{2}, \quad \sin(x) = 2\sin\frac{x}{2}\cos\frac{x}{2}$$

en la identidad anterior obtenemos

$$\begin{aligned}\sum_{j=0}^{M}e^{i(l'\pm l)jh} &= \frac{(-1)^{(l'\pm l)}-1}{-2\sin^2\frac{(l'\pm l)h}{2} + 2i\sin\frac{(l'\pm l)h}{2}\cos\frac{(l'\pm l)h}{2}}\\ &= \frac{(-1)^{(l'\pm l)}-1}{2i\sin\frac{(l'\pm l)h}{2}(\cos\frac{(l'\pm l)h}{2} + i\sin\frac{(l'\pm l)h}{2})}\\ &= \frac{\left[(-1)^{(l'\pm l)}-1\right]e^{-i\frac{(l'\pm l)h}{2}}}{2i\sin\frac{(l'\pm l)h}{2}}\\ &= -\frac{i}{2}\left[(-1)^{(l'\pm l)}-1\right]\cot\frac{(l'\pm l)h}{2} - \frac{1}{2}\left[(-1)^{(l'\pm l)}-1\right].\end{aligned}$$

Resulta que

$$Re\left(\sum_{j=0}^{M}e^{i(l'-l)jh}\right) = -\frac{1}{2}\left[(-1)^{(l'-l)}-1\right],$$

y

$$Re\left(\sum_{j=0}^{M}e^{i(l'+l)jh}\right) = -\frac{1}{2}\left[(-1)^{(l'+l)}-1\right]$$

y por ello

$$h\sum_{j=1}^{M}\sin(ljh)\sin(l'jh) = \frac{h}{2}\left[-\frac{1}{2}(-1)^{(l'-l)} + \frac{1}{2} + \frac{1}{2}(-1)^{(l'+l)} - \frac{1}{2}\right] = 0.$$

Para $1 \leq l' = l \leq M$ se tiene

$$h\sum_{j=1}^{M}\sin^2(ljh) = \frac{h}{2}\sum_{j=0}^{M}(1-\cos(2ljh)) = \frac{h(M+1)}{2} = \frac{\pi}{2},$$

puesto que

$$\sum_{j=0}^{M}\cos(2ljh) = Re\left(\frac{e^{i2lh(M+1)}-1}{e^{i2lh}-1}\right) = 0.$$

A partir de estas dos identidades se deducen autoamaticamente las propiedades de los autovectores de la matriz A_h. Esto concluye la prueba del Lema.

■

Este hecho permite desarrollar fácilmente las soluciones de (1.2.27) en series de Fourier, es decir, en la base $\{\vec{W}_l(h)\}_{l=1,\ldots,M}$:

$$\vec{u}(t) = \vec{u_h}(t) = \sum_{l=1}^{M}\hat{\varphi}_l e^{-\lambda_l(h)t}\vec{W}_l(h), \tag{1.2.41}$$

donde $\hat{\varphi}_l$ son los coeficientes del vector de datos iniciales $\vec{\varphi}$ en la base de autovectores $\{\vec{W}_l\}$, i.e.

$$\hat{\varphi}_l = \langle\vec{\varphi}, \vec{W}_l(h)\rangle_h, \tag{1.2.42}$$

de modo que

$$\vec{\varphi} = \sum_{\ell=1}^{M}\hat{\varphi}_l\vec{W}_l(h). \tag{1.2.43}$$

Las analogías entre la fórmula de representación (1.2.41) y el desarrollo en serie de Fourier (1.2.10) de las soluciones del problema continuo son evidentes. En realidad sólo hay dos diferencias dignas de ser reseñadas:

(a) En (1.2.41) se tiene una suma finita de M términos en lugar de la serie infinita de (1.2.10). Ahora bien $M \to \infty$ cuando $h \to 0$.

(b) Las exponenciales temporales de (1.2.41) y (1.2.10) no son exactamente las mismas pues en ellas intervienen los autovalores de uno y otro problema, si bien, en virtud de (1.2.31), ambas son semejantes.

En vista de esta similitud existente entre las expresiones de las soluciones continuas y discretas, estas últimas presentan propiedades semejantes a las de las primeras. En particular, en lo referente a decaimiento de la solución tenemos:

$$\| \vec{u}(t) \|_h^2 = \sum_{l=1}^{M} | \hat{\varphi}_l |^2 \, e^{-2\lambda_l(h)t} \leq e^{-2\lambda_1(h)t} \| \vec{\varphi} \|_h^2 \, . \tag{1.2.44}$$

Así, en el límite cuando $h \to 0$, recuperamos la tasa de decaimiento en tiempo del problema continuo (1.2.17) puesto que $\lambda_1(h) \to 1$ cuando $h \to 0$.

El primer resultado importante de esta sección es el siguiente y garantiza la convergencia de las soluciones del problema semi-discreto (1.2.15) a las soluciones del problema continuo (1.2.8) cuando $h \to 0$, bajo una elección adecuada de los datos iniciales del problema discreto.

Desde el punto de vista del desarrollo en serie de Fourier de las soluciones que estamos barajando, a la hora de elegir una aproximación del dato inicial φ del problema semi-discreto parece adecuado proceder del siguiente modo.

Dado $\varphi \in L^2(0,\pi)$, consideramos su desarrollo en serie de Fourier

$$\varphi(x) = \sum_{l=1}^{\infty} \hat{\varphi}_l w_l(x), \tag{1.2.45}$$

donde

$$\hat{\varphi}_l = \int_0^{\pi} \varphi(x) w_l(x) dx, \tag{1.2.46}$$

de modo que, por la identidad de Parseval, se tiene

$$\| \varphi \|_{L^2(0,\pi)} = \left[\sum_{l=1}^{\infty} (\hat{\varphi}_l)^2 \right]^{1/2}. \tag{1.2.47}$$

Elegimos entonces el dato inicial $\vec{\varphi}$ de la ecuación discreta de modo que tenga los mismos coeficientes de Fourier que los M primeros de $\varphi(x)$, i.e.

$$\vec{\varphi} = \vec{\varphi}(h) = \sum_{l=1}^{M} \hat{\varphi}_l \vec{W}_l(h). \tag{1.2.48}$$

En este caso los coeficientes de Fourier del desarrollo (1.2.41) de la solución del problema semi-discreto coinciden con los coeficientes φ_l de la función $\varphi(x)$.

Con esta elección de los datos iniciales es fácil probar la convergencia del esquema numérico. Tenemos el siguiente resultado:

Theorem 1.2.1 *Supongamos que $\varphi \in L^2(0,\pi)$ y consideremos los datos iniciales del problema semi-discreto* (1.2.27) *como en* (1.2.48).

Entonces, las soluciones $\vec{u}_h = \vec{u}_h(t)$ del problema semi-discreto (1.2.27), *cuando $h \to 0$, convergen a la solución $u = u(x,t)$ del problema continuo en el sentido que*

$$\| \vec{u}_h(t) - \underline{\vec{u}}(t) \|_h \to 0, \text{ para todo } t > 0, \tag{1.2.49}$$

cuando $h \to 0$, donde $\underline{\vec{u}}(t)$ es la restricción de la solución de la ecuación del calor a los nodos del mallado: $\underline{u}_j(t) = u(x_j, t)$.

Conviene explicar la noción de convergencia adoptada en (1.2.49). La cantidad que aparece en (1.2.49), para cada $t > 0$, representa la norma $\| \cdot \|_h$ de la diferencia entre la solución discreta $\vec{u}_h(t)$ y la continua $u(\cdot, t)$. Ahora bien, como $u(\cdot, t)$ depende continuamente de x, a la hora de compararla con la solución discreta, sólo interviene la restricción de u a los puntos x_j del mallado que denotamos mediante $\underline{\vec{u}}(t)$.

Demostración del Teorema 2.1.

A la hora de estudiar la diferencia $\vec{u}_h(t) - \underline{\vec{u}}(t)$ distinguimos las bajas y las altas frecuencias, es decir los rangos $\ell \leq M_0$ y $\ell \geq M_0 + 1$ respectivamente:

$$\begin{aligned} \vec{u}_h(t) - \underline{\vec{u}}(t) = & \sum_{l=1}^{M_0} \hat{\varphi}_l [e^{-\lambda_l(h)t} - e^{-\lambda_l t}] \vec{W}_l(h) \\ & + \sum_{l=M_0+1}^{M} \hat{\varphi}_l e^{-\lambda_l(h)t} \vec{W}_l(h) - \sum_{l=M_0+1}^{\infty} \hat{\varphi}_l e^{-\lambda_l t} \vec{w}_l. \end{aligned} \tag{1.2.50}$$

El valor del parámetro de corte M_0 será fijado más adelante.

Conviene observar que en el tercer sumatorio I_3, mediante la expresión $\vec{w}_l$ denotamos las restricción de la autofunción continua $w_l = w_l(x)$ a los puntos del mallado. Se trata por tanto de una notación puesto que, para $\ell > M$, $\vec{w}_\ell$ no corresponde a un autovector de la matriz A_h. Por el contrario, cuando $\ell \leq M$, $_\ell = \vec{w}_\ell(h)$. En el caso que nos ocupa (diferencias finitas en una dimensión espacial), las expresiones son en este caso particularmente simples pues, como ya dijimos, los autovectores del problema discreto son restricciones al mallado de las autofunciones continuas.

A la hora de estimar los tres términos en los que hemos descompuesto la diferencia (I_1 para las bajas frecuencias, I_2, I_3 para las altas) el Lema 2.1 nos será de gran utilidad.

Tomando normas $\| \cdot \|_h$ en (1.2.50) obtenemos

$$\| \vec{u}_h(t) - \underline{\vec{u}}(t) \|_h \leq \| I_1 \|_h + \| I_2 \|_h + \| I_3 \|_h . \tag{1.2.51}$$

Estimamos ahora por separado las tres normas $\| I_j \|_h, j = 1, 2, 3$.

Observamos en primer lugar que

$$||\vec{w}_l||_h \leq \sqrt{\pi} ||\vec{w}_l||_\infty = \sqrt{\pi} max_{j=1,\ldots,M} |\vec{w}_l(x_j)| = \sqrt{2}. \tag{1.2.52}$$

De este modo deducimos que

$$||I_3||_h = \sqrt{2} \sum_{j \geq M_0+1} |\hat{\varphi}_j| \, e^{-\lambda_j t} \leq \sqrt{2} \left[\sum_{j \geq M_0+1} |\hat{\varphi}_j|^2 \right]^{1/2} \left[\sum_{j \geq M_0+1} e^{-2\lambda_j t} \right]^{1/2} . \tag{1.2.53}$$

Evidentemente, este cálculo está justificado por la convergencia de la última de las series que interviene en esta desigualdad, lo cual está garantizado para todo $t > 0$.

De (1.2.53) tenemos

$$\left|I_3\right|_h^{1/2} \leq C(t)\Big(\sum_{j\geq M_0+1}\left|\hat{\varphi}_j\right|^2\Big)^{1/2}, \tag{1.2.54}$$

con

$$C(t) = \left[\sum_{j\geq 1} e^{-2\lambda_j t}\right]^{1/2}. \tag{1.2.55}$$

Como el dato inicial $\varphi \in L^2(0,\pi)$, sus coeficientes de Fourier satisfacen

$$\int_0^\pi \varphi^2(x)dx = \sum_{j=1}^{\infty}|\hat{\varphi}_j|^2 \tag{1.2.56}$$

y, por lo tanto, dado $\varepsilon > 0$ arbitrariamente pequeño, existe $M_0 \in \mathbf{N}$ tal que

$$\sum_{j\geq M_0+1}|\hat{\varphi}_j|^2 \leq \varepsilon^2, \quad \forall M \geq M_0. \tag{1.2.57}$$

Dado $t > 0$, este permite por tanto fijar el valor de M_0 de modo que

$$\left|I_3(t)\right|_h \leq \varepsilon. \tag{1.2.58}$$

El término I_2 puede estimarse de manera idéntica con la misma elección de M_0. Pero en este caso puede incluso encontrarse una estimación uniforme para todo $t \geq 0$ gracias a las propiedades de ortonormalidad de los autovectores $\vec{w}_l(h)$. En efecto,

$$||I_2||_h^2 = \sum_{j=M_0+1}^{M}\hat{\varphi}_j^2 e^{-2\lambda_j(h)t} \leq \sum_{j=M_0+1}^{M}\hat{\varphi}_j^2 \leq \varepsilon^2, \quad \forall t \geq 0. \tag{1.2.59}$$

Por tanto, de los desarrollos anteriores se deduce que, dado $\varepsilon > 0$ arbitrariamente pequeño podemos hallar M_0 tal que

$$||I_2(t)||_h + ||I_3(t)||_h \leq 2\varepsilon, \quad \forall t \geq 0. \tag{1.2.60}$$

Procedemos ahora a la estimación de $\left|I_1\right|_h$. Tenemos

$$\begin{aligned}\left|I_1\right|_h^2 &= \sum_{\ell=1}^{M_0}(\hat{\varphi}_l)^2\left(e^{-\lambda_l(h)t} - e^{-\lambda_l t}\right)^2\left|\vec{w}_l(h)\right|_h^2 \\ &= \sum_{\ell=1}^{M_0}(\hat{\varphi}_l)^2\left(e^{-\lambda_l(h)t} - e^{-\lambda_l t}\right)^2.\end{aligned} \tag{1.2.61}$$

Nótese que en esta ocasión el valor de M_0 está ya fijado, en virtud de la elección hecha antes en función de ε. En esta ocasión es el parámetro h el que tiende a cero. Obsérvese que, para cada $l \in \{1, \dots, M_0\}$ y $T > 0$ fijo, en vista de (1.2.31), el sumando $(\varphi_l)^2 \left(e^{-\lambda_l(h)t} - e^{-\lambda_l t}\right)$ tiende a cero cuando $h \to 0$ uniformemente en $t \in [0, T]$. Por lo tanto, en vista de que el número M_0 de sumandos está fijado, deducimos que

$$\left|I_1\right|_h \to 0,\ h \to 0,\ \text{ uniformemente en } t \in [0, T].$$

En particular, eligiendo h suficientemente pequeño se puede asegurar que $\left|I_1(t)\right|_h \leq \varepsilon$, para todo $t \geq 0$.

Combinando estas estimaciones deducimos, que para cualquier $t > 0$ y $\varepsilon > 0$ existe h suficientemente pequeño tal que, según (1.2.51),

$$\left|\vec{u}_h(t) - \underline{\vec{u}}(t)\right|_h \leq 3\varepsilon.$$

Esto concluye la demostración del Teorema 1.2.1.

■

Observación 1.2.2 En (1.2.49) hemos establecido la convergencia de las soluciones del problema semi-discreto $\vec{u}_h$ a la del problema continuo en el sentido de la norma $\|\cdot\|_h$. Pero ésto no es más que una de las posibles maneras de establecer la proximidad entre las soluciones de ambos problemas. A continuación presentamos algunas variantes.

■

Algunas variantes del Teorema de convergencia 2.1.

- *Variante 1. Datos iniciales en $H_0^1(0, \pi)$.*

  Supongamos, por ejemplo, que el dato inicial φ es un poco más regular:

  $$\varphi \in H_0^1(0, \pi) = \left\{\varphi \in L^2(0, \pi) :\ \varphi' \in L^2(0, \pi),\ \varphi(0) = \varphi(\pi) = 0\right\}.$$

  En este caso, obviamente, tenemos el resultado de convergencia del Teorema 2.1. Pero, bajo esta hipótesis adicional sobre el dato inicial, se puede dar una versión más precisa y cuantitativa de este resultado.

En este caso los coeficientes de Fourier $\{\hat{\varphi}_l\}_{\ell\in\mathbf{N}}$ de φ satisfacen:[5]

$$\| \varphi \|^2_{H_0^1(0,\pi)} = \sum_{l\geq 1} l^2 \left|\hat{\varphi}_l\right|^2 < \infty. \tag{1.2.62}$$

Gracias a esta hipótesis adicional la elección del M_0 que es preciso para que se cumpla (1.2.57) se puede realizar explícitamente. En efecto

$$\sum_{l\geq M_0+1} \left|\hat{\varphi}_l\right|^2 \leq \frac{1}{(M_0+1)^2} \sum_{l\geq M_0+1} l^2 \left|\hat{\varphi}_l\right| \leq \frac{\| \varphi \|^2_{H_0^1(0,\pi)}}{(M_0+1)^2} \tag{1.2.63}$$

y, por tanto, basta con elegir

$$M_0 = \Big[\frac{\| \varphi \|_{H_0^1(0,\pi)}}{\varepsilon}\Big] - 1, \tag{1.2.64}$$

siendo $\Big[\cdot\Big]$ la función entera, para que (1.2.57) se cumpla.

Una vez que M_0 ha sido fijado de este modo, también h puede ser elegido de forma que $\| I_1 \|_h$ sea menor que ε.

En efecto,

$$\begin{aligned}
\Big|I_1\Big|_h^2 &= \sum_{\ell=1}^{M_0} \left|\hat{\varphi}_l\right|^2 \left(e^{-\lambda_l(h)t} - e^{-\lambda_l t}\right) && (1.2.65)\\
&= \sum_{l=1}^{M_0} \left|\hat{\varphi}_l\right|^2 t\left(\lambda_l - \lambda_l(h)\right) e^{-\mu_\ell(h)t} && (1.2.66)\\
&\leq \sum_{\ell=1}^{M_0} \left|\hat{\varphi}_l\right|^2 t\left(\lambda_l - \lambda_l(h)\right),
\end{aligned}$$

donde, mediante $\mu_\ell(h)$, hemos denotado un número real entre $\lambda_l(h)$ y λ_l, obtenido en la aplicación del Teorema del Valor Medio.

Por otra parte,

$$\begin{aligned}
\lambda_l - \lambda_l(h) &= l^2 - \frac{4}{h^2}\operatorname{sen}^2\left(l\frac{h}{2}\right) = \frac{1}{h^2}\left[(hl)^2 - 4\operatorname{sen}^2\left(l\frac{h}{2}\right)\right]\\
&= \frac{1}{h^2}\left[hl + 2\operatorname{sen}(lh/2)\right]\left[hl - 2\operatorname{sen}(lh/2)\right]\\
&\leq \frac{2l}{h}\left[hl - 2\operatorname{sen}(lh/2)\right] \leq \frac{l}{h} C(hl)^3\\
&= C\,h^2 l^4 \leq C M_0^4 h^2 && (1.2.67)
\end{aligned}$$

[5] Obsérvese que, en virtud de la desigualdad de Poincaré, la norma inducida por $H^1(0,\pi)$ sobre $H_0^1(0,\pi)$ y la norma $\| \varphi \|^2_{H_0^1(0,\pi)} = \left[\int_0^\pi |\varphi'|^2 dx\right]^{1/2}$ son normas equivalentes. A nivel de los coeficientes de Fourier esta última se reduce a $\| \varphi \|_{H_0^1(0,\pi)} = \left[\sum_{l\geq 1} l^2 \left|\hat{\varphi}_l\right|^2\right]^{1/2}$ que es a su vez equivalente a la inducida por la norma $\| \varphi \|_{H_0^1(0,\pi)} = \left[\sum_{l\geq 1}(1+l^2)\left|\hat{\varphi}_l\right|^2\right]^{1/2}$.

donde C es una constante que puede ser calculada explícitamente. El mayor valor de dicha constante viene dado por

$$C = 2 \max_{|\tau| \leq \pi} \left[\tau - 2\,\mathrm{sen}(\tau/2)\right] / \tau^3 \tag{1.2.68}$$

siempre y cuando $h > 0$ sea lo suficientemente pequeño de modo que $lh \leq M_0 h \leq (M+1)h = \pi$, es decir que $M + 1 \geq M_0$.

En virtud de (1.2.67) y en vista del valor explícito de M_0 dado en (1.2.64) el valor de h para que, según (1.2.65), $\| I_1 \|_h \leq \varepsilon$ puede calcularse explícitamente. Esto permite cuantificar el resultado de convergencia del Teorema 2.1. Obsérvese sin embargo que este tipo de argumentos necesita que el dato inicial este en un espacio más pequeño que el espacio $L^2(0,\pi)$ donde el resultado de convergencia ha sido probado.

Pero hay otras variantes del resultado de convergencia del Teorema 1.2.1 que pueden resultar incluso más elocuentes.

■

- *Variante 2. Convergencia en la norma del máximo*

  Por ejemplo, puede resultar más natural estimar la distancia entre la solución $\vec{u}_h$ del problema semi-discreto y la solución u de (1.2.8) en la norma del máximo:

$$\left|\vec{u}_h(t) - \underline{\vec{u}}(t)\right|_\infty = \max_{j \in \{1,\ldots,M\}} |u_j(t) - u(x_j,t)|\,. \tag{1.2.69}$$

  Para la estimación de esta cantidad procedemos del modo siguiente. Descomponiendo la norma de la diferencia como en la prueba del Teorema 1.2.1 tenemos:

$$\begin{aligned}
\left|\vec{u}_h(t) - \underline{\vec{u}}(t)\right|_\infty &= \Big| \sum_{l=1}^{M} \hat{\varphi}_l e^{-\lambda_l(h)t} \vec{W}_l(h) - \sum_{l=1}^{\infty} \hat{\varphi}_l e^{-\lambda_l t} \vec{w}_l \Big|_\infty \\
&\leq \sqrt{\frac{2}{\pi}} \sum_{l=1}^{M_0} |\hat{\varphi}_l| \left| e^{-\lambda_l(h)t} - e^{-\lambda_l t} \right| \\
&\quad + \sqrt{\tfrac{2}{\pi}} \sum_{l=M_0+1}^{M} |\hat{\varphi}_l|\, e^{-\lambda_l(h)t} + \sqrt{\frac{2}{\pi}} \sum_{l \geq M_0+1} |\hat{\varphi}_l|\, e^{-\lambda_l t} \\
&= I_1 + I_2 + I_3.
\end{aligned}$$

  En esta desigualdad hemos tenido en cuenta que $\left|\vec{w}_l\right|_\infty \leq \sqrt{2/\pi}$, para todo $l \geq 1$.

Estimamos ahora el último término:

$$I_3 = \sqrt{\frac{2}{\pi}} \sum_{l \geq M_0+1} |\hat{\varphi}_l| \, e^{-\lambda_l t} \leq \sqrt{\frac{2}{\pi}} \left[\sum_{l \geq M_0+1} |\hat{\varphi}_l|^2 \right]^{1/2} \left[\sum_{l \geq M_0+1} e^{-2\lambda_l t} \right]^{1/2}.$$

Como $\lambda_l = l^2$, la última serie de esta desigualdad converge para cada $t > 0$, i.e.

$$\sum_{l=1}^{\infty} e^{-2\lambda_l t} = \sum_{l=1}^{\infty} e^{-2l^2 t} < \infty$$

y, por lo tanto, con una elección adecuada de $M_0 = M_0(\varepsilon)$, puede asegurarse que

$$I_3 \leq \varepsilon \parallel \varphi \parallel_{L^2(0,\pi)} .$$

La misma acotación es válida para I_2.

Fijado el valor de M_0 de modo que estas cotas de I_2 y I_3 sean válidas procedemos a acotar I_1:

$$I_1 \leq \sqrt{\frac{2}{\pi}} \sum_{l=1}^{M_0} |\hat{\varphi}_l| \, \left| e^{-\lambda_l(h) t} - e^{-\lambda_l t} \right| .$$

Este último término tiende a cero cuando $h \to 0$ puesto que se trata de una suma finita en la que cada uno de los términos tiende a cero.

De este modo concluimos que, cuando el dato inicial $\varphi \in L^2(0,\pi)$, para todo $t > 0$ se tiene

$$\left| \vec{u}_h(t) - \underline{\vec{u}}(t) \right|_{\infty} \to 0, \; h \to 0. \tag{1.2.70}$$

Obsérvese que en este caso no se tiene una convergencia uniforme en tiempo $t \in [0,T]$. En efecto, la convergencia en $t = 0$ exigiría estimar los términos I_2 y I_3 de otro modo, y necesitaríamos de la hipótesis

$$\sum_{j=1}^{\infty} |\hat{\varphi}_l| < \infty,$$

cosa que no está garantizada por la condición $\varphi \in L^2(0,\pi)$. Sin embargo, sí que bastaría con suponer que $\varphi \in H_0^1(0,\pi)$, como en la variante anterior, puesto que

$$\sum_{j=1}^{\infty} |\hat{\varphi}_l| \leq \left[\sum_{j=1}^{\infty} |\hat{\varphi}_l|^2 \, j^2 \right]^{1/2} \left[\sum_{j=1}^{\infty} j^{-2} \right]^{1/2} = C \parallel \varphi \parallel_{H_0^1(0,\pi)} . \tag{1.2.71}$$

Volviendo al caso general $\varphi \in L^2(0,\pi)$, acabamos de comprobar que, a pesar de que el dato inicial φ se supone únicamente en $L^2(0,\pi)$, la convergencia de la solución semi-discreta a la solución continua se produce en la norma del máximo (que corresponde a la norma de $L^\infty(0,\pi)$). Esto es así gracias al efecto regularizante que, como hemos mencionado, caracteriza a la ecuación del calor y que todos los problemas semi-discretos (1.2.24) comparten. En efecto, la solución de la ecuación del calor es de la forma

$$u(x,t) = \sum_{l=1}^{\infty} \hat{\varphi}_l e^{-l^2 t} w_l(x)$$

de modo que

$$\begin{aligned} | \, u(x,t) \, | &\leq \sum_{l=1}^{\infty} |\hat{\varphi}_l| \, e^{-l^2 t} \, |w_l(x)| \leq \sqrt{\frac{2}{\pi}} \sum_{l=1}^{\infty} |\hat{\varphi}_l| \, e^{-l^2 t} \qquad (1.2.72) \\ &\leq \sqrt{\frac{2}{\pi}} \left[\sum_{l=1}^{\infty} |\hat{\varphi}_l|^2 \right]^{1/2} \left[\sum_{l=1}^{\infty} e^{-2l^2 t} \right]^{1/2} = C(t) \parallel \varphi \parallel_{L^2(0,\pi)} \end{aligned}$$

con $C(t) < \infty$ para todo $t > 0$. Esto garantiza el efecto regularizante $L^2(0,\pi) \to L^\infty(0,\pi)$ en la ecuación del calor para cualquier instante de tiempo $t > 0$. Obviamente $C(0) = \infty$, lo cual corresponde a la existencia de funciones de $L^2(0,\pi)$ que no pertenecen a $L^\infty(0,\pi)$.

Este efecto regularizante es compartido por todas las soluciones del problema semi-discreto. Para comprobarlo basta observar que existe $c > 0$ tal que

$$\lambda_l(h) \geq cl^2, \, \forall h > 0, \, \forall l = 1, \ldots, M, \qquad (1.2.73)$$

lo cual garantiza un control uniforme (con respecto a h) de las series que intervienen en la cuantificación de dicho efecto.

■

- *Variante 3. Datos iniciales en Fourier dependientes de h.*

  En el Teorema 1.2.1 hemos optado por elegir en el problema semi-discreto (1.2.24) (o (1.2.27)) como dato inicial la truncamiento de la serie de Fourier del dato inicial de la ecuación del calor. Podría decirse pues que el dato inicial es independiente de h, en el sentido en que sus coeficientes de Fourier (los que se pueden imponer en el problema semi-discreto) lo son.

  En realidad, el mismo método de demostración del Teorema 1.2.1 permite probar otro tipo de resultados en los que el dato inicial del problema

semi-discreto no es necesariamente ése. En particular, permite establecer la convergencia de las soluciones a partir de informaciones mínimas sobre la convergencia de los datos iniciales.

Por ejemplo, supongamos que en el problema semi-discreto (1.2.24) (o (1.2.27)) tomamos como dato inicial

$$\vec{\varphi}(h) = \sum_{l=1}^{M} \hat{\varphi}_l(h)\vec{w}_l(h),$$

es decir, definimos el dato inicial mediante coeficientes de Fourier que dependen de h. Si extendemos estos coeficientes de Fourier $\hat{\varphi}_l(h)$ por cero para $l \geq M+1$, podemos identificar, para cada $h > 0$ los coeficientes de Fourier de $\vec{\varphi}(h)$ con una sucesión de ℓ^2 (el espacio de Hilbert de las sucesiones de números reales de cuadrado sumable). El enunciado del Teorema 1.2.1 puede entonces generalizarse del siguiente modo:

"El resultado del Teorema 1.2.1 es aún cierto si, cuando $h \to 0$, $\{\hat{\varphi}_l(h)\}_{l\geq 1}$ *converge en* ℓ^2 *a* $\{\hat{\varphi}_l\}_{l\geq 1}$, *donde* $\{\hat{\varphi}_l(h)\}_{l\geq 1}$ *(resp.* $\{\hat{\varphi}_l\}_{l\geq 1}$*) representa el elemento de* ℓ^2 *constituido por los coeficientes de Fourier del dato* $\vec{\varphi}(h)$ *del problema discreto (resp. del dato* $\varphi \in L^2(0,\pi)$ *del problema continuo)".*

En realidad, si hacemos uso del efecto regularizante, tal y como comentabamos en la variante anterior, se puede probar la convergencia del esquema bajo hipótesis más débiles sobre los datos iniciales:[6]

"Supongamos que los datos iniciales $\vec{\varphi}(h)$ *del problema semi-discreto* (1.2.24) (o (1.2.27)) *son tales que* $\{\hat{\varphi}_l(h)\}_{l\geq 1}$ *converge débilmente en* ℓ^2 *a* $\{\hat{\varphi}_l\}_{l\geq 1}$ *cuando* $h \to 0$. *Entonces, la convergencia* (1.2.49) *es cierta uniformemente en* $t \geq \delta$, *para cualquier* $\delta > 0$*".*

■

[6] En un espacio de Hilbert H se dice que una sucesión $\{h_k\}_{k\geq 1}$ converge débilmente a un elemento $h \in H$, si $(h_k, g)_H \to (h, g)_H$ para todo $g \in H$. Obviamente, la convergencia clásica en el sentido de la norma (también denominada convergencia fuerte) implica la convergencia débil. Por otra parte, si una sucesión converge débilmente y además se tiene que $||h_k||_H \to ||h||_H$, entonces se tiene también la convergencia en norma. Una de las propiedades más utilizadas de la convergencia débil es que de toda sucesión acotada en un espacio de Hilbert se puede extraer una subsucesión que converge débilmente.

Variante 4. Datos iniciales por restricción al mallado.

Pero en todos estos resultados la convergencia de la solución del problema discreto al continuo se establece en función del comportamiento de los coeficientes de Fourier de los datos iniciales cuando $h \to 0$. Sin embargo, desde un punto de vista estríctamente numérico, éste no es el modo más natural de proceder puesto que sería deseable disponer de un método más sencillo que no pase por el cálculo de las series de Fourier continuas y/o discretas para realizar la elección de los datos iniciales en la ecuación semi-discreta.

Supongamos por ejemplo que el dato inicial φ de la ecuación del calor es un poco más regular: $\varphi \in C([0,\pi])$. En este caso la elección más natural del dato inicial en el método numérico (1.2.24) (o (1.2.27)) es

$$u_j(0) = \varphi_j = \varphi(x_j).$$

En efecto, al elegir este dato inicial no necesitamos calcular los coeficientes de Fourier de φ (lo cual exige realizar una integral y, desde un punto de vista numérico la utilización de fórmulas de cuadratura) sino que basta evaluar el dato inicial sobre los puntos del mallado.

Si bien ésto supone una elección distinta de los datos iniciales, su valor efectivo no dista mucho del que utilizamos mediante series de Fourier. En efecto en el Teorema 1.2.1 hicimos la elección

$$\vec{\varphi} = \sum_{l=1}^{M} \hat{\varphi}_l \vec{W}_l(h)$$

que denotaremos mediante $\underline{\vec{\varphi}}$ para distinguirla de la anterior. Teniendo en cuenta que la k-ésima componente del vector $\vec{w}_l(h)$ coincide con el valor de la autofunción $w_l(x)$ en el punto $x = x_k = kh$, tenemos entonces

$$\underline{\varphi}_k = \sum_{l=1}^{M} \hat{\varphi}_l w_l(x_k)$$

mientras que

$$\varphi_k = \varphi(x_k) = \sum_{l=1}^{\infty} \hat{\varphi}_l w_l(x_k).$$

Por tanto

$$\left|\varphi_k - \underline{\varphi}_k\right| = \left|\sum_{l=M+1}^{\infty} \hat{\varphi}_l w_l(x_k)\right| \leq \sqrt{\frac{2}{\pi}} \sum_{l \geq M+1} |\, \hat{\varphi}_l \,|\,.$$

Así

$$\left|\vec{\varphi} - \underline{\vec{\varphi}}\right|_h^2 = h\sum_{k=1}^{M}\left|\varphi_k - \underline{\varphi}_k\right|^2 \leq 2\left(\sum_{l\geq M+1} \mid \hat{\varphi}_l \mid\right)^2.$$

Con el objeto de garantizar que

$$\left|\vec{\varphi} - \underline{\vec{\varphi}}\right|_h \to 0, \text{ cuando } h \to 0$$

basta por tanto con saber que

$$\sum_{l=1}^{\infty} \mid \hat{\varphi}_l \mid < \infty,$$

lo cual, como vimos en la variante 1, está garantizado, por ejemplo, si $\varphi \in H_0^1(0,\pi)$, lo que supone una hipótesis ligeramente más fuerte que la continuidad de φ.

Es fácil ver entonces que el resultado del Teorema 1.2.1 se preserva con esta elección del dato inicial del problema semi-discreto.

■

Variante 5. Datos iniciales en media.

Hay otras elecciones posibles del dato inicial en el problema semi-discreto (1.2.24) (o (1.2.27)) que no pasan por el cálculo de los coeficientes de Fourier del dato inicial. Por ejemplo, para cualquier $\varphi \in L^2(0,\pi)$, podemos elegir

$$\varphi_j = \frac{1}{h}\int_{x_j-h/2}^{x_j+h/2} \varphi(x)dx.$$

No es difícil ver que esta elección de los datos iniciales conduce a resultados semejantes de convergencia.

En realidad, y esto es comentario interesante y útil, una vez que se dispone de un resultado de convergencia para una determinada elección de los datos iniciales, no es difícil probar la convergencia para otras posibles opciones puesto que basta con estimar la diferencia de las soluciones del problema discreto o semi-discreto para las dos elecciones de los datos, sin necesidad de volver a comprobar la proximidad con el modelo continuo.

Esto es particularmente sencillo en el caso que nos ocupa puesto que si $\vec{u}_h$ y $\vec{v}_h$ son dos soluciones del problema semi-discreto (1.2.24) correspondientes a datos iniciales $\vec{\varphi}$ y $\vec{\psi}$, se tiene

$$||\vec{u}_h(t) - \vec{v}_h(t)||_h \leq ||\vec{\varphi} - \vec{\psi}||_h, \quad \forall t > 0, \forall h > 0, \tag{1.2.74}$$

tal y como se desprende de (1.2.44).

Por lo tanto, el método que hemos desarrollado en esta sección, basado en series de Fourier, permite probar la convergencia del método no sólo cuando los datos iniciales han sido adaptados en función del desarrollo en serie de Fourier del dato inicial de la ecuación del calor, sino en cualquier circunstancia en la que el dato inicial de la ecuación semi-discreta haya sido elegido de manera coherente.

■

Una interpretación global del resultado de convergencia nodal.

En el Teorema 2.1 hemos probado la convergencia de la solución del problema semi-discreto a la del problema continuo en el sentido de la norma discreta $||\cdot||_h$, i.e. sobre los nodos del mallado. Sin embargo, en la medida en que la solución de la ecuación del calor está definida en todo el intervalo $(0,\pi)$ cabría esperar que pueda también establecerse un resultado de carácter global. Con el objeto de probar dicho resultado lo primero que tenemos que hacer es extender la función discreta $\vec{u}$ a una función definida en todo el intervalo $(0,\pi)$. Hay diversas maneras de realizar ésto. La más sencilla es tal vez considerar la función constante a trozos:

$$[E\vec{u}_h(t)](x) = \sum_{j=1}^{M} u_j(t)\chi_{[x_j-h/2,x_j+h/2]}, \tag{1.2.75}$$

donde $\chi_{[x_j-h/2,x_j+h/2]}$ denota la función característica del intervalo $[x_j-h/2, x_j+h/2]$. Esta función extendida está definida para todo $x \in (0,\pi)$ y para todo t. Cabe por tanto plantearse su convergencia hacia la solución de la ecuación del calor cuando $h \to 0$.

Como consecuencia inmediata del Teorema 2.1 tenemos el siguiente resultado:

Corollary 1.2.1 *Bajo las hipótesis del Teorema 2.1 se tiene*

$$E\vec{u}_h(t) \to u(t) \quad en \quad L^2(0,\pi), \quad \forall t > 0. \tag{1.2.76}$$

Demostración del Corolario 2.1.

Para probar este Corolario basta observar que

$$\begin{aligned} ||E\vec{u}_h(t) - u(t)||^2_{L^2(0,\pi)} = & \sum_{j=1}^{M} \int_{x_j-h/2}^{x_j+h/2} |u_j(t) - u(x,t)|^2 dx \\ & + \int_0^{h/2} |u(x,t)|^2 dx + \int_{\pi-h/2}^{\pi} |u(x,t)|^2 dx. \end{aligned} \tag{1.2.77}$$

Las dos últimas integrales, evidentemente, tienden a cero cuando $h \to 0$. Cada una de las otras integrales que interviene en el sumatorio puede ser estimada del siguiente modo:

$$\begin{aligned}\int_{x_j-h/2}^{x_j+h/2} |u_j(t)-u(x,t)|^2 dx \quad &\leq 2h|u_j(t)-u(x_j,t)|^2 \\ &+2\int_{x_j-h/2}^{x_j+h/2} |u(x,t)-u(x_j,t)|^2 dx.\end{aligned} \tag{1.2.78}$$

Por tanto,

$$\begin{aligned}&\sum_{j=1}^{M}\int_{x_j-h/2}^{x_j+h/2} |u_j(t)-u(x,t)|^2 dx \\ &\leq 2h\sum_{j=1}^{M}|u_j(t)-u(x_j,t)|^2 + 2\sum_{j=1}^{M}\int_{x_j-h/2}^{x_j+h/2}|u(x,t)-u(x_j,t)|^2 dx \\ &= 2||\vec{u}(t)-\underline{\vec{u}}(t)||_h^2 + 2\sum_{j=1}^{M}\int_{x_j-h/2}^{x_j+h/2}|u(x,t)-u(x_j,t)|^2 dx\end{aligned}$$

El primero de estos dos términos tiende a cero en virtud del resultado de convergencia del Teorema 2.1.

Basta entonces analizar el último de ellos.

Para ello observamos que, por la desigualdad de Poincaré, cada uno de los términos que interviene en el sumatorio satisface:

$$\int_{x_j-h/2}^{x_j+h/2}|u(x,t)-u(x_j,t)|^2 dx \leq \frac{h^2}{\pi^2}\int_{x_j-h/2}^{x_j+h/2}|u_x(x,t)|^2 dx. \tag{1.2.79}$$

Por lo tanto,

$$\sum_{j=1}^{M}\int_{x_j-h/2}^{x_j+h/2}|u(x,t)-u(x_j,t)|^2 dx \leq \frac{h^2}{\pi^2}\int_0^{\pi}|u_x(x,t)|^2 dx, \tag{1.2.80}$$

que, evidentemente, tiende a cero cuando $h \to 0$ cuando $u(t) \in H_0^1(0,\pi)$. Pero esto es cierto para todo $t > 0$ a causa del efecto regularizante de la ecuación del calor, tal y como se puede observar en el desarrollo de Fourier de la solución u.

■

En el Corolario 2.1 hemos optado por la extensión constante a trozos de la solución numérica al intervalo $(0,\pi)$ pero el mismo resultado se cumple para cualquier otra extensión razonable, por ejemplo las funciones continuas y lineales a trozos.

1.2.3. Semi-discretización espacial: El método de la energía

Hemos probado, mediante métodos basados en el desarrollo en series de Fourier de las soluciones, la convergencia del problema semi-discreto (1.2.24) a la ecuación del calor. En lo sucesivo lo haremos mediante el *método de la energía*.

El método de la energía está basado en la identidad de energía (1.2.18) que las soluciones de la ecuación del calor satisfacen.

Para el problema semi-discreto se satisface una identidad de energía semejante. En efecto multiplicamos la j-ésima ecuación (1.2.24) por u_j y sumamos con respecto al índice $j = 1, \cdots, M$. Obtenemos así

$$0 = \sum_{j=1}^{M} u_j u_j' - \frac{1}{h^2} \sum_{j=1}^{M} \left[u_{j+1} + u_{j-1} - 2u_j \right] u_j.$$

Observamos en primer lugar que

$$\sum_{j=1}^{M} u_j u_j' = \frac{1}{2} \frac{d}{dt} \left[\sum_{j=1}^{M} |u_j|^2 \right]$$

y, por otra parte,

$$\begin{aligned}
&\sum_{j=1}^{M} \left[u_{j+1} + u_{j-1} - 2u_j \right] u_j = \sum_{j=1}^{M} \left[(u_{j+1} - u_j) + (u_{j-1} - u_j) \right] u_j \\
&= \sum_{j=1}^{M} (u_{j+1} - u_j)\, u_j + \sum_{j=1}^{M} (u_{j-1} - u_j)\, u_j \\
&= \sum_{j=1}^{M} (u_{j+1} - u_j)\, u_j - \sum_{j=0}^{M-1} (u_{j+1} - u_j)\, u_{j+1} \\
&= -\sum_{j=0}^{M} (u_{j+1} - u_j)^2 .
\end{aligned}$$

Obtenemos así la siguiente identidad de energía para el sistema semi-discreto (1.2.24):

$$\frac{d}{dt} \left[\frac{h}{2} \sum_{j=1}^{M} |u_j|^2 \right] = -h \sum_{j=0}^{M} \left| \frac{u_{j+1} - u_j}{h} \right|^2 . \tag{1.2.81}$$

Las similitudes entre las identidades de energía (1.2.18) del modelo continuo y la identidad (1.2.81) del caso semi-discreto son obvias. En el término de la izquierda de (1.2.81) se observa la derivada temporal de una suma discreta que

aproxima el cuadrado de la norma en $L^2(0,\pi)$ de la identidad (1.2.18). Por otro lado, en el miembro de la derecha de (1.2.81) encontramos una versión discreta de la norma de u_x en $L^2(0,\pi)$.

A partir de estas identidades podemos dar una demostración alternativa de la convergencia del problema semi-discreto (1.2.24) al continuo.

En primer lugar observamos que la solución del problema continuo $u = u(x,t)$ es una solución aproximada del problema discreto. En efecto, sea

$$\underline{u}_j(t) = u(x_j,t), \quad j = 1,\dots,M,\ t>0, \tag{1.2.82}$$

siendo $u = u(x,t)$ la solución exacta de (1.2.8).

En efecto, $\left\{\ \underline{u}_j(t)\right\}_{j=1,\dots,M}$ satisface

$$\begin{cases} \underline{u}_j' + \frac{\left[2\ \underline{u}_j - \ \underline{u}_{j+1} - \ \underline{u}_{j-1}\right]}{h^2} = u_{xx}(x_j,t) \quad + \quad \frac{\left[2\ \underline{u}_j - \ \underline{u}_{j+1} - \ \underline{u}_{j-1}\right]}{h^2} = \varepsilon_j(t), \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad j = 1,\dots,M, t>0 \\ \underline{u}_0 = \ \underline{u}_{M+1} = 0, \quad t>0 \\ \underline{u}_j(0) = \varphi_j,\ j = 1,\dots,\ M. \end{cases} \tag{1.2.83}$$

con $\varphi_j = \varphi(x_j)$. En el segundo miembro de (1.2.83) aparece el *residuo* o *error de truncación* asociado al método que, como se observa en su propia definición, se trata del error que se comete al considerar la solución del problema continuo como solución aproximada del problema discreto. Conviene subrayar este hecho: las demostraciones de convergencia están basadas en el análisis de la solución continua como solución aproximada del problema discreto y no al revés.

Consideramos ahora la diferencia entre la solución continua $\underline{u}_j$ sobre el mallado y la solución del problema semi-discreto:

$$\vec{v} = \{v_j\}_{j=1,\dots,M}; \quad v_j = \ \underline{u}_j - u_j. \tag{1.2.84}$$

En vista de (1.2.24) y (1.2.83), $\{v_j\}_{j=0,\dots,M}$ satisface

$$\begin{cases} v_j' + \frac{[2v_j - v_{j+1} - v_{j-1}]}{h^2} = \varepsilon_j, & j = 1,\dots,M, \quad t>0 \\ v_0 = v_{M+1} = 0, & t>0 \\ v_j(0) = 0, & j = 1,\dots,M. \end{cases} \tag{1.2.85}$$

El sistema (1.2.85) no es homogéneo y su dato inicial es nulo pues hemos supuesto que en el sistema semi-discreto el dato inicial del sistema continuo se toma de manera exacta sobre los puntos del mallado. Pero sería fácil adaptar las estimaciones que vamos a realizar al caso en que también hubiese un cierto error en los datos iniciales y, en particular, a las demás situaciones discutidas en

el apartado anterior. Los términos ε_j del segundo miembro (el error de truncación) de (1.2.85) representan, tal y como se observa en su definición (1.2.83), la diferencia entre el laplaciano continuo y el discreto evaluado en la solución real u de (1.2.8). El método de la energía se adapta con facilidad a esta situación.

Retomamos la estimación de energía en el sistema (1.2.85). Multiplicando cada ecuación de (1.2.85) por v_j y sumando con respecto a $j = 1, \ldots, M$, se obtiene

$$\frac{d}{dt}\left[\frac{h}{2}\sum_{j=1}^{M}|v_j|^2\right] = -h\sum_{j=0}^{M}\left|\frac{v_{j+1}-v_j}{h}\right|^2 + h\sum_{j=1}^{M}\varepsilon_j v_j. \tag{1.2.86}$$

Necesitamos ahora la siguiente desigualdad elemental:

Lemma 1.2.2 *Para todo $\delta > 0$ existe $h_0 > 0$ de modo que para todo $0 < h < h_0$ y para toda función discreta $\{a_0, a_1, \ldots, a_M, a_{M+1}\}$ tal que $a_0 = a_{M+1} = 0$ se tiene*

$$h\sum_{j=0}^{M}\left|\frac{a_{j+1}-a_j}{h}\right|^2 \geq (1-\delta)h\sum_{j=1}^{M}|a_j|^2. \tag{1.2.87}$$

Observación 1.2.3 Nótese que (1.2.87) es la versión discreta de la clásica *desigualdad de Poincaré* (1.2.21) que ya comentamos en la Observación 2.1. En (1.2.87) se establece la versión discreta de esta desigualdad con una constante $(1-\delta)$ arbitrariamente próxima a la constante unidad de la desigualdad (1.2.21).

Tal y como mencionamos en aquella Observación, la mejor constante de la desigualdad de Poincaré venía dada por el principio de minimalidad que involucra en el cociente de Rayleigh.

La demostración del Lema está basada en analizar el correspondiente principio variacional en el caso discreto en función del paso del mallado h.

■

Demostración del Lema 1.2.2. Tal y como se indicó en (1.2.29) el mínimo autovalor de la matriz A_h definida en (1.2.26) es $\lambda_1(h) = 4\,\mathrm{sen}^2(h/2)/\,h^2$. Como A_h es simétrica, λ_1 está caracterizado por

$$\frac{4}{h^2}\,\mathrm{sen}^2\left(\frac{h}{2}\right) = \lambda_1(h) = \min_{\vec{a}\in\mathbf{R}^M}\frac{\langle A_h\vec{a},\,\vec{a}\rangle_h}{\|\,\vec{a}\,\|_h^2}. \tag{1.2.88}$$

Es fácil comprobar que

$$\langle A_h\vec{a},\,\vec{a}\rangle_h \Big/ \|\,\vec{a}\,\|_h^2 = h\sum_{j=0}^{M}\left|(a_{j+1}-a_j)/h\right|^2 \Big/ h\sum_{j=1}^{M}|a_j|^2.$$

Deducimos por tanto que

$$h\sum_{j=0}^{M}\left|\frac{a_{j+1}-a_j}{h}\right|^2 \geq \operatorname{sen}^2\left(\frac{h}{2}\right) h\sum_{j=1}^{M}|a_j|^2,$$

para todo $h>0$ y toda función discreta $\{a_0,\ldots,a_{M+1}\}$, con $a_0=a_{M+1}=0$.

Basta por último observar que

$$\frac{4}{h^2}\operatorname{sen}^2\left(\frac{h}{2}\right)\longrightarrow 1,\ h\to 0,$$

de modo que para todo $\delta>0$ existe $h_0>0$ tal que

$$\frac{4}{h^2}\operatorname{sen}^2\left(\frac{h}{2}\right)\geq 1-\delta,\quad \forall 0<h<h_0.$$

■

Aplicando (1.2.87) con $\delta=1/2$ en (1.2.86) obtenemos

$$\frac{d}{dt}\left[\frac{h}{2}\sum_{j=1}^{M}|\ v_j\ |^2\right]\leq -\frac{h}{2}\sum_{j=1}^{M}|\ v_j\ |^2+h\sum_{j=1}^{M}\varepsilon_j v_j\leq \frac{h}{2}\sum_{j=1}^{M}|\ \varepsilon_j\ |^2. \tag{1.2.89}$$

Por lo tanto

$$h\sum_{j=1}^{M}|v_j(t)|^2\leq h\sum_{j=1}^{M}\int_0^T|\varepsilon_j|^2\,dt,\ \forall 0<h<h_0,\ 0\leq t\leq T. \tag{1.2.90}$$

Basta por tanto con que estimemos el error producido por los términos ε_j (el error de truncación). Recordemos que

$$\varepsilon_j=u_{xx}(x_j,t)+\frac{\left[2\ \underline{u}_j-\ \underline{u}_{j+1}-\ \underline{u}_{j-1}\right]}{h^2}. \tag{1.2.91}$$

Como es bien sabido, el esquema en diferencias finitas de tres puntos proporciona una aproximación de orden dos del operador derivada segunda. Por lo tanto

$$|\varepsilon_j(t)|^2\leq C\,h^4\parallel u(t)\parallel^2_{C^4([0,\pi])},\ \forall j=1,\ldots,M,\ \forall 0<h<h_0,\ \forall 0\leq t\leq T. \tag{1.2.92}$$

Combinando (1.2.90) y (1.2.92) deducimos que

$$h\sum_{j=1}^{M}|v_j(t)|^2\leq C\,h^4\int_0^T\parallel u(t)\parallel^2_{C^4([0,\pi])}\,dt. \tag{1.2.93}$$

Hemos por tanto probado el siguiente resultado.

Theorem 1.2.2 *Supongamos que el dato inicial $\varphi = \varphi(x)$ es tal que la solución $u = u(x,t)$ de la ecuación del calor* (1.2.8) *verifica, para todo $T < \infty$,*

$$\int_0^T \| u(t) \|^2_{C^4([0,\pi])} \, dt < \infty. \tag{1.2.94}$$

Entonces, para todo $0 < T < \infty$ existe una constante $C_T > 0$ tal que

$$h \sum_{j=1}^{M} \left| \underline{u}_j(t) - u_j(t) \right|^2 = \| \ \underline{\vec{u}}(t) - \vec{u}_h(t) \ \|_h^2 \leq C_T h^4, \tag{1.2.95}$$

para todo $0 \leq t \leq T$ y para todo $h > 0$, donde $\underline{\vec{u}} = \{ \underline{u}_j \}_{j=1,\ldots,M}$ denota la restricción a los puntos del mallado de la solución de la ecuación del calor (1.2.8) y $\vec{u}_h = \{u_j\}_{j=1,\ldots,M}$ representa la solución del sistema semi-discreto (1.2.24).

Observación 1.2.4 En (1.2.95) hemos establecido que el sistema semi-discreto (1.2.24) proporciona una estimación de orden dos de la ecuación del calor (1.2.8). Este resultado cabía ser esperado pues la única discretización que ha sido realizada es la del laplaciano espacial, al ser sustituido por el esquema de tres puntos que, como es bien sabido, es una aproximación de orden dos.

El Teorema 1.2.2 ha sido establecido mediante el método de energía que ha sido aplicado en una de sus versiones más simples. Muchas otras variantes son posibles. En realidad, por cada estimación de energía de la que dispongamos para la ecuación del calor (1.2.8) se puede establecer un resultado de convergencia distinto que, bajo hipótesis de regularidad adecuadas sobre la solución de la ecuación del calor, confirmará que el método semi-discreto es de segundo orden. Recordemos que la estimación de energía (1.2.18) en la que nos hemos inspirado para probar el Teorema 1.2.2 se obtiene multiplicando la ecuación del calor (1.2.8) por u. Si en lugar de multiplicar por u, multiplicamos la ecuación por $\partial^{2m} u / \partial x^{2m}$, es decir por una derivada espacial de u de orden par, se obtiene una nueva identidad de energía de la forma [7]

$$\frac{d}{dt} \left[\frac{1}{2} \int_0^\pi \left| \frac{\partial^m u}{\partial x^m} \right|^2 dx \right] = - \int_0^\pi \left| \frac{\partial^{m+1} u}{\partial x^{m+1}} \right|^2 dx.$$

Como hemos mencionado anteriormente, esta identidad también tiene su análogo semi-discreto sobre el que podría establecerse un resultado del tipo del Teorema 1.2.2.

[7] En este caso se usa el hecho de que, como u, se anula en los extremos del intervalo espacial, para todo t, lo mismo ocurre con las derivadas temporales sucesivas. Esto conduce a que las derivadas de orden par de u con respecto a la variable espacial también se anulen para todo t.

Los comentarios realizados en el Teorema 1.2.1 son también aplicables en el Teorema 1.2.2. Por ejemplo, si bien en (1.2.95) hemos establecido una estimación en norma L^2, también el método de la energía nos habría permitido probar estimaciones en otras normas, por ejemplo, en la norma del máximo.

En el Teorema 1.2.2 hemos supuesto que la solución u de la ecuación del calor es suficientemente regular como para que (1.2.94) se satisfaga, i.e. que

$$u \in L^1\left(0,T;C^4([0,\pi]\right).$$

Esto, evidentemente, exige que el dato inicial sea también suficientemente regular. Bastaría por ejemplo con que el dato inicial fuese de clase C^3, aunque esta hipótesis se podría debilitar.

Si comparamos el Teorema 1.2.1 y el Teorema 1.2.2 se observa inmediatamente que si bien en el primero, usando series de Fourier, obteníamos un resultado de convergencia bajo condiciones mínimas sobre el dato inicial $\left(\varphi \in L^2(0,\pi)\right)$, en el método de la energía hemos utilizado hipótesis más fuertes sobre el dato pero, como contrapartida, hemos obtenido un resultado más fuerte puesto que hemos probado que el método es de orden dos. El método de Fourier también permite obtener ordenes de convergencia pero, tal y como se señaló en la Observación 1.2.2, eso exige hipótesis adicionales sobre el dato inicial, también en este caso.

Por consiguiente, el tipo de resultados que se puede obtener por ambos métodos es semejante, si bien el método de la energía es más flexible pues se puede aplicar en situaciones en las que, por la presencia de coeficientes que dependen del tiempo o de no-linealidades, las soluciones no pueden descomponerse en series de Fourier mediante el método de separación de variables.

■

1.2.4. Consistencia + estabilidad = Convergencia

Es habitual que en los textos dedicados al Análisis Numérico de EDP se incluya un Teorema, habitualmente atribuido a P. Lax, que garantiza que

Consistencia + Estabilidad = Convergencia.

Sin embargo, tanto al interpretar el concepto de estabilidad como el de convergencia y hacer uso de este resultado, lo mismo que ocurre en el marco de las EDP, nos enfrentamos a genuinos problemas de dimensión infinita y la elección que se hace de las normas es por tanto fundamental. Así, estos tres conceptos han de ser manipulados en un mismo contexto, una vez establecidas con claridad

las normas en las que trabajamos, lo cual, en realidad, consiste en determinar el criterio o distancia en la que se va a comprobar la convergencia del método.

Es por eso que, en estas notas, en lugar de incluir este enunciado como Teorema, lo presentamos simplemente como un principio general, de validez universal una vez que se han elegido con prudencia las normas, pero que conviene también utilizar con cuidado en cada caso. Es decir, para cada EDP y cada aproximación numérica habrá que elegir de manera precisa la norma en la que se va a medir la convergencia del método.

Lo más habitual es utilizar este principio general con el objeto de probar la convergencia. De este modo el problema se reduce a verificar dos propiedades básicas del esquema: su consistencia y estabilidad.

En las secciones anteriores estos dos conceptos han surgido ya y el principio general ha sido implícitamente utilizado en la demostración de los dos Teoremas de convergencia. El objeto de esta sección es comentar y clarificar el modo en que estos conceptos han intervenido y se han combinado en las pruebas de convergencia y de este modo ilustrar este principio general fundamental del Análisis Numérico de las EDP.

Si bien estos conceptos y principios están también presentes en la prueba del Teorema 1.2.1 realizada mediante desarrollos en series de Fourier, son tal vez más claros en la demostración del Teorema 1.2.2 realizada mediante el método de la energía. Nos centraremos pues en este segundo caso, dejando para el lector la reflexión sobre la prueba del primer resultado mediante el método de Fourier que comentaremos muy brevemente al final de la sección.

- *Consistencia:*

  La consistencia del método numérico hace referencia a su coherencia a la hora de aproximar la EDP. Se trata simplemente de comprobar si el esquema numérico utilizado es un esquema razonable para aproximar la EDP en cuestión o, si por el contrario, corre el riesgo de aproximar a otra EDP. Lo más habitual es comprobar la consistencia mediante el desarrollo de Taylor. El problema se reduce entonces a verificar si, cuando el paso del mallado tiende a cero, (en el caso que nos ocupa: $h \to 0$), las soluciones regulares del problema continuo son soluciones aproximadas del problema discreto en la medida en que el error de truncación tiende a cero. Cuando ésto es así con un error del orden de una potencia p del tamaño del mallado se dice que el método es de orden p.

  En nuestro caso particular, como (1.2.24) es una aproximación semi-discreta de (1.2.8) en la que la variable tiempo no ha sido discretizada, la propiedad

de consistencia del esquema se reduce meramente a la consistencia de la aproximación de tres puntos del operador de derivada segunda en espacio que, como sabemos y recordamos en (1.2.91) y (1.2.92), es efectivamente consistente de orden dos.

Conviene subrayar que, aunque pueda resultar paradójico, a la hora de comprobar la consistencia no verificamos hasta qué punto las soluciones del problema discreto son soluciones del problema continuo módulo un cierto error, sino que hacemos precisamente lo contrario: comprobamos si la solución del problema continuo es una solución aproximada del problema discreto.

Evidentemente ésto se hace exclusivamente a la hora de verificar la bondad del método numérico a priori puesto que, en la práctica, no disponemos de la solución del problema continuo: el objeto del cálculo numérico es precisamente aproximar la solución real de la EDP.

■

- *Estabilidad:*

  Pero la consistencia por sí sola no basta para garantizar la convergencia del método. Es preciso analizar su estabilidad.

  La propiedad de estabilidad consiste en asegurarse de que los esquemas discretos o semi-discretos, en su evolución temporal (discreta o continua), no amplifiquen los errores iniciales o, al menos, no lo hagan de manera creciente y descontrolada a medida que el paso del mallado tiende a cero.

  En el marco del esquema (1.2.24) esta propiedad de estabilidad queda debidamente recogida en (1.2.81) donde hemos probado que, para cualquier $h > 0$, la norma L^2 de las soluciones discretas (la norma $\|\cdot\|_h$) decrece con el tiempo. Se trata pues de una situación ideal en la que los errores iniciales no sólo no se amplifican en exceso sino que decrecen en tiempo.

  Si analizamos detenidamente la prueba del Teorema 1.2.2 mediante el método de la energía observaremos que ésta no es más que una combinación adecuada de las propiedades de consistencia y estabilidad y que por tanto obedece fielmente al principio de P. Lax antes mencionado. En efecto, con el objeto de probar la convergencia, hemos establecido en primer lugar que la solución del problema continuo es una solución aproximada del problema discreto (véase (1.2.83)). Es decir, hemos empezado usando la consistencia del método. Después, usando la linealidad del esquema

discreto, hemos introducido la variable $\vec{v}$ diferencia entre la solución del problema continuo y discreto, y hemos establecido que se trata de una solución aproximada del problema discreto en el que el dato inicial es nulo pero la dinámica está forzada por un segundo miembro $(\varepsilon_j,\, j = 1, \ldots, M)$ que tiende a cero cuando $h \to 0$ (véase (1.2.85)). Finalmente, usando la estabilidad, hemos establecido que esta diferencia se mantiene pequeña cuando el tiempo avanza y tiende a cero cuando $h \to 0$.

Puede resultar paradójico que el concepto de estabilidad haga referencia a la propagación de errores en el dato inicial y que, sin embargo, se aplique en el sistema (1.2.85) donde el dato inicial es exactamente cero pero en el que está presente una fuerza externa $\vec{\varepsilon}(t)$ continua en tiempo. Pero, en vista del principio de Duhamel o de la fórmula de variación de las constantes, esto no debería de resultar extraño pues el efecto de un segundo miembro continuo en una ecuación de evolución no es otro que el de la integral temporal de los efectos que ese segundo miembro tendría, en cada instante de tiempo, como perturbación del dato inicial.

Este hecho queda claramente reflejado en la fórmula de representación de la solución de ecuaciones diferenciales no homogéneas:

$$\begin{cases} \dot{x}(t) = Ax(t) + F(t), & t > 0 \\ x(0) = x_0. \end{cases}$$

La *fórmula de variación de las constantes* asegura en este caso que

$$x(t) = e^{At}x_0 + \int_0^t e^{A(t-s)}F(s)ds.$$

En esta fórmula se observa con claridad que el efecto de un segundo miembro en la ecuación es una integral temporal de los efectos que este segundo miembro tendría en el dato inicial.

■

Hemos descrito cómo los conceptos de consistencia y estabilidad han jugado un papel esencial en la prueba del resultado convergencia del Teorema 1.2.2. Estos conceptos están también presentes en la prueba realizada del Teorema 1.2.1 mediante el método de Fourier. En efecto, la consistencia del esquema garantiza su convergencia para las soluciones que involucran un solo modo de Fourier, mientras que la estabilidad asegura que basta con probar la convergencia para datos iniciales que involucran únicamente un número finito de componentes.

Nuevamente, la consistencia junto con la estabilidad proporcionan la convergencia del método.

Conviene sin embargo subrayar que los sistemas de EDO que obtenemos al realizar discretizaciones espaciales tienen un carácter *stiff*. Para ello basta observar que el ratio $\lambda_M(h)/\lambda_1(h)$ entre el máximo y mínimo autovalor de la matriz A_h es del orden de $1/h^2$. En virtud del análisis de los sistemas stiff realizados en capítulos previos, se trata de un hecho relevante que habremos de tener en cuenta a la hora de proceder a la discretización temporal del sistema, con el objeto de garantizar la convergencia de las discretizaciones completas espacio-tiempo.

En la sección 2.6 veremos algunos ejemplos de métodos semi-discretos y completamente discretos que, siendo consistentes, no son estables y, por tanto, no convergen. Este hecho confirma la necesidad tanto de la propiedad de consistencia como de estabilidad de un método numérico para garantizar su convergencia.

1.2.5. Aproximaciones completamente discretas

En las secciones 2.2 y 2.3 hemos analizado la convergencia del esquema semi-discreto (1.2.24) al modelo continuo (1.2.8). Los resultados de convergencia que hemos probado han legitimado la utilización del esquema (1.2.24) puesto que sus soluciones convergen a las de la ecuación del calor (1.2.8) cuando $h \to 0$. Desde un punto de vista práctico, estos resultados permiten concentrar nuestros esfuerzos en el cálculo de las soluciones (1.2.24) con h suficientemente pequeño pero fijo.

El sistema (1.2.24) es un sistema de M ecuaciones diferenciales acopladas con M incógnitas. Parece por tanto natural utilizar los métodos numéricos desarrollados para la resolución aproximada de ecuaciones diferenciales. Al hacerlo discretizando la variable temporal, acabamos obteniendo esquemas completamente discretos para la resolución de (1.2.8).

Esta sección está destinada al estudio de las dos aproximaciones completamente discretas más naturales en este contexto que son las que se obtienen al aplicar el método explícito e implícito de Euler para la resolución aproximada de EDO de primer orden. Por supuesto, la variedad de los métodos posibles es muy grande puesto que cualquier método de aproximación del laplaciano (de la segunda derivada espacial: esquema en diferencias finitas de orden superior, método espectral o de elementos finitos, ...) junto con cualquier método de aproximación de la derivada temporal (método del trapecio, de Runge-Kutta, o multi-paso, por ejemplo) dan lugar a un método completamente discreto nuevo.

Pero el análisis de estos dos métodos más básicos permite estudiar los aspectos fundamentales de la teoría que bastaría para analizar cualquier otro tipo de esquemas.

En esta ocasión el paso espacial será denotado mediante Δx (en lugar de h), mientras que el paso temporal estará denotado por Δt. Mediante u_j^k denotaremos la aproximación de la solución u de (1.2.8) en el punto $x = x_j = j\Delta x$ en el instante de tiempo $t = t_k = k\Delta t$.

Con esta nueva notación, si en el sistema semi-discreto (1.2.24) aplicamos el *esquema explícito de Euler* en la variable temporal obtenemos el sistema:

$$\begin{cases} \frac{u_j^{k+1}-u_j^k}{\Delta t} + \frac{\left[2u_j^k - u_{j+1}^k - u_{j-1}^k\right]}{(\Delta x)^2} = 0, & j = 1, \dots, M; \quad k \geq 0, \\ u_0^k = u_{M+1}^k = 0, & k \geq 0, \\ u_j^0 = \varphi_j, & j = 1, \dots, M. \end{cases} \tag{1.2.96}$$

Sin embargo, si aplicamos el método de Euler implícito obtenemos

$$\begin{cases} \frac{u_j^{k+1}-u_j^k}{\Delta t} + \frac{\left[2u_j^{k+1} - u_{j+1}^{k+1} - u_{j-1}^{k+1}\right]}{(\Delta x)^2} = 0, & j = 1, \dots, M; \quad k \geq 0, \\ u_0^k = u_{M+1}^k = 0, & k \geq 0, \\ u_j^0 = \varphi_j, & j = 1, \dots, M. \end{cases} \tag{1.2.97}$$

La diferencia principal entre (1.2.96) y (1.2.97) es la habitual entre esquemas explícitos e implícitos. Mientras que (1.2.97) permite "leer" explícitamente el valor de la solución discreta en el paso temporal $k+1$ a partir de la solución en el paso temporal k, el método implícito (1.2.97) exige, en cada paso temporal, la resolución de un sistema tridiagonal de M ecuaciones lineales con M incógnitas.

Con el objeto de estudiar la convergencia de estos métodos conviene utilizar el *número de Courant*

$$\mu = \Delta t/(\Delta x)^2, \tag{1.2.98}$$

en el que queda claramente reflejada la diferencia de homogeneidad en la variable espacial y temporal del operador en derivadas parciales involucrado en la ecuación del calor, en la que una derivada temporal juega el mismo papel que dos derivadas espaciales.

Con esta nueva notación los esquemas (1.2.96) y (1.2.97) pueden reescribirse del modo siguiente:

- *Método de Euler explícito:*

$$u_j^{k+1} = u_j^k + \mu\left[u_{j+1}^k + u_{j-1}^k - 2u_j^k\right]. \tag{1.2.99}$$

- *Método de Euler implícito:*

$$u_j^{k+1} + \mu \left[2u_j^{k+1} - u_{j+1}^{k+1} - u_{j-1}^{k+1}\right] = u_j^k. \tag{1.2.100}$$

La consistencia de ambos métodos está claramente garantizada.

En efecto, hemos utilizado el esquema de tres puntos para la aproximación de la segunda derivada espacial que es consistente de orden dos, mientras que la discretización de la derivada temporal mediante la diferencia finita que involucra dos niveles temporales (k y $k+1$) proporciona una aproximación consistente de orden uno en tiempo.

El análisis de la convergencia de los métodos se hará a valor de μ fijo. En otras palabras, describiremos el rango de valores del parámetro de Courant μ para el que los métodos discretos en consideración son convergentes. Fijado este valor del parámetro de Courant tenemos $\Delta t = \mu(\Delta x)^2$. Por tanto un orden de consistencia temporal corresponde a un orden dos de consistencia espacial. Es por eso que diremos que ambos métodos (1.2.99) y (1.2.100) son consistentes de orden dos (teniendo como referencia el paso espacial). De manera más precisa, si, dada una solución u del problema continuo original (1.2.8), introducimos la proyección de dicha solución sobre el mallado

$$\underline{u}_j^k = u(x_j,\, t_k) = u(j\Delta x,\, k\Delta t), \tag{1.2.101}$$

entonces $\underline{u}$ es solución aproximada de los esquemas discretos. En efecto se tiene

$$\underline{u}_j^{k+1} - \left[\, \underline{u}_j^k + \mu \left[u_{j+1}^k + \, \underline{u}_{j-1}^k - 2u_j^k\right]\right] = O\left(\Delta t \left((\Delta x)^2 + \Delta t\right)\right) \tag{1.2.102}$$

y

$$\underline{u}_j^{k+1} - \, \underline{u}_j^k + \mu \left[2 \, \underline{u}_j^{k+1} - \, \underline{u}_{j+1}^{k+1} - \, \underline{u}_{j-1}^{k+1}\right] = O\left(\Delta t \left((\Delta x)^2 + \Delta t\right)\right) \tag{1.2.103}$$

que corresponde efectivamente a la consistencia de orden dos de los métodos siempre y cuando u tenga dos derivadas temporales y cuatro derivadas espaciales acotadas.

Es natural que obtengamos el mismo tipo de hipótesis sobre dos derivadas temporales y cuatro derivadas espaciales de la solución de la ecuación del calor. En efecto, la ecuación garantiza que $u_t = u_{xx}$, lo cual implica también que $u_{tt} = u_{xxxx}$.

Conviene observar la presencia de un término adicional Δt en la estimación del error de truncación, con respecto al $(\Delta x)^2$ propio de la aproximación espacial

y al Δt de la aproximación temporal puesto que (1.2.96) y (1.2.97) han sido ambas multiplicadas por Δt para obtener (1.2.99) y (1.2.100).

Pero, lo mismo que ocurría en el caso de los esquemas semi-discretos, la consistencia no basta para garantizar la convergencia del método. De hecho estos dos esquemas tienen un comportamiento bien distinto en relación con el rango de valores de μ para los que se tiene la convergencia:

Theorem 1.2.3 *El método explícito* (1.2.96) *es convergente si* $0 < \mu \leq 1/2$ *mientras que el método implícito* (1.2.97) *lo es para todo* $\mu > 0$.

En ambos casos, los métodos son convergentes de orden dos, lo que significa que, si la solución de (1.2.8) *es suficientemente regular, se tiene, para todo* $0 < T < \infty$,

$$\max_{k=0,\ldots,[T/\Delta t]} \left| \vec{u^k} - \vec{\underline{u}^k} \right|_{\Delta x} = O\left((\Delta x)^2 \right) \tag{1.2.104}$$

cuando $\Delta x \to 0$.

Observación 1.2.5 En (1.2.104) mediante $\| \cdot \|_{\Delta x}$ denotamos la norma L^2 en el mallado de paso Δx, i.e.

$$\| a \|_{\Delta x} = \left[\Delta x \sum_{j=1}^{M} |a_j|^2 \right]^{1/2},$$

de modo que

$$\left| \vec{u^k} - \vec{\underline{u}^k} \right|_{\Delta x} = \left[\Delta x \sum_{j=1}^{M} \left| u_j^k - \underline{u}_j^k \right|^2 \right]^{1/2}.$$

Conviene también observar que $\vec{u^k}$ (resp. $\vec{\underline{u}^k}$) denota el vector de componentes $u_j^k, j = 1, \ldots, M$, (resp. $\underline{u}_j^k, j = 1, \ldots, M$) que, a su vez, representa la solución del problema discreto (resp. continuo) en el paso temporal k.

En (1.2.104) estimamos la norma de la diferencia entre la solución continua y discreta para los pasos $k = 0, 1, \ldots, [T/\Delta t]$, que son los necesarios para cubrir el intervalo temporal $[0, T]$.

Por otra parte, en (1.2.104), se enuncia la convergencia a cero del error cuando $\Delta x \to 0$. Evidentemente, aunque no se diga explícitamente, también $\Delta t \to 0$ puesto que en el enunciado se supone que el parámetro μ de Courant está fijado dentro del rango en el que se tiene la convergencia ($\mu \in (0, 1/2)$ para el método explícito y $\mu \in (0, \infty)$ para el implícito).

En virtud de este resultado de convergencia se observa la superioridad del método implícito frente al explícito puesto que su convergencia está garantizada para un valor arbitrario del parámetro de Courant μ. Esto permite tomar pasos temporales más grandes que para el método explícito.

■

Demostración del Teorema 2.3.

En la demostración supondremos que la solución u de la ecuación continua (1.2.8) es de clase C^2 en tiempo y C^4 en espacio. Esto garantiza que las identidades (1.2.102) y (1.2.103) sean válidas, uniformemente en $k = 0, \ldots, [T/\Delta t]$ y $j = 1, \ldots, M$, a medida que $\Delta t, \Delta x \to 0$.

Introducimos ahora el error

$$e_j^k = \underline{u}_j^k - u_j^k \tag{1.2.105}$$

que mide la diferencia entre las soluciones del problema continuo y del problema discreto.

El error e_j^k es solución del siguiente sistema dinámico discreto:

- *Método explícito:*

$$\begin{aligned} e_j^{k+1} &= e_j^k + \mu\left[e_{j+1}^k + e_{j-1}^k - 2e_j^k\right] + O\left(\Delta t\left((\Delta x)^2 + \Delta t\right)\right) \\ &= (1-2\mu)e_j^k + \mu\left(e_{j+1}^k + e_{j-1}^k\right) + O\left(\Delta t(\Delta x)^2 + \Delta t^2\right) \end{aligned} \tag{1.2.106}$$

- *Método implícito:*

$$\begin{aligned} e_j^{k+1} + \mu\left[2e_j^{k+1} - e_{j+1}^{k+1} - e_{j-1}^{k+1}\right] &= (1+2\mu)\, e_j^{k+1} - \mu\left(e_{j+1}^{k+1} + e_{j-1}^{k+1}\right) \\ &= e_j^k + O\left(\Delta t\left((\Delta x)^2 + \Delta t\right)\right). \end{aligned} \tag{1.2.107}$$

En la medida en que hemos supuesto que el dato inicial del problema discreto coincide exactamente con el del problema continuo tenemos

$$e_j^0 = 0,\ j = 1, \ldots, M. \tag{1.2.108}$$

Procedemos ahora a estimar el error distinguiendo los métodos explícito e implícito:

- *Método explícito*

A partir de (1.2.106) observamos que si

$$\varepsilon^k = \max_{j=1,\ldots,M} \left|e_j^k\right|, \tag{1.2.109}$$

se tiene

$$\varepsilon^{k+1} \leq \left[|1-2\mu| + 2\mu\right]\varepsilon^k + C\left(\Delta t\left((\Delta x)^2 + \Delta t\right)\right). \tag{1.2.110}$$

Iterando esta desigualdad y haciendo uso de (1.2.108) obtenemos

$$\varepsilon^k \leq C\left(\Delta t\left((\Delta x)^2+\Delta t\right)\right)\left[1+\alpha_\mu+\alpha_\mu^2+\cdots+\alpha_\mu^{k-1}\right] \tag{1.2.111}$$

donde

$$\alpha_\mu = \left[|1-2\mu|+2\mu\right]. \tag{1.2.112}$$

Claramente hay dos casos que distinguir: (i) $\mu \leq 1/2$; y (ii) $\mu > 1/2$. En el primero

$$\alpha_\mu = 1$$

y por tanto (1.2.111) se reduce a

$$\varepsilon^k \leq C\left(\Delta t\left((\Delta x)^2+\Delta t\right)\right)k \leq C_T\left((\Delta x)^2+\Delta t\right) = O\left((\Delta x)^2\right) \tag{1.2.113}$$

puesto que $k\Delta t \leq T$ y que $\Delta t = \mu(\Delta x)^2$. La desigualdad (1.2.113) proporciona el resultado de convergencia enunciado en el Teorema para $\mu \leq 1/2$.

La situación es distinta cuando $\mu > 1/2$. En este caso

$$\alpha_\mu = 4\mu - 1 > 1$$

y por tanto el miembro de la derecha de (1.2.111) diverge exponencialmente. En efecto, en el último paso temporal en el que $k \sim T/\Delta t$ tenemos que

$$\begin{aligned}
&C\left(\Delta t\left((\Delta x)^2+\Delta t\right)\right)\left(1+\alpha_\mu+\cdots+\alpha_\mu^{k-1}\right)\\
&\geq C\left(\Delta t\left((\Delta x)^2+\Delta t\right)\right)\alpha_\mu^{k-1} \geq C\left(\Delta t\left((\Delta x)^2+\Delta t\right)\right)e^{\beta_\mu T/\Delta t}\\
&\to \infty, \text{ cuando } \Delta x \to 0 \text{ con } \Delta t = \mu(\Delta x)^2.
\end{aligned} \tag{1.2.114}$$

Esta estimación del error no permite concluir la convergencia del método. En realidad en este caso el método es inestable y por tanto no converge. Este último aspecto será analizado con más en detalle en la próxima sección mediante el método de von Neumann.

- *Método implícito*

En este caso, de (1.2.107), con independencia del valor de $\mu > 0$ se deduce que

$$\varepsilon^{k+1} \leq \varepsilon^k + C\left(\Delta t\left((\Delta x)^2+\Delta t\right)\right). \tag{1.2.115}$$

En efecto, para comprobar este hecho basta con analizar la desigualdad (1.2.107) para el valor j^* de j para el que

$$\left|e_{j*}^{k+1}\right| = \varepsilon^{k+1} = \max_{j=1,\ldots,M}\left|e_j^{k+1}\right|.$$

De (1.2.107) con $j = j^*$ deducimos que

$$\begin{aligned}
\varepsilon^{k+1} &= (1+2\mu)\varepsilon^{k+1} - 2\mu\varepsilon^{k+1} \\
&\leq (1+2\mu)\left|e_{j^*}^{k+1}\right| - \mu\left(\left|e_{j^*+1}^{k+1}\right| + \left|e_{j^*-1}^{k+1}\right|\right) \\
&\leq \left|(1+2\mu)e_{j^*}^{k+1} - \mu\left(e_{j^*+1}^{k+1} + e_{j^*-1}^{k+1}\right)\right| \\
&\leq \left|e_{j^*}^{k}\right| + C\left(\Delta t\left((\Delta x)^2 + \Delta t\right)\right) \\
&\leq \varepsilon^k + C\left(\Delta t\left((\Delta x)^2 + \Delta t\right)\right).
\end{aligned}$$

En este argumento el valor de j^* puede depender de k pero esto es irrrelevante pues finalmente acabamos estableciendo una relación entre ε^k y ε^{k+1} que en nada depende del valor de j^*.

Por (1.2.115) observamos que, en el caso del método implícito, con independencia del valor de $\mu > 0$, nos encontramos exactamente en las mismas condiciones que en el método explícito cuando $\mu \in (0,\, 1/2)$. La convergencia del método implícito y, en particular, (1.2.104) quedan por tanto probados.

■

La demostración del resultado de convergencia ha sido realizada trabajando directamente y de manera elemental en las ecuaciones discretas (1.2.109) y (1.2.110) que describen cómo el error numérico se propaga. El mismo tipo de resultados podría haberse obtenido utilizando desarrollos en serie de Fourier. Este método será desarrollado con más detalle en la próxima sección donde describiremos el método de von Neumann para el estudio de la estabilidad y convergencia de métodos numéricos para problemas en toda la recta real o en un intervalo acotado con condiciones de contorno periódicas.

Pero, comentemos brevemente cómo el método de la sección 1.2.2 puede ser adaptado a este contexto de las aproximaciones completamente discretas, lo cual permite, en particular, comprobar el comportamiento patológico del método explícito cuando $\mu > 1/2$. En efecto, en (1.2.10) dimos ya el desarrollo en serie de Fourier de la solución de la ecuación del calor continua (1.2.8). Procedemos ahora al desarrollo de la solución del problema discreto explícito, en la base de los autovectores de la matriz $A_{\Delta x}(= A_h$ definida en (1.2.26) con $h = \Delta x)$. Tenemos que

$$\vec{u}^k = \sum_{\ell=1}^{M} p_\ell^k \vec{W}_\ell(\Delta x), \tag{1.2.116}$$

donde $\vec{W}_\ell(\Delta x)$ denota el $\ell-$ésimo autovector de la matriz $A_{\Delta x}$ y p_ℓ^k el coeficiente de Fourier de este $\ell-$ésimo autovector en el paso temporal k. Teniendo en cuenta

que

$$A_{\Delta x}\vec{W}_\ell(\Delta x) = \lambda_\ell(\Delta x)\vec{W}_\ell(\Delta x), \tag{1.2.117}$$

la función discreta definida en (1.2.116) satisface la ecuación discreta (1.2.96) o (1.2.99) si y sólamente si

$$p_\ell^{k+1} = p_\ell^k - \mu\lambda_\ell(\Delta x)(\Delta x)^2 p_\ell^k = \left[1 - \mu(\Delta x)^2\lambda_\ell(\Delta x)\right] p_\ell^k. \tag{1.2.118}$$

Por lo tanto

$$p_\ell^k = \left[1 - \mu(\Delta x)^2\lambda_\ell(\Delta x)\right]^k p_\ell^0 = (\alpha_\ell)^k\, p_\ell^0. \tag{1.2.119}$$

Es decir, cada componente de Fourier de la solución del problema discreto evoluciona de manera exponencial según la ley (1.2.119). Evidentemente, el comportamiento de p_ℓ^k, i.e. su evolución a medida que k aumenta, depende de que $|\alpha_\ell| \leq 1$ o $|\alpha_\ell| > 1$. Analicemos pues el valor de $|\alpha_\ell|$. Tenemos

$$\alpha_\ell = 1 - \mu(\Delta x)^2\lambda_\ell(\Delta x).$$

Por otra parte,

$$\lambda_\ell(\Delta x) = \frac{4}{(\Delta x)^2}\,\mathrm{sen}^2\left(\frac{\ell\Delta x}{2}\right).$$

Por tanto,

$$\alpha_\ell = 1 - 4\mu\,\mathrm{sen}^2\left(\frac{\ell\Delta x}{2}\right).$$

Cuando $0 \leq \mu \leq 1/2$, $|\alpha_\ell| \leq 1$ para todos los valores de ℓ. Esto asegura la estabilidad del método puesto que todos las componentes de Fourier de la solución discreta se mantienen acotadas en k, con independencia del valor Δx.

La situación es completamente distinta cuando $\mu > 1/2$. En efecto, en este caso, cuando $\ell = M = \frac{\pi}{\Delta x} - 1$, tenemos

$$\alpha_\ell = 1 - 4\mu\,\mathrm{sen}^2\left(\frac{\pi}{2} - \frac{\Delta x}{2}\right) = 1 - 4\mu\cos^2\left(\frac{\Delta x}{2}\right) \to 1 - 4\mu, \text{ cuando } \Delta x \to 0.$$

Obviamente $|1 - 4\mu| > 1$ cuando $\mu > 1/2$. Vemos por tanto que, cuando $\mu > 1/2$, para Δx suficientemente pequeño, existen índices ℓ para los que $|\ \alpha_\ell\ | > 1$, lo cual implica la inestabilidad del método.

El método explícito (1.2.96) (o (1.2.99)) con $\mu > 1/2$ es por tanto un primer y buen ejemplo de método numérico que, a pesar de ser consistente, no es convergente por no ser estable.

Esto no ocurre en el método implícito (1.2.97) (o (1.2.100)) en el que la ley de evolución de los coeficientes de Fourier es de la forma:

$$p_\ell^{k+1} + \mu(\Delta x)^2\lambda_\ell(\Delta x)p_\ell^{k+1} = p_\ell^k,$$

es decir

$$p_\ell^k = \left(1 + \mu(\Delta x)^2 \lambda_\ell(\Delta x)\right)^{-k} p_\ell^0.$$

En este caso, evidentemente, la estabilidad queda garantizada para todo valor del número de Courant $\mu > 0$ puesto que

$$\left|\left(1 + \mu(\Delta x)^2 \lambda_\ell(\Delta x)\right)^{-1}\right| = \frac{1}{1 + \mu(\Delta x)^2 \lambda_\ell(\Delta x)} < 1,$$

para todo $\Delta x > 0$ y todo $\ell = 1, \ldots, M$.

No sería difícil desarrollar este método de Fourier y completarlo con las técnicas desarrolladas en la sección 1.2.2 para dar una demostración alternativa del Teorema 1.2.3.

Conviene también subrayar que la inestabilidad que hemos detectado en el método explícito cuando $\mu > 1/2$ y, en particular, la interpretación que acabamos de hacer mediante el desarrollo en series de Fourier se asemeja en gran medida al tipo de análisis que hacíamos en el contexto de los problemas "stiff". En efecto, en aquella ocasión introdujimos el concepto $A-$estabilidad en el que exigíamos que el esquema numérico a $h > 0$ fijo reprodujese, al avanzar el tiempo, las propiedades de estabilidad de la ecuación diferencial continua. En el caso que nos ocupa la ecuación continua en cuestión es el sistema semi-discreto (1.2.24) (o (1.2.27)) en el que sabemos que, a h fijo, todas las soluciones tienden a cero exponencialmente cuando $t \to \infty$ (para comprobarlo basta observar el desarrollo (1.2.41) en serie de Fourier de la solución del problema semi-discreto). Cuando $\mu > 1/2$ hemos comprobado que el esquema numérico explícito deja de reproducir este comportamiento asintótico puesto que surgen componentes de Fourier que se amplifican exponencialmente cuando $k \to \infty$.

A pesar de ello, hemos visto que el esquema explícito converge cuando $0 \leq \mu \leq 1/2$. Esto, sin embargo, no está en contradicción con el resultado que asegura la falta de $A-$estabilidad de los métodos de Runge-Kutta explícitos (el método de Euler explícito es efectivamente un método de Runge-Kutte explícito). En efecto, en este caso estamos analizando el sistema (1.2.24) (o (1.2.27)) en el que la matriz involucrada, $A_{\Delta x}$, si bien es sumamente dispersa pues su autovalor mínimo es $\lambda_1(\Delta x) = 4\,\mathrm{sen}^2(\Delta x/2)/\,(\Delta x)^2$ (muy próximo a 1 cuando $\Delta x \sim 0$) y el máximo es sin embargo $\lambda_M(\Delta x) = 4\cos^2(\Delta x/2)/\,(\Delta x)^2$ (muy próximo a $4/(\Delta x)^2$ cuando $\Delta x \sim 0$), su grado de dispersión está limitado. Esto queda de manifiesto en el hecho que

$$(\Delta x)^2 \lambda_l(\Delta x) \leq 4,\ \forall \Delta x > 0,\ \forall j = 1, \ldots, M$$

y es lo que permite garantizar la convergencia del método explícito, pero sólo en el rango $0 < \mu \leq 1/2$.

Vemos pues que el mal comportamiento del esquema explícito responde exactamente a las mismas patologías habituales de los métodos explícitos a la hora de abordar los problemas stiff, si bien, en el caso particular de la ecuación del calor que nos ocupa, ésto no es incompatible con la convergencia del método para $0 < \mu \leq 1/2$.

Sin embargo, esta limitación sobre μ obliga a tomar pasos temporales muy pequeños, lo cual hace que el coste computacional de la implementación de este método sea muy elevado.

Conviene por último observar que los métodos semi-discretos pueden verse como límite cuando $\mu \to 0$ de los métodos discretos. En efecto, si fijado Δx pasamos al límite de un esquema completamente discreto cuando $\Delta t \to 0$ recuperamos un esquema semi-discreto. Esto corresponde a considerar el número de Courant $\mu = 0$. El hecho que el método explícito sea convergente para $0 < \mu \leq 1/2$, a pesar de no serlo para $\mu > 1/2$, es pues perfectamente compatible (y lo explica) con el hecho de que el método semi-discreto (1.2.24) sea convergente, tal y como vimos en las secciones 1.2.2 y 1.2.3.

1.2.6. El análisis de von Neumann

En esta sección presentamos el análisis de von Neumann para el estudio de la estabilidad de esquemas numéricos, que es especialmente útil en problemas definidos en todo el espacio con mallados regulares. El método, basado en la utilización de la transformada discreta de Fourier, es muy semejante al desarrollado en la sección 1.2.2 y 1.2.5 mediante el desarrollo en series de Fourier.

Con el objeto de ilustrar las ideas básicas de este método consideramos el problema de Cauchy para la ecuación del calor en toda la recta real:

$$\begin{cases} u_t - u_{xx} = 0, & x \in \mathbf{R}, \quad t > 0 \\ u(x,0) = \varphi(x), & x \in \mathbf{R}. \end{cases} \tag{1.2.120}$$

Suponemos que

$$\varphi \in L^2(\mathbf{R}) \tag{1.2.121}$$

de modo que (1.2.120) admite una única solución

$$u \in C\left([0,\infty); L^2(\mathbf{R})\right). \tag{1.2.122}$$

En este caso la identidad de energía garantiza que

$$\frac{d}{dt}\left[\frac{1}{2}\int_{\mathbf{R}} u^2(x,t)dx\right] = -\int_{\mathbf{R}} |u_x|^2 \, dx. \tag{1.2.123}$$

Como indicamos en la sección 1.2.1, la solución de (1.2.120) puede representarse por convolución con la solución fundamental

$$G(x,t) = (4\pi t)^{1/2} \exp\left(-\mid x \mid^2 / 4t\right), \tag{1.2.124}$$

de modo que

$$u(x,t) = \left[G(\cdot,t) * \varphi\right](x) = (4\pi t)^{-1/2} \int_{\mathbf{R}} \exp\left(-\frac{\mid x-y \mid^2}{4t}\right) \varphi(y) dy. \tag{1.2.125}$$

Dado $h > 0$ introducimos la partición

$$x_j = jh,\ j \in \mathbf{Z} \tag{1.2.126}$$

y consideramos la semi-discretización

$$\begin{cases} u_j' + \frac{[2u_j - u_{j+1} - u_{j-1}]}{h^2} = 0, & j \in \mathbf{Z}, \quad t > 0 \\ u_j(0) = \varphi_j, & j \in \mathbf{Z}. \end{cases} \tag{1.2.127}$$

En este caso (1.2.127) es un sistema de una infinidad de ecuaciones diferenciales de primer orden acopladas de tres en tres. Son varias las maneras en las que se puede resolver el sistema (1.2.127). Una de las más naturales consiste en truncar el sistema y considerarlo en un rango de índices finito $[-M,\ M]$ con condiciones de contorno de Dirichlet:

$$\begin{cases} u_j' + \frac{[2u_j - u_{j+1} - u_{j-1}]}{h^2} = 0, & -M < j < M, \quad t > 0 \\ u_{-M} = u_M = 0, & t > 0 \\ u_j(0) = \varphi_j. \end{cases} \tag{1.2.128}$$

Es fácil comprobar que la solución de (1.2.128), cuando $M \to \infty$, converge a la solución de (1.2.127).

Pero, admitiendo que el sistema (1.2.127) ha sido ya resuelto (problema sobre el que volveremos más adelante) analicemos la convergencia de las soluciones de (1.2.127) a las de (1.2.120) cuando $h \to 0$.

Para el sistema (1.2.127) se verifica la siguiente identidad de energía

$$\frac{d}{dt}\left[\frac{h}{2}\sum_{j\in\mathbf{Z}} |u_j|^2\right] = -h\sum_{j\in\mathbf{Z}} \left|\frac{u_{j+1}-u_j}{h}\right|^2, \tag{1.2.129}$$

que es claramente una versión discreta de (1.2.123). En (1.2.129) estamos implícitamente asumiendo que $\vec{u}(t) = \{u_j(t)\}_{j\in\mathbf{Z}} \in \ell^2 = \ell^2(\mathbf{Z})$ lo cual equivale a suponer que $\vec{\varphi} = \{\varphi_j\}_{j\in\mathbf{Z}} \in \ell^2$. Esta última condición se verifica puesto que el dato inicial φ del problema continuo (1.2.120) pertenece a $L^2(\mathbf{R})$, como indicamos en (1.2.121).

El esquema (1.2.127) es consistente de orden dos. La prueba de este hecho, basada en el desarrollo de Taylor, es independiente de que el intervalo en el que se verifique la ecuación sea acotado o no. El esquema será por tanto convergente si comprobamos que es estable. Es en la verificación de la propiedad de estabilidad donde el método de von Neumann basado en la utilización de la transformada discreta de Fourier resulta sumamente útil. En este caso sin embargo la estabilidad es también inmediata a partir de la identidad de energía (1.2.129).

Pero existen otros esquemas semi-discretos sumamente naturales que, siendo consistentes, carecen de la propiedad de estabilidad y, por tanto, no son convergentes. Para observar este hecho basta considerar la siguiente familia de métodos:

$$\alpha u'_{j+1} + (1-2\alpha)u'_j + \alpha u'_{j-1} + \frac{[2u_j - u_{j+1} - u_{j-1}]}{h^2} = 0, \tag{1.2.130}$$

donde $0 \leq \alpha \leq 1/2$. Cuando $\alpha = 0$ recuperamos el sistema anterior que, como vimos, es convergente.

Para esta familia de métodos la consistencia es fácil de verificar. Sin embargo la estabilidad no está garantizada y, de hecho, sólo se verifica cuando $0 \leq \alpha \leq 1/4$. Como consecuencia de este hecho, el método es divergente cuando $\alpha > 1/4$. Para comprobar este hecho lo más sencillo es utilizar el método de von Neumann que introducimos en el caso completamente discreto. Volveremos sobre el ejemplo (1.2.130) más adelante.

Introducimos ahora los dos esquemas completamente discretos más naturales en la aproximción de la ecuación del calor, correspondientes a utilizar el método de Euler explícito e implícito para la discretización de la derivada temporal.

El *método explícito* se escribe en este caso

$$\begin{cases} \frac{u_j^{k+1}-u_j^k}{\Delta t} + \frac{\left[2u_j^k - u_{j+1}^k - u_{j-1}^k\right]}{(\Delta x)^2} = 0, & j \in \mathbf{Z}, \quad t > 0, \\ u_j^0 = \varphi_j, & j \in \mathbf{Z}, \end{cases} \tag{1.2.131}$$

mientras que el *método implícito* es

$$\begin{cases} \frac{u_j^{k+1}-u_j^k}{\Delta t} + \frac{\left[2u_j^{k+1} - u_{j+1}^{k+1} - u_{j-1}^{k+1}\right]}{(\Delta x)^2} = 0, & j \in \mathbf{Z}, \quad t > 0, \\ u_j^0 = \varphi_j, & j \in \mathbf{Z}. \end{cases} \tag{1.2.132}$$

Nuevamente u_j^k representa una aproximación de $u\,(j\Delta x,\, k\Delta t)$.

Los esquemas (1.2.131) y (1.2.132) son ambos consistentes de orden dos tal y como vimos en la sección 1.2.5. La solución en el paso temporal k es una sucesión $\left\{u_j^k\right\}_{j\in\mathbf{Z}}$ que denotamos de manera simplificada mediante la notación vectorial $\vec{u}^k$.

En el estudio de la estabilidad utilizamos la *transformada discreta de Fourier*. Así a cada sucesión

$$\vec{w} = \{w_l\}_{j\in\mathbf{Z}} \in \ell^2(\mathbf{Z}) \tag{1.2.133}$$

le asignamos la función continua

$$\check{w}(\theta) = \sum_{j=-\infty}^{\infty} w_j e^{-ij\theta}. \tag{1.2.134}$$

Recíprocamente, tenemos la fórmula de inversión

$$w_j = \frac{1}{2\pi}\int_0^{2\pi} \check{w}(\theta) e^{ij\theta} d\theta. \tag{1.2.135}$$

La aplicación que a $\vec{w}$ le asocia $\check{w}$ es una isometría de $\ell^2(\mathbf{Z})$ en $L^2(0,2\pi)$ si dotamos a estos espacios con las normas

$$\| \vec{w} \|_{\ell^2(\mathbf{Z})} = \sum_{j=-\infty}^{\infty} |w_j|^2 \tag{1.2.136}$$

y

$$||\check{w}||_{L^2(0,\pi)} = \left[\frac{1}{2\pi}\int_0^{2\pi} |\check{w}(\theta)|^2 d\theta\right]^{1/2}. \tag{1.2.137}$$

Para comprobar estas propiedades basta con tener en cuenta las identidades:

$$\int_0^{2\pi} e^{ij\theta} e^{-ik\theta} d\theta = \begin{cases} 0 & \text{si} \quad j \neq k \\ 2\pi & \text{si} \quad j = k. \end{cases} \tag{1.2.138}$$

Mediante esta transformación de Fourier, a cada solución $\left\{\vec{u}^k\right\}_{k\geq 0}$ de (1.2.131) o (1.2.132) le podemos asignar funciones continuas $\left\{\check{u}^k(\theta)\right\}_{k\geq 0}$ donde

$$\check{u}^k(\theta) = \sum_{j\in\mathbf{Z}} u_j^k e^{-ij\theta}. \tag{1.2.139}$$

La clave en el estudio de la estabilidad de los esquemas numéricos mediante esta transformada de Fourier reside en la siguiente identidad:

$$\begin{aligned} \sum_{j\in\mathbf{Z}} a_{j+1} e^{-ij\theta} &= \sum_{j\in\mathbf{Z}} a_{j+1} e^{-i(j+1)\theta} e^{i\theta} = e^{i\theta} \sum_{j\in\mathbf{Z}} a_{j+1} e^{-i(j+1)\theta} \\ &= e^{i\theta} \sum_{j\in\mathbf{Z}} a_j e^{-ij\theta} = e^{i\theta} \check{a}(\theta). \end{aligned} \tag{1.2.140}$$

De modo análogo tenemos

$$\sum_{j\in\mathbf{Z}} a_{j-1} e^{-ij\theta} = e^{-i\theta} \check{a}(\theta). \tag{1.2.141}$$

Así, tomando la transformada de Fourier en las ecuaciones (1.2.131) y (1.2.132) obtenemos las siguientes ecuaciones para la transformada de Fourier $\check{u}^k(\theta)$.

- *Método explícito:*

$$\frac{\check{u}^{k+1}(\theta)-\check{u}^k(\theta)}{\Delta t}+\frac{a\left(e^{i\theta}\right)}{(\Delta x)^2}\check{u}^k(\theta)=0,\quad k\geq 0,\,\theta\in[0,2\pi); \tag{1.2.142}$$

- *Método implícito:*

$$\frac{\check{u}^{k+1}(\theta)-\check{u}^k(\theta)}{\Delta t}+\frac{a\left(e^{i\theta}\right)}{(\Delta x)^2}\check{u}^{k+1}(\theta)=0,\quad k\geq 0,\,\theta\in[0,2\pi), \tag{1.2.143}$$

donde

$$a\left(e^{i\theta}\right)=2-e^{i\theta}-e^{-i\theta}. \tag{1.2.144}$$

En vista de la equivalencia de las normas de $\ell^2(\mathbf{Z})$ y de $L^2(0,2\pi)$ para las sucesiones de $\ell^2(\mathbf{Z})$ y sus transformadas de Fourier, para analizar la evolución de la norma de $\vec{u}^k$ en $\ell^2(\mathbf{Z})$ basta con analizar la norma $L^2(0,2\pi)$ de $\check{u}^k(\theta)$. Ahora bien, en vista de (1.2.142) y (1.2.143) tenemos que

$$\check{u}^{k+1}(\theta)=\left[1-\mu a\left(e^{i\theta}\right)\right]\check{u}^k(\theta) \tag{1.2.145}$$

y

$$\check{u}^{k+1}(\theta)=\frac{1}{\left[1+\mu a\left(e^{i\theta}\right)\right]}\,\check{u}^k(\theta) \tag{1.2.146}$$

en el método explícito e implícito respectivamente. Aquí y en lo que sigue μ denota el número de Courant $\mu=\Delta t/(\Delta x)^2$.

Iterando esta regla de recurrencia obtenemos para cada uno de estos esquemas

$$\check{u}^k(\theta)=\left[1-\mu a\left(e^{i\theta}\right)\right]^k\widehat{\varphi}(\theta) \tag{1.2.147}$$

y

$$\check{u}^k(\theta)=\left[1+\mu a\left(e^{i\theta}\right)\right]^{-k}\widehat{\varphi}(\theta). \tag{1.2.148}$$

Tenemos por tanto

$$\sum_{j\in\mathbf{Z}}\left|u_j^k\right|^2=\frac{1}{2\pi}\int_0^{2\pi}\left|\check{u}^k(\theta)\right|^2d\theta=\frac{1}{2\pi}\int_0^{2\pi}\left|1-\mu a\left(e^{i\theta}\right)\right|^{2k}\left|\widehat{\varphi}(\theta)\right|^2d\theta \tag{1.2.149}$$

y

$$\sum_{j\in\mathbf{Z}}\left|u_j^k\right|^2=\frac{1}{2\pi}\int_0^{\pi}\left|\check{u}^k(\theta)\right|^2d\theta=\frac{1}{2\pi}\int_0^{2\pi}\left|1+\mu a\left(e^{i\theta}\right)\right|^{-2k}\left|\widehat{\varphi}(\theta)\right|^2d\theta. \tag{1.2.150}$$

La comprobación de la estabilidad del esquema numérico consiste en probar cotas sobre el crecimiento de la norma $L^2(0,2\pi)$ de $\check{u}^k(\theta)$ de manera controlada, independientemente del paso del mallado. En la medida que para recorrer un intervalo temporal $[0,T]$ el número de pasos temporales que hemos de dar es del orden de $k \sim T/\Delta t$, es fácil comprobar que la condición necesaria y suficiente para la estabilidad de cada uno de los métodos es que

$$\left|1-\mu a\left(e^{i\theta}\right)\right| \leq 1,\ \forall \theta \in [0,2\pi) \tag{1.2.151}$$

y

$$\left|1+\mu a\left(e^{i\theta}\right)\right|^{-1} \leq 1,\ \forall \theta \in [0,2\pi) \tag{1.2.152}$$

respectivamente.

La suficiencia de las condiciones (1.2.151) y (1.2.152) para la estabilidad del método explícito e implícito respectivamente es evidente pues, bajo estas condiciones, en virtud de (1.2.149) y (1.2.150), se tiene

$$\left|\vec{u^k}\right|_{\ell^2(\mathbf{Z})} \leq \left|\vec{\varphi}\right|_{\ell^2(\mathbf{Z})},\ \forall k \geq 0. \tag{1.2.153}$$

El hecho de que las condiciones (1.2.151) o (1.2.152) sean también necesarias exige un poco más de reflexión. Pero para entenderlo basta observar que si para algún $\theta_0 \in [0,2\pi)$ el módulo de alguna de estas cantidades es estrictamente mayor que uno, por continuidad lo es en un intervalo de la forma $[\theta_0-\delta,\,\theta_0+\delta]$ para algún $\delta > 0$. Basta entonces considerar funciones φ tales que el soporte de $\widehat{\varphi}$ esté contenido en $[\theta_0-\delta,\,\theta_0+\delta]$ para ver que la norma de la solución $\vec{u^k}$ correspondiente crece exponencialmente en k. En este punto conviene observar que, en vista de la fórmula de inversión (1.2.135), a cada elección $\widehat{\varphi}$ le corresponde una del dato inicial discreto $\vec{\varphi} \in \ell^2(\mathbf{Z})$.

Por lo tanto para analizar la estabilidad de los métodos discretos en consideración basta estudiar el rango de números de Courant μ para el que se satisface (1.2.151) y (1.2.152), respectivamente.

• *Método explícito*

Conviene observar que

$$a\left(e^{i\theta}\right) = 2 - e^{i\theta} - e^{-i\theta} = 2(1-\cos\theta). \tag{1.2.154}$$

Por tanto,

$$\left|1-\mu a\left(e^{i\theta}\right)\right| = \left|1-2\mu(1-\cos\theta)\right|.$$

Es fácil comprobar que

$$\left|1-2\mu\left(1-\cos\theta\right)\right| \leq 1,\ \forall \theta \in [0,2\pi) \Leftrightarrow 0 \leq \mu \leq 1/2.$$

Deducimos por tanto que el método explícito (1.2.131) es estable y por tanto convergente (puesto que, como dijimos, el método es consistente de orden dos) si y sólo si $0 < \mu \leq 1/2$.

• *Método implícito*

En vista de (1.2.154) es evidente que

$$1 + \mu a\left(e^{i\theta}\right) \geq 1,\ \forall \theta \in [0, 2\pi),\ \forall \mu > 0.$$

Por lo tanto (1.2.152) se satisface para todo $\mu > 0$ y por tanto el método implícito (1.2.132) es estable (y, por consiguiente, convergente) para todo $\mu > 0$.

De este modo vemos que en el caso del problema de Cauchy en toda la recta real se tienen los mismos resultados que en el caso de un intervalo acotado con condiciones de contorno de Dirichlet. Volvamos ahora sobre el sistema semi-discreto (1.2.130). Recordemos que, tal y como se comprueba con facilidad mediante el desarrollo de Taylor, el método es consistente si $0 \leq \alpha \leq 1/2$.

Con el objeto de estudiar la estabilidad observamos que si $\vec{u}$ satisface (1.2.130), entonces, $\check{u}$ verifica:

$$(1 - 2(\alpha - \cos\theta))\check{u}'(\theta, t) + \frac{a(e^{i\theta})}{h^2}\check{u}(\theta, t) = 0, \tag{1.2.155}$$

o, de otro modo,

$$\check{u}'(\theta, t) + \frac{b(\theta)}{h^2}\check{u}(\theta, t) = 0, \tag{1.2.156}$$

donde

$$b(\theta) = \frac{a(e^{i\theta})}{1 - 2(\alpha - \cos\theta)}. \tag{1.2.157}$$

En este caso, por tanto, el esquema es estable, sí y sólo sí, $b(\theta) \geq 0$ para todo $\theta \in [0, 2\pi)$ y esto ocurre sólo cuando $0 < \alpha < 1/4$.

Vemos por tanto que, gracias al análisis de von Neumann, es fácil estudiar también la estabilidad de los métodos semi-discretos y que, tal y como indicabamos al inicio de esta sección, permite comprobar la existencia métodos consistentes y no estables y, por tanto, no convergentes. A la hora de elegir los datos iniciales en el sistema semi-discreto (1.2.127) o en los sistemas completamente discretos (1.2.131) y (1.2.132) tenemos varias posibilidades:

(a) Cuando $\varphi \in L^2(\mathbf{R}) \cap C(\mathbf{R})$ podemos elegir simplemente

$$\varphi_j = \varphi(jh);$$

(b) Cuando $\varphi \in L^2(\mathbf{R})$ podemos sustituir la evaluación puntual de φ por el conjunto de sus medias

$$\varphi_j = \frac{1}{h}\int_{(j-1/2)h}^{(j+1/2)h} \varphi(x)dx;$$

(c) Cuando $\varphi \in L^2(\mathbf{R})$ podemos también definir los datos iniciales utilizando la transformada de Fourier. Sea $\mathcal{F}(\varphi) \in L^2(\mathbf{R})$ la transformada de Fourier continua de φ, i.e.

$$\mathcal{F}(\varphi)(\xi) = \int_{\mathbf{R}} \varphi(x)e^{-ix\xi}dx.$$

Es natural aproximar esta transformada de Fourier por una función de banda limitada

$$\psi_h = \mathcal{F}(\varphi)1_{[-\pi/h,\,\pi/h]},$$

donde $1_{[-\pi/h,\,\pi/h]}(\xi)$ denota la función característica del intervalo $\xi \in [-\pi/h,\,\pi/h]$. Basta entonces elegir el dato inicial φ_j del problema semi-discreto o completamente discreto como el valor de la antitransformada de Fourier de ψ_h en el punto jh (o $j\Delta x$ en el caso completamente discreto).

Conviene por último señalar que la transformada discreta de Fourier puede ser utilizada para probar la existencia y unicidad de las soluciones del problema semi-discreto y discreto.

Consideremos por ejemplo el problema semi-discreto (1.2.127). Se trata de un sistema de una infinidad de ecuaciones diferenciales de orden uno acopladas de tres en tres. Por lo tanto, los resultados clásicos de la teoría de EDO no pueden ser aplicados por tratarse de un problema en dimensión infinita. La transformada discreta de Fourier puede, efectivamente, utilizarse para resolver (1.2.127). En efecto, $\{\vec{u}(t)\} = \{u_j(t)\}_{j\in\mathbf{Z}}$ es la solución de (1.2.127) si y sólo si la transformada de Fourier correspondiente $\check{u}(\theta,t)$ verifica

$$\begin{cases} \check{u}_t(\theta,t) + \frac{1}{h^2}a(e^{i\theta})\check{u}(\theta,t) = 0, & \theta \in [0,2\pi), \quad t>0, \\ \check{u}(\theta,0) = \widehat{\varphi}(\theta), & \theta \in [0,2\pi). \end{cases} \tag{1.2.158}$$

De (1.2.158) deducimos que

$$\check{u}(\theta,t) = \exp\left[-\frac{1}{h^2}a\left(e^{i\theta}\right)\right]\widehat{\varphi}(\theta). \tag{1.2.159}$$

Como la transformada de Fourier define una isometría de $\ell^2(\mathbf{Z})$ en $L^2(0,\pi)$ sabemos que, como $\vec{\varphi} \in \ell^2(\mathbf{Z})$ entonces $\widehat{\varphi}(\theta) \in L^2(0,2\pi)$. Como $a\left(e^{i\theta}\right) \geq 0$ para todo $\theta \in [0,2\pi)$ tenemos que

$$|\check{u}(\theta,t)| \leq |\widehat{\varphi}(\theta)|\,,\ \forall\theta \in [0,2\pi),\ \forall t>0.$$

Por tanto, $\breve{u}(\theta,t) \in L^2(0,2\pi)$, para todo $t>0$. Por consiguiente, la transformada inversa de Fourier $\vec{u}(t) \in \ell^2(\mathbf{Z})$. De este modo deducimos que para todo $\vec{\varphi} \in \ell^2(\mathbf{Z})$ el sistema (1.2.127) admite una única solución $\vec{u}(t) \in C\left([0,\infty);\, \ell^2(\mathbf{Z})\right)$ que es la transformada de Fourier inversa de la solución de (1.2.158).

Vemos por tanto que la transformada discreta de Fourier no sólo es un método para analizar la estabilidad y convergencia de los métodos numéricos sino incluso para resolver las ecuaciones semi-discreta y discreta que en ellos intervienen.

1.2.7. El método de elementos finitos

Tal y como vimos en el capítulo anterior en el estudio de la ecuación de Laplace, uno de los métodos más frecuentemente utilizados en la actualidad es el *método de elementos finitos* (MEF).

El MEF, si bien en una dimensión espacial y en problemas con coeficientes constantes como los que estamos estudiando proporciona esquemas numéricos de aproximación muy semejantes a los que hemos barajado mediante diferencias finitas, conceptualmente es bien distinto. Como vimos en el contexto de la ecuación de Laplace, el MEF consiste en buscar "soluciones.en un espacio de dimensión finita. No adoptamos pues el punto "de vista nodal"sino el del *método de Galerkin*: habiendo elegido una base de funciones y definido subespacios de dimensión finita, elegimos en él la función que es solución de la EDP en el sentido más preciso posible. Para ello comprobamos si el producto escalar del operador diferencial con todas las funciones de dicho espacio es nulo (para lo cual es preciso adoptar una formulación variacional del problema que se obtiene integrando por partes). De este modo obtenemos una proyección de la solución real de la EDP sobre el espacio finito-dimensional considerado.

Esta metodología puede también ser aplicada en el marco de las ecuaciones de evolución y, en particular, en el de la ecuación del calor analizada en esta sección.

Presentemos pues brevemente la adaptación del MEF al problema que nos ocupa. Lo haremos en el caso de las aproximaciones semi-discretas si bien la adaptación al caso de aproximaciones completamente discretas (explícitas o implícitas en tiempo) es automática.

Consideramos la misma partición $\{x_j\}_{j=1,\ldots,M}$, $x_j = jh$ del intervalo $(0,\pi)$. A cada nodo interno x_j, $j=1,\ldots,M$ le asociamos una función de base $\phi_j(x)$, continua y lineal a trozos tal que $\phi_j(x_l) = \delta_{jl}$, $j=1,\ldots,M; l=0,\ldots,M+1$, siendo δ la delta de Kronecker.

Introducimos ahora el subespacio vectorial de dimensión M generado por las funciones $\{\phi_j\}_{j=1,...,M}$:

$$V_h = \text{span}\,\{\phi_j : j = 1, ..., M\}. \tag{1.2.160}$$

Buscamos ahora soluciones aproximadas de la ecuación del calor en el subespacio $C([0,T]; V_h)$ de modo que

$$u_h(x,t) = \sum_{j=1}^{M} u_j(t)\phi_j(x). \tag{1.2.161}$$

Observese que se trata de una función que depende tanto de la variable espacial x como de la temporal. Por tanto, en este caso, el método numérico proporciona automáticamente una función continua. Es sin embargo evidente que la función u_h así obtenida, para cada instante de tiempo t, varía con respecto a x como una función continua y lineal a trozos. Es pues imposible que pueda reproducir exactamente la dinámica de cualquier solución $u = u(x,t)$ de la ecuación del calor. Sin embargo, como veremos, se puede demostrar con bastante facilidad que esta función puede ser elegida de modo que converja a la solución del calor cuando $h \to 0$.

Pero para que la función u_h esté unívocamente determinada es imprescindible definir con precisión sus coeficientes $u_j(t)$. Conviene señalar en este punto que, debido a la elección de las funciones de base $\{\phi_j\}$ realizada, $u_j(t)$ es también el valor en el punto $x = x_j$ de la función u_h. Sin embargo, en el marco de los elementos finitos la interpretación más natural es que $u_j(t)$ son los coeficientes de u_h como función que, en cada instante t, pertenece al subespacio V_h.

Conviene también señalar que el método de elementos finitos no es sólo una herramienta para la aproximación numérica de EDP sino que en realidad permite modelizar fenómenos de la Mecánica de Medios continuos a través de modelos discretos, que pueden ser manipulados mucho más fácilmente tanto de un punto de vista conceptual como computacional. Es por eso que el MEF está tan extendido en la literatura de las diversas ramas de la Ingeniería y que fue introducido paralelamente tanto en el ámbito de la modelización como de las aproximaciones numéricas hasta convertirse en una disciplina unificada y bien establecida.[8]

Debemos por tanto obtener las ecuaciones más naturales que gobiernan la dinámica de los coeficientes $u_j(t)$ de la solución aproximada. Como decíamos lo hacemos a través del método de Galerkin. Para ello es necesario, en primer lugar,

[8] El lector interesado en una descripción de los orígenes del método de elementos finitos podrá consultar ([32]).

introducir la formulación variacional de la ecuación del calor (1.2.8). Recordemos que la solución de esta ecuación pertenece al espacio $u \in C([0,T];L^2(0,\pi)) \cap L^2(0,T;H_0^1(0,\pi)$. La formulación variacional de (1.2.8) es por tanto la siguiente: *Hallar* $u \in C([0,T];L^2(0,\pi)) \cap L^2(0,T;H_0^1(0,\pi))$[9] *que satisfaga*

$$\frac{d}{dt}\int_0^\pi u(x,t)\Phi(x)dx = -\int_0^\pi u_x(x,t)\Phi_x(x)dx, \text{ p. c. t. } t>0,\ \forall \Phi \in H_0^1(0,\pi), \tag{1.2.162}$$

junto con la condición inicial

$$u(x,0)=\varphi(x), \quad \text{en} \quad (0,\pi). \tag{1.2.163}$$

Conviene comentar brevemente el sentido de esta formulación variacional. La condición inicial de (1.2.8) ha sido, evidentemente, tenida en cuenta en (1.2.163). Esta última tiene sentido puesto que u depende continuamente en tiempo a valores en el espacio $L^2(0,\pi)$. Su evaluación en un instante t dado y, en particular, en $t=0$ está pues plenamente justificada. Por otra parte, en (1.2.162) se da una versión débil o distribucional de la ecuación del calor (1.2.8). Los dos términos de (1.2.162) tienen sentido. El segundo miembro es una función de $L^2(0,T)$ puesto que la solución u pertenece al espacio $L^2(0,T;H_0^1(0,\pi))$. Sin embargo, a causa del efecto regularizante de la ecuación del calor, se trata también de una función de $C((0,T];H_0^1(0,\pi))$ por lo que podría exigirse que (1.2.163) se verifique no sólo para casi todo t en el intervalo $[0,T]$ sino para todo $t>0$. Por otra parte, como $u \in C([0,T];L^2(0,\pi))$ tenemos que la función $\int_0^\pi u(x,t)\Phi(x)dx$ es continua en tiempo. Su derivada temporal ha de ser por tanto entendida en el sentido de las distribuciones. Sin embargo, la solución de la ecuación calor también pertenece al espacio $H^1(0,T;H^{-1}(0,\pi))$, siendo $H^{-1}(0,\pi)$ el dual de $H_0^1(0,\pi)$, por lo que esta derivada temporal tiene también sentido en $L^2(0,T)$ o, por el efecto regularizante, un sentido puntual para cada $t>0$. Por último señalar que la condición de contorno que se impone en (1.2.8) está incorporada en la formulación variacional al haber exigido a la función u que pertenezca al espacio $L^2(0,T;H_0^1(0,\pi))$. Pero no vamos a profundizar más en este terreno.

[9] El espacio $C([0,T];L^2(0,\pi)) \cap L^2(0,T;H_0^1(0,\pi))$ es un espacio de Banach, intersección de $C([0,T];L^2(0,\pi))$ y de $L^2(0,T;H_0^1(0,\pi))$, en el que la norma viene dada por la suma de las normas en cada uno de los espacios. Las funciones de $C([0,T];L^2(0,\pi))$, dependen continuamente de $t \in [0,T]$ y toman valores en $L^2(0,\pi)$, y su norma viene dada por $||f||_{C([0,T];L^2(0,\pi))} = \text{máx}_{t\in[0,T]}\,||f(t)||_{L^2(0,\pi)}$ mientras que las de $L^2(0,T;H_0^1(0,\pi))$ son funciones mediables y de cuadrado integrable de $t \in 0,T]$ a valores en el espacio de Sobolev $H_0^1(0,\pi)$. La norma correspondiente es entonces $||f||_{L^2(0,T;H_0^1(0,\pi))} = \left[\int_0^T \int_0^\pi f_x^2(x,t)dxdt\right]^{1/2}$.

El lector interesado en un estudio más detallado del punto de vista variacional para la resolución de EDP de evolución podrá consultar los textos [7] y [16].

Inspirándonos en la formulación variacional de la ecuación del calor es fácil introducir el análogo discreto que caracteriza a la solución que el método de elementos finitos proporciona. La formulación variacional para el cálculo de la solución aproximada u_h es pues la siguiente: *Hallar* $u_h \in C^1([0,T];V_h)$ *tal que*

$$\frac{d}{dt}\int_0^\pi u_h(x,t)\Phi_j(x)dx = -\int_0^\pi u_{h,x}(x,t)\Phi_{j,x}(x)dx,\ \forall\, t>0,\ \forall j=1,...,M, \tag{1.2.164}$$

junto con la condición inicial

$$u_h(x,0)=\varphi_h(x), \quad \text{en} \quad (0,\pi). \tag{1.2.165}$$

Las analogías entre la formulación variacional del problema continuo (1.2.162)-(1.2.163) y la del problema discreto (1.2.164)-(1.2.165) son evidentes . En esta ocasión, como $u_h(t)$ "vive.[en] cada instante de t en el espacio de dimensión finita V_h, suponemos que u_h es de clase C^1 en tiempo y eso permite escribir (1.2.165) para cada instante t. Por otra parte, en (1.2.164) exigimos que la versión débil de la ecuación del calor discretizada se verifique para todas las funciones de base Φ_j, $j=1,...,M$. Esto es totalmente equivalente a suponer que se verifica para cualquier función test Φ del espacio V_h. En (1.2.165) hemos tomado un dato inicial $u_{h,0}$. Tal y como comentamos en el marco de las diferencias finitas, son varias las maneras en las que este dato inicial se puede tomar de manera que aproxime el dato inicial u_0 de la ecuación del calor. La manera más sencilla en este contexto es tal vez elegir φ_h como la proyección ortogonal de u_0 sobre V_h.

El sistema (1.2.164)-(1.2.165) es un sistema de M ecuaciones diferenciales lineales acopladas en las incógnitas $u_j(t)$, $j=1,...,M$.

Este sistema se puede escribir de manera mucho más explícita usando las matrices de masa $\mathcal{M}_h$ y de rigidez $\mathcal{R}_h$ del método de elementos finitos.

Recordemos que los elementos m_{jk} de la matriz de masa $\mathcal{M}_h$ son precisamente

$$m_{jk}=\int_0^\pi \Phi_j(x)\Phi_k(x)dx, \tag{1.2.166}$$

y lo elementos (r_{jk}) de la matriz de rigidez $\mathcal{R}_h$:

$$r_{jk}=\int_0^\pi \Phi_{j,x}(x)\Phi_{k,x}(x)dx. \tag{1.2.167}$$

Todos estos coeficientes pueden calcularse explícitamente con facilidad y constituyen en ambos casos matrices simétricas tridiagonales, con coeficientes

constantes a lo largo de las diagonales y de las sub y super-diagonales. De manera más precisa se tiene:

$$\mathcal{M}_h = h \begin{pmatrix} 2/3 & 1/6 & 0 & 0 \\ 1/6 & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & 1/6 \\ 0 & 0 & 1/6 & 2/3 \end{pmatrix}, \tag{1.2.168}$$

y

$$\mathcal{R}_h = \frac{1}{h} \begin{pmatrix} 2 & -1 & 0 & 0 \\ -1 & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & -1 \\ 0 & 0 & -1 & 2 \end{pmatrix}. \tag{1.2.169}$$

Observese que, en particular,

$$\mathcal{R}_h = hA_h, \tag{1.2.170}$$

donde A_h es la matriz (1.2.26) que interviene en la discretización de tres puntos del Laplaciano mediante elementos finitos.

El sistema (1.2.164)-(1.2.165), con la notación vectorial empleada en el método de diferencias finitas puede entonces escribirse en la forma:

$$\begin{cases} \mathcal{M}_h \vec{u}'(t) + \mathcal{R}_h \vec{u}(t) = 0, \quad t > 0 \\ \vec{u}(0) = \vec{\varphi}. \end{cases} \tag{1.2.171}$$

Ambas matrices $\mathcal{M}_h$ y $\mathcal{R}_h$ comparten los autovectores de la matriz A_h. Así los autovectores de estas matrices son $\{\vec{W}_\ell(h)\}_{\ell=1,\dots,M}$ definidos (1.2.30). Esto permite desarrollar con facilidad las soluciones de (1.2.171) en la base de estos autovectores. En efecto, teniendo en cuenta que la ecuación en (1.2.171) puede también escribirse en la forma

$$\vec{u}'(t) + \mathcal{M}_h^{-1} \mathcal{R}_h \vec{u}(t) = 0, \tag{1.2.172}$$

calculando los autovalores de $\mathcal{M}_h^{-1}\mathcal{R}_h$ que vienen dados por

$$\mu_j(h) = \frac{6}{h^2} \frac{1 - \cos(jh)}{2 + \cos(jh)}, j = 1, \cdots, M \tag{1.2.173}$$

Obtenemos así la siguiente expresión en series de Fourier de las soluciones del problema semi-discreto:

$$\vec{u}(t) = \vec{u_h}(t) = \sum_{j=1}^{M} \hat{\varphi}_j e^{-\mu_j(h)t} \vec{w}_l(h). \tag{1.2.174}$$

A partir de esta expresión no es difícil adaptar los resultados de convergencia que hemos probado para la aproximación mediante diferencias finitas al caso presente del MEF. Nuevamente, en este caso simple unidimensional nos encontramos con que los autovectores del esquema discreto coinciden en los nodos del mallado con los del problema continuo. Y nuevamente también los autovalores $\mu_j(h)$ del problema discreto convergen a los del continuo cuando $h \to 0$. En efecto, de (1.2.173) se deduce inmediatamente que

$$\mu_j(h) \to j^2, \quad \text{cuando} \quad h \to 0, \tag{1.2.175}$$

para cada j fijo.

También el método de energía puede ser desarrollado en este caso sin dificultad. En efecto, multiplicando la ecuación (1.2.171) componente a componente por $u_j(t)$, obtenemos la identidad:

$$\frac{d}{dt}\mathcal{E}_h(t) = -h\sum_{j=0}^{M}\left|\frac{u_{j+1}-u_j}{h}\right|^2 \tag{1.2.176}$$

donde en esta ocasión la energía viene dada por

$$\mathcal{E}_h(t) = \frac{h}{6}\sum_{j=1}^{M} |\ u_j\ |^2 + \frac{h}{12}\sum_{j=0}^{M} |u_j + u_{j+1}|^2 . \tag{1.2.177}$$

A partir de esta ley de disipación de la energía discreta (que es un análogo de la energía L^2 de la ecuación del calor continua), se establecen con facilidad resultados de convergencia similares a los probados en el caso del método de diferencias finitas.

1.3. La ecuación de ondas

Tal y como hemos mencionado en la introducción, la ecuación de ondas es otro de los modelos más relevantes que se escribe en términos de EDP puesto que interviene, de uno u otro modo, en infinidad de problemas de la Mecánica, de la Física y de la Ingeniería. Así, se trata de un modelo ubicuo en elasticidad y vibraciones de estructuras pero también en el ámbito de la propagación de ondas acústicas o electromagnéticas.

Desde un punto de vista matemático la ecuación de ondas es el opuesto exacto de la del calor pues se trata de un sistema reversible en tiempo, conservativo, carente de efectos regularizantes y en el que la velocidad de propagación es finita.

En esta sección seguiremos el guión de la anterior. En primer lugar recordaremos algunas propiedades básicas de la ecuación de ondas que nos servirán de guía a la hora de analizar los modelos discretizados correspondientes. Después estudiaremos la convergencia de las aproximaciones semi-discretas y por último las completamente discretas.

Por una cuestión de espacio nos ceñiremos a la ecuación de ondas en una sola dimensión espacial. Sin embargo, muchos de los conceptos y resultados que veremos y desarrollaremos se adaptan con relativa facilidad a la ecuación de ondas en varias dimensiones espaciales, al sistema de Lamé en elasticidad, a las ecuaciones de placas que involucran habitualmente el operador biharmónico, las ecuaciones de Schrödinger de la Mecánica Cuántica o incluso a otros modelos como las ecuaciones de Korteweg-de-Vries para las olas en canales poco profundos. Todos estos temas quedan para desarrollos posteriores.

1.3.1. Propiedades básicas de la ecuación de ondas $1-d$

En primer lugar consideramos el problema de Cauchy en toda la recta:

$$\begin{cases} u_{tt} - u_{xx} = 0, & x \in \mathbf{R}, \quad t > 0 \\ u(x,0) = \varphi(x), u_t(x,0) = \psi(x), & x \in \mathbf{R}. \end{cases} \tag{1.3.1}$$

El operador en derivadas parciales involucrado $\partial_t^2 - \partial_x^2$ se denota frecuentemente mediante el símbolo $\Box$ y se denomina d'Alembertiano. Su versión multidimensional es

$$\Box = \partial_t^2 - \Delta_x = \partial_t^2 - \sum_{i=1}^{N} \partial_{x_i}^2.$$

Pero en esta sección nos limitaremos al caso unidimensional.

En una dimensión espacial la ecuación de ondas es un modelo simplificado para las vibraciones de pequeña amplitud de una cuerda y mediante $u = u(x,t)$ se describen las deformaciones verticales de la misma.

La solución de (6.3.1) puede calcularse de forma explícita. En efecto, es fácil comprobar que la solución de (6.3.1) es

$$u(x,t) = \frac{1}{2}\left[\varphi(x+t) + \varphi(x-t)\right] + \frac{1}{2}\int_{x-t}^{x+t} \psi(s)ds. \tag{1.3.2}$$

Conviene también observar que u es de la forma

$$u(x,t) = \mathcal{F}(x+t) + \mathcal{G}(x-t) \tag{1.3.3}$$

donde

$$\mathcal{F}(s) = \frac{1}{2}\varphi(s) + \frac{1}{2}\int_0^s \psi(\sigma)d\sigma, \tag{1.3.4}$$

$$\mathcal{G}(s) = \frac{1}{2}\varphi(s) + \frac{1}{2}\int_s^0 \psi(\sigma)d\sigma. \tag{1.3.5}$$

Es también digno de mención que cualquier función de la forma (1.3.3) es solución de (6.3.1). Este hecho corresponde a la siguiente factorización del *operador de d'Alembert*

$$\Box = \partial_t^2 - \partial_x^2 = (\partial_t - \partial_x)(\partial_t + \partial_x) = (\partial_t + \partial_x)(\partial_t - \partial_x), \tag{1.3.6}$$

según la cual el *operador de ondas* es la composición de los dos *operadores de transporte* de orden uno $\partial_t \pm \partial_x$, cuyas soluciones son efectivamente *ondas viajeras* de la forma $\mathcal{F}(x+t)$ o $\mathcal{G}(x-t)$.

En la fórmula (1.3.2) se observa también otra de las propiedades fundamentales de la ecuación de ondas: *la velocidad finita de propagación.* Así el valor de la solución u en el punto (x,t) depende exclusivamente del valor de los datos iniciales en el *intervalo de dependencia* $[x-t,\, x+t]$. Por otra parte, una perturbación de los datos iniciales en el instante $t=0$ en el punto x_0 sólo afecta al valor de la solución en el cono de influencia $|x-x_0| < |t|$.

Otra de las propiedades que se deduce de la fórmula de representación (1.3.2) es la *ausencia de efecto regularizante.* En efecto, de (1.3.2) se deduce que la solución u, en cualquier instante $t > 0$, es tan regular como el dato inicial φ para la posición y gana una derivada con respecto a la velocidad inicial ψ. Del mismo modo, la velocidad u_t tiene la misma regularidad que ψ y pierde una derivada con respecto a φ.

Al tratarse de una ecuación de orden dos en tiempo, las genuinas incógnitas del problema son tanto u como u_t. Es por eso que en (6.3.1) hemos de proporcionar los datos iniciales de ambas incógnitas para garantizar la existencia y unicidad de soluciones. Desde este punto de vista es habitual escribir la ecuación (6.3.1) como un sistema de la forma

$$\begin{cases} u_t = v \\ v_t = u_{xx} \end{cases}$$

o bien como un sistema de leyes de conservación hiperbólica

$$\begin{cases} u_t = w_x \\ w_t = u_x \end{cases}$$

Como hemos dicho anteriormente algunas de estas propiedades de la ecuación se preservan en más de una dimensión espacial. En particular, se tienen fórmulas de representación explícitas semejantes a (1.3.2) aunque algo más complejas ([7], [30]).

Otra de las propiedades importantes de la ecuación de ondas es la *ley de conservación de la energía*. En este caso la energía correspondiente es

$$E(t) = \frac{1}{2}\int_{\mathbf{R}} \left[|u_x(x,t)|^2 + |u_t(x,t)|^2\right] dx \tag{1.3.7}$$

y se tiene

$$\frac{dE}{dt}(t) = 0,\ \forall t \geq 0, \tag{1.3.8}$$

lo cual es fácil de comprobar multiplicando la ecuación de (6.3.1) por u_t e integrando en $\mathbf{R}$.

Esta ley de conservación de la energía sugiere que el espacio $H^1(\mathbf{R}) \times L^2(\mathbf{R})$ es un marco funcional adecuado para la resolución de la ecuación de ondas. Y, efectivamente, es así. Por tanto, para todo dato inicial $(\varphi, \psi) \in H^1(\mathbf{R}) \times L^2(\mathbf{R})$ existe una única solución de (6.3.1) en la clase $u \in C\left([0,\infty);\, H^1(\mathbf{R})\right) \cap C^1\left([0,\infty);\, L^2(\mathbf{R})\right)$. Consiguientemente vemos que el vector incógnita preserva, con continuidad en tiempo, la regularidad de los datos iniciales.

Son diversos los métodos que permiten probar este tipo de resultados de existencia y unicidad: método de Galerkin, teoría de semigrupos, transformada de Fourier, etc. Pero en el caso que nos ocupa puede deducirse inmediatamente de la fórmula de d'Alembert (1.3.2).

De la fórmula de representación (1.3.2) se deduce que la ecuación (6.3.1) está bien puesta en infinidad de otros espacios. Pero el más natural para resolverlo y el que se extiende de manera natural a otras situaciones como problemas de frontera o situaciones multidimensionales es precisamente el marco hilbertiano $H^1(\mathbf{R}) \times L^2(\mathbf{R})$.

Consideramos ahora la ecuación de ondas en un intervalo acotado:

$$\begin{cases} u_{tt} - u_{xx} = 0, & 0 < x < \pi, \quad t > 0 \\ u(0,t) = u(\pi,t) = 0, & t > 0 \\ u(x,0) = \varphi(x),\ u_t(x,0) = \psi(x), & 0 < x < \pi. \end{cases} \tag{1.3.9}$$

Se trata de un modelo simplificado para las vibraciones de una cuerda de longitud π. En este caso hemos impuesto condiciones de contorno de Dirichlet que indican que la cuerda está fija en sus extremos, si bien los resultados que aquí describimos se adaptan con facilidad a otras.

Las soluciones de (1.3.9) se pueden representar fácilmente en series de Fourier. En efecto, si los datos iniciales admiten el desarrollo en serie de Fourier

$$\varphi(x) = \sum_{l\geq 1} \hat{\varphi}_l w_l(x);\ \psi(x) = \sum_{l\geq 1} \hat{\psi}_l w_l(x),\ 0 < x < \pi, \tag{1.3.10}$$

donde $w_l(x)$ viene dado por (1.2.9), la solución de (1.3.9) se escribe del siguiente modo

$$u(x,t) = \sum_{l=1}^{\infty} \left(\hat{\varphi}_l \cos(lt) + \frac{\hat{\psi}_l}{l} \operatorname{sen}(lt) \right) w_l(x). \tag{1.3.11}$$

Esta expresión puede ser simplificada utilizando exponenciales complejas:

$$u(x,t) = \sum_{l=-\infty}^{+\infty} \hat{\theta}_l e^{ilt} w_l(x), \tag{1.3.12}$$

donde

$$w_{-l}(x) = w_l(x),\ l \geq 1 \text{ y } \hat{\theta}_l = \frac{l\hat{\varphi}_l - i\hat{\psi}_l}{2l},\ \hat{\theta}_{-l} = \hat{\theta}_l + \frac{i\hat{\psi}_l}{l}, \quad \forall l \geq 1. \tag{1.3.13}$$

Como veremos más adelante, las soluciones de los problemas semi-discretos y completamente discretos que consideramos admiten desarrollos en serie de Fourier semejantes.

Nuevamente, la energía de las soluciones de (1.3.9) se conserva en tiempo.

En efecto

$$E(t) = \frac{1}{2} \int_0^{\pi} \left[|u_t(x,t)|^2 + |u_x(x,t)|^2 \right] dx, \tag{1.3.14}$$

satisface

$$\frac{dE}{dt}(t) = 0,\ \forall t \geq 0, \tag{1.3.15}$$

lo cual puede comprobarse fácilmente de dos maneras. La primera, por el método de la energía, multiplicando la ecuación (1.3.9) por u_t e integrando con respecto a $x \in (0,\pi)$. La segunda mediante simple inspección del desarrollo en serie de Fourier (1.3.11). El espacio natural para resolver (1.3.9) es $H_0^1(0,\pi) \times L^2(0,\pi)$. De ese modo, cuando $(\varphi,\psi) \in H_0^1(0,\pi) \times L^2(0,\pi)$, (1.3.9) admite una única solución $u \in C\left([0,\infty);\ H_0^1(0,\pi)\right) \cap C^1\left([0,\infty);\ L^2(0,\pi)\right)$.

La energía equivale al cuadrado de la norma canónica del espacio $H_0^1(0,\pi) \times L^2(0,\pi)$. El hecho de que la energía permanezca constante en tiempo equivale a que la trayectoria de la solución permanece indefinidamente en una esfera del espacio $H_0^1(0,\pi) \times L^2(0,\pi)$, la que corresponde a los datos iniciales del sistema.

El hecho de que los datos iniciales del problema pertenezcan a $H_0^1(0,\pi) \times L^2(0,\pi)$ equivale a que los coeficientes $\{\vec{\varphi}_l\}_{l\geq 1}$ y $\left\{\vec{\psi}_l\right\}_{l\geq 1}$ del desarrollo en serie

de Fourier (1.3.11) de u satisfagan

$$\sum_{l\geq 1}\left[l^2\left|\vec{\varphi}_l\right|^2+\left|\vec{\psi}_l\right|^2\right]<\infty, \tag{1.3.16}$$

o, equivalentemente,

$$\sum_{l=-\infty}^{+\infty}\left|\vec{\theta}_l\right|^2 l^2<\infty. \tag{1.3.17}$$

1.3.2. Semi-discretización espacial: El método de Fourier

Con las notaciones de la sección 1.2 dedicada a la ecuación del calor, la semi-discretización espacial más natural de la ecuación de ondas (1.3.9) es la siguiente:

$$\begin{cases} u_j''+\frac{[2u_j-u_{j+1}-u_{j-1}]}{h^2}=0, & j=1,\ldots,M, \quad t>0\\ u_0=u_{M+1}=0, & t>0\\ u_j(0)=\varphi_j,\ u_j'(0)=\psi_j, & j=1,\ldots,M. \end{cases} \tag{1.3.18}$$

Se trata de un sistema acoplado de M ecuaciones diferenciales de orden dos con M incógnitas. Las M incógnitas $u_1(t),\ldots,u_M(t)$ proporcionan aproximaciones de la solución continua $u=u(x,t)$ de la ecuación de ondas (1.3.9) en los puntos $x_1,\ldots,x_M$ del mallado. En vista de las condiciones de contorno de Dirichlet es natural imponer que los desplazamientos u_0 y u_{M+1} que proporcionan aproximaciones de u en los extremos $x_0=0$ y $x_{M+1}=\pi$ sean nulos.

Los datos iniciales $(\varphi_j,\,\psi_j)_{j=1}^M$ son típicamente una aproximación de los datos iniciales $(\varphi(x),\,\psi(x))$ de la ecuación de ondas (1.3.9) sobre el mallado. Tal y como vimos en la sección anterior, son diversas las elecciones posibles. En cada uno de los resultados de convergencia que probaremos indicaremos explícitamente la elección de los datos iniciales realizada.

Con la notación vectorial de la sección 1.2.2 el sistema (1.3.18) puede escribirse en la forma

$$\begin{cases} \vec{u}''(t)+A_h\vec{u}(t)=0, \quad t>0,\\ \vec{u}(0)=\vec{\varphi}, \quad \vec{u}'(0)=\vec{\psi}. \end{cases} \tag{1.3.19}$$

Las soluciones de (1.3.19) también pueden desarrollarse en serie de Fourier. Así, los datos iniciales admiten el desarrollo

$$\vec{\varphi}=\sum_{l=1}^{M}\hat{\varphi}_l(h)\vec{W}_l(h);\ \vec{\psi}=\sum_{l=1}^{M}\hat{\psi}_l(h)\vec{W}_l(h), \tag{1.3.20}$$

donde

$$\hat{\varphi}_l(h)=\langle\vec{\varphi},\,\vec{W}_l(h)\rangle_h;\ \hat{\psi}_l(h)=\langle\vec{\psi},\,\vec{W}_l(h)\rangle_h. \tag{1.3.21}$$

Entonces, la solución de (1.3.19) puede escribirse del siguiente modo

$$\vec{u}_h(t) = \sum_{l=1}^{M} \left(\hat{\varphi}_l(h) \cos(\mu_l(h)t) + \frac{\hat{\psi}_l(h)}{\mu_l(h)} \operatorname{sen}(\mu_l(h)t) \right) \vec{W}_l(h), \tag{1.3.22}$$

donde

$$\mu_l(h) = \sqrt{\lambda_l(h)}. \tag{1.3.23}$$

Aquí y en lo sucesivo $\vec{W}_l(h)$ y $\lambda_l(h)$ denotan los autovectores y autovalores de la matriz A_h introducidos en (1.2.29)-(1.2.30).

Nuevamente la expresión puede simplificarse:

$$\vec{u}_h(t) = \sum_{l=-M}^{M} \hat{\theta}_l(h) e^{i\mu_l(h)t} \vec{W}_l(h), \tag{1.3.24}$$

con la convención

$$\vec{W}_{-l}(h) = \vec{W}_l(h),\ l = 1, \ldots, M;\ \mu_{-l}(h) = -\mu_l(h),\ l = 1, \ldots, M, \tag{1.3.25}$$

donde

$$\hat{\theta}_l(h) = \frac{\mu_l(h)\hat{\varphi}_l(h) - i\hat{\psi}_l(h)}{2\mu_l(h)},\ \hat{\theta}_{-l}(h) = \hat{\theta}_l(h) + \frac{i\hat{\psi}_l(h)}{2\mu_l(h)}, \quad \forall l \geq 1. \tag{1.3.26}$$

Si tenemos en cuenta que, para l fijo, $\mu_l(h) \to l$ cuando $h \to 0$, a la vez la analogía entre las expresiones (1.3.13) y (1.3.24) son evidentes.

El sistema (1.3.18) y (1.3.19) tiene también la propiedad de conservación de la energía. En este caso la energía conservada admite la expresión

$$E_h(t) = \frac{h}{2} \sum_{k=0}^{M} \left[\left| \frac{u_{j+1} - u_j}{h} \right|^2 + \left| u_j' \right|^2 \right]. \tag{1.3.27}$$

La ley de conservación de la energía puede obtenerse al menos de dos maneras: a) A partir del desarrollo en serie de Fourier de las soluciones (1.3.22); b) Multiplicando cada una de las ecuaciones de (1.3.18) por u_j' y sumando con respecto al índice j.

La energía discreta E_h es evidentemente una aproximación de la energía continua E de (1.3.14) y las dos formas que hemos indicado de probar su carácter conservativo son también versiones discretas de las pruebas habituales en la ecuación de ondas continua.

De todas estas observaciones se deduce que el sistema semi-discreto (1.3.18) es una aproximación natural del sistema continuo (1.3.9).

Con el objeto de establecer la convergencia cuando $h \to 0$ de las solcuiones del sistema semi-discreto a las de la ecuación de ondas hemos de introducir

normas que midan la distancia entre ambas. Tenemos por una parte la norma L^2 discreta introducida en la sección 1.2:

$$\left|\vec{a}\right|_h = \left[h\sum_{j=1}^{M}|a_j|^2\right]^{1/2}. \tag{1.3.28}$$

Introducimos también la norma H^1-discreta correspondiente

$$\left|\vec{a}\right|_{1,h} = \left[h\sum_{j=0}^{M}\left|\frac{a_{j+1}-a_j}{h}\right|^2\right]^{1/2}, \tag{1.3.29}$$

siempre con la convención $a_0 = a_{M+1} = 0$.

Pero es también natural medir la proximidad de las soluciones en función de sus coeficientes de Fourier. Tal y como vimos en la sección 1.3.1, las soluciones de energía finita de la ecuación de ondas pueden ser identificadas, en virtud de (1.3.17), con sus coeficientes de Fourier pertenecientes al espacio $\mathcal{H}^1$:

$$\mathcal{H}^1 = \left\{\{\hat{a}_l\}_{l\in\mathbf{Z}} : \sum_{l\in\mathbf{Z}} l^2\left|\hat{a}_l\right|^2 < \infty\right\} \tag{1.3.30}$$

que es un espacio de Hilbert con la norma

$$\left|\{\hat{a}_l\}_{l\in\mathbf{Z}}\right|_{\mathcal{H}^1} = \left[\sum_{l\in\mathbf{Z}} l^2\left|\hat{a}_l\right|^2\right]^{1/2}. \tag{1.3.31}$$

En efecto, tal como observamos en (1.3.17), las soluciones de energía finita de la ecuación de ondas corresponden a datos iniciales que se representan a través de los coeficientes $\left\{\hat{\theta}_l\right\}_{l\in\mathbf{Z}} \in \mathcal{H}^1$.

En vista de la representación de Fourier (1.3.12) de las soluciones de la ecuación de ondas, éstas pueden escribirse del siguiente modo

$$u(x,t) = \sum_{l=-\infty}^{+\infty}\hat{u}_l(t)w_l(x), \tag{1.3.32}$$

donde

$$\hat{u}_l(t) = \hat{\theta}_l e^{ilt}. \tag{1.3.33}$$

De este modo se observa que la regularidad propia de las soluciones de energía finita, i.e. el hecho de que

$$u \in C\left([0,\infty);\, H_0^1(0,\pi)\right) \cap C^1\left([0,\infty);\, L^2(0,\pi)\right), \tag{1.3.34}$$

equivale a que sus coeficientes de Fourier $\{\hat{u}_l(t)\}_{l\in\mathbf{Z}}$ pertenezcan al espacio

$$\{\hat{u}_l(t)\}_{l\in\mathbf{Z}} \in C\left([0,\infty);\,\mathcal{H}^1\right)\cap C^1\left([0,\infty);\,\ell^2\right). \tag{1.3.35}$$

Asimismo, las soluciones del problema semi-discreto, en vista de (1.3.24), pueden escribirse como

$$\vec{u}(t)=\sum_{l=-M}^{M}\hat{u}_{l,h}(t)\vec{W}_l(h) \tag{1.3.36}$$

donde

$$\hat{u}_{l,h}(t)=\hat{\theta}^l(h)e^{i\mu_l(h)t},\ l=-M,\dots,M. \tag{1.3.37}$$

Si extendemos por cero estos coeficientes de Fourier para los índices $|l|>M$, nos encontramos nuevamente con que, para las soluciones del problema semi-discreto, se tiene

$$\{\hat{u}_{l,h}(t)\}_{l=-M}^{M}\in C\left([0,\infty);\,\mathcal{H}^1\right)\cap C^1\left([0,\infty);\,\ell^2\right). \tag{1.3.38}$$

Es por tanto natural medir la convergencia de las soluciones del problema semi-discreto al continuo a través de la convergencia de sus coeficientes de Fourier en el espacio $C\left([0,\infty);\,\mathcal{H}^1\right)\cap C^1\left([0,\infty);\,\ell^2\right)$.

Con estas notaciones y con el objeto de establecer la convergencia de las soluciones del problema semi-discreto al continuo realizamos la siguiente elección de los datos iniciales de (1.3.19):

$$\vec{\varphi}=\sum_{l=-M}^{M}\hat{\varphi}_l\vec{W}_l(h);\ \vec{\psi}=\sum_{l=-M}^{M}\hat{\psi}_l\vec{W}_l(h), \tag{1.3.39}$$

donde $\{\hat{\varphi}_l\}_{l\in\mathbf{Z}}$ y $\left\{\hat{\psi}_l\right\}_{l\in\mathbf{Z}}$ son los coeficientes de Fourier del dato inicial continuo $(\varphi(x),\,\psi(x))\in H_0^1(0,\pi)\times L^2(0,\pi)$. Es decir, elegimos los datos iniciales del problema semi-discreto de modo que coincidan con los del problema continuo para los índices $-M\le k\le M$. De este modo, la solución de (1.3.19) admite el desarrollo (1.3.24) (o (1.3.36)-(1.3.37)) con los coeficientes de Fourier

$$\hat{\theta}_l(h)=\frac{\mu_l(h)\hat{\varphi}_l-i\hat{\psi}_l}{2\mu_l(h)},\ l=-M,\dots,M. \tag{1.3.40}$$

Con esta elección de los datos iniciales en el sistema semi-discreto tenemos el siguiente resultado de convergencia.

Theorem 1.3.1 *Supongamos que $(\varphi,\psi)\in H_0^1(0,\pi)\times L^2(0,\pi)$ y que elegimos los datos iniciales del problema semi-discreto como en* (1.3.39).

Entonces, las soluciones del problema semi-discreto convergen a las del continuo en el sentido que

$$\begin{cases} \{\hat{u}_{\ell,h}(t)\}_{\ell=-M}^{M} \to \{\hat{u}_{\ell}(t)\}_{\ell=-\infty}^{\infty} \ \text{en } C\left([0,T];\mathcal{H}^1\right) \cap C^1\left([0,T];\ell^2\right) \\ \text{cuando } h \to 0. \end{cases} \tag{1.3.41}$$

para todo $0 < T < \infty$.

Observación 1.3.1

- Hemos enunciado la convergencia de las soluciones en términos de sus coeficientes de Fourier. En efecto, en (1.3.41) se establece que los coeficientes de Fourier de las soluciones semi-discretas, junto con sus derivadas temporales, convergen, cuando $h \to 0$, uniformemente en tiempo, a los de la solución continua en el espacio $\mathcal{H}^1 \times \ell^2$. Como hemos dicho antes, este es el espacio natural al que pertenecen los coeficientes de Fourier de las soluciones de energía finita de la ecuación de ondas, por lo que el resultado es óptimo.

- Este resultado puede ser también interpretado a través de la convergencia de las soluciones en el espacio de la energía. Volveremos sobre este punto más adelante.

- Tal y como habíamos anunciado en (1.3.41) hemos abusado de la notación puesto que hemos interpretado que $\{\hat{u}_{\ell,h}\}_{\ell=-M}^{M}$, que es un vector finito, es una sucesión. Como mencionábamos, ha de sobreentenderse que extendemos el valor de $\hat{u}_{\ell,h}$ por cero para todos los índices $|\ell| > M$.

■

Demostración del Teorema 3.1.

Probamos exclusivamente la convergencia en $C\left([0,T];\mathcal{H}^1\right)$. La convergencia en $C^1\left([0,T];\ell^2\right)$ puede ser probada de manera análoga.

Sea

$$\hat{v}_{\ell,h}(t) = \hat{u}_{\ell}(t) - \hat{u}_{\ell,h}(t) = \hat{\theta}_{\ell} e^{i\ell t} - \hat{\theta}_{\ell}(h) e^{i\mu_{\ell}(h)t}, \tag{1.3.42}$$

donde los coeficientes $\hat{\theta}_{\ell}$ y $\hat{\theta}_{\ell}(h)$ vienen dados por (1.3.13) y (1.3.40) respectivamente.

Tenemos

$$\left| \{\hat{v}_{\ell,h}(t)\} \right|_{\mathcal{H}^1}^2 = \sum_{\ell=-\infty}^{\infty} |\hat{v}_{\ell,h}(t)|^2 \ell^2 = \sum_{\ell=-\infty}^{\infty} |\hat{u}_{\ell}(t) - \hat{u}_{\ell,h}(t)|^2 \ell^2 \tag{1.3.43}$$

$$= \sum_{\ell=-M}^{M} \left| \hat{\theta}_{\ell} e^{i\ell t} - \hat{\theta}_{\ell}(h) e^{i\mu_{\ell}(h)t} \right|^2 \ell^2 + \sum_{|\ell|>M} \left| \hat{\theta}_{\ell} \right|^2 \ell^2 = I_1(h) + I_2(h)$$

El término $I_2(h)$ es fácil de estimar. En efecto, en virtud de (1.3.17), y en la medida en que los coeficientes que intervienen en la serie $I_2(h)$ no dependen de h es fácil ver que, dado $\varepsilon > 0$ arbitrario, existe $M > 0$ de modo que

$$|I_2(h)| \leq \varepsilon, \quad \forall h > 0.$$

Una vez fijado el valor de M es fácil ver que $I_1(h) \to 0$. De hecho, esta convergencia es uniforme en $t \in [0, T]$. En efecto, como, una vez fijado el valor de M, $I_1(h)$ es una suma finita, basta comprobar que, para cada ℓ fijo,

$$\hat{\theta}_\ell(h) e^{i\mu_\ell(h)t} \to \hat{\theta}_\ell e^{i\ell t},\ h \to 0, \quad \text{uniformemente en } t \in [0, T],$$

lo cual es evidentemente cierto puesto que

$$\mu_\ell(h) \to \ell,\ h \to 0$$

y de que, en vista de la elección de los datos iniciales para el sistema discreto realizado en el enunciado del Teorema 1.3.1,

$$\hat{\theta}_\ell(h) \to \hat{\theta}_\ell,\ h \to 0.$$

■

Tal y como hemos mencionado anteriormente, este Teorema no proporciona un resultado muy intuitivo puesto que garantiza la convergencia de los coeficientes de Fourier pero no de las soluciones en el espacio físico. En este sentido, sería natural analizar la convergencia de la solución discreta $\vec{u}_h$ hacia los valores de la solución continua $u = u(x, t)$ sobre los puntos del mallado (que denotamos mediante $\vec{\underline{u}}$) tal como hicimos en el Teorema 1.2.1 en el caso de la ecuación del calor.

En este caso, en vista de la ausencia de efectos disipativos necesitamos hipótesis adicionales de regularidad sobre los datos iniciales de la ecuación de ondas:

Theorem 1.3.2 *Supongamos que* $(\varphi, \psi) \in \left[H^2 \cap H_0^1(0, \pi)\right] \times H_0^1(0, \pi)$ *y que elegimos los datos iniciales del problema semi-discreto como en* (1.3.39).

Entonces

$$\left|\vec{u}_h(t) - \vec{\underline{u}}(t)\right|_{1,h} + \left|\vec{u}_h'(t) - \vec{\underline{u}}'(t)\right|_h \to 0 \tag{1.3.44}$$

cuando $h \to 0$ *uniformemente en* $0 \leq t \leq T$ *para todo* $0 < T < \infty$.

Observación 1.3.2

- El espacio $H^2 \cap H_0^1(0,\pi)$ es el constituido por las funciones $\varphi \in H_0^1(0,\pi)$ tales que $\varphi_{xx} \in L^2(0,\pi)$. Dotado de la norma[10]

$$\| \varphi \|_{H^2 \cap H_0^1(0,\pi)} = \left[\int_0^\pi |\varphi_{xx}|^2 \, dx \right]^{1/2}$$

constituye un espacio de Hilbert.

- El espacio $\left[H^2 \cap H_0^1(0,\pi)\right] \times H_0^1(0,\pi)$ es un marco funcional natural para resolver la ecuación de ondas. En efecto, si $(\varphi,\psi) \in \left[H^2 \cap H_0^1(0,\pi)\right] \times H_0^1(0,\pi)$, la ecuación de ondas (6.3.1) admite una única solución

$$u \in C\left([0,\infty);\, H^2 \cap H_0^1(0,\pi)\right) \cap C^1\left([0,\infty);\, H_0^1(0,\pi)\right). \tag{1.3.45}$$

Además la energía

$$E_+(t) = \frac{1}{2} \int_0^\pi \left[|u_{xx}(x,t)|^2 + |u_{tx}(x,t)|^2 \right] dx \tag{1.3.46}$$

que es equivalente al cuadrado de la norma en este espacio se conserva en el tiempo (para comprobarlo basta con multiplicar la ecuación de ondas por $-u_{xxt}$ e integrar por partes en x).

- Cuando $\varphi \in H^2 \cap H_0^1(0,\pi)$ sus coeficientes de Fourier satisfacen

$$\sum_{\ell=1}^{\infty} |\hat{\varphi}_\ell|^2 \, \ell^4 < \infty. \tag{1.3.47}$$

Este hecho jugará un papel decisivo en la prueba del Teorema.

■

Demostración del Teorema 1.3.2.

Analizamos el primer término de (1.3.44) puesto que el segundo puede ser tratado del mismo modo.

Procediendo como en la demostración del Teorema 2.1 y distinguiendo bajas y altas frecuencias descomponemos $\vec{u}_h - \underline{\vec{u}}$ del siguiente modo:

$$\begin{aligned} \vec{u}_h(t) - \underline{\vec{u}}(t) &= \sum_{\ell=-M_0}^{M_0} \left[\hat{\theta}_\ell(h) e^{i\mu_\ell(h)t} - \hat{\theta}_\ell e^{i\ell t} \right] \vec{W}_\ell(h) \\ &\quad + \sum_{M_0+1 \le |\ell| \le M} \hat{\theta}_\ell(h) e^{i\mu_\ell(h)t} \vec{W}_\ell(h) - \sum_{|\ell|>M_0} \hat{\theta}_\ell e^{i\ell t} \vec{W}_\ell = I_1 + I_2 + I_3. \end{aligned} \tag{1.3.48}$$

Analizamos ahora la norma $\| I_k \|_{1,h}^2$, para cada $k = 1, 2, 3$.

[10] Obsérvese que en el subespacio de H^2 considerado esta semi-norma es en realidad una norma, equivalente a la norma canónica de H^2.

- *Término I_1.*

  Tomando normas $\| \cdot \|_{1,h}$ en I_1 y utilizando las propiedades de los autovectores $\vec{W}_\ell(h)$ enunciados en el Lema 2.1 y la desigualdad (1.2.32) para los autovalores discretos se tiene:

$$\begin{aligned} \| I_1 \|_{1,h}^2 &= \sum_{\ell=-M_0}^{M_0} \lambda_\ell(h) \left| \hat{\theta}_\ell(h) e^{i\mu_\ell(h)t} - \hat{\theta}_\ell e^{i\ell t} \right|^2 \\ &\leq \sum_{\ell=-M_0}^{M_0} \ell^2 \left| \hat{\theta}_\ell(h) e^{i\mu_\ell(h)t} - \hat{\theta}_\ell e^{i\ell t} \right|^2 . \end{aligned}$$

- *Término I_2.*

  Tenemos, por el mismo argumento,

$$\| I_2 \|_{1,h}^2 \leq \sum_{M_0+1<|\ell|\leq M} \lambda_\ell(h) |\hat{\theta}_\ell(h)|^2, \tag{1.3.49}$$

  en virtud de la elección (1.3.40) de los coeficientes de Fourier de los datos iniciales del problema semi-discreto deducimos por tanto que

$$\| I_2 \|_{1,h}^2 \leq C \sum_{M_0+1<|\ell|\leq M} \left[\ell^2 |\hat{\varphi}_\ell|^2 + |\hat{\psi}_\ell|^2 \right]. \tag{1.3.50}$$

  La convergencia de esta serie está garantizada por la hipótesis de que los datos iniciales considerados son de energía finita, i.e. $(\varphi, \psi) \in H_0^1(0,\pi) \times L^2(0,\pi)$.

- *Término I_3.*

  Este término ha de ser analizado con algo más de cuidado. En efecto, $\vec{W}_\ell$ representa en este caso la restricción a los puntos del mallado de las autofunciones continuas que no se obtienen en el espectro de la matriz. Por lo tanto se pierden las propiedades de ortogonalidad enunciadas en el Lema 2.1. Por tanto, en este caso aplicamos nuevamente la desigualdad triangular y obtenemos

$$\| I_3 \|_{1,h} \leq \sum_{|\ell|>M_0} \left| \hat{\theta}_\ell \right| \| \vec{W}_\ell \|_{1,h} \leq \sum_{|\ell|>M_0} \left| \hat{\theta}_\ell \right| \| W_{\ell,x} \|_{L^2(0,\pi)} = \sum_{|\ell|>M_0} |\ell| \left| \hat{\theta}_\ell \right|. \tag{1.3.51}$$

  En esta última desigualdad hemos utilizado la desigualdad elemental

$$h \sum_{j=0}^{M} \frac{|f(x_{j+1}) - f(x_j)|^2}{h^2} \leq \int_0^\pi f_x^2(x) dx \tag{1.3.52}$$

que es válida para cualquier función de $H_0^1(0,\pi)$ y cualquier $h>0$, y por otra parte que

$$\| W_{\ell,x} \|^2_{L^2(0,\pi)} = \int_0^\pi |W_{\ell,x}|^2 \, dx = \ell^2. \tag{1.3.53}$$

Con el objeto de poder estimar la serie que se obtiene en esta estimación es indispensable suponer que

$$\sum_\ell |\ell| \left|\hat{\theta}_\ell\right| < \infty, \tag{1.3.54}$$

lo cual queda plenamente garantizado si los datos iniciales (φ,ψ) pertenecen a $[H^2 \cap H_0^1(0,\pi)] \times H_0^1(0,\pi)$ puesto que, en virtud de (1.3.16), (1.3.17) y (1.3.47), se tiene

$$\sum_\ell |\ell| \left|\hat{\theta}_\ell\right| \leq \left(\sum_\ell \ell^2 \left|\hat{\theta}\right|^2\right)^{1/2} \left(\sum_\ell \ell^{-2}\right)^{1/2} < \infty.$$

De estas desigualdades es fácil concluir la demostración del Teorema 3.2. En efecto, de (1.3.50), (1.3.51) y (1.3.54) vemos que dado $\varepsilon>0$ arbitrario existe $M_0>0$ tal que

$$\| I_2 \|_{1,h} + \| I_3 \|_{1,h} \leq \varepsilon, \; \forall h>0.$$

Una vez fijado el valor de M_0 el hecho que $\| I_1 \|_{1,h} \to 0$ cuando $h \to 0$ resulta evidente de (1.3.49), puesto que se trata de una suma de un número finito de términos en los que cada uno tiende a cero puesto que $\hat{\theta}_\ell(h) \to \hat{\theta}_\ell$ y $\mu_\ell(h) \to \ell$ respectivamente.

Se observa asimismo que la convergencia es uniforme en intervalos compactos de tiempo $[0,T]$.

Esto concluye la demostración de Teorema 3.2.

■

Observación 1.3.3

- Un análisis cuidadoso de la demostración del Teorema 3.2 permite establecer una estimación sobre el orden de convergencia.
- El Teorema 3.2 sólo demuestra un posible resultado de convergencia entre los varios que podrían probarse mediante argumentos semejantes. Muchas otras variantes, similares a las descritas en la sección 2.2 en el contexto de la ecuación del calor, son posibles.

■

Los resultados de convergencia que hemos presentado en los dos Teoremas anteriores hacen referencia a los coeficientes de Fourier o a la restricción de las soluciones a los puntos del mallado. Sin embargo, tal y como ocurría en el caso de la ecuación del calor, estos resultados admiten también una interpretación global. Para ello es necesario introducir un operador de extensión que a una función discreta definida sobre el mallado asocie una función continua de la variable x. Para ello, tal y como se hizo en la sección 2.7 dedicada el estudio de la ecuación del calor mediante el método de elementos finitos, dada una solución discreta $\vec{u}_h$ de (1.3.18), podemos definir su extensión

$$[E\vec{u}_h](x,t) = \sum_{j=1}^{N} u_j(t)\phi_j(x), \tag{1.3.55}$$

donde $\{\phi_j\}_{j=1,\ldots,N}$ son precisamente las funciones de base del método de elementos finitos que, tal y como indicamos en la sección 2.7, son continuos, lineales a trozos y satisfacen $\phi_j(x_\ell) = \delta_{j\ell}$.

Es fácil comprobar que, dada una solución discreta $\vec{u}_h$ de (1.3.18), $E\vec{u}_h$ es una función perteneciente al espacio $C\left([0,\infty);\, H_0^1(0,\pi)\right) \cap C^1\left([0,\infty);\, L^2(0,\pi)\right)$. Cabe por tanto plantearse la convergencia de $E\vec{u}_h$ hacia la solución $u = u(x,t)$ del problema continuo (1.3.9) en el espacio de la energía.

El siguiente resultado responde a esta cuestión:

Theorem 1.3.3 *Bajo las hipótesis el Teorema 3.2 se tiene*

$$E\vec{u}_h \to u \text{ en } C\left([0,T]; H_0^1(0,\pi)\right) \cap C^1\left([0,T];\, L^2(0,\pi)\right) \tag{1.3.56}$$

para todo intervalo compacto $[0.T]$

Demostración

Procedemos en dos etapas.

Etapa 1. Convergencia en $C\left([0,T];\, H_0^1(0,\pi)\right)$.

Observamos que

$$\begin{aligned}
\left|E\vec{u}_h(t)-u(t)\right|^2_{H_0^1(0,\pi)} &= \int_0^\pi \left|\partial_x\left(E\vec{u}_h\right)(x,t)-u_x(x,t)\right|^2 dx \qquad (1.3.57)\\
&= \sum_{j=0}^{N}\int_{x_j}^{x_{j+1}}\left|\frac{u_{j+1}(t)-u_j(t)}{h}-u_x(x,t)\right|^2 dx\\
&\leq 2\sum_{j=0}^{N}\int_{x_j}^{x_{j+1}}\left|\frac{u_{j+1}(t)-u_j(t)}{h}-\frac{u(x_{j+1},t)-u(x_j,t)}{h}\right|^2 dx\\
&+ 2\sum_{j=0}^{N}\int_{x_j}^{x_{j+1}}\left|u_x(x,t)-\frac{u(x_{j+1},t)-u(x_j,t)}{h}\right|^2 dx\\
&= 2h\sum_{j=0}^{N}\left|\frac{[u_{j+1}(t)-u(x_{j+1},t)]-[u_j(t)-u(x_j,t)]}{h}\right|^2\\
&+ 2\sum_{j=0}^{N}\int_{x_j}^{x_{j+1}}\left|u_x(x,t)-\frac{1}{h}\int_{x_j}^{x_{j+1}}u_x(s,t)ds\right|^2 dx = I_1+I_2.
\end{aligned}$$

Por otra parte, $I_1 = 2\left|\vec{u}_h(t)-\underline{\vec{u}}(t)\right|^2_{1,h}$ que, tal y como vimos en el Teorema 3.2, converge a cero uniformemente en $0\leq t\leq T$ cuando $h\to 0$.

Además,

$$\begin{aligned}
I_2 &= 2\sum_{j=0}^{N}\int_{x_j}^{x_{j+1}}\left|\frac{1}{h}\int_{x_j}^{x_{j+1}}\left(u_x(x,t)-u_x(s,t)\right)ds\right|^2 dx\\
&\leq \frac{2}{h^2}\sum_{j=0}^{N}\int_{x_j}^{x_{j+1}}\left|\int_{x_j}^{x_{j+1}}\left(u_x(x,t)-u_x(x,t)\right)ds\right|^2 dx\\
&= \frac{2}{h^2}\sum_{j=0}^{N}\int_{x_j}^{x_{j+1}}\left|\int_{x_j}^{x_{j+1}}\int_{s}^{x}u_{xx}(\sigma,t)d\sigma dx\right|^2 dx\\
&\leq \frac{2}{h^2}\sum_{j=0}^{N}\int_{x_j}^{x_{j+1}}\left|\int_{x_j}^{x_{j+1}}\int_{x_j}^{x_{j+1}}\left|u_{xx}(\sigma,t)\right|d\sigma ds\right|^2 dx\\
&= 2h\sum_{j=0}^{N}\left|\int_{x_j}^{x_{j+1}}\left|u_{xx}(\sigma,t)\right|d\sigma\right|^2\\
&\leq 2h^2\sum_{j=0}^{N}\int_{x_j}^{x_{j+1}}\left|u_{xx}(\sigma,t)\right|^2 d\sigma = 2h^2\int_0^\pi\left|u_{xx}(x,t)\right|^2 dx\\
&= 2h^2\parallel u(t)\parallel^2_{H^2\cap H_0'(0,\pi)}.
\end{aligned}$$

Este último término tiende a cero con un orden $O(h^2)$ cuando $h\to 0$ puesto que,

tal y como se indicó en la Observación 3.2, bajo la hipótesis de regularidad de los datos iniciales del Teorema 3.2, la solución u del problema continuo (1.3.9) pertenece a la clase $C\left([0,\infty);\, H^2 \cap H_0^1(0,\pi)\right)$. En realidad la convergencia es uniforme para todo $t \geq 0$.

Etapa 2. Convergencia en $C^1\left([0,T];\, L^2(0,\pi)\right)$.

En este caso

$$\begin{aligned}
&\left|E\vec{u}_h'(t) - u_t(t)\right|^2_{L^2(0,\pi)} = \int_0^\pi \left|\sum_{k=1}^N u_k'(t)\phi_k(x) - u_t(x,t)\right|^2 dx \\
\leq\ & 2\int_0^\pi \left|\sum_{k=1}^N \left(u_k'(t) - u_t(x_k,t)\right)\phi_k(x)\right|^2 dx \\
&+ 2\int_0^\pi \left|\sum_{k=1}^N u_t(x_k,t)\phi_k(x) - u_t(x,t)\right|^2 dx = I_1 + I_2.
\end{aligned}$$

El primer término I_1, en virtud de la expresión de la matriz de masa que aparece en el método de elementos finitos, admite la expresión

$$\begin{aligned}
I_1 &= 2\sum_{j,k=1}^N \left(u_k'(t) - u_t(x_k,t)\right)\left(u_j'(t) - u_t(x_j,t)\right)\int_0^\pi \phi_k(x)\phi_j(x)dx \\
&= 2h\sum_{k=1}^N \left[\frac{2}{3}\left|u_k'(t) - u_t(x_k,t)\right|^2 + \frac{1}{3}\left(u_k'(t) - u_t(x_k,t)\right)\left(u_{k+1}'(t) - u_t(x_{k+1},t)\right)\right] \\
&\leq C\left|\vec{u}_h'(t) - \underline{\vec{u}}'(t)\right|^2_h
\end{aligned}$$

que tiende a cero uniformemente en $[0,T]$ cuando $h \to 0$, por el Teorema 3.2.

Por otra parte,

$$\begin{aligned}
I_2 &= 2\sum_{j=0}^{N}\int_{x_j}^{x_{j+1}}\left|u_t(x,t)-\frac{\left(u_t\left(x_{j+1},t\right)-u_t\left(x_j,t\right)\right)}{h}(x-x_j)-u_t(x_j,t)\right|^2 dx\\
&= 2\sum_{j=0}^{N}\int_{x_j}^{x_{j+1}}\left|\int_{x_j}^{x}\left[u_{tx}(s,t)-\frac{\left(u_t\left(x_{j+1},t\right)-u_t\left(x_j,t\right)\right)}{h}\right]ds\right|^2 dx\\
&= 2\sum_{j=0}^{N}\int_{x_j}^{x_{j+1}}\left|\int_{x_j}^{x}\left[u_{tx}(s,t)-\frac{1}{h}\int_{x_j}^{x_{j+1}}u_{tx}(\sigma,t)d\sigma\right]ds\right|^2 dx\\
&= 2\sum_{j=0}^{N}\int_{x_j}^{x_{j+1}}\left|\int_{x_j}^{x}\left[\frac{1}{h}\int_{x_j}^{x_{j+1}}\left[u_{tx}(s,t)-u_{tx}(\sigma,t)\right]d\sigma ds\right]\right|^2 dx\\
&\leq 2\sum_{j=0}^{N}\int_{x_j}^{x_{j+1}}\left|2\int_{x_j}^{x_{j+1}}\left|u_{tx}(s,t)\right|ds\right|^2 dx\\
&\leq 8h\sum_{j=0}^{N}\int_{x_j}^{x_{j+1}}\int_{x_j}^{x_{j+1}}\left|u_{tx}(s,t)\right|^2 dsdx\\
&= 8h^2\int_0^{\pi}\left|u_{tx}(s,t)\right|^2 ds.
\end{aligned}$$

Este término tiende a cero con un orden $O(h^2)$ puesto que $u_t \in C\left([0,T];\, H_0^1(0,\pi)\right)$, tal y como señalamos en la Observación 3.2.

■

1.3.3. Semi-discretización espacial: El método de la energía

La prueba de la convergencia de las soluciones del problema discreto a las del continuo está basada en el hecho que la energía (1.3.12) (resp. (1.3.27)) se conserva para las soluciones del problema continuo (1.3.9) (rep. para las del problema discreto (1.3.18)).

Procedemos como en la sección 1.2.3 en el caso de la ecuación del calor y por tanto consideramos la solución $u = u(x,t)$ de la ecuación de ondas continua (1.3.9) como una solución aproximada de la ecuación discreta (1.3.18). Tenemos

entonces

$$\begin{cases} \underline{u}_j''(t) + \frac{\left[2\ \underline{u}_j(t) - \underline{u}_{j+1}(t) - \underline{u}_{j-1}(t)\right]}{h^2} & = u_{xx}(x_{j,t}) + \frac{\left[2\ \underline{u}_j(t) - \underline{u}_{j+1}(t) - \underline{u}_{j-1}(t)\right]}{h^2} \\ & = \varepsilon_j(t),\ j = 1, \dots, M,\ t > 0 \\ \underline{u}_0 = \ \underline{u}_{M+1} = 0,\ t > 0 \\ \underline{u}_j(0) = \varphi_j,\ \underline{u}_j'(0) = \psi_j,\ j = 1, \dots, M \end{cases} \tag{1.3.58}$$

con $\varphi_j = \varphi(x_j)$ y $\psi_j = \psi(x_j)$ si, por ejemplo, los datos (φ, ψ) son continuos, lo cual está garantizado en las hipótesis de los Teoremas 3.2 y 3.3 cuando $(\varphi, \psi) \in \left[H^2 \cap H_0^1(0,\pi)\right] \times H_0^1(0,\pi)$.

En el segundo miembro de (1.3.58) aparece un *residuo* o *error de truncación* $\vec{\varepsilon}(t) = (\varepsilon_j(t))_{j=1}, \dots, M$.

Consideramos ahora la diferencia de las soluciones continua y discreta

$$\vec{v}_h(t) = \ \vec{\underline{u}}(t) - \vec{u}_h(t), \tag{1.3.59}$$

que satisface

$$\begin{cases} v_j''(t) + \frac{[2v_j(t) - v_{j+1}(t) - v_{j-1}(t)]}{h^2} = \varepsilon_j(t),\ j = 1, \dots, M,\ t \geq 0 \\ v_0 = v_{M+1} = 0,\ t \geq 0 \\ v_j(0) = v_j'(0) = 0,\ j = 1, \dots, M. \end{cases} \tag{1.3.60}$$

Multiplicando en (1.3.60) por $v_j'(t)$ y sumando en j, como es propio para la obtención de la identidad de energía para las soluciones del sistema semi-discreto, obtenemos que

$$\frac{d}{dt}\left[\frac{h}{2}\sum_{j=1}^{N}\left|v_j'(t)\right|^2 + \frac{h}{2}\sum_{j=0}^{N}\left|\frac{v_{j+1}(t) - v_j(t)}{h}\right|^2\right] = h\sum_{j=1}^{N}\varepsilon_j(t)v_j'(t).$$

y por tanto

$$\left|\frac{d}{dt}\left[\frac{h}{2}\sum_{j=1}^{N}\left[\left|v_j'(t)\right|^2 + \left|\frac{v_{j+1}(t) - v_j(t)}{h}\right|^2\right]\right]\right| \leq \frac{h}{2}\sum_{j=1}^{N}\left[\left|\varepsilon_j(t)\right|^2 + \left|v_j'(t)\right|^2\right].$$

Aplicando el Lema de Gronwall deducimos que

$$\frac{h}{2}\sum_{j=1}^{N}\left[\left|v_j'(t)\right|^2 + \left|\frac{v_{j+1}(t) - v_j(t)}{h}\right|^2\right] \leq \frac{Te^T}{2}\max_{0\leq t\leq T}\left|\vec{\varepsilon}(t)\right|_h^2,\ \forall t \in [0,T].$$

Basta por tanto que estimemos el error de truncación $\vec{\varepsilon}(t)$. Tenemos

$$|\varepsilon_j(t)| \leq Ch^2 \parallel u(t) \parallel_{C^4([0,\pi])},\ \forall j = 1, \dots, M,\ \forall 0 < h < h_0,\ \forall t \geq 0.$$

De estas dos últimas desigualdades deducimos que

$$\frac{1}{2}\left[\| \vec{v}(t) \|_{1,h}^2 + \| \vec{v}'(t) \|_h^2\right] \leq Ch^4 \| u(t) \|_{L^\infty(0,T;C^4(0,\pi))}, \forall t \in [0,T].$$

Hemos probado el siguiente resultado:

Theorem 1.3.4 *Supongamos que los datos iniciales* (φ, ψ) *de la ecuación de ondas* (1.3.9) *son tales que la solución* $u = u(x,t)$ *satisface*

$$u \in C\left([0,T]; C^4([0,\pi])\right). \tag{1.3.61}$$

Entonces, para todo $0 < T < \infty$ *existe una constante* $C_T > 0$ *tal que*

$$\left|\vec{u}_h(t) - \underline{\vec{u}}(t)\right|_{1,h}^2 + \left|\vec{u}_h'(t) - \underline{\vec{u}}'(t)\right|_h^2 \leq C_T h^4 \tag{1.3.62}$$

para todo $0 \leq t \leq T$ *y todo* $h > 0$*, donde* $\underline{\vec{u}}$ *denota la restricción a los puntos del mallado de la solución de la ecuación de ondas* (1.3.9) *y* $\vec{u}_h$ *representa la solución del sistema semi-discreto (1.3.18).*

Observación 1.3.4

- La estimación de error (1.3.62) garantiza que la ecuación semi-discreta (1.3.18) proporciona una aproximación de orden 2 de la ecuación de ondas (1.3.9). Se trata de un resultado natural puesto que la única discretización realizada en el esquema (1.3.18) es la del laplaciano a través del esquema de tres puntos que, como hemos visto en capítulos anteriores, es un esquema de aproximación de orden 2.
- El resultado del Teorema 1.3.4 es sólo uno de los posibles que pueden ser obtenidos mediante el método de la energía. En particular, el método de la energía puede ser utilizado para probar la convergencia del método bajo hipótesis de regularidad de la solución mucho más débiles que (1.3.61).
- Tal y como se observó en la sección 2.3 la ventaja del método de la energía frente al de descomposición en series de Fourier es que puede ser aplicado en un contexto mucho más amplio: coeficientes variables dependientes del espacio y del tiempo, problemas no-lineales,....

■

Observación 1.3.5 *"Consistencia+Estabilidad=Convergencia".*

Aunque no se haya mencionado explícitamente, las demostraciones de convergencia de los resultados anteriores están inspiradas en el principio de Lax

discutido en la sección 2.4 según el cual la propiedad de convergencia equivale a la de consistencia más la de estabilidad. Comentemos este aspecto brevemente.

Las demostraciones de los Teoremas 1.3.1 y 1.3.2 mediante series de Fourier reflejan perfectamente este principio. La consistencia garantiza que los autovectores y autovalores del esquema discreto se aproximan a los del continuo. La estabilidad permite truncar las series de Fourier y reducir el problema a la convergencia de una suma finita (garantizada a su vez por la consistencia) manteniendo un control uniforme en tiempo de la cola de la serie.

La demostración del Teorema 1.3.3, basada en el método de la energía, también refleja esta filosofía. La consistencia del esquema garantiza que una solución regular de la ecuación continua es una solución aproximada del sistema semi-discreto con una cota (del orden de $O(h^2)$) sobre el error de truncamiento. La estabilidad del esquema se ve reflejada en la propiedad de conservación de la energía para el sistema semi-discreto.

1.3.4. Aproximaciones completamente discretas

Utilizamos las mismas notaciones que en la sección 2.5 dedicada al estudio de las aproximaciones completamente discretas de la ecuación del calor.

El esquema completamente discreto más habitual para la aproximación numérica de la ecuación de ondas es el conocido como "leap-frog":

$$\begin{cases} \frac{u_j^{k+1}-2u_j^k+u_j^{k-1}}{(\Delta t)^2} = \frac{u_{j+1}^k-2u_j^k+u_{j-1}^k}{(\Delta x)^2},\ j=1,\dots,M;\ k\geq 0 \\ u_0^k = u_{M+1}^k = 0,\ k\geq 0 \\ u_j^0 = \varphi_j,\ u_j^1 = \xi_j,\ j=1,\dots,M. \end{cases} \tag{1.3.63}$$

Obviamente, el esquema (1.3.63) se obtiene aplicando el clásico esquema de tres puntos centrado para aproximar tanto la segunda derivada temporal como espacial.

Se trata de un esquema de dos pasos por lo que es indispensable para su inicialización dar el valor de la solución discreta $\{u_j^k\}$ en los dos primeros niveles temporales $k=0,1$. Como u_j^0 es una aproximación de la solución de la ecuación de ondas $u=u(x,t)$ en el punto $(x,t)=(x_j,0)$ es natural tener como dato inicial en el nivel $k=0$ para el problema discreto

$$u_j^0 = \varphi_j = \varphi(x_j). \tag{1.3.64}$$

Por otra parte u_j^1 es una aproximación de $u=u(x,t)$ en el punto $(x,t)=(x_j,\Delta t)$. Por lo tanto, por el desarrollo de Taylor, es natural tomar como dato inicial en el esquema discreto para $h=1$

$$\xi_j = \varphi_j + \Delta t\psi_j = \varphi(x_j) + \Delta t\psi(x_j). \tag{1.3.65}$$

En efecto, si u es suficientemente regular se tiene

$$u(x_j, \Delta t) = u(x_j, 0) + \Delta t u_t(x_j, 0) + O\left((\Delta t)^2\right),$$

lo cual justifica la elección (1.3.65).

Obviamente, para que (1.3.65) sea posible es indispensable que tanto φ como ψ sean funciones continuas. Cuando se eligen datos iniciales (φ, ψ) de energía finita, i.e. $(\varphi, \psi) \in H_0^1(0, \pi) \times L^2(0, \pi)$, la continuidad de φ está garantizada pero no la de ψ. En ese caso, el valor de $\psi_j = \psi(x_j)$ puede ser sustituido, por ejemplo, por una media de ψ entorno al punto $x = x_j$.

El esquema (1.3.63) es claramente explícito y permite calcular fácilmente el valor de la solución discreta en el paso $k + 1$ a partir de los valores en los dos pasos anteriores $k - 1$ y k.

Por otra parte, el esquema es consistente de orden dos puesto que, como indicabamos anteriormente, nos hemos limitado a aplicar el esquema centrado de tres puntos a la hora de discretizar la derivada segunda en tiempo y en espacio.

El número de Courant en este caso viene dado por

$$\mu = \Delta t / \Delta x. \tag{1.3.66}$$

Con esta notación el esquema (1.3.63) puede ser escrito de la siguiente manera:

$$u_j^{k+1} = 2u_j^k - u_j^{k-1} + \mu^2 \left[u_{j+1}^k - 2u_j^k + u_{j-1}^k\right]. \tag{1.3.67}$$

Conviene señalar que el modo en que los pasos de espacio y tiempo intervienen en la definición del número de Courant μ en (1.3.66) es muy distinto a cómo lo hacen en el caso de la ecuación del calor ((1.2.95)). En este caso en el número de Courant se refleja el hecho de que en la ecuación de ondas las derivadas temporales y espaciales juegan un papel completamente simétrico, lo cual queda claramente de manifiesto en la propia expresión de la ecuación de ondas (dos derivadas temporales coinciden con dos derivadas espaciales) o de la energía de la misma.

Como decíamos anteriormente, el método es consistente de orden y, por consiguiente, si $u = u(x, t)$ es una solución suficientemente regular (de clase C^4) de la ecuación de ondas y denotamos mediante $\underline{u}_j^k$ sus restricciones a los puntos del mallado, al insertar esos valores en el esquema discreto obtenemos un error de truncatura del orden de $O\left((\Delta t)^2\left((\Delta x)^2 + (\Delta t)^2\right)\right)$. Es decir $\underline{u}_j^k$ es una solución aproximada de (1.3.67) que verifica

$$\underline{u}_j^{k+1} = 2\ \underline{u}_j^k - \ \underline{u}_j^{k-1} + \mu^2 \left[\ \underline{u}_{j+1}^k - 2\ \underline{u}_j^k + \ \underline{u}_{j-1}^k\right] + O\left((\Delta t)^2\left((\Delta x)^2 + (\Delta t)^2\right)\right). \tag{1.3.68}$$

Sin embargo, la consistencia del esquema no garantiza su convergencia, sino que es necesaria también una propiedad de estabilidad.

La estabilidad de los métodos multi-paso ha sido ya estudiada en el marco de los sistemas de ecuaciones diferenciales ordinarias. En aquél contexto veíamos que la estabilidad necesaria podía garantizarse a través de la *condición de la raíz*: El polinomio característico del método lineal multi-paso ha de tener todas las raíces de módulo menor o igual que uno y, en caso de tener alguna raíz de módulo unidad, ésta ha de ser simple.

Con el objeto de adaptar este concepto al marco del sistema discreto (1.3.68) conviene utilizar el desarrollo en serie de Fourier. Así, la solución discreta puede descomponerse como

$$\vec{u}^k = \sum_{\ell=1}^{M} \rho_\ell^k \vec{W}_\ell(\Delta x), \tag{1.3.69}$$

donde, como es habitual, $\vec{W}_\ell(\Delta x)$ denota el ℓ-ésimo autovector de la matriz $A_{\Delta x}$ de la discretización mediante el esquema de tres puntos del laplaciano ((1.2.26) con $h = \Delta x$) y ρ_ℓ^k el coeficiente de Fourier correspondiente en el paso temporal k.

Aplicando el esquema discreto en la expresión (1.3.69) obtenemos

$$\rho_\ell^{k+1} = 2\rho_\ell^k - \rho_\ell^{k-1} - \mu^2(\Delta x)^2\lambda_\ell(\Delta x)\rho_\ell^k. \tag{1.3.70}$$

Cada una de las expresiones (1.3.70) es un esquema de evolución discreto en tiempo en dos pasos para cada valor de $\ell = 1, \ldots, M$. Su polinomio característico es en este caso

$$P_\ell(\lambda) = \lambda^2 - \left[2 - \mu^2(\Delta x)^2\lambda_\ell(\Delta x)\right]\lambda + 1. \tag{1.3.71}$$

Sus raíces son

$$\lambda_\ell^{\pm} = \frac{2 - \mu^2(\Delta x)^2\lambda_\ell(\Delta x) \pm \sqrt{\left(2 - \mu^2(\Delta x)^2\lambda_\ell(\Delta x)\right)^2 - 4}}{2}. \tag{1.3.72}$$

En vista de esta expresión es fácil comprobar que la condición de estabilidad se satisface si y sólo si

$$\mu \leq 1. \tag{1.3.73}$$

Es lo que se denomina la condición de Courant-Friedrichs-Lewy.

En efecto, para comprobarlo conviene distinguir dos casos:

Caso 1: Autovalores complejos.

Esto ocurre cuando

$$\left(2 - \mu^2(\Delta x)\lambda_\ell(\Delta x)\right)^2 - 4 \leq 0$$

i.e.

$$\left|2-\mu^2(\Delta x)^2\lambda_\ell(\Delta x)\right|\leq 2. \tag{1.3.74}$$

Teniendo en cuenta que

$$c(\Delta x)^2\leq(\Delta x)^2\lambda_\ell(\Delta x)\leq 4 \tag{1.3.75}$$

para $c>0$ adecuado, es obvio que la condición (1.3.74) se verifica siempre y cuando $\mu\leq 1$. Cuando $\mu>1$, como la cota superior en (1.3.75) es óptima, i.e. a medida que $\Delta x\to 0$, $\lambda_M(\Delta x)\to 4$, se deduce la existencia de coeficientes ℓ para los cuales la condición (1.3.74) es violada. Estos serán analizados en el segundo caso.

Volviendo al caso $\mu\leq 1$, en el que (1.3.74) se cumple, observamos que

$$\left(\lambda_\ell^{\pm}\right)^2=\frac{1}{4}\left[\left|2-\mu^2(\Delta x)^2\lambda_\ell(\Delta x)\right|^2+4-\left|2-\mu^2(\Delta x)^2\lambda_\ell(\Delta x)\right|^2\right]=1.$$

Además, las raíces son simples siempre y cuando $\mu>0$, cosa que, evidentemente, siempre se cumple.

Caso 2: Autovalores reales:

Consideremos ahora el caso en que alguno de los autovalores es real. Esto ocurre cuando

$$\left(2-\mu^2(\Delta x)^2\lambda_\ell(\Delta x)\right)^2\geq 4$$

lo cual exige que

$$\mu^2(\Delta x)^2\lambda_\ell(\Delta x)\geq 4. \tag{1.3.76}$$

Tal y como veíamos en el caso 1, esto ocurre, por ejemplo, en cuanto $\mu>1$, para el último autovalor $\lambda_M(\Delta x)$, a condición que $h>0$ sea suficientemente pequeño.

En este caso los dos autovalores $\lambda_\ell^{\pm}$ son reales y el de mayor módulo es el que corresponde al signo negativo para el cual se tiene

$$\left|\lambda_\ell^-\right|=\frac{\mu^2(\Delta x)^2\lambda_\ell(\Delta x)-2+\sqrt{\left(2-\mu^2(\Delta x)^2\lambda_\ell(\Delta x)\right)^2-4}}{2}$$

que es, evidentemente, estrictamente mayor que 1 en virtud de (1.3.76).

De este análisis deducimos que el método lineal multipaso que el método de leap-frog genera en cada uno de los componentes de Fourier del problema discreto es estable si y sólo se verifica la condición $\mu\leq 1$.

Como consecuencia de este análisis deducimos que

Theorem 1.3.5 *El esquema "leap-frog"* (1.3.63) *es un esquema convergente de orden 2 para la ecuación de ondas si y sólo si* $\mu\leq 1$.

Volveremos más adelante sobre la demostración de este resultado.

Un aspecto muy importante de la condición de estabilidad (1.3.73) es el relacionado con los dominios de dependencia.

En la sección 1.3.1 vimos, por medio de la fórmula de d'Alembert, que la solución de la ecuación de ondas continua en toda la recta real depende en el punto (x,t) de los valores de los datos iniciales en el intervalo de dependencia $[x-t,\, x+t]$. Para el esquema discreto (1.3.67) es fácil comprobar que la solución discreta en el punto $(x,t) = (j\Delta x,\, k\Delta t)$ depende del valor de los datos iniciales en intervalo $[(j-k)\Delta x,\, (j+k)\Delta x)] = [x - k\Delta x,\, x + k\Delta x] = \left[x - \frac{k\Delta t}{\mu},\, x + \frac{k\Delta t}{\mu}\right] = \left[x - \frac{t}{\mu},\, x + \frac{t}{\mu}\right]$. Vemos por tanto que la condición de estabilidad $\mu \leq 1$ del esquema numérico garantiza simplemente que "*el dominio de dependencia en el esquema discreto contenga al dominio de dependencia de la ecuación de ondas continua.*" Esto es una condición perfectamente natural e indispensable para que el esquema numérico sea convergente. En efecto, en caso de que la condición no se cumpla, el esquema numérico ignora información esencial de los datos iniciales para la determinación de la solución del problema continuo y esto hace que la convergencia sea imposible.

Volvamos ahora sobre la demostración de Teorema 1.3.5. En primer lugar es fácil comprobar que cuando $\mu > 1$ el esquema diverge. En efecto, para verlo basta con utilizar que cuando $\mu > 1$, existen componentes de Fourier ℓ para los que el esquema multipaso correspondiente viola la condición de estabilidad. Esto permite construir fácilmente soluciones en variables separadas para el sistema discreto que, a medida que $h \to 0$, divergen.

La prueba de la convergencia del esquema cuando $\mu \leq 1$ puede realizarse de diversas maneras. La primera prueba es la basada en el desarrollo en serie de Fourier. No es difícil adaptar la prueba del Teorema 1.3.1. Basta com proceder del mismo modo y utilizar la propiedad de estabilidad del esquema para mantener controladas uniformemente en tiempo las colas de las series truncadas.

La segunda prueba de la convergencia está basada en el método de la energía. En este caso la energía de las soluciones en el nivel temporal k viene dada por

$$E^k = \frac{\Delta x}{2} \sum_{j=0}^{M} \left[\left[\frac{u_j^{k+1} - u_j^k}{\Delta t} \right]^2 + \left[\frac{u_{j+1}^{k+1} - u_j^{k+1}}{\Delta x} \right] \left[\frac{u_{j+1}^k - u_j^k}{\Delta x} \right] \right]. \qquad (1.3.77)$$

No es difícil comprobar que la energía E^k se conserva en tiempo para las soluciones del problema discreto (1.3.63), i.e.

$$E^{k+1} = E^k,\ \forall k \geq 0. \qquad (1.3.78)$$

Para ello basta con multiplicar en la ecuación discreta por $\frac{1}{2\Delta t}\left(u_j^{k+1} - u_j^{k-1}\right)$

y sumar con respecto al índice espacial j. Esta prueba es análoga a la de la conservación de la energía en la ecuación de ondas continua. En efecto, en la ecuación de ondas multiplicamos la ecuación por u_t e integramos con respecto a $x \in (0, \pi)$. El procedimiento presentado por el sistema discreto es la versión discreta del método continuo. Conviene sin embargo subrayar que, en vista de la estructura simétrica del esquema de "leap-frog" la discretización introducida de u_t es también centrada y simétrica.

Pero la conservación de la energía por si sola no garantiza la estabilidad del método. Para poder garantizar que se cumple esta propiedad es preciso comprobar que la energía definida una cantidad definida positiva, cosa que no es obvio en su definición. No se difícil comprobar que esto último es cierto si y sólo si $\mu \leq 1$.

1.3.5. El análisis de von Neumann

En la sección 1.2.6 veíamos como el análisis de von Neumann permitía analizar con facilidad la estabilidad de los esquemas numéricos de aproximación, sobre todo en ausencia de condiciones de contorno.

Consideremos pues el problema de Cauchy en toda la recta real para la ecuación de ondas (6.3.1) y sus aproximaciones semi-discreta y completamente discretas siguientes:

•*Aproximación semi-discreta:*

$$\begin{cases} u_j'' + \dfrac{[2u_j - u_{j+1} - u_{j-1}]}{h^2} = 0,\ j \in \mathbf{Z},\ \mathbf{t} \geq \mathbf{0} \\ u_j(0) = \varphi_j,\ u_j'(0) = \psi_j,\ j \in \mathbf{Z} \end{cases} \tag{1.3.79}$$

•*Aproximación completamente discreta:*

$$\begin{cases} \dfrac{u_j^{k+1} - 2u_j^k + u_j^{k-1}}{(\Delta t)^2} = \dfrac{u_{j+1}^k - 2u_j^k + u_{j-1}^k}{(\Delta x)^2},\ j \in \mathbf{Z},\ \mathbf{k} \geq \mathbf{0} \\ u_j^0 = \varphi_j,\ u_j^1 = \xi_j,\ j \in \mathbf{Z}. \end{cases} \tag{1.3.80}$$

No es difícil adaptar los desarrollos de la sección anterior para probar que:

- El método semi-discreto (1.3.79) converge y es de orden 2, cuando $\Delta x \to 0$.
- El método completamente discreto (1.3.80) es convergente si y sólo si $\mu = \Delta t/\Delta x \leq 1$

Comprobemos como, efectivamente, el análisis de von Neumann permite detectar las propiedades de estabilidad que estos resultados exigen.

Con las notacions de la sección 1.2.6, dada la solución de (1.3.79) o (1.3.80) definimos sus transformada de Fourier $\check{w}(\theta,t)$ o $\check{w}^k(\theta)$ respectivamente. Las ecuaciones que gobiernan su evolución son entonces:

• *Método semi-discreto:*

$$\check{u}''(\theta,t)+\frac{a(e^{i\theta})}{(\Delta x)^2}\check{u}(\theta,t)=0,\ t\geq 0,\ \theta\in[0,2\pi) \tag{1.3.81}$$

donde, $a(\cdot)$ está definida como en (1.2.140), i.e.

$$a(e^{i\theta})=2-e^{i\theta}-e^{-i\theta}. \tag{1.3.82}$$

En (1.3.81) y en lo que sigue $'$ denota la derivada con respecto al tiempo.

• *Método completamente discreto:*

$$\frac{\check{u}^{k+1}(\theta)-2\check{u}^k(\theta)+\check{u}^{k-1}(\theta)}{(\Delta t)^2}+\frac{a(e^{i\theta})}{(\Delta x)^2}\check{u}^k(\theta)=0,\ k\geq 0,\ \theta\in[0,2\pi). \tag{1.3.83}$$

En el caso de la ecuación semi-discreta, multiplicando por $\check{u}'(\theta,t)$ se obtiene que

$$\frac{d}{dt}\left[\frac{1}{2}\left|\check{u}'(\theta,t)\right|^2+\frac{a(e^{i\theta})}{2(\Delta x)^2}\left|\check{u}(\theta,t)\right|^2\right]=0.$$

Esto garantiza la estabilidad del método.

En el caso completamente discreto, para cada valor de θ obtenemos un esquema discreto de dos pasos semejante al que obteníamos en (1.3.70) al utilizar la separación de variables para estudiar la estabilidad del método en el intervalo finito con condiciones de contorno de Dirichlet. En este caso la propiedad de estabilidad se reduce a comprobar que las raíces del polinomio característico de (1.3.83), para cada valor de θ, son de módulo menor o igual a uno y, en caso de ser de módulo unidad, se trata de una raíz simple. No es difícil comprobar que esto ocurre si y sólo si $\mu\leq 1$.

1.3.6. El método de elementos finitos

No es difícil de adaptar el método de elementos finitos, tal y como lo desarrollamos en la sección 1.2.7 para la ecuación del calor, al caso de la ecuación de ondas.

Para hacerlo, conviene en primer lugar dar una formulación variacional de la ecuación de ondas (1.3.9).

En este caso, tal y como indicamos en la sección 1.3.1, el espacio de energía natural es

$$u\in C\left([0,\infty);\,H_0^1(0,\pi)\right)\cap C^1\left([0,\infty);\,L^2(0,\pi)\right) \tag{1.3.84}$$

y la formulación variacional correspondiente es

$$\frac{d^2}{dt^2}\int_0^\pi u(x,t)\phi(x)dx + \int_0^\pi u_x(x,t)\phi_x(x)dx = 0,\ \forall \phi \in H_0^1(0,\pi). \tag{1.3.85}$$

Conviene interpretar (1.3.85) con cautela puesto que la regularidad indicada en (1.3.84) no garantiza que el primer término de (1.3.85) tenga un sentido clásico puesto que de (1.3.84) no se deduce que la función $\int_0^\pi u(x,t)\phi(x)dx$ sea de clase C^2 con respecto al tiempo. Sin embargo lo es puesto que las soluciones de la ecuación de ondas con la propiedad de regularidad (1.3.84) verifican también que

$$u \in C^2\left([0,\infty);\ H^{-1}(0,\pi)\right), \tag{1.3.86}$$

donde $H^{-1}(0,\pi)$ es el espacio de Sobolev dual de $H_0^1(0,\pi)$. Esto permite interpretar (1.3.85) como una ecuación diferencial clásica que se verifica para cada tiempo $t \geq 0$ y cada función test $\phi \in H_0^1(0,\pi)$.

Obviamente (1.3.84) y (1.3.85) han de ser completadas con las condiciones iniciales

$$u(x,0) = \varphi(x),\ u_t(x,0) = \psi(x),\ 0 < x < \pi. \tag{1.3.87}$$

De esta formulación débil de la ecuación de ondas es fácil obtener las ecuaciones de la aproximación por elementos finitos.

Para ello introducimos el espacio V_h de funciones lineales a trozos y continuas de $H_0^1(0,\pi)$, de dimensión M, asociado al mallado de paso h.

La formulación variacional más natural es entonces:

Hallar $u_h \in C^2([0,\infty);\ V_h)$ *tal que*

$$\frac{d^2}{dt^2}\int_0^\pi u_h(x,t)\phi_j(x)dx + \int_0^\pi u_{h,x}(x,t)\phi_{j,x}(x)dx = 0,\ \forall t > 0,\ \forall j = 1,\ldots,M \tag{1.3.88}$$

junto con las condiciones iniciales.

$$u_h(x,0) = \varphi_h(x),\ u_h'(x,0) = \psi_h(x),\ 0 < x < \pi. \tag{1.3.89}$$

Aquí y en lo sucesivo, como en la sección 1.2.7, $\phi_j = \phi_j(x)$ denota la j-ésima función de base del espacio V_h, con centro en el punto $x_j = jh$ del mallado. Las funciones φ_h y ψ_h son por otra parte aproximaciones del dato inicial continuo (φ, ψ) en V_h.

Utilizando la notación vectorial habitual y las matrices de masa y rigidez introducidas en el marco de la ecuación del calor, es fácil reescribir el problema

como un sistema de M ecuaciones diferenciales lineales de orden 2, acopladas, con M incógnitas:

$$\begin{cases} \mathcal{M}_h \vec{u}''(t) + \mathcal{R}_h \vec{u}(t) = 0,\ t > 0 \\ \vec{u}(0) = \vec{\varphi}_h,\ \vec{u}'(0) = \vec{\psi}_h. \end{cases} \tag{1.3.90}$$

Los métodos desarrollados en las secciones anteriores, tanto los basados en las descomposiciones espectrales como en el método de la energía, permiten comprobar que se trata de un método convergente de orden 2.

Como es habitual en el método de elementos finitos, en una dimensión espacial, (1.3.90) proporciona un esquema discreto que también puede interpretarse como una variación del método semi-discreto en diferencias finitas (1.3.18). En efecto, (1.3.90) puede reescribirse como

$$h\left[\frac{2}{3}u_j''(t) + \frac{1}{6}u_{j+1}''(t) + \frac{1}{6}u_{j-1}''(t)\right] = \frac{1}{h}\left[2u_j(t) - u_{j+1}(t) - u_{j-1}(t)\right]. \tag{1.3.91}$$

Pero, la ventaja del método de elementos finitos reside precisamente no en este hecho sino en su versatilidad para adaptarse a situaciones más complejas donde el método de diferencias finitas es de difícil utilización (problemas no-lineales, geometrías complejas en varias variables espaciales, coeficientes variables, etc.).

La identidad de energía es también fácil de obtener. De (1.3.90) y teniendo en cuenta que tanto la matriz de masa domo de rigidez son simétricas y definidas positivas la energía correspondiente viene dada por

$$E_h(t) = \frac{1}{2}\langle \mathcal{M}_h \vec{u}'(t),\ \vec{u}'(t)\rangle + \frac{1}{2}\langle \mathcal{R}_h \vec{u}_h(t),\ \vec{u}_h(t)\rangle, \tag{1.3.92}$$

donde $< \cdot, \cdot >$ denota le producto escalar euclideo.

La energía puede también escribirse componente a componente. Obtendríamos entonces:

$$E_h(t) = \frac{h^2}{2}\sum_{j=0}^{M}\left[\frac{2}{3}\left|u_j'(t)\right|^2 + \frac{1}{3}u_{j+1}'(t)u_j'(t) + \left|\frac{u_{j+1}(t) - u_j(t)}{h}\right|^2\right], \tag{1.3.93}$$

cuya conservación es también fácil deducir de la escritura del sistema componente a componente como en (1.3.91). En efecto, multiplicando en (1.3.91) por $u_j'(t)$ y sumando en j no es difícil verificar que la energía (1.3.93) se conserva.

En la expresión (1.3.93) de la energía y con el objeto de comprobar su carácter definido-positivo es conveniente observar que los dos primeros términos del sumatorio pueden agruparse dando lugar al cuadrado perfecto

$$\frac{1}{3}\left|u_{j+1}'(t) + u_j'(t)\right|^2.$$

Una vez que se ha obtenido el esquema de aproximación semi-discreto (1.3.90) y probado su convergencia, no es difícil introducir esquemas completamente discretos basados en el método de elementos finitos. Bastaría por ejemplo utilizar diferencias finitas centradas en la discretización de la segunda derivada temporal de (1.3.90). Pero el análisis de la convergencia de este método y, en particular, de la condición de estabilidad en función del número de Courant queda fuera del contenido de este texto y constituye un excelente ejercicio para que el lector evalúe su dominio de las materias abordadas en estas notas.

Capítulo 2

Movimiento armónico en una dimensión

El modelo más simple de vibraciones es el correspondiente al de una masa puntual desplazándose a lo largo de una linea recta con una aceleración orientada hacia un punto fijo y proporcional a la distancia a ese punto. Este es precisamente el movimiento asociado a un simple sistema masa-muelle, en el que el muelle es el responsable de la aceleración de la masa sujeta al mismo.

El movimiento descrito por la masa es lo que se denomina *movimiento armónico simple*.

Las ecuaciones que gobiernan este movimiento son

$$mx'' = -kx \tag{2.0.1}$$

o,

$$mx'' + kx = 0. \tag{2.0.2}$$

En estas ecuaciones $x = x(t)$ representa la distancia de la masa al punto fijo, m es la masa de la partícula y k es la constante de rigidez del muelle.

En (2.0.1) y en todo lo que sigue x' denota la derivada de x con respecto al tiempo. Ocasionalmente utilizaremos también otras notaciones $x' = dx/dt$.

Introduciendo la constante

$$\omega_0 = \sqrt{k/m} \tag{2.0.3}$$

el sistema (2.0.2) puede ser reescrito como

$$x'' + \omega_0^2 x = 0, \tag{2.0.4}$$

cuya solución general es

$$x(t) = A\cos(\omega_0 t + \phi). \tag{2.0.5}$$

en esta expresión en la que A es la amplitud de oscilación, ω_0 su frecuencia y ϕ la fase inicial del movimiento, se observa que el movimiento descrito por la masa es puramente oscilante.

Habida cuenta que se trata de una ecuación de orden dos en tiempo, las genuinas variables del sistema no son solamente la posición $x = x(t)$ de la masa sino también su velocidad

$$v = x' = -\omega_0 A\sin(\omega_0 t + \phi). \tag{2.0.6}$$

Obviamente, la trayectoria $t \to (x, x')$ describe una elipse de ecuación

$$\mid x' \mid^2 + \omega_0^2 x^2 = cte., \tag{2.0.7}$$

en el plano de fases.

El hecho de que la trayectoria quede atrapada en la elipse (2.0.7) puede obtenerse fácilmente a través de un argumento de conservación de energía. En efecto, multiplicando en (2.0.4) por x' deducimos que

$$\left(x'' + \omega_0^2 x\right) x' = \frac{d}{dt}\left[\frac{1}{2} \mid x' \mid^2 + \frac{\omega_0^2}{2} \mid x \mid^2\right] = 0. \tag{2.0.8}$$

Esta identidad confirma que la energía total de la vibración

$$e(t) = \frac{1}{2} \mid x' \mid^2 + \frac{\omega_0^2}{2} \mid x \mid^2 \tag{2.0.9}$$

se conserva en tiempo y permite determinar la elipse (2.0.7) en la que la trayectoria permanece.

Es evidente que dos oscilaciones armónicas con la misma frecuencia ω_0 que se superponen generan una nueva oscilación armónica de la misma frecuencia. Por otra parte es fácil calcular la amplitud y fase de la nueva oscilación a partir de las dos originales. Pero esto deja de ser cierto cuando las frecuencias no son las mismas, dando lugar a un fenómeno que, como veremos más adelante, jugará un papel importante en el análisis numérico de las ondas.

Con el objeto de analizar este nuevo fenómeno de superposición conviene reescribir las soluciones en la forma de exponenciales complejas

$$x_1 = A_1 e^{i(\omega_1 t + \phi_1)};\ x_2 = A_2 e^{i(\omega_2 t + \phi_2)}. \tag{2.0.10}$$

Cuando el cociente de las dos frecuencias ω_1 y ω_2 es un número racional, la superposición de estos dos movimientos

$$x = x_1 + x_2 = A_1 e^{i(\omega_1 t + \phi_1)} + A_2 e^{i(\omega_2 t + \phi_2)}$$

da lugar a un movimiento periódico de frecuencia igual al máximo común divisor de ω_1 y ω_2. Cuando el ratio ω_1/ω_2 es irracional la superposición de x_1 y x_2 no tiene ninguna propiedad de periodicidad temporal.

El caso en que ambas frecuencias sean muy próximas es particularmente interesante. Supongamos que

$$\omega_2 = \omega_1 + \Delta\omega. \tag{2.0.11}$$

Entonces

$$\begin{aligned} x = x_1 + x_2 &= A_1 e^{i(\omega_1 t + \phi_1)} + A_2 e^{i(\omega_2 t + \phi_2)} \\ &= \left[A_1 e^{i\phi_1} + A_2 e^{i(\phi_2 + \Delta\omega t)} \right] e^{i\omega_1 t} \\ &= A(t) e^{i(\omega_1 t + \phi(t))}, \end{aligned} \tag{2.0.12}$$

donde

$$A(t) = \sqrt{A_1^2 + A_2^2 + 2A_1 A_2 \cos(\phi_1 - \phi_2 - \Delta\omega t)} \tag{2.0.13}$$

y

$$\operatorname{tg} \phi(t) = \frac{A_1 \sin\phi_1 + A_2 \sin(\phi_2 + \Delta\omega t)}{A_1 \cos\phi_1 + A_2 \cos(\phi_2 + \Delta\omega t)}. \tag{2.0.14}$$

Esto permite interpretar la oscilación obtenida por superposición como un movimiento armónico simple aproximado con frecuencia ω_1 y con una amplitud y fase variando lentamente con frecuencia $\Delta\omega/2\pi$.

El resultado de esta vibración es semejante al de una vibración de frecuencia $\Delta\omega/2\pi$ modulada a través de la función de amplitud (2.0.13).

La dinámica analizada hasta ahora es puramente conservativa. Pero en la mayoría de sistemas de origen físico la disipación está presente. El rozamiento debido al desplazamiento sobre una superficie, o la resistencia producida por el movimiento en el seno de un fluido viscoso son dos ejemplos claros. En este último caso, por ejemplo, el efecto disipativo consiste en que el movimiento se ve afectado por una fuerza proporcional a la velocidad pero de signo contrario. Obtenemos asi el sistema

$$mx'' + Rx' + kx = 0, \tag{2.0.15}$$

donde R es la constante de resistencia mecánica.

Es fácil calcular la solución general de (2.0.15) como superposición de las dos soluciones fundamentales obtenidas resolviendo el polinomio característico de (2.0.15):

$$m\lambda^2 + R\lambda + k = 0. \tag{2.0.16}$$

Obtenemos las dos raíces

$$\lambda_{\pm} = \frac{-R \pm \sqrt{R^2 - 4mk}}{2m} \tag{2.0.17}$$

De (2.0.17) es fácil deducir que:

* Cuando $0 < R < 2\sqrt{mk}$, es decir, para tasas de disipación suficientemente pequeñas, los autovalores $\lambda_{\pm}$ son complejos con parte real $-R/2m$. De este modo las soluciones de (2.0.15) admiten la expresión

  $$x(t) = e^{-Rt/2m}\left[\alpha_+ e^{i\sqrt{R^2-4mk}t/2m} + \alpha_- e^{-i\sqrt{R^2-4mk}t/2m}\right].$$

  Las soluciones son por tanto oscilaciones armónicas exponencialmente amortiguadas.

  Conviene también observar que en este rango de valores de R, la tasa de decaimiento exponencial $R/2m$, depende de manera lineal y creciente de R.

* Cuando $R > 2\sqrt{mk}$ los dos autovalores $\lambda_{\pm}$ son reales y por tanto las soluciones no oscilan. En este caso la tasa exponencial de decaimiento de la solución general (2.0.15) viene dada por el autovalor λ_+ al que corresponde la solución fundamental con menor decaimiento.

* De este modo la función $\gamma(R)$ que establece la tasa exponencial de decaimiento toma los valores:

  $$\gamma(R) = \begin{cases} \frac{R}{2m}, & \text{cuando} \quad 0 < R < 2\sqrt{mk} \\ \frac{R}{2m} - \frac{\sqrt{R^2-4mk}}{2m}, & \text{cuando} \quad R > 2\sqrt{mk}. \end{cases}$$

  Esta función es creciente cuando $0 < R < 2\sqrt{mk}$ y decreciente cuando $R > 2\sqrt{mk}$ y alcanza su máximo cuando $R = 2\sqrt{mk}$. En este caso $\lambda_+ = \lambda_-$ y por tanto las soluciones fundamentales de (2.0.15) son $x_1(t) = e^{-Rt/2m}$ y $x_2(t) = te^{-Rt/2m}$.

  Por tanto, la elección de la constante disipativa R que maximiza la tasa de decaimiento exponencial es

  $$R = 2\sqrt{mk},$$

y la tasa óptima correspondiente

$$\gamma = \frac{R}{2m} = \sqrt{\frac{k}{m}},$$

si bien ésta no se alcanza, en un sentido estricto, puesto que la solución fundamental correspondiente presenta un factor multiplicativo t.

* De este análisis se deduce que, contrariamente a lo que podría indicarnos una primera intuición, la tasa exponencial de decaimiento no es una función creciente del coeficiente de disipación R. De hecho, a medida que $R \to \infty$ la tasa de decaimiento tiende a cero. El hecho que cuando R supera el valor critico $2\sqrt{mk}$ la tasa de decaimiento empieza a decrecer se denomina fenómeno de sobredisipación ("overdamping").

En esta sección hemos estudiado algunos de los aspectos más sencillos del movimiento armónico lineal. Evidentemente, las ecuaciones de ondas y sus aproximaciones numéricas, objetivo de este curso, son de naturaleza mucho más compleja. Pero puede decirse que la ecuación de ondas es en realidad el análogo de la ecuación (2.0.1) del oscilador armónico en un espacio de Hilbert en dimensión infinita.

Por otra parte, la aproximación numérica de las ecuaciones de ondas conduce de manera natural a versiones vectoriales de la ecuación (2.0.1) en las que los fenómenos aqui descritos son también relevantes. Surge sin embargo una nueva problemática relativa a la interacción de los diferentes componentes del sistema que se hace más y más compleja a medida que el sistema aumenta de dimensión y se aproxima a la ecuación de ondas original.

En la próxima sección analizamos la ecuación de transporte lineal en la que algunas de estas dificultades pueden ya ser vislumbradas.

2.1. La ecuación de ondas y sus variantes

En esta sección indicamos algunos ejemplos de ecuaciones y

sistemas en Derivadas Parciales de la Física, Mecánica y otras Ciencias en las que, de un modo u otro, intervienen los mismos fenómenos ondulatorios que la ecuación de ondas describe.

Recordemos que, normalmente, cuando nos referimos a la ecuación de ondas, la incógnita es una función escalar $u = u(x,t)$ donde $x = (x_1, \cdots, x_n) \in \mathbf{R}^n$ denota la variable espacial y $t \in \mathbf{R}$ la temporal. En las aplicaciones físicas la dimensión espacial es normalmente $n = 1, 2, 3$. *La ecuación de ondas* se escribe

entonces

$$u_{tt} - \Delta u = 0, \tag{2.1.1}$$

donde $u_t = \partial u/\partial t$ denota la derivación temporal con respecto al tiempo y Δ es al clásico operador de Laplace:

$$\Delta = \sum_{i=1}^{n} \frac{\partial^2}{\partial x_i^2}. \tag{2.1.2}$$

La ecuación de ondas en dimensiones espaciales $n = 1$ y 2 permite modelizar las pequeñas vibraciones de cuerdas y membranas, mientras que en tres dimensiones espaciales interviene en la propagación del potencial de un campo acústico.

Para simplificar la presentación en esta sección introduciremos las ecuaciones en su forma más sencilla. En particular, supondremos que los coeficientes son constantes (lo cual equivale a suponer que el medio considerado es homogéneo) y los normalizamos al valor unidad, lo cual en este caso no supone ninguna pérdida de generalidad como se puede comprobar mediante una simple dilatación/contracción de la variable temporal o espacial.

En el ámbito de las frecuencias, como es habitual en acústica y en el estudio de las vibraciones, la ecuación de ondas puede también reducirse a *la ecuación de Helmholtz*

$$-\Delta u = \lambda u. \tag{2.1.3}$$

La ecuación de transporte lineal

$$u_t + \sum_{i=1}^{n} b_i u_{x_i} = 0, \tag{2.1.4}$$

y *la ecuación de Liouville*

$$u_t - \sum_{i=1}^{n} (b_i u)_{x_i} = 0 \tag{2.1.5}$$

están también intimamente ligadas a la ecuación de ondas. En efecto, en una dimensión espacial, la ecuación de ondas

$$u_{tt} - u_{xx} = 0 \tag{2.1.6}$$

puede también escribirse como

$$(\partial_t + \partial_x)(\partial_t - \partial_x) u = 0, \tag{2.1.7}$$

o

$$(\partial_t - \partial_x)(\partial_t + \partial_x) u = 0, \tag{2.1.8}$$

o, lo que es lo mismo, el operador de d'Alembert

$$\partial_t^2 - \partial_x^2 \tag{2.1.9}$$

puede factorizarse de las dos siguientes maneras

$$\partial_t^2 - \partial_x^2 = (\partial_t + \partial_x)(\partial_t - \partial_x) = (\partial_t - \partial_x)(\partial_t + \partial_x). \tag{2.1.10}$$

Vemos pues que el operador de d'Alembert es la composición de dos operadores de transporte.

Conviene también señalar que, cuando los coeficientes b_i son constantes, la ecuación de transporte y de Liouville sólo difieren en un signo, diferencia que puede ser eliminada invirtiendo el sentido del tiempo.

Utilizando las notaciones habituales

$$\nabla u = (\partial_1 u, \cdots, \partial_n u) \tag{2.1.11}$$

$$\operatorname{div} \vec{u} = \sum_{i=1}^{n} \frac{\partial u_i}{\partial x_i} \tag{2.1.12}$$

para los operadores gradiente y divergencia y denotando mediante $\cdot$ el producto escalar en $\mathbf{R}^n$ las ecuaciones de transporte y Liouville se pueden escribir respectivamente como

$$u_t + \vec{b} \cdot \nabla u = 0 \tag{2.1.13}$$

y

$$u_t - \operatorname{div}\left(\vec{b} u\right) = 0. \tag{2.1.14}$$

La *ecuación de Schrödinger* de la Mecánica Cuántica, que también interviene en el estudio de fibras ópticas es también una ecuación que, en muchos sentidos, se asemeja a la ecuación de ondas:

$$i u_t + \Delta u = 0. \tag{2.1.15}$$

En este caso, la incógnita u toma valores complejos.

La *ecuación de las placas vibrantes*

$$u_{tt} + \Delta^2 u = 0 \tag{2.1.16}$$

es también muy similar a la ecuación de ondas. Además puede factorizarse en dos operadores de Schrödinger conjugados

$$\partial_t^2 + \Delta^2 = -(i\partial_t + \Delta)(i\partial_t - \Delta). \tag{2.1.17}$$

En una dimensión espacial la ecuación

$$\partial_t^2 + \partial_x^4 u = 0 \tag{2.1.18}$$

describe las vibraciones de una viga.

Las siguientes son también variantes de la ecuación de ondas:

$$u_{tt} - u_{xx} + d\, u_t = 0 \quad \textit{(ecuación del telégrafo)}, \tag{2.1.19}$$

$$u_t + u_{xxx} = 0 \quad \textit{(ecuación de Airy)}, \tag{2.1.20}$$

$$u_{tt} - \Delta u + u = 0 \quad \textit{(ecuación de Klein-Gordon)}, \tag{2.1.21}$$

El *sistema de Lamé* para las vibraciones de un cuerpo tridimensional elástico puede también entenderse como un sistema de ecuaciones de ondas acopladas:

$$u_{tt} - \lambda \Delta u - (\lambda + \mu) \nabla \operatorname{div} u = 0. \tag{2.1.22}$$

En este caso la incógnita u es un vector de tres componentes $u = (u_1, u_2, u_3)$ que describe las deformaciones del cuerpo elástico.

Las ecuaciones que hemos descrito son *lineales* y provienen de ecuaciones y sistemas más complejos de la Mecánica, de carácter no-lineal, a través de linealizaciones, lo cual las hace válidas sólo para pequeños valores de la incógnita u.

El *sistema de Maxwell* para las ondas electromagnéticas posee también muchas de las características de las ecuaciones de ondas:

$$\begin{cases} E_t = \operatorname{rot} B \\ B_t = -\operatorname{rot} E \\ \operatorname{div} B = \operatorname{div} E = 0. \end{cases} \tag{2.1.23}$$

Aquí rot denota el rotacional de un campo de vectores.

Con el objeto de entender la semejanza de este sistema con la ecuación de ondas (2.1.6) conviene observar que esta última también puede escribirse en la forma de un sistema hiperbólico de ecuaciones de orden uno:

$$\begin{cases} u_t = v_x \\ v_t = u_x. \end{cases} \tag{2.1.24}$$

Sin embargo, muchas ecuaciones relevantes que intervienen en el estudio de las ondas tienen un carácter no-lineal. Por ejemplo, *la ecuación eikonal,*

$$| \nabla u | = 1 \tag{2.1.25}$$

interviene en el cálculo de soluciones de ecuaciones de ondas mediante métodos de la Óptica Geométrica.

Lo mismo ocurre con *ecuación de Hamilton-Jacobi*:

$$u_t + H(\nabla u,\, x) = 0. \tag{2.1.26}$$

La ecuación de Korteweg-de Vries (KdV) es una versión no-lineal de la ecuación de Airy que permite analizar la propagación de ondas en canales:

$$u_t + uu_x + u_{xxx} = 0 \tag{2.1.27}$$

y da lugar a los célebres *solitones*.

En el contexto de la Mecánica de Fluidos los dos ejemplos más relevantes son sin duda *las ecuaciones de Navier-Stokes* para un fluido viscoso homogéneo e incompresible

$$\begin{cases} \rho u_t - \nu\Delta u + u\cdot\nabla u = \nabla p \\ \operatorname{div} u = 0 \end{cases} \tag{2.1.28}$$

y *las ecuaciones de Euler* para fluidos perfectos

$$\begin{cases} \rho u_t + u\cdot\nabla u = \nabla p \\ \operatorname{div} u = 0. \end{cases} \tag{2.1.29}$$

En estos sistemas u denota el campo de velocidades del fluido y p es la presión.

Las ecuaciones de Burgers viscosa e inviscida son, en algún sentido, versiones unidimensionales de estas ecuaciones

$$u_t + uu_x - u_{xx} = 0, \tag{2.1.30}$$

$$u_t + uu_x = 0. \tag{2.1.31}$$

En esta última las soluciones desarrollan ondas de choque en tiempo finito.

Las ecuaciones que hemos citado, aunque numerosas, no son más que algunos de los ejemplos más relevantes de ecuaciones en las que intervienen de un modo u otro fenómenos de propagación de ondas y en las que los contenidos que desarrollaremos en este curso resultarán de utilidad.

2.2. La fórmula de D'Alembert

Consideramos la ecuación de ondas unidimensional $(1-d)$ en toda la recta real

$$u_{tt} - u_{xx} = 0,\ x \in \mathbf{R},\ t > 0. \tag{2.2.1}$$

D'Alembert observó que las soluciones de (2.2.1) pueden escribirse como superposición de dos ondas de transporte

$$u(x,t) = f(x+t) + g(x-t). \tag{2.2.2}$$

Es fácil comprobar que toda función de la forma (2.2.2) es solución de (2.2.1).

La fórmula (2.2.2) muestra que la velocidad de propagación en el modelo (2.2.1) es uno. En efecto, según (2.2.2), las soluciones de (2.2.1) son superposición de ondas de transporte que viajan en el espacio $\mathbf{R}$ a velocidad uno a izquierda y derecha.

Para comprobar que toda solución de (2.2.1) es de la forma (2.2.2) basta observar que el operador de d'Alembert $\partial_t^2 - \partial_x^2$ puede descomponerse del modo siguiente:

$$u_{tt} - u_{xx} = (\partial_t - \partial_x)(\partial_t + \partial_x)\, u = 0. \tag{2.2.3}$$

Introduciendo la variable auxiliar

$$v = (\partial_t + \partial_x)\, u, \tag{2.2.4}$$

la ecuación se escribe como

$$(\partial_t - \partial_x)\, v = v_t - v_x = 0, \tag{2.2.5}$$

de modo que

$$v = h(x+t). \tag{2.2.6}$$

La ecuación (2.2.4) se reduce entonces a

$$u_t + u_x = h(x+t). \tag{2.2.7}$$

Para su resolución observamos que la función

$$w(t) = u(t + x_0, t)$$

verifica

$$w'(t) = h(2t + x_0)$$

cuya solución es

$$w(t) = \frac{H(2t+x_0)}{2} + w(0) = \frac{H(2t+x_0)}{2} + u(x_0, 0), \tag{2.2.8}$$

donde H es una primitiva de h.

Por lo tanto, como

$$u(x,t) = w(t)$$

con $x_0 = x - t$ obtenemos

$$u(x,t) = \frac{H(x+t)}{2} + u(x-t,\,0) \tag{2.2.9}$$

lo cual confirma la expresión (2.2.2).

Esta fórmula permite calcular explícitamente la solución del problem de Cauchy:

$$\begin{cases} u_{tt} - u_{xx} = 0, & x \in \mathbf{R}, \quad t > 0 \\ u(x,0) = \varphi(x),\ u_t(x,0) = \psi(x), & x \in \mathbf{R}. \end{cases} \tag{2.2.10}$$

En efecto, en vista de la expresión (2.2.2), e identificando los perfiles f y g en función de los datos iniciales φ y ψ obtenemos que

$$u(x,t) = \frac{\varphi(x+t) + \varphi(x-t)}{2} + \frac{1}{2}\int_{x-t}^{x+t} \psi(y)dy \tag{2.2.11}$$

es la única solución de (2.2.10).

2.3. Resolución de la ecuación de ondas mediante series de Fourier

Consideramos la ecuación de ondas unidimensional $(1-d)$:

$$\begin{cases} u_{tt} - u_{xx} = 0, & 0 < x < \pi, \quad t > 0 \\ u(0,t) = u(\pi,t) = 0, & t > 0 \\ u(x,0) = u_0(x), u_t(x,0) = u_1(x), & 0 < x < \pi. \end{cases} \tag{2.3.1}$$

Se trata de un modelo sencillo para las vibraciones de una cuerda unidimensional flexible de longitud π, fijada en sus extremos $x = 0, \pi$.

Es fácil representar las soluciones de (2.3.1) mediante series de Fourier. Para ello escribimos el desarrollo de Fourier de los datos iniciales:

$$u_0(x) = \sum_{k=1}^{\infty} a_k \operatorname{sen}(kx),\ u_1(x) = \sum_{k=1}^{\infty} b_k \operatorname{sen}(kx), \tag{2.3.2}$$

donde los coeficientes de Fourier vienen dados por las clásicas fórmulas:

$$a_k = \frac{2}{\pi}\int_0^{\pi} u_0(x)\operatorname{sen}(kx)dx;\ b_k = \frac{2}{\pi}\int_0^{\pi} u_1(x)\operatorname{sen}(kx)dx, \quad k \geq 1. \tag{2.3.3}$$

La solución de (2.3.1) viene entonces dada por la fórmula

$$u(x,t) = \sum_{k=1}^{\infty}\left(a_k\cos(kt) + \frac{b_k}{k}\operatorname{sen}(kt)\right)\operatorname{sen}(kx). \tag{2.3.4}$$

Conviene observar que la evolución temporal de cada uno de los coeficientes de Fourier

$$u_k(t) = a_k \cos(kt) + \frac{b_k}{k} \operatorname{sen}(kt), \tag{2.3.5}$$

obedece la ecuación del muelle

$$u_k'' + k^2 u_k = 0. \tag{2.3.6}$$

Para cada una de estas ecuaciones la energía

$$e_k(t) = \frac{1}{2}\left[\mid u_k'(t) \mid^2 + k^2 \mid u_k(t) \mid^2 \right] \tag{2.3.7}$$

se conserva en tiempo[1]

Superponiendo cada una de las leyes de conservación de las energías e_k, $k \geq 1$, de las diferentes componentes de Fourier de la solución obtenemos la ley de conservación de la energía de las soluciones de (2.3.1):

$$E(t) = \frac{1}{2} \int_0^\pi \left[|u_x(x,t)|^2 + |u_t(x,t)|^2 \right] dx. \tag{2.3.8}$$

Se trata de la energía total de la vibración, suma de la energía potencial y de la energía cinética.

Se cumple efectivamente que

$$E(t) = E(0), \, \forall t \geq 0 \tag{2.3.9}$$

para las soluciones de (2.3.1).

Esta ley de conservación de energía puede probarse de, al menos, dos modos distintos:

• *Series de Fourier:*

Si utilizamos las propiedades clásicas de ortogonalidad de las funciones trigonométricas

$$\int_0^\pi \operatorname{sen}(kx) \operatorname{sen}(jx) dx = \frac{\pi}{2} \delta_{jk}, \, \int_0^\pi \cos(kx) \cos(jx) = \frac{\pi}{2} \delta_{jk}, \tag{2.3.10}$$

donde δ_{jk} denota la delta de Kronecker, la ley de conservación (2.3.9) se deduce efectivamente de la conservación de las energías e_k de (2.3.7) para cada $k \geq 1$.

• *Método de la energía:*

[1] Para comprobarlo basta multiplicar (2.3.6) por u_k' y observar que $u_k'' u_k' = \frac{1}{2}\left((u_k')^2\right)'$ y $u_k u_k' = \frac{1}{2}\left(u_k^2\right)'$.

La ley de conservación (2.3.9) puede también obtenerse directamente de (2.3.1). Basta para ello multiplicar por u_t e integrar con respecto a $x \in (0,\pi)$. Tenemos entonces

$$\int_0^\pi (u_{tt} - u_{xx})\, u_t dx = 0.$$

Por otra parte,

$$\int_0^\pi u_{tt} u_t dx = \frac{1}{2}\frac{d}{dt}\int_0^\pi |u_t(x,t)|^2\, dx$$

y

$$-\int_0^\pi u_{xx} u_t dx = \int_0^\pi u_x u_{xt} dx = \frac{1}{2}\frac{d}{dt}\int_0^\pi |u_x(x,t)|^2\, dx.$$

En la última identidad hemos utilizado la fórmula de integración por partes y las condiciones de contorno de modo que, como $u(\cdot,t) = 0$ para $x = 0,\pi$, necesariamente también se tiene $u_t(\cdot,t) = 0$ para $x = 0,\pi$.

Los argumentos anteriores son formales pero la ley de conservación y la estructura de la energía E en (2.3.8) indican en realidad cuál es el espacio natural para resolver la ecuación de ondas. En efecto, se trata del espacio de Hilbert, también denominado espacio de energía,

$$H = H_0^1(0,\pi) \times L^2(0,\pi). \tag{2.3.11}$$

La norma natural en este espacio es

$$|(f,g)|_H = \left[\| f \|^2_{H_0^1(0,\pi)} + \| g \|^2_{L^2(0,\pi)}\right]^{1/2} = \left[\int_0^\pi \left(f_x^2 + g^2\right) dx\right]^{1/2}. \tag{2.3.12}$$

Conviene observar que, salvo un factor multiplicativo $1/2$ la energía E coincide con el cuadrado de la norma H de (u, u_t).

Deducimos que la norma en H de la solución[2] (u, u_t) se conserva a lo largo del tiempo. Esto sugiere que H es el espacio natural para resolver el sistema (2.3.1). Esto es así y se tiene el siguiente resultado de existencia y unicidad:

"Para todo par de datos iniciales $(u_0, u_1) \in H$, i.e. $u_0 \in H_0^1(0,\pi)$ y $u_1 \in L^2(0,\pi)$, existe una única solución $(u, u_t) \in C([0,\infty); H)$ de (2.3.1). Esta solución pertenece por tanto a la clase

$$u \in C\left([0,\infty);\, H_0^1(0,\pi)\right) \cap C^1\left([0,\infty);\, L^2(0,\pi)\right) \tag{2.3.13}$$

[2]En este punto abusamos un tanto de la terminología. En efecto, la solución de (2.3.1) es la función $u = u(x,t)$. Ahora bien, como (2.3.1) es una ecuación de orden dos en tiempo es natural escribirla como un sistema de dos ecuaciones de orden uno en tiempo, con dos incógnitas. En este caso el par (u, u_t) puede ser considerado como la solución, lo cual es coherente con el hecho de haber introducido dos datos iniciales en el sistema (2.3.1).

y la energía correspondiente $E(t)$ de (2.3.8) *se conserva en el tiempo".*

En lo que respecta al desarrollo en serie de Fourier (2.3.2)-(2.3.3) de los datos iniciales, el hecho de que estos pertenezcan a $H_0^1(0,\pi) \times L^2(0,\pi)$ significa que

$$\sum_{k=1}^{\infty} [k^2|a_k|^2 + |b_k|^2] < \infty. \tag{2.3.14}$$

De hecho

$$E(0) = \frac{1}{2}\int_0^\pi [|u_{0,x}|^2 + |u_1|^2]dx = \frac{\pi}{4}\sum_{k=1}^{\infty}[k^2|a_k|^2 + |b_k|^2] < \infty. \tag{2.3.15}$$

Este resultado de existencia y unicidad puede probarse de al menos dos maneras adicionales, además del método de series de Fourier que acabamos de desarrollar:

- la teoría de semigrupos;
- el método de Galerkin.

El mismo tipo de análisis puede ser desarrollado con muy pocas modificaciones en el caso de varias dimensiones espaciales. Basta para ello utilizar los resultados clásicos sobre la descomposición espectral de la ecuación de Laplace.

Con el objeto de presentar los resultados fundamentales en el caso de varias dimensiones consideramos un *abierto* Ω de $\mathbf{R}^n$, $n \geq 1$. En este punto la regularidad de Ω no es relevante. Con el objeto de desarrollar las soluciones en series de Fourier es, sin embargo, importante suponer que Ω es *acotado*.

Consideramos entonces la ecuación de ondas

$$\begin{cases} u_{tt} - \Delta u = 0, & x \in \Omega, \quad t > 0 \\ u = 0, & x \in \partial\Omega, \quad t > 0 \\ u(x,0) = u_0(x),\ u_t(x,0) = u_1(x), & x \in \Omega. \end{cases} \tag{2.3.16}$$

Aquí y en lo sucesivo Δ denota el clásico operador de Laplace

$$\Delta u = \sum_{i=1}^{n} \frac{\partial^2 u}{\partial x_i^2}. \tag{2.3.17}$$

Para $n \geq 1$, (2.3.16) es claramente una generalización de la ecuación de la cuerda vibrante (2.3.1). Cuando $n = 2$, (2.3.16) es un modelo para las vibraciones de una membrana que, en reposo, ocupa el dominio Ω del plano. Cuando $n = 3$, (2.3.16) describe la propagación de la presión de las ondas acústicas. Sin embargo, desde un punto de vista matemático, la ecuación (2.3.16) puede tratarse de modo semejante en cualquier dimensión espacial.

Consideramos ahora el problema espectral:

$$\begin{cases} -\Delta\varphi = \lambda\varphi & \text{en} \quad \Omega \\ \varphi = 0 & \text{en} \quad \partial\Omega. \end{cases} \tag{2.3.18}$$

Es bien sabido (véase [2] o [8], por ejemplo) que los autovalores $\{\lambda_j\}_{j\geq 1}$ de (2.3.18) constituyen una sucesión creciente de números positivos que tiende a infinito

$$0 < \lambda_1 < \lambda_2 \leq \lambda_3 \leq \cdots \leq \lambda_n \leq \cdots \to \infty.$$

El primero de los autovalores es simple. Es habitual repetir el resto de acuerdo a su multiplicidad. De este modo, existe una sucesión de autofunciones $\{\varphi_j\}_{j\geq 1}$, donde φ_j es una autofunción asociada al autovalor λ_j, que constituye una base ortonormal de $L^2(\Omega)$. Es decir, se tiene, en particular,

$$\int_\Omega \varphi_j\varphi_k dx = \delta_{jk}. \tag{2.3.19}$$

De acuerdo a (2.3.19), multiplicando la ecuación (2.3.18) correspondiente a λ_k por φ_j e integrando en Ω, gracias a la fórmula de Green obtenemos que

$$\int_\Omega \nabla\varphi_j \cdot \nabla\varphi_k dx = \lambda_j \int_\Omega \varphi_j\varphi_k dx = \lambda_j\delta_{jk} = \lambda_k\delta_{jk}. \tag{2.3.20}$$

De este modo se deduce que las autofunciones son también ortogonales en $H_0^1(\Omega)$. Más concretamente, la sucesión $\left\{\varphi_j/\sqrt{\lambda_j}\right\}_{j\geq 1}$ constituye una base ortonormal de $H_0^1(\Omega)$.

Utilizando esta base de funciones propias del Laplaciano podemos resolver la ecuación de ondas (2.3.16) como lo hicimos en una variable espacial. Para ello desarrollamos los datos iniciales (u_0, u_1) de (2.3.16) del modo siguiente

$$u_0(x) = \sum_{k=1}^{\infty} a_k\varphi_k(x);\ u_1(x) = \sum_{k=1}^{\infty} b_k\varphi_k(x). \tag{2.3.21}$$

Buscamos entonces la solución u de (2.3.16) en la forma

$$u(x,t) = \sum_{k=1}^{\infty} u_k(t)\varphi_k(x). \tag{2.3.22}$$

Observamos entonces que los coeficientes $\{u_k\}$ han de resolver la ecuación diferencial:

$$u_k''(t) + \lambda_k u_k(t) = 0,\ t > 0,\ u_k(0) = a_k,\ u_k'(0) = b_k, \tag{2.3.23}$$

de modo que

$$u_k(t) = a_k \cos\left(\sqrt{\lambda_k}t\right) + \frac{b_k}{\sqrt{\lambda_k}} \operatorname{sen}\left(\sqrt{\lambda_k}t\right). \tag{2.3.24}$$

De este modo obtenemos que la solución u de (2.3.16) admite la expresión

$$u(x,t) = \sum_{k=1}^{\infty} \left(a_k \cos\left(\sqrt{\lambda_k}t\right) + \frac{b_k}{\sqrt{\lambda_k}} \operatorname{sen}\left(\sqrt{\lambda_k}t\right) \right) \varphi_k(x). \tag{2.3.25}$$

La similitud de la expresión (2.3.4) del caso de una dimensión espacial con la fórmula (2.3.25) del caso general es evidente. En realidad (2.3.4) es un caso particular de (2.3.25). Basta observar que cuando $\Omega = (0, \pi)$, el problema de autovalores para el Laplaciano se convierte en un problema clásico de Sturm-Liouville. El espectro es por tanto explícito:

$$\lambda_k = k^2,\ k \geq 1;\ \varphi_k(x) = \sqrt{\frac{2}{\pi}} \operatorname{sen}(kx),\ k \geq 1. \tag{2.3.26}$$

Con estos datos las expresiones (2.3.4) y (2.3.25) coinciden efectivamente.

La energía de las soluciones de (2.3.16) es en este caso

$$E(t) = \frac{1}{2} \int_{\Omega} \left[|\ \nabla u(x,t)\ |^2 + |u_t(x,t)|^2 \right] dx \tag{2.3.27}$$

y también se conserva en tiempo. Nuevamente la energía es proporcional al cuadrado de la norma en el espacio de la energía $H = H_0^1(\Omega) \times L^2(\Omega)$.

En este caso el resultado básico de existencia y unicidad de soluciones dice que:

"Si $(u_0, u_1) \in H_0^1(\Omega) \times L^2(\Omega)$*, existe una única solución* $(u, u_t) \in C([0,\infty);\ H)$*, i.e.*

$$u \in C\left([0,\infty);\ H_0^1(\Omega)\right) \cap C^1\left([0,\infty);\ L^2(\Omega)\right), \tag{2.3.28}$$

de (2.3.16)*. La energía* $E(t)$ *en* (2.3.27) *es constante en tiempo".*

Conviene también señalar que, si bien la regularidad (2.3.28) de las soluciones débiles permite interpretar la ecuación de ondas en un sentido débil, el hecho que u sea solución con la regularidad (2.3.28), junto con la propiedad del operador de Laplace con condiciones de contono de Dirichlet de constituir un isomorfismo de $H_0^1(\Omega)$ en $H^{-1}(\Omega)$, permite deducir que $u \in C^2\left([0,\infty);\ H^{-1}(\Omega)\right)$. De este modo se concluye que la ecuación (2.3.16) tiene sentido para cada $t > 0$ en el espacio $H^{-1}(\Omega)$.

Acabamos de ver cómo se puede aplicar el método de Fourier para la resolución de la ecuación de ondas. Basta para ello conocer la descomposición espectral del Laplaciano con condiciones de Dirichlet (2.3.18).

El método de Fourier puede ser adaptado a muchas otras situaciones:

- Condiciones de contorno de Neumann, o mixtas en las que la condición de Dirichlet y Neumann se satisfacen en subconjuntos complementarios de la frontera.

- Ecuaciones más generales con coeficientes dependientes de x de la forma:

$$\rho(x)u_{tt} - \operatorname{div}(a(x)\nabla u) + q(x)u = 0,$$

donde ρ, a y q son funciones medibles y acotadas y ρ y a son uniformemente positivas, i.e. existen ρ_0, $a_0 > 0$ tales que

$$\rho(x) \geq \rho_0,\ a(x) \geq a_0,\ p.c.t.\ x \in \Omega.$$

Pero es cierto también que el método de Fourier tiene sus limitaciones. En particular no permite abordar ecuaciones no lineales, con coeficientes que dependen de x y t, etc. En estos últimos casos, los métodos de Galerkin y la teoría de semi-grupos se muestran mucho más flexibles y útiles.

2.4. Series de Fourier como método numérico

En la sección anterior hemos visto que la ecuación de ondas puede ser resuelta mediante series de Fourier obteniéndose la expresión

$$u(x,t) = \sum_{k=1}^{\infty} \left[a_k \cos\left(\sqrt{\lambda_k}t\right) + \frac{b_k}{\sqrt{\lambda_k}} \operatorname{sen}\left(\sqrt{\lambda_k}t\right) \right] \varphi_k(x), \tag{2.4.1}$$

siendo $\{\varphi_k\}_{k\geqslant 1}$ y $\{\lambda_k\}_{k\geqslant 1}$ las autofunciones y autovalores del Laplaciano. Como vimos, es conveniente elegir $\{\varphi_k\}_{k\geqslant 1}$ de modo que constituyan una base ortonormal de $L^2(\Omega)$.

Vimos asimismo que la energía

$$E(t) = \frac{1}{2}\int_{\Omega} \left[\mid \nabla u(x,t) \mid^2 + \mid u_t(x,t) \mid^2 \right] dx, \tag{2.4.2}$$

se conserva a lo largo de las trayectorias.

La energía inicial de las soluciones viene dada por

$$E(0) = \frac{1}{2}\sum_{k=1}^{\infty} \left[\mid \lambda_k a_k \mid^2 + \mid b_k \mid^2 \right]. \tag{2.4.3}$$

Así, la hipótesis de que los datos iniciales sean de energía finita

$$(u_0, u_1) \in H_0^1(\Omega) \times L^2(\Omega), \tag{2.4.4}$$

es equivalente a que las sucesiones $\left\{a_k\sqrt{\lambda_k}\right\}_{k\geqslant 1}$, $\{b_k\}_{k\geqslant 1}$ pertenezcan al espacio de las sucesiones de cuadrado sumable ℓ^2.

En vista del desarrollo en serie (2.4.1) de las soluciones, parece natural construir un método numérico en el que la aproximación venga dada, simplemente, por las sumas parciales de la serie:

$$u_N(x,t)=\sum_{k=1}^{N}\left[a_k\cos\left(\sqrt{\lambda_k}t\right)+\frac{b_k}{\sqrt{\lambda_k}}\operatorname{sen}\left(\sqrt{\lambda_k}t\right)\right]\varphi_k(x). \tag{2.4.5}$$

La suma finita de u_N en (2.4.5) proporciona, efectivamente, una aproximación de la solución u representada en la serie de Fourier (2.4.1). Para comprobarlo consideremos el resto

$$\varepsilon_N=u-u_N=\sum_{k\geqslant N+1}\left[a_k\cos\left(\sqrt{\lambda_k}t\right)+\frac{b_k}{\sqrt{\lambda_k}}\operatorname{sen}\left(\sqrt{\lambda_k}t\right)\right]\varphi_k(x). \tag{2.4.6}$$

Teniendo en cuenta que

$$\int_\Omega \nabla\varphi_k\cdot\nabla\varphi_j dx=\begin{cases}0, & \text{si} \quad k\neq j\\ \lambda_k, & \text{si} \quad k=j,\end{cases}$$

es fácil comprobar que

$$\begin{aligned}\left|\left|\nabla\varepsilon_N(t)\right|\right|^2_{L^2(\Omega)} &= \sum_{k\geqslant N+1}\lambda_k\left[a_k\cos\left(\sqrt{\lambda_k}t\right)+\frac{b_k}{\sqrt{\lambda_k}}\operatorname{sen}\left(\sqrt{\lambda_k}t\right)\right]^2 \\ &\leqslant 2\sum_{k\geqslant N+1}\left[\lambda_k\mid a_k\mid^2+\mid b_k\mid^2\right].\end{aligned} \tag{2.4.7}$$

Como la serie (2.4.3) de la energía inicial es convergente, en virtud de (2.4.7) deducimos que

$$u_N(t)\to u(t)\text{ en }C\left([0,\infty);\,H_0^1(\Omega)\right), \tag{2.4.8}$$

cuando $N\to\infty$.

El mismo argumento permite probar que

$$u_{N,t}\to u_t(t)\text{ en }C\left([0,\infty);\,L^2(\Omega)\right). \tag{2.4.9}$$

De (2.4.8)-(2.4.9) deducimos que, cuando los datos iniciales están en el espacio de la energía $H_0^1(\Omega)\times L^2(\Omega)$, las sumas parciales (2.4.5) proporcionan una aproximación eficaz de la solución en dicho espacio, uniformemente en tiempo $t\geqslant 0$.

Cabe por tanto preguntarse sobre la tasa o velocidad de la convergencia. El argumento anterior no proporciona ninguna información en este sentido puesto

que la mera convergencia de la serie (2.4.3) no permite decir nada sobre la velocidad de convergencia de sus sumas parciales.

Con el objeto de obtener tasas de convergencia es necesario hacer hipótesis adicionales sobre los datos iniciales. Supongamos por ejemplo que

$$(u_0,\, u_1) \in \Big[H^2 \cap H_0^1(\Omega)\Big] \times H_0^1(\Omega). \tag{2.4.10}$$

En este caso tenemos

$$\sum_{k\geqslant 1} \Big[\lambda_k^2 \mid a_k \mid^2 + \lambda_k \mid b_k \mid^2 \Big] < \infty. \tag{2.4.11}$$

En efecto, tal y como veíamos anteriormente en el caso de u_0, el que $u_1 \in H_0^1(\Omega)$ se caracteriza porque sus coeficientes de Fourier $(b_k)_{k\geqslant 1}$ satisfacen

$$\sum_{k\geqslant 1}^{\infty} \lambda_k \mid b_k \mid^2 < \infty. \tag{2.4.12}$$

Por otra parte, $\parallel \Delta\varphi \parallel_{L^2(\Omega)}$ define una norma equivalente a la inducida por $H^2(\Omega)$ en el subespacio [3] $H^2 \cap H_0^1(\Omega)$. Por otra parte, se tiene

$$\int_\Omega \Delta\varphi_k \Delta\varphi_j dx = \begin{cases} 0 & \text{si} \quad k \neq j \\ \lambda_k^2 & \text{si} \quad k = j. \end{cases} \tag{2.4.13}$$

Deducimos por tanto que

$$\Big|\Big|\Delta u_0\Big|\Big|^2_{L^2(\Omega)} = \sum_{k\geqslant 1} \lambda_k^2 \, |a_k|^2 \tag{2.4.14}$$

y, de este modo, observamos que, efectivamente, la serie (2.4.11) converge.

La información adicional que (2.4.11) proporciona sobre los coeficientes de Fourier permite obtener tasas de convergencia de u_N hacia u en el espacio de la energía. Por ejemplo, volviendo a (2.4.7) tenemos, usando el hecho de que $\{\lambda_j\}$ es creciente,

$$\begin{aligned} \Big|\Big|\nabla\varepsilon_N(t)\Big|\Big|^2_{L^2(\Omega)} &\leqslant 2 \sum_{k\geqslant N+1} \Big[\lambda_k \, |a_k|^2 + |b_k|^2 \Big] \\ &\leqslant 2 \sum_{k\geqslant N+1} \frac{1}{\lambda_k} \Big[\lambda_k^2 \, |a_k|^2 + \lambda_k \, |b_k|^2 \Big] \\ &\leqslant \frac{2}{\lambda_{N+1}} \sum_{k\geqslant N+1} \Big[\lambda_k^2 \, |a_k|^2 + \lambda_k \, |b_k|^2 \Big] \\ &\leqslant \frac{C}{\lambda_{N+1}} \Big|\Big| \, (u_0, u_1) \, \Big|\Big|^2_{H^2\cap H_0^1(\Omega)\times H_0^1(\Omega)}. \end{aligned}$$

[3] En este punto ultilizamos el resultado clásico de regularidad de las soluciones del problema de Dirichlet para el Laplaciano que garantiza que, si el dominio es de clase C^2 y el segundo miembro está en $L^2(\Omega)$, entonces la solución pertenece a $H^2 \cap H_0^1(\Omega)$.

El mismo argumento puede ser utilizado para estimar la norma de $\varepsilon_{N,t}$ en $L^2(\Omega)$. De este modo deducimos que

$$\| u-u_N \|_{L^\infty\left(0,\infty;\, H_0^1(\Omega)\right)\cap W^{1,\infty}(0,\infty;\, L^2(\Omega))} \leq \frac{C}{\sqrt{\lambda_{N+1}}} \Big|\Big| (u_0,\, u_1) \Big|\Big|_{H^2\cap H_0^1(\Omega)\times H_0^1(\Omega)}. \tag{2.4.15}$$

Esta desigualdad proporciona estimaciones explícitas sobre la velocidad de convergencia. En efecto, el clásico Teorema de Weyl sobre la distribución asintótica de los autovalores del Laplaciano asegura que

$$\lambda_N \sim c(\Omega) N^{2/n},\ N \to \infty \tag{2.4.16}$$

donde $c(\Omega)$ es una constante positiva que depende del dominio y n es la dimensión espacial[4].

Combinando (2.4.15) y (2.4.16) obtenemos que u_N converge a u en el espacio de la energía, uniformemente en tiempo $t \geqslant 0$, con un orden de $O\left(N^{-1/n}\right)$.

La hipótesis $(u_0, u_1) \in H^2 \cap H_0^1(\Omega) \times H_0^1(\Omega)$ realizada sobre los datos iniciales es sólo una de las posibles. De manera general puede decirse que, cuando los datos iniciales son más regulares que lo que el espacio de la energía exige y verifican las condiciones de compatibilidad adecuadas en relación a las condiciones de contorno, entonces, se puede establecer una estimación sobre la velocidad de convergencia de la aproximación que las sumas parciales del desarrollo en serie de Fourier proporcionan a la solución de la ecuación de ondas.

Este método de aproximación lo denominaremos *método de Fourier*. Se trata de un método de aproximación sumamente útil en una dimensión espacial puesto que, al disponer de la expresión explícita de las autofunciones φ_k y autovalores de λ_k, la aproximación u_N puede calcularse de manera totalmente explícita. Bastaría para ello con utilizar una fórmula de cuadratura para aproximar el valor (2.3.3) de los coeficientes de Fourier.

El método de Fourier es sin embargo mucho menos eficaz en varias dimensiones espaciales. En efecto, en ese caso no disponemos de la expresión explícita de las autofunciones y autovalores y su aproximación numérica es un problema tan complejo como el de la propia aproximación de la ecuación de ondas.

Otro de los inconvenientes del método de Fourier es que, cuando la ecuación es no-lineal o tiene coeficientes que dependen de (x, t), ya no se puede obtener una expresión explícita de la solución en serie de Fourier y por tanto tampoco de sus aproximaciones.

[4]Es obvio que, por ejemplo, en una dimensión espacial $n = 1$, la expresión asintótica en (2.4.16) coincide con lo que se observa en la expresión explícita del espectro. En efecto, recordemos que, cuando $\Omega = (0, \pi)$, $\lambda_k = k^2$.

Es por eso que el método de Fourier tiene una utilidad limitada y que precisamos de métodos más sistemáticos y robustos que funcionen no sólo en casos particulares sino para amplias clases de ecuaciones. En este marco destacan los métodos de diferencias y elementos finitos, que serán el objeto central de este curso.

2.5. La ecuación de ondas disipativa

Hemos visto cómo el método de Fourier permite representar explícitamente las soluciones de la ecuación de ondas y que proporciona en sí un método numérico de aproximación de las mismas. En esta sección vamos a describir cómo se puede utilizar este método para analizar propiedades cualitativas de ecuaciones de ondas perturbadas. Para ello consideremos el caso de la ecuación de ondas disipativa:

$$\begin{cases} u_{tt} - \Delta u + a u_t = 0 & \text{en} \quad \Omega \times (0,\infty) \\ u = 0 & \text{en} \quad \partial\Omega \times (0,\infty) \\ u(x,0) = u_0(x),\ u_t(x,0) = u_1(x) & \text{en} \quad \Omega. \end{cases} \tag{2.5.1}$$

Suponemos que Ω es un abierto acotado de $\mathbb{R}^n$ con $n \geqslant 1$ y que la constante a es positiva: $a > 0$.

La ecuación de ondas (2.5.1) incorpora un término disipativo. La manera más natural de interpretar el efecto del término añadido au_t es reescribir la ecuación como

$$u_{tt} - \Delta u = -a u_t. \tag{2.5.2}$$

En esta expresión se observa que $-au_t$ representa una fuerza que actúa en el dominio Ω en cada instante de tiempo. La fuerza aplicada es proporcional a la velocidad u_t, con una constante de proporcionalidad a que supondremos positiva. Por último, se observa que la fuerza aplicada es de signo contrario a la velocidad de modo que cuando $u_t > 0$ (resp. $u_t < 0$) la fuerza aplicada es negativa (resp. positiva).

Para representar las soluciones en series de Fourier desarrollamos en primer lugar los datos iniciales:

$$u_0(x) = \sum_{k=1}^{\infty} a_k \varphi_k(x),\ u_1(x) = \sum_{k=1}^{\infty} b_k \varphi_k(x),\ x \in \Omega. \tag{2.5.3}$$

Aquí y en lo que sigue $\{\varphi_k\}_{k\geqslant 1}$:

$$-\Delta \varphi_k = \lambda_k \varphi_k \text{ en } \Omega;\ \varphi_k = 0 \text{ en } \partial\Omega. \tag{2.5.4}$$

Buscamos entonces una solución de (2.5.1) de la forma

$$u(x,t) = \sum_{k=1}^{\infty} u_k(t) \varphi_k(x) \tag{2.5.5}$$

con $u_k = u_k(t)$ solución de

$$\begin{cases} u_k'' + \lambda_k u_k + a u_k' = 0, & t > 0 \\ u_k(0) = a_k,\ u_k'(0) = b_k. \end{cases} \tag{2.5.6}$$

La solución de (2.5.6) puede calcularse explícitamente. Para ello basta calcular las raíces del polinomio característico

$$\mu^2 + \lambda_k + a\mu = 0, \tag{2.5.7}$$

que vienen dadas por

$$\mu_\pm = \frac{-a \pm \sqrt{a^2 - 4\lambda_k}}{2}. \tag{2.5.8}$$

La solución de (2.5.6) es de la forma

$$u_k(t) = \alpha_+ e^{\frac{-a+\sqrt{a^2-4\lambda_k}}{2}t} + \alpha_- e^{\frac{-a-\sqrt{a^2-4\lambda_k}}{2}t} \tag{2.5.9}$$

donde las constantes α_- y α_+ son tales que los datos iniciales de (2.5.6) se verifican, i.e.

$$\begin{cases} \alpha_+ + \alpha_- = a_k \\ \frac{-a+\sqrt{a^2-4\lambda_k}}{2}\alpha_+ - \frac{\left(a+\sqrt{a^2-4\lambda_k}\right)}{2}\alpha_- = b_k. \end{cases} \tag{2.5.10}$$

Esto es así cuando

$$a^2 \neq 4\lambda_k. \tag{2.5.11}$$

En caso contrario, cuando $a^2 = 4\lambda_k$, la solución es de la forma

$$u_k(t) = \alpha e^{-at/2} + \beta t e^{-at/2}, \tag{2.5.12}$$

donde las constantes α y β son tales que se verifican los datos iniciales:

$$\alpha = a_k,\ -\frac{a}{2}\alpha + \beta = b_k. \tag{2.5.13}$$

A partir de estas expresiones es fácil deducir cuál es el comportamiento cualitativo de las soluciones (2.5.5) de (2.5.1). Para ello es conveniente analizar la evolución temporal de la energía:

$$E(t) = \frac{1}{2}\int_\Omega \Big[|\nabla u(x,t)|^2 + |u_t(x,t)|^2\Big] dx. \tag{2.5.14}$$

Es fácil comprobar que la energía E es decreciente. En efecto, multiplicando por u_t en la ecuación (2.5.1) obtenemos la fórmula de disipación de la energía:

$$\frac{dE}{dt}(t) = -a\int_\Omega |u_t(x,t)|^2 dx. \tag{2.5.15}$$

De esta identidad deducimos que, efectivamente, la energía decrece en el tiempo.

Pero la identidad (2.5.15) en sí misma no proporciona información precisa sobre el modo en que las soluciones decrecen. Este análisis exige la utilización de las series de Fourier.

Recordemos que, por las propiedades de ortogonalidad de las autofunciones $\{\varphi_k\}_{k\geqslant 1}$ en $L^2(\Omega)$ y $H_0^1(\Omega)$, tenemos

$$E(t) = \frac{1}{2}\sum_{k=1}^{\infty}\Big[|u_k'(t)|^2 + \lambda_k|u_k(t)|^2\Big]. \tag{2.5.16}$$

Introducimos la notación

$$e_k(t) = \frac{1}{2}\Big[|u_k'(t)|^2 + \lambda_k|u_k(t)|^2\Big] \tag{2.5.17}$$

para la energía de cada una de las componentes de Fourier.

En efecto:

$$E(t) = \sum_{k=1}^{\infty} e_k(t) = \sum_{k=1}^{\infty}\frac{1}{2}\Big[|u_k'(t)|^2 + \lambda_k|u_k(t)|^2\Big]. \tag{2.5.18}$$

Ahora bien, en el caso genérico en el que (2.5.9) se cumple (obsérvese que $\{\lambda_k\}_{k\geqslant 1}$ es un conjunto numerable y que, por tanto, para casi todo $a > 0$ la condición (2.5.11) se cumple) de la expresión (2.5.9) deducimos que $e_k(t)$ es una función con un decaimiento exponencial que satisface

$$e_k(t) \leqslant Ce_k(0)e^{-\omega_k t} \tag{2.5.19}$$

donde C es una constante positiva independiente de k y de la solución y ω_k es la tasa exponencial de decaimiento de la $k-$ésima componente de Fourier que viene dada por

$$\omega_k = \begin{cases} \frac{a-\sqrt{a^2-4\lambda_k}}{2} & , \quad \text{cuando} \quad a^2 > 4\lambda_k \\ \\ \frac{a}{2} & , \quad \text{cuando} \quad a^2 < 4\lambda_k. \end{cases} \tag{2.5.20}$$

El caso crítico $a^2 = 4\lambda_k$ será considerado más adelante.

En el caso en que la condición (2.5.11) no se cumple tenemos un resultado ligeramente distinto

$$e_k(t) \leqslant Ce_k(0)te^{-\omega_k t}, \tag{2.5.21}$$

con

$$\omega_k = a/2. \tag{2.5.22}$$

Combinando estos resultados sobre el decaimiento de cada componente de Fourier y (2.5.18) deducimos que

$$E(t) \leqslant CE(0)e^{-\omega t}, \tag{2.5.23}$$

con

$$\omega = \omega(a) = \begin{cases} \frac{a}{2} & \text{si} \quad a^2 < 4\lambda_1, \\ \frac{a-\sqrt{a^2-4\lambda_1}}{2} & \text{si} \quad a^2 > 4\lambda_1, \end{cases} \tag{2.5.24}$$

teniendo en cuenta también que en el caso crítico en que

$$a^2 = 4\lambda_1, \tag{2.5.25}$$

tenemos un decaimiento ligeramente inferior

$$E(t) \leqslant CE(0)te^{-\frac{a}{2}t}. \tag{2.5.26}$$

En cualquier caso vemos que la función

$$a \to \omega(a) \tag{2.5.27}$$

que al potencial disipativo a le asocia la tasa exponencial de decaimiento de la energía de las soluciones tiene las siguientes propiedades:

- $\omega(a)$ crece linealmente para $a \in [0,\, 2\sqrt{\lambda_1}]$;
- $\omega(a)$ decrece cuando $a > 2\sqrt{\lambda_1}$;
- $\omega(a) \searrow 0$ cuando $a \nearrow \infty$;
- El máximo de $\omega(a)$ se alcanza cuando (2.5.25) se cumple, es decir, para el potencial disipativo

$$a = 2\sqrt{\lambda_1}. \tag{2.5.28}$$

En particular vemos que, contrariamente a lo que una primera intuición podría sugerir, la tasa de decaimiento de las soluciones, $\omega(a)$, no es una función monótona creciente del potencial disipativo a, ni tiende a infinito cuando $a \nearrow \infty$ sino que, la cantidad de disipación que el sistema (2.5.1) puede soportar se satura cuando se alcanza el valor crítico (2.5.28) del potencial disipativo y a partir de ese momento, para mayores valores de a, la tasa de decaimiento comienza a decrecer. Esto es lo que se conoce como fenómeno de sobredisipación ("overdamping"en inglés). A partir del valor (2.5.28) del potencial disipativo, el decaimiento empeora.

Pero ésto es así cuando se consideran globalmente todas las posibles soluciones de (2.5.1) o, lo que es lo mismo, se tienen en cuenta todas las componentes de Fourier de la solución. Las expresiones (2.5.9) y (2.5.20) muestran que, si se consideran únicamente las altas frecuencias de Fourier correspondientes a autovalores que satisfacen

$$4\lambda_k \geqslant a^2, \tag{2.5.29}$$

entonces la energía de las soluciones decrece con una tasa exponencial $a/2$.

Por tanto, a pesar de que de manera global el fenómeno de sobredisipación se produce, las altas frecuencias si que presentan un comportamiento más acorde con la intuición de modo que su tasa de decaimiento aumenta linealmente con el potencial disipativo a.

Este fenómeno de sobredisipación no ocurre en otros modelos más sencillos. Por ejemplo, si consideramos la ecuación del calor

$$\begin{cases} u_t - \Delta u + au = 0 & \text{en} \quad \Omega \times (0,\infty), \\ u = 0 & \text{en} \quad \partial\Omega \times (0,\infty) \\ u(x,0) = u_0(x) & \text{en} \quad \Omega, \end{cases} \tag{2.5.30}$$

utilizando su desarrollo en serie de Fourier es muy fácil probar que

$$\| u(t) \|_{L^2(\Omega)} \leqslant e^{-\frac{(\lambda_1+a)}{2}t} \| u_0 \|_{L^2(\Omega)}, \, \forall t > 0 \tag{2.5.31}$$

para toda solución y todo potencial disipativo a. Vemos pues que en este caso la tasa de decaimiento aumenta de manera lineal con el potencial disipativo.

Qué es lo que distingue la ecuación de ondas de la del calor y hace que en la primera se produzca un fenómeno de sobredisipación? La respuesta es sencilla: La ecuación de ondas es de orden dos en tiempo y sus genuinas incógnitas son u y u_t. Un solo potencial disipativo es incapaz de aumentar arbitrariamente la tasa de decaimiento.

Esto se pone claramente de manifiesto cuando escribimos la ecuación de ondas (2.5.1) en forma de sistema. Tenemos

$$\begin{cases} u_t = v \\ v_t = \Delta u - av. \end{cases} \tag{2.5.32}$$

En la segunda ecuación de (2.5.32) vemos que el potencial a disipa efectivamente la segunda componente $v = u_t$ del sistema. Uno podría pensar que de (2.5.32) se desprende que la primera componente u no se disipa. Pero esto no es así, ambas lo hacen a través del acoplamiento del sistema pero sin que se pueda evitar el fenómeno de sobredisipación.

El remedio parece entonces sencillo. Utilizamos dos potenciales distintos $a > 0$ y $b > 0$ que afecten tanto la componente u_t como u. Llegamos así al sistema:

$$\begin{cases} u_{tt} - \Delta u + a u_t + b u = 0 & \text{en} \quad \Omega \times (0, \infty) \\ u = 0 & \text{en} \quad \partial\Omega \times (0, \infty) \\ u(0) = u_0,\ u_t(0) = u_1 & \text{en} \quad \Omega. \end{cases} \tag{2.5.33}$$

En este caso la energía del sistema es

$$E_b(t) = \frac{1}{2} \int_\Omega \Big[|u_t(x,t)|^2 + |\nabla u(x,t)|^2 + b u^2(x,t) \Big] dx \tag{2.5.34}$$

y satisface

$$\frac{dE_b}{dt}(t) = -a \int_\Omega u_t^2(x,t) dx. \tag{2.5.35}$$

El análisis de Fourier proporciona una expresión de las soluciones de (2.5.33) de la forma (2.5.5) donde, ahora, cada coeficiente de Fourier es solución de

$$u_k'' + (\lambda_k + b) u + a u' = 0. \tag{2.5.36}$$

las raíces del polinomio característico son ahora

$$\mu_\pm^b = \frac{-a \pm \sqrt{a^2 - 4(\lambda_k + b)}}{2}. \tag{2.5.37}$$

De este modo vemos que, para cualquier $a > 0$, si tomamos $b > 0$ suficientemente grande de modo que

$$a^2 < 4(\lambda_1 + b) \tag{2.5.38}$$

cada componente de Fourier decae con una tasa exponencial

$$\omega_k^b(a) = -\frac{a}{2}.$$

Vemos pues que eligiendo b de acuerdo a (2.5.38) se puede garantizar que la energía E_b satisface

$$E_b(t) \leqslant C E_b(0) e^{-\frac{a}{2} t}$$

evitándose así el fenómeno de sobredisipación.

Un análisis análogo permite describir el modo en que el espectro de la ecuación de ondas se convierte en el del calor a lo largo de la familia uniparamétrica de ecuaciones:

$$\varepsilon u_{tt} - \Delta u + u_t = 0. \tag{2.5.39}$$

En efecto, se observa que, en el límite cuando $\varepsilon \to 0$, se obtiene la ecuación del calor:

$$u_t - \Delta u = 0. \tag{2.5.40}$$

Es interesante analizar cómo el espectro de la ecuación de ondas disipativa, esencialmente localizado a lo largo de una recta vertical del plano complejo se convierte en un espectro localizado en el semieje real negativo. El hecho de que el orden del sistema pase de ser dos a ser uno también queda de manifiesto en este proceso límite puesto que la mitad de los autovalores de la ecuación de ondas se desvanecen tendiendo a menos infinito.

2.6. Teoría de Semigrupos

En las secciones anteriores hemos descrito cómo la ecuación de ondas puede ser resuelta mediante series de Fourier. Sin embargo, tal y como señalamos, este método carece de la generalidad que desearíamos puesto que no permite analizar ecuaciones con coeficientes dependientes de (x,t), ecuaciones no-lineales, etc.

En esta sección vamos a indicar el modo en que la ecuación de ondas puede enmarcarse en el contexto de la teoría de semigrupos que se muestra mucho más flexible a la hora de abordar sus variantes.

Consideremos pues la ecuación de ondas

$$\left\{\begin{array}{ccc} u_{tt}-\Delta u=0 & \text{en} & \Omega\times(0,\infty) \\ u=0 & \text{en} & \partial\Omega\times(0,\infty) \\ u(x,0)=u_0,\, u_t(x,0)=u_1(x) & \text{en} & \Omega, \end{array}\right. \tag{2.6.1}$$

donde Ω es un abierto de $\mathbb{R}^n$, $n\geqslant 1$, que supondremos acotado para simplificar la presentación, si bien esta hipótesis no es en absoluto esencial.

Conviene escribir la ecuación de ondas como un sistema de orden uno:

$$\left\{\begin{array}{rcl} u_t &=& v \\ v_t &=& \Delta u. \end{array}\right. \tag{2.6.2}$$

De este modo la incógnita genuina del sistema es el par $U=(u,v)=(u,\,u_t)$, lo cual coincide con nuestra intuición según la cual la verdadera incógnita no es sólamente la *posición* u sino también la *velocidad* u_t. Por otra parte, esto explica que en (4.1) tomemos dos datos iniciales u_0 y u_1 para u y u_t respectivamente.

En la variable vectorial U el sistema (2.6.2) puede escribirse formalmente como[5]

$$U_t=AU \tag{2.6.3}$$

donde A es el operador lineal

$$A=\left(\begin{array}{cc} 0 & I \\ \Delta & 0 \end{array}\right), \tag{2.6.4}$$

siendo I el operador identidad y Δ el operador de Laplace.

Pero la escritura (2.6.3)-(2.6.4) es puramente formal. En efecto, como es bien sabido, en el marco de los espacios de Hilbert (o de Banach) de dimensión infinita, una definición rigurosa de operador exige no solamente que indiquemos el modo en que actúa sino también su dominio.

[5]En este punto abusamos de la notación, pues U se trataría del vector columna $\binom{u}{u_t}$ si bien, para simplificar la escritura a veces lo escribiremos como vector fila.

Como habíamos indicado anteriormente, el espacio natural para resolver la ecuación de ondas es el espacio de Hilbert

$$H = H_0^1(\Omega) \times L^2(\Omega). \tag{2.6.5}$$

La elección de este espacio es efectivamente natural en vista de los siguientes hechos:

- La energía

$$E(t) = \frac{1}{2}\int_\Omega \Big[\mid \nabla u(x,t) \mid^2 + \mid u_t(x,t) \mid^2 \Big] dx \tag{2.6.6}$$

  se conserva en tiempo, lo cual puede ser comprobado formalmente multiplicando la ecuación de ondas por u_t e integrando en Ω.

  La conservación de la energía sugiere que, efectivamente, es natural buscar soluciones tales que $u \in H^1(\Omega)$ y $u_t \in L^2(\Omega)$.

- La condición de contorno de Dirichlet, $u = 0$ en $\partial\Omega$, sugiere la necesidad de buscar soluciones que se anulen en la frontera. Es bien conocido que, en el marco del espacio de Sobolev H^1, la manera más natural de interpretar esta condición es exigir que $u \in H_0^1(\Omega)$.

El espacio de la energía H es *un espacio de Hilbert* dotado de la norma:

$$\| (f,g) \|_H = \Big[\Big|\Big| f \Big|\Big|^2_{H_0^1(\Omega)} + \Big|\Big| g \Big|\Big|^2_{L^2(\Omega)} \Big]^{1/2}. \tag{2.6.7}$$

Por otra parte, las normas $\| \cdot \|_{L^2(\Omega)}$, $\| \cdot \|_{H_0^1(\Omega)}$ están definidas de la manera usual[6]:

$$\Big|\Big| f \Big|\Big|_{H_0^1(\Omega)} = \Big[\int_\Omega \mid \nabla f \mid^2 dx \Big]^{1/2}; \; \Big|\Big| g \Big|\Big|_{L^2(\Omega)} = \Big[\int_\Omega g^2 dx \Big]^{1/2}. \tag{2.6.8}$$

Definimos el operador A como un operador lineal no-acotado en H. Para ello establecemos que el dominio del operador A es precisamente el subespacio de los elementos $V \in H$ para los que $AV \in H$. En vista de la estructura de A ésto da como resultado el dominio:

$$\begin{aligned} D(A) &= \big\{ (u,v) \in H_0^1(\Omega) \times L^2(\Omega) : \, v \in H_0^1(\Omega), \, \Delta u \in L^2(\Omega) \big\} \\ &= \big\{ (u,v) \in H_0^1(\Omega) \times H_0^1(\Omega) : \, \Delta u \in L^2(\Omega) \big\}. \end{aligned} \tag{2.6.9}$$

[6] En este punto utilizamos implícitamente el hecho que Ω sea *acotado*. En efecto, si no lo fuese (o si, de manera más general, si Ω no fuese acotado en una dirección) no se podría garantizar que la desigualdad de Poincaré se verifica, lo cual a su vez no permitiría garantizar que la norma definida en (2.6.8) fuese equivalente a la inducida por $H^1(\Omega)$ sobre el subespacio $H_0^1(\Omega)$.

Cuando el dominio Ω es de clase C^2 el resultado clásico de regularidad elíptica que garantiza que las funciones $u \in H_0^1(\Omega)$ tales que $\Delta u \in L^2(\Omega)$ pertenecen en realidad a $H^2(\Omega)$, permite reescribir el dominio de la manera siguiente

$$D(A) = \Big[H^2(\Omega) \cap H_0^1(\Omega)\Big] \times H_0^1(\Omega). \tag{2.6.10}$$

En este punto conviene subrayar que la hipótesis de que Ω sea regular de clase C^2 no es en absoluto esencial. Todo lo que vamos a decir en lo sucesivo identificando el dominio con (2.6.10) es también cierto, sin la hipótesis de regularidad del abierto Ω, tomando (2.6.9) como definición del dominio del operador.

En lo sucesivo supondremos por tanto que Ω, además de ser acotado, es de clase C^2.

Es fácil comprobar que A es un operador anti-adjunto, i.e.

$$A^* = -A. \tag{2.6.11}$$

Basta para ello utilizar el hecho de que el operador de Laplace A con dominio $H^2(\Omega) \cap H_0^1(\Omega)$ en el espacio de Hilbert $L^2(\Omega)$ es un operador autoadjunto.

Pero, de hecho, para comprobar la antisimetría que (2.6.11) indica basta con realizar el siguiente cálculo elemental:

$$\begin{aligned} & \Big(AU,\, \widetilde{U}\Big)_H = \Big(v,\, \widetilde{u}\Big)_{H_0^1(\Omega)} + \Big(\Delta u,\, \widetilde{v}\Big)_{L^2(\Omega)} \\ = & \int_\Omega \Big[\nabla v \cdot \nabla \widetilde{u} + \Delta u \widetilde{v}\Big] dx = -\int_\Omega \Big[v \Delta \widetilde{u} + \nabla u \cdot \nabla \widetilde{v}\Big] dx \\ = & -\Big(U,\, A\widetilde{U}\Big)_H \end{aligned} \tag{2.6.12}$$

para todo U, $\widetilde{U} \in D(A)$.

En (2.6.12) y en lo sucesivo mediante $(\cdot,\cdot)_H$ denotamos el producto escalar en H. En vista de la estructura de H como espacio producto, el producto escalar en H es la suma de los productos escalares en $H_0^1(\Omega)$ y $L^2(\Omega)$ de las primeras y segundas componentes de vector V.

Con esta definición del operador A podemos ahora escribir la ecuación de ondas (2.6.1) en la forma del siguiente problema de Cauchy abstracto

$$\begin{cases} U_t = AU, \quad t > 0 \\ U(0) = U_0, \end{cases} \tag{2.6.13}$$

donde el dato inicial U_0 es, evidentemente, el vector columna $(u_0,\, u_1)$ de los datos iniciales de (2.6.1).

Tenemos dos tipos de soluciones de (2.6.13). Aquéllas que denominaremos *soluciones fuertes* tales que[7]

$$U \in C\Big([0,\infty);\, D(A)\Big) \cap C^1([0,\infty);\, H). \tag{2.6.14}$$

En este caso tanto el término de la izquierda como de la derecha de (2.6.13) son funciones bien definidas que pertenecen al espacio $C([0,\infty);\, H)$ y, por tanto, la ecuación de (2.6.13) tiene sentido en el espacio H para todo valor de $t > 0$. La segunda ecuación de (2.6.13) relativa al dato inicial tiene también sentido pues, por la continuidad de U en tiempo a valores en $D(A)$, $U(0)$ está bien definida en $D(A)$. Es por este hecho precisamente que sólo cabe esperar la existencia de soluciones fuertes cuando el dato inicial U_0 de (2.6.13) pertenece a $D(A)$.

En términos de la posición u y velocidad u_t de la solución de la ecuación de ondas (2.6.1), la regularidad (2.6.14) equivale a

$$u \in C\Big([0,\infty);\, H^2 \cap H_0^1(\Omega)\Big) \cap C^1\Big([0,\infty);\, H_0^1(\Omega)\Big) \cap C^2\Big([0,\infty);\, L^2(\Omega)\Big). \tag{2.6.15}$$

Es también claro que (2.6.15) permite dar un sentido a todas las ecuaciones de (2.6.1). En particular, la ecuación de ondas se verifica, para cada $t > 0$, en $L^2(\Omega)$ y, por tanto, en particular, para casi todo $x \in \Omega$.

Las *soluciones débiles* de (2.6.13) son menos regulares. Son en realidad aquéllas que pertenecen al espacio de la energía, i.e.

$$U \in C([0,\infty);\, H) \tag{2.6.16}$$

o bien

$$u \in C\Big([0,\infty);\, H_0^1(\Omega)\Big) \cap C^1\Big([0,\infty);\, L^2(\Omega)\Big). \tag{2.6.17}$$

Cabe preguntarse por el sentido de (2.6.13) bajo las condiciones de regularidad (2.6.16). En efecto, este sentido no está a priori claro pues (2.6.16) no permite definir, en principio, AU, al no pertenecer U a $D(A)$ ni permite calcular la derivada temporal de U.

A pesar de ello, tiene efectivamente sentido hablar de soluciones débiles de (2.6.1) o (2.6.13) y esto se puede ver con más claridad en el contexto de (2.6.1) y bajo la condición de regularidad (2.6.17). En efecto, es bien sabido que el operador $-\Delta$ define un isomorfismo de $H_0^1(\Omega)$ en su dual $H^{-1}(\Omega)$. Por tanto, como $u \in C\Big([0,\infty);\, H_0^1(\Omega)\Big)$, tenemos también que $\Delta u \in C\Big([0,\infty);\, H^{-1}(\Omega)\Big)$. Por otra parte, como $u \in C\Big([0,\infty);\, H_0^1(\Omega)\Big)$, se trata en particular de una

[7] El dominio $D(A)$ de un operador se puede dotar de estructura Hilbertiana a través de la norma $||u||_{D(A)} = [||u||_H^2 + ||Au||_H^2]^{1/2}$.

distribución por lo que su derivada segunda temporal u_{tt} está bien definida en el espacio de las distribuciones $\mathcal{D}'(\Omega \times (0,\infty))$. La ecuación de ondas (2.6.1) tiene por tanto sentido en el marco de las distribuciones. Ahora bien, como $\Delta u \in C\Big([0,\infty);\, H^{-1}(\Omega)\Big)$, de la propia ecuación de ondas deducimos que $u_{tt} \in C\Big([0,\infty);\, H^{-1}(\Omega)\Big)$ y entonces la ecuación de ondas tiene sentido, para todo $t > 0$, en $H^{-1}(\Omega)$. Vemos por tanto que las soluciones débiles de la ecuación de ondas, en la clase (2.6.17), por ser soluciones de la ecuación de ondas, tienen la propiedad de regularidad adicional

$$u \in C^2\Big([0,\infty);\, H^{-1}(\Omega)\Big). \tag{2.6.18}$$

La teoría de semi-grupos garantiza la existencia y unicidad de soluciones de (2.6.13) (y por tanto de la ecuación de ondas original) tanto fuertes como débiles. Basta para ello aplicar el Teorema de Hille-Yosida en su versión más elemental (véase, por ejemplo, el capítulo VII del libro [2]).

Con el objeto de enunciar este importante Teorema conviene recordar la noción de operador maximal disipativo.

Definition 2.6.1 *Un operador $A : D(A) \subset H \to H$ lineal, no acotado, en el espacio de Hilbert H se dice disipativo si*

$$(Av,\, v)_H \leqslant 0,\ \forall v \in D(A). \tag{2.6.19}$$

Se dice que es maximal-disipativo si, además, satisface la siguiente condición de maximalidad:

$$R(I - A) = H \Leftrightarrow \forall f \in H,\ \exists u \in D(A)\ \ t.q.\ u - Au = f. \tag{2.6.20}$$

Bajo esta condición se verifica el siguiente importante Teorema:

Theorem 2.6.1 *(de Hille-Yosida)*

Sea A un operador maximal-disipativo en un espacio de Hilbert H. Entonces, para todo $u_0 \in D(A)$ existe una función

$$u \in C\Big([0,\infty);\, D(A)\Big) \cap C^1\Big([0,\infty);\, H\Big) \tag{2.6.21}$$

única tal que

$$\begin{cases} \frac{du}{dt} = Au \quad \ en \quad [0,\infty) \\ u(0) = u_0. \end{cases} \tag{2.6.22}$$

Además se tiene

$$\| u(t) \| \leqslant \| u_0 \|_H,\ \left|\left| \frac{du}{dt}(t) \right|\right|_H = \| Au(t) \|_H \leqslant \| Au_0 \|_H,\ \forall t > 0. \tag{2.6.23}$$

Es fácil comprobar que el operador A asociado a la ecuación de ondas (2.6.1) definido anteriormente es maximal disipativo. El hecho de que A sea anti-adjunto ($A^* = -A$) garantiza que tanto A como $-A$ son disipativos en el sentido de la Definición 2.6.1[8].

En efecto, de (2.6.12) deducimos que

$$(AU,\, U)_H = -(AU,\, U)_H$$

y por tanto

$$(AU,\, U)_H = 0 \tag{2.6.24}$$

lo cual garantiza la disipatividad de A y $-A$.[9]

Por otra parte, el operador A de la ecuación de ondas verifica también la condición de maximalidad (2.6.20). En efecto, para comprobarlo basta ver que para todo par $(f, g) \in H$, *i.e.* $f \in H_0^1(\Omega)$, $g \in L^2(\Omega)$, existe al menos una solución de la ecuación

$$(I - A)\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} f \\ g \end{pmatrix} \tag{2.6.25}$$

con $(u, v) \in D(A) = \left[H^2 \cap H_0^1(\Omega)\right] \times H_0^1(\Omega)$. Esto es efectivamente cierto. Dada la forma explícita del operador A el sistema (2.6.25) se escribe del siguiente modo

$$u - v = f,\; v - \Delta u = g. \tag{2.6.26}$$

La primera ecuación de (2.6.26) puede reescribirse como

$$v = u - f \tag{2.6.27}$$

y entonces la segunda adquiere la forma

$$u - \Delta u = g + f. \tag{2.6.28}$$

[8]En el contexto de los sistemas de la Mecánica la palabra "disipativo" tiene un sentido preciso: Se dice que un sistema de evolución es disipativo si la energía de las soluciones decrece en tiempo. Es este precisamente el sentido del término en el marco abstracto en la Teoría de Operadores que se desprende de (2.6.19). En efecto, multiplicando escalarmente en H la primera ecuación (2.6.22) por u, en virtud de (2.6.19) deducimos que $\left(\frac{d}{dt}u(t),\, u(t)\right)_H = \frac{1}{2}\frac{d}{dt} \parallel u(t) \parallel_H^2 =< Au(t),\, u(t),\, u(t)) >\leqslant 0$.

[9]Conviene observar que cuando A satisface (2.6.24) las soluciones de la ecuación abstracta

$$\frac{du}{dt}(t) = Au(t)$$

conservan la energía puesto que

$$\frac{1}{2}\frac{d}{dt}\Big|\Big|u(t)\Big|\Big|_H^2 = (Au(t),\, u(t)) = 0.$$

Como $g + f \in L^2(\Omega)$, la segunda ecuación (2.6.28), que puede escribirse de manera más precisa como

$$\begin{cases} u - \Delta u = f + g & \text{en} \quad \Omega \\ u = 0 & \text{en} \quad \partial\Omega, \end{cases} \tag{2.6.29}$$

admite una única solución $u \in H^2 \cap H_0^1(\Omega)$ por los resultados clásicos de existencia, unicidad y regularidad para el problema de Dirichlet. Como $f \in H_0^1(\Omega)$ y $u \in H^2 \cap H_0^1(\Omega)$ la solución v de (2.6.27) satisface entonces $v \in H_0^1(\Omega)$. Deducimos entonces que (2.6.25) admite una única solución en $D(A)$, lo cual garantiza la maximalidad de A.

El Teorema 2.6.1, aplicado a la versión abstracta (2.6.13) de la ecuación de ondas (2.6.1) proporciona de manera inmediata la existencia y unicidad de soluciones fuertes. En efecto, se tiene:

Theorem 2.6.2 *Si Ω es un dominio acotado de clase C^2, para cada par de datos iniciales $(u_0, u_1) \in \Big[H^2 \cap H_0^1(\Omega)\Big] \times H_0^1(\Omega)$, la ecuación de ondas posee una única solución fuerte en la clase*

$$u \in C\Big([0,\infty);\, H^2 \cap H_0^1(\Omega)\Big) \cap C^1\Big([0,\infty);\, H_0^1(\Omega)\Big) \cap C^2\Big([0,\infty);\, L^2(\Omega)\Big). \tag{2.6.30}$$

Sólo nos resta deducir la existencia y unicidad de soluciones en el espacio de la energía. Tenemos para ello varias opciones. Una de ellas consiste en analizar el operador de ondas como operador no acotado en el espacio de Hilbert $\widetilde{H} = L^2(\Omega) \times H^{-1}(\Omega)$ con dominio $H \subset \widetilde{H}$. Es fácil comprobar que el operador A antes definido es también un operador maximal disipativo en este marco funcional. De este modo, como consecuencia del Teorema de Hille-Yosida deducimos que:

Theorem 2.6.3 *En las hipótesis del Teorema 2.6.2, si los datos iniciales $(u_0, u_1) \in H_0^1(\Omega) \times L^2(\Omega)$ la ecuación de ondas* (2.6.1) *admite una única solución en*

$$u \in C\Big([0,\infty); H_0^1(\Omega)\Big) \cap C^1\Big[0,\infty);\, L^2(\Omega)\Big) \cap C^2\Big([0,\infty);\, H^{-1}(\Omega)\Big). \tag{2.6.31}$$

Conviene observar que ambos teoremas de existencia y unicidad (Teoremas 2.6.2 y 2.6.3) proporcionan resultados semejantes pero en espacios que difieren en una derivada en su regularidad.

La posibilidad de obtener soluciones débiles a partir de soluciones fuertes puede también explicarse en el marco del problema abstracto (2.6.22). En efecto, suponiendo que A es un operador maximal disipativo, consideremos el problema abstracto y definamos la función

$$v(t) = \int_0^t u(s)ds + v_0. \tag{2.6.32}$$

Integrando a su vez la ecuación de (2.6.22) con respecto al tiempo obtenemos

$$u(t) - u_0 = A \int_0^t u ds$$

que podemos reescribir de la siguiente manera:

$$u(t) = A \int_0^t u ds + u_0 \Leftrightarrow v_t = Av - Av_0 + u_0.$$

Por lo tanto, para poder garantizar que también v es una solución del problema abstracto (2.6.22) basta con elegir v_0 de modo que

$$Av_0 = u_0. \tag{2.6.33}$$

Supongamos que, dado $u_0 \in H$, (2.6.33) admite una única solución $v_0 \in D(A)$. Entonces la función v definida en (2.6.32) satisface

$$\begin{cases} v_t = Av, \quad t > 0 \\ v(0) = v_0. \end{cases} \tag{2.6.34}$$

En virtud del Teorema de Hille-Yosida, como $v_0 \in D(A)$, la ecuación (2.6.34) admite una única solución fuerte

$$v \in C\Big([0,\infty);\, D(A)\Big) \cap C^1\Big([0,\infty);\, H\Big). \tag{2.6.35}$$

De (2.6.35) deducimos que

$$u = v_t \in C\Big([0,\infty);\, H\Big). \tag{2.6.36}$$

Vemos de este modo que, cuando $u_0 \in H$, la ecuación abstracta admite una única solución débil en la clase (2.6.36).

Comentemos brevemente la ecuación (2.6.33). En la definición de operador maximal disipativo se garantiza que $I - A$ es un operador con rango pleno. Pero nada se dice del operador A. Conviene sin embargo señalar que ésto es irrelevante a la hora de resolver el problema abstracto (2.6.22). En efecto, introduzcamos el clásico cambio de variables

$$w(t) = e^{\lambda t} u(t), \tag{2.6.37}$$

donde $\lambda \in \mathbb{R}$.

Entonces $w_t = e^{\lambda t}\left[u_t + \lambda u\right]$. Por tanto, u es solución de (2.6.22) si y sólo si w es solución de

$$\begin{cases} w_t = Aw + \lambda w, \quad t > 0 \\ w(0) = u_0 \end{cases} \tag{2.6.38}$$

Esto indica que el cambio de variable permite transformar soluciones fuertes (resp. débiles) de (2.6.22) en soluciones fuertes (resp. débiles) de (2.6.38) y viceversa.

Por otra parte, cuando A es maximal-disipativo, para $\lambda = -1$, el operador $A - I$ de (2.6.38) es de rango pleno. Esto permite utilizar el argumento anterior de integración en tiempo para obtener soluciones débiles a partir de las soluciones fuertes directamente en (2.6.38) cuando $\lambda = -1$ (porque el problema correspondiente a (2.6.33) podría efectivamente garantizarse que tiene una única solución $v_0 \in D(A)$ para cada $u_0 \in H$).

En el caso de la ecuación de ondas, (2.6.33) puede resolverse directamente sin apelar al cambio de variables. En efecto, en este caso, el problema (2.6.33) puede reescribirse de la siguiente manera: Dado $(u_0,\, u_1) \in H_0^1(\Omega) \times L^2(\Omega)$ hallar $(v_0,\, v_1) \in \left[H^2 \cap H_0^1(\Omega)\right] \times H_0^1(\Omega)$ tal que

$$v_1 = u_0;\ \Delta v_0 = u_1. \tag{2.6.39}$$

La primera ecuación de (2.6.39) proporciona inmediatamente la solución $v_1 \in H_0^1(\Omega)$. Por otra parte, como $u_1 \in L^2(\Omega)$, sabemos que el problema elíptico

$$-\Delta v_0 = -u_1 \text{ en } \Omega;\ v_0 = 0 \text{ en } \partial\Omega, \tag{2.6.40}$$

admite una única solución $v_0 \in H^2 \cap H_0^1(\Omega)$.

Por lo tanto, en el marco de la ecuación de ondas, (2.6.33) admite una única solución, la cual permite obtener soluciones débiles de la ecuación de ondas a partir de las soluciones fuertes, a través del cambio de variable (2.6.32).

Como ya hemos indicado anteriormente, en el marco de la ecuación de ondas, el operador de ondas A es antiadjunto, y ésto equivale a la ley de conservación de energía (2.6.6). Vemos por tanto cómo la Teoría de semigrupos permite recuperar todos los resultados obtenidos mediante series de Fourier, pero con la ventaja de ofrecer un marco mucho más flexible para abordar otras ecuaciones.

Hemos visto que los resultados clásicos de la Teoría de semigrupos permiten construir soluciones fuertes para datos iniciales en $D(A) = H^2 \cap H_0^1(\Omega) \times H_0^1(\Omega)$ y soluciones débiles para datos en $H = H_0^1(\Omega) \times L^2(\Omega)$. Pero estos no son más que dos de los posibles ejemplos de marcos funcionales en los que la ecuación de ondas está bien puesta. Otro ejemplo interesante es el de las soluciones ultradébiles con datos iniciales $(u_0, u_1) \in L^2(\Omega) \times H^{-1}(\Omega)$. En este caso el espacio natural para las soluciones es $C([0,T]; L^2(\Omega)) \times C^1([0,T]; H^{-1}(\Omega))$. Los argumentos anteriores permiten probar de dos maneras distintas este resultado de existencia y unicidad de soluciones ultradébiles. En efecto:

- El cambio de variables (2.6.32) establece una relación biunívoca entre soluciones ultradébiles u y soluciones débiles v. Como corolario del Teorema 6.3, mediante este cambio de variable, se deduce la existencia y unicidad de soluciones ultradébiles.

- El teorema de Hille-Yosida puede también aplicarse directamente en este marco funcional para obtener la existencia y unicidad de soluciones ultradébiles. Basta para ello considerar el operador A en el espacio $H^{-1}(\Omega) \times [H^2 \cap H_0^1(\Omega)]'$ con dominio $L^2(\Omega) \times H^{-1}(\Omega) \subset H^{-1}(\Omega) \times [H^2 \cap H_0^1(\Omega)]'$. Las soluciones que el Teorema de Hille-Yosida proporciona pertenecen entonces a la clase

$$u \in C([0,T]; L^2(\Omega)) \cap C^1([0,T]; H^{-1}(\Omega)) \cap C^2([0,T]; [H^2 \cap H_0^1(\Omega)]'). \tag{2.6.41}$$

Observación. Los diferentes marcos funcionales y grados de regularidad de las diversas soluciones que hemos construido y considerado pueden también entenderse en el marco de la representación de las soluciones en series de Fourier. Como ya hemos mencionado anteriormente, la Teoría de semigrupos permite sin embargo construir estas soluciones para una familia mucho más amplia de ecuaciones.

Por ejemplo, si desarrollamos las soluciones de la ecuación de ondas como en (2.3.25) las soluciones débiles de energía finita corresponden a coeficientes de Fourier tales que:

$$\sum_{k=1}^{\infty} [\lambda_k |a_k|^2 + |b_k|^2] < \infty. \tag{2.6.42}$$

Las soluciones fuertes exigen sin embargo condiciones más fuertes sobre los coeficientes de Fourier:

$$\sum_{k=1}^{\infty} [\lambda_k^2 |a_k|^2 + \lambda_k |b_k|^2] < \infty. \tag{2.6.43}$$

Por último, las soluciones ultradébiles exigen simplemente que

$$\sum_{k=1}^{\infty} [|a_k|^2 + \lambda_k^{-1} |b_k|^2] < \infty. \tag{2.6.44}$$

■

En virtud del Teorema de Hille-Yosida (Teorema 2.6.1), cuando A es un operador maximal disipativo, es el *generador de un semigrupo* $S(t) : H \to H$ que a cada

$u_0 \in H$ asocia el valor $u(t) = S(t)u_0$ de la solución del problema abstracto (2.6.22) en cada instante de tiempo $t > 0$.

El semigrupo $S(t)$ también se denota habitualmente como e^{At}, en vista de la analogía del sistema abstracto (2.6.22) con el clásico sistema lineal de ecuaciones diferenciales lineales con coeficientes constantes en el que A es una matriz.

El semigrupo $\{S(t)\}_{t\geqslant 0} = \{e^{At}\}_{t\geqslant 0}$ es una familia uniparamétrica de operadores lineales acotados en H. En realidad, en virtud de (2.6.23), $S(t)$ es una contracción para cada $t \geqslant 0$. Por otra parte, el semigrupo verifica las siguientes propiedades:

- $S(0) = I$,
- $t \to S(t)u_0$ es continua de $[0, \infty)$ en H para cada $u_0 \in H$,
- $S(t) \circ S(s) = S(t + s)$.

La última propiedad, denominada propiedad de semigrupo, es debida al carácter autónomo (o invariante por traslaciones temporales) de la ecuación (2.6.26).

Consideramos por último la ecuación de ondas no-homogénea:

$$\begin{cases} u_{tt} - \Delta u = f & \text{en} \quad \Omega \times (0, \infty) \\ u = 0 & \text{en} \quad \partial\Omega \times (0, \infty) \\ u(0) = u_0,\ u_t(0) = u_1 & \text{en} \quad \Omega. \end{cases} \tag{2.6.45}$$

En este caso (2.6.45) describe las vibraciones del cuerpo Ω sometido a una fuerza exterior $f = f(x, t)$.

El problema (2.6.45) también puede ser escrito en el marco de los problemas abstractos que se pueden abordar en el contexto de la Teoría de Semigrupos. En efecto, la primera ecuación de (2.6.45) puede escribirse como el sistema

$$\begin{cases} u_t &= v \\ v_t &= \Delta u + f, \end{cases} \tag{2.6.46}$$

que puede también reformularse como el problema abstracto

$$\begin{cases} U_t = AU + F, \quad t > 0 \\ U(0) = U_0 \end{cases} \tag{2.6.47}$$

donde $U = (u,\, u_t)$, A es el generador del semigrupo de la ecuación de ondas que acabamos de estudiar y

$$F(t) = \begin{pmatrix} 0 \\ f(t) \end{pmatrix}. \tag{2.6.48}$$

Vemos por tanto que la fuerza externa F aplicada en la versión abstracta (2.6.47) del sistema (2.6.45) tiene una primera componente nula mientras que la función f de (2.6.45) interviene sólo en su segunda componente.

Inspirándonos en la fórmula de variación de las constantes para la resolución de ecuaciones diferenciales no homogéneas, el problema abstracto (2.6.47) puede escribirse en la forma integral siguiente

$$U(t) = S(t)U_0 + \int_0^t S(t-s)F(s)ds = e^{At}U_0 + \int_0^t e^{A(t-s)}F(s)ds, \quad (2.6.49)$$

siendo $S(t) = e^{At}$ el semigrupo generado por el operador maximal disipativo A.

En virtud de los resultados anteriores sobre las soluciones fuertes y débiles del sistema abstracto (2.6.22) asociado al operador A, es fácil deducir que:

- Si $F \in L^2(0,T;\, D(A))$, entonces $e^{A(t-s)}F(s) \in L^1(0,t;\, D(A))$.

  Basta para ello utilizar las estimaciones (2.6.23) que, con las notaciones presentes, garantizan que

  $$\left|\left|e^{A(t-s)}F(s)\right|\right|_H \leqslant \left|\left|F(s)\right|\right|_H, \quad \left|\left|Ae^{A(t-s)}F(s)\right|\right|_H \leqslant \left|\left|AF(s)\right|\right|_H,$$

  para todo $t \geqslant s$ y casi todo $s \in [0,T]$.

  Deducimos entonces que

  $$\int_0^t e^{A(t-s)}F(s)ds \in C([0,T];\, D(A)).$$

  Sin embargo, para que podamos garantizar que se tiene una solución fuerte en la clase (2.6.21) es necesario también que

  $$\int_0^t e^{A(t-s)}F(s)ds \in C^1([0,T];\, H)$$

  para lo que es también necesario que $F \in C([0,T];\, H)$.

- Si $F \in L^1(0,\, T;\, H)$, entonces $e^{A(t-s)}F(s) \in L^1(0,\, t;\, H)$ para todo $0 \leqslant t \leqslant T$ y por tanto

  $$\int_0^t e^{A(t-s)}F(s)ds \in C([0,T];\, H).$$

De estos hechos deducimos los siguientes resultados de existencia y unicidad para el sistema abstracto no homogéneo (2.6.47):

- Si $U_0 \in D(A)$ y $F \in C([0,T];\, H) \cap L^1(0,T;\, D(A))$ entonces (2.6.47) admite una única solución fuerte en la clase

$$U \in C([0,T];\, D(A)) \cap C^1([0,T];\, H).$$

  El mismo resultado es válido bajo la hipótesis de que $F \in W^{1,1}(0,T;\, H)$.

- Si $U_0 \in H$ y $F \in L^1(0,T;\, H)$, entonces (2.6.47) admite una única solución débil. $U \in C([0,T];\, H)$.

Aplicando estos resultados a la ecuación de ondas no-homogénea (2.6.45) obtenemos los siguientes resultados de existencia y unicidad:

- Si $(u_0,\, u_1) \in H^2 \cap H_0^1(\Omega) \times H_0^1(\Omega)$ y $f \in C([0,T];\, L^2(\Omega)) \cap L^1(0,T;\, H_0^1(\Omega))$, entonces (2.6.45) admite una única solución fuerte u en la clase (2.6.30).

- Si $(u_0,\, u_1) \in H_0^1(\Omega) \times L^2(\Omega)$ y $f \in L^1(0,T;\, L^2(\Omega))$, entonces (2.6.45) admite una solución débil

$$u \in C([0,T];\, H_0^1(\Omega)) \cap C^1([0,T]; L^2(\Omega)). \tag{2.6.50}$$

En este punto conviene subrayar que, salvo que impongamos condiciones adicionales al segundo miembro f, no podemos garantizar que

$$u \in C^2([0,T];\, H^{-1}(\Omega)). \tag{2.6.51}$$

En efecto, como $u \in C([0,T];\, H_0^1(\Omega))$ y $-\Delta$ es un isomorfismo de $H_0^1(\Omega)$ en $H^{-1}(\Omega)$, tenemos $-\Delta u \in C([0,T];\, H^{-1}(\Omega))$. Por tanto, para que (2.6.51) pueda cumplirse, en vista de la ecuación $f = u_{tt} - \Delta u$, es imprescindible que $f \in C([0,T];\, H^{-1}(\Omega))$.

El cambio de variable (2.6.32) también puede ser aplicado en el marco de la ecuación abstracta (2.6.47) y permite nuevamente establecer una correspondencia biunívoca entre soluciones fuertes y débiles.

Las mismas técnicas que las desarrolladas en el caso homogéneo pueden ser también utilizadas en el no homogéneo. Esto nos permite, por ejemplo, construir soluciones ultradébiles de (2.6.45). De este modo obtenemos que si $(u_0, u_1) \in L^2(\Omega) \times H^{-1}(\Omega)$ y $F \in L^1(0,T;\, H^{-1}(\Omega))$, la ecuación (2.6.45) admite una única solución ultra-débil en la clase

$$u \in C([0,T];\, L^2(\Omega)) \cap C^1([0,T];\, H^{-1}(\Omega)).$$

Además, si $f \in C\Big([0,T];\, \Big[H^2 \cap H_0^1(\Omega)\Big]'\Big)$, esta solución pertenece a

$$u \in C^2\Big([0,T];\, \Big(H^2 \cap H_0^1(\Omega)\Big)'\Big).$$

Pero, hasta ahora, todos los resultados que hemos obtenido sobre la ecuación de ondas mediante técnicas de teoría de semigrupos, pueden también ser obtenidos mediante series de Fourier. Sin embargo, como habíamos mencionado anteriormente, la teoría de semigrupos es indispensable si deseamos abordar ecuaciones más generales con coeficientes variables dependientes de (x,t), no lineales, etc. Ilustramos este hecho analizando el ejemplo de una ecuación de ondas con un potencial $p = p(x,t) \in L^\infty(\Omega \times (0,T))$, i.e.

$$\begin{cases} u_{tt} - \Delta u + p(x,t)u = 0 & \text{en} \quad \Omega \times (0,T) \\ u = 0 & \text{en} \quad \partial\Omega \times (0,T) \\ u(x,0) = u_0(x),\ u_t(x,0) = u_1(x) & \text{en} \quad \Omega. \end{cases} \tag{2.6.52}$$

Nuevamente la ecuación (2.6.52) puede ser escrita en la forma de un sistema

$$\begin{cases} u_t = v \\ v_t = \Delta u - p(x,t)u, \end{cases} \tag{2.6.53}$$

o, en su versión abstracta,

$$U_t = AU + B(t)U \tag{2.6.54}$$

donde A es el operador maximal-disipativo asociado a la ecuación de ondas y $B(t) : H \to H$ es un operador lineal acotado dependiente del tiempo:

$$B(t)U = B(t)\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} 0 \\ -p(x,t)u \end{pmatrix} \tag{2.6.55}$$

la ecuación abstracta (2.6.54) puede escribirse como una ecuación integral

$$U(t) = e^{At}U_0 + \int_0^t e^{A(t-s)}B(s)U(s)ds. \tag{2.6.56}$$

Introduciendo la aplicación

$$[\phi(U)](t) = e^{At}U_0 + \int_0^t e^{A(t-s)}B(s)U(s)ds,\ 0 \leqslant t \leqslant T \tag{2.6.57}$$

la ecuación integral (2.6.56) puede también ser reescrita como un problema de punto fijo

$$U(t) = [\phi(U)](t),\ 0 \leqslant t \leqslant T \tag{2.6.58}$$

que puede ser resuelto mediante la aplicación del Teorema de punto fijo de Banach para aplicaciones contractivas.

En efecto, si utilizamos que $B(t)$ es un operador lineal y acotado de H en H, con una cota independiente de $0 \leqslant t \leqslant T$, es fácil comprobar que la

aplicación (2.6.57) constituye una contracción estricta en $C([0,\tau);\, H)$ para un τ suficientemente pequeño ($0 \leqslant \tau \leqslant T$). De este modo obtenemos una única solución $U \in C([0,\tau];\, H)$ que, mediante un argumento de continuación puede ser extendido a una solución global única $U \in C([0,T];\, H)$[10].

Aplicando este resultado abstracto en el caso de la ecuación de ondas (2.6.52) con potencial obtenemos inmediatamente que: *Si* $(u_0, u_1) \in H_0^1(\Omega) \times L^2(\Omega)$, *y* $p \in L^\infty(\Omega \times (0,T))$, *la ecuación de ondas con potencial (2.6.52) admite una única solución*

$$u \in C\Big([0,T];\, H_0^1(\Omega)\Big) \cap C^1\Big([0,T];\, L^2(\Omega)\Big).$$

En realidad la estructura (2.6.55) del operador permite debilitar la hipótesis sobre el potencial p para que el resultado anterior sea válido. En efecto, en la práctica es suficiente que el operador de multiplicación $u \to p(t)u$ envie de manera acotada $H_0^1(\Omega)$ en $L^2(\Omega)$. Si utilizamos las inclusiones de Sobolev es fácil comprobar que esto es así cuando:

- Si $n = 1$, $p \in L^\infty(0,T;\, L^2(\Omega))$;
- Si $n = 2$, $p \in L^\infty(0,T;\, L^r(\Omega))$, para algún $r > 2$;
- Si $n \geqslant 3$, $p \in L^\infty(0,T;\, L^n(\Omega))$.

Más aún, basta analizar con un poco más de cuidado la prueba del carácter contractivo de la aplicación Φ para observar que las hipótesis L^∞ en la variable t pueden ser debilitadas y sustituidas por hipótesis L^1. Así, el resultado anterior de existencia y unicidad de soluciones débiles para la ecuación de ondas con potencial (2.6.52) es cierto en cuanto el potencial p satisface las condiciones:

- $p \in L^1(0,T;\, L^2(\Omega))$, si $n = 1$.
- $p \in L^1(0,T;\, L^r(\Omega))$, con $r > 2$, si $n = 2$.
- $p \in L^1(0,T;\, L^n(\Omega))$, si $n \geqslant 3$.

Los mismos argumentos permiten obtener soluciones fuertes. Pero en este caso habremos de comprobar si el operador abstracto $B(t)$ envia $D(A)$ en $D(A)$. En el marco de la ecuación de ondas con potencial esto supone imponer hipótesis sobre el potencial $p = p(x,t)$ de modo que, para cada t, el operador de

[10] Esto es así puesto que la amplitud de $\tau > 0$ del intervalo temporal en el que podemos aplicar el Teorema de punto fijo a Φ para deducir la existencia local de soluciones, depende exclusivamente de la cota de la que dispongamos sobre la norma del operador B

multiplicación mediante $p(t)$ envie $H^2 \cap H_0^1(\Omega)$ en $H_0^1(\Omega)$ y que haga ésto de modo que la cota resultante pertenezca a $L^1(0,T)$. Esto, evidentemente, exige hipótesis adicionales sobre la regularidad del potencial p.

Estos argumentos permiten en realidad obtener resultados de existencia y unicidad tanto de soluciones fuertes como débiles para ecuaciones más generales con potenciales de la forma

$$u_{tt} - \Delta u + a(x,t) \cdot \nabla u + b(x,t)u_t + p(x,t)u = 0. \tag{2.6.59}$$

Consideremos ahora brevemente una ecuación de ondas semilineal

$$\begin{cases} u_{tt} - \Delta u = f(u) & \text{en} \quad \Omega \times (0,\infty) \\ u = 0 & \text{en} \quad \partial\Omega \times (0,\infty) \\ u(x,0) = u_0(x),\ u_t(x,0) = u_1(x) & \text{en} \quad \Omega. \end{cases} \tag{2.6.60}$$

En esta ocasión $f : \mathbb{R} \to \mathbb{R}$ es una función no lineal. Nuevamente, la ecuación (2.6.60) puede ser reescrita en la forma de un sistema

$$\begin{cases} u_t = v \\ v_t = \Delta u + f(u) \end{cases} \tag{2.6.61}$$

que, a su vez, puede ser enmarcado en un sistema semilineal abstracto

$$U_t = AU + F(U) \tag{2.6.62}$$

donde

$$F(U) = F\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} 0 \\ f(u) \end{pmatrix}. \tag{2.6.63}$$

El problema puede entonces ser reducido a la ecuación integral

$$U(t) = e^{At}U_0 + \int_0^t e^{A(t-s)}F(U(s))ds \tag{2.6.64}$$

que, a su vez, es equivalente al problema de punto fijo

$$U(t) = [\phi(U)](t), \tag{2.6.65}$$

para la función

$$[\phi(U)](t) = e^{At}U_0 + \int_0^t e^{A(t-s)}F(U(s))ds. \tag{2.6.66}$$

Sea $R = \| U_0 \|_H$ y B_{2R} la bola de radio $2R$ en H. Supongamos que la no-linealidad F envía H en H de modo que se trate de una función Lipschitziana sobre conjuntos acotados de H, es decir: para todo $k > 0$, existe $L_k > 0$ tal que

$$\begin{aligned} \| F(U_1) - F(U_2) \|_H &\leqslant L_k \| U_1 - U_2 \|_H \\ &\forall U_1, U_2 \in H : \| U_1 \|_H, \| U_2 \|_H \leqslant k. \end{aligned} \tag{2.6.67}$$

Bajo estas hipótesis es fácil comprobar que si $\tau > 0$ es suficientemente pequeño, Φ es una contracción estrictamente en $C([0,\tau];\, B_{2R})$. Esto permite deducir la existencia y unicidad de una solución local (en tiempo) de (2.6.64) en $C([0,\tau];\, B_{2R})$.

Veamos lo que la hipótesis (2.6.67) supone sobre la no-linealidad de la ecuación de ondas (2.6.60). En vista de la forma particular (2.6.63) de la no-linealidad del modelo abstracto correspondiente basta en realidad con comprobar que f envía $H_0^1(\Omega)$ en $L^2(\Omega)$ de manera Lipschitz sobre conjuntos acotados. Supongamos que la función f se comporta esencialmente como una potencia $p \geqslant 1$. Es decir supongamos que

$$\Big| f(x) - f(y) \Big| \leqslant C(1+ \mid x \mid^{p-1} + \mid y \mid^{p-1}) \mid x - y \mid, \, \forall x,\, y \in \mathbb{R} \tag{2.6.68}$$

para algún $p \geqslant 1$ y $C > 0$[11].

Necesitamos comprobar si para todo $k > 0$ existe $L_k > 0$ tal que

$$\begin{aligned} \| f(u_1) - f(u_2) \|_{L^2(\Omega)} &\leqslant L_k \, \| u_1 - u_2 \|_{H_0^1(\Omega)}, \\ &\forall u_1, u_2 \in H_0^1(\Omega) : \, \| u_1 \|_{H_0^1(\Omega)}, \| u_2 \|_{H_0^1(\Omega)} \leqslant k. \end{aligned} \tag{2.6.69}$$

En vista de la hipótesis (2.6.68) y usando las inclusiones de Sobolev es fácil comprobar que (2.6.69) se cumple bajo las siguientes restricciones sobre p:

$$\begin{cases} \bullet \text{ Para todo} \quad 1 \leqslant p < \infty, & \text{si} \quad n = 1, 2. \\ \bullet \text{ Para todo} \quad 1 \leqslant p \leqslant \frac{n}{n-2}, & \text{si} \quad n \geqslant 3. \end{cases} \tag{2.6.70}$$

Deducimos por tanto que: "Bajo estas condiciones sobre el exponente p, si la no-linealidad f satisface la condición de Lipschitz (2.6.68), para cada par de datos iniciales $(u_0, u_1) \in H_0^1(\Omega) \times L^2(\Omega)$ existe un $\tau > 0$ y una única solución $u \in C([0,\tau];\, H_0^1(\Omega)) \cap C^1([0,\tau];\, L^2(\Omega))$".

Una vez que la solución local en tiempo ha sido obtenida, mediante los mismos argumentos de prolongación que se utilizan en el marco de las Ecuaciones Diferenciales Ordinarias (EDO), esta solución local puede ser prolongada al máximo intervalo de existencia $T_{\text{máx}}$ de modo que la solución única de (2.6.60) se obtiene finalmente en la clase

$$C([0, T_{\text{máx}});\, H_0^1(\Omega)) \cap C^1([0, T_{\text{máx}});\, L^2(\Omega)).$$

Además, para el tiempo máximo de existencia se verifica la siguiente alternativa: O bien $T_{\text{máx}} = \infty$ (*existencia global*) y por lo tanto la solución está definida para

[11] Esta hipótesis se cumple, por ejemplo, si $f \in C^1(\mathbb{R};\mathbb{R})$ y $\limsup_{|x|\to\infty} \dfrac{\mid f'(x) \mid}{\mid x \mid^{p-1}} < \infty$.

todo tiempo, o bien $T_{\text{máx}} < \infty$ (*explosión en tiempo finito*) y en este caso

$$\lim_{t \nearrow T_{\text{máx}}} \| u(t) \|_{H_0^1(\Omega)} + \| u_t(t) \|_{L^2(\Omega)} = \infty. \tag{2.6.71}$$

El fenómeno que subyace a esta alternativa es fácil de entender. Mientras que la solución se mantiene acotada puede ser prolongada en el tiempo, con un paso temporal que depende continuamente de la cota de la solución. Por lo tanto, la única manera en que la solución pueda no ser prolongada a todos los tiempos es si explota en tiempo finito.

Mediante una mera hipótesis de crecimiento del tipo (2.6.68) sobre la no-linealidad es imposible determinar si se produce explosión en tiempo finito o no y para ésto son necesarias hipótesis adicinales sobre el "signo" de la no-linealidad.

Consideremos en primer lugar el caso en que la no-linealidad f tiene el "buen signo":

$$\begin{cases} u_{tt} - \Delta u + \mid u \mid^{p-1} u = 0 & \text{en} \quad \Omega \times (0, \infty) \\ u = 0 & \text{en} \quad \partial\Omega \times (0, \infty) \\ u(x,0) = u_0(x),\ u_t(x,0) = u_1(x) & \text{en} \quad \Omega. \end{cases} \tag{2.6.72}$$

En este caso, evidentemente, la condición (2.6.68) se cumple y bajo las hipótesis (2.6.70) se deduce la existencia y unicidad local (en tiempo) de soluciones de energía finita de (2.6.72). Además, mientras la solución existe, su energía se conserva. En este caso la energía viene dada por

$$E(t) = \frac{1}{2} \int_\Omega \Big[\mid \nabla u(x,t) \mid^2 + \mid u_t(x,t) \mid^2 \Big] dx + \frac{1}{p+1} \int_\Omega \mid u(x,t) \mid^{p+1} dx. \tag{2.6.73}$$

Como la energía $E(t)$ se conserva y claramente mayora al cuadrado de la norma de (u, u_t) en $H_0^1(\Omega) \times L^2(\Omega)$ deducimos inmediatamente que (2.6.71) es imposible. De este modo concluimos que, bajo la condición (2.6.70), para cada par de datos iniciales $(u_0, u_1) \in H_0^1(\Omega) \times L^2(\Omega)$ existe una única solución global

$$u \in C([0, \infty);\, H_0^1(\Omega)) \cap C^1([0, \infty);\, L^2(\Omega)) \tag{2.6.74}$$

y que la energía $E(t)$ de la solución definida en (2.6.73) se conserva para todo $t \geqslant 0$.

La situación cambia completamente para no-linealidades con "mal-signo":

$$\begin{cases} u_{tt} - \Delta u = \mid u \mid^{p-1} u & \text{en} \quad \Omega \times (0, \infty) \\ u = 0 & \text{en} \quad \partial\Omega \times (0, \infty) \\ u(x,0) = u_0(x),\ u_t(x,0) = u_1(x) & \text{en} \quad \Omega. \end{cases} \tag{2.6.75}$$

La existencia y unicidad de soluciones locales (en tiempo) es igualmente cierta en este caso. Pero no se puede decir lo mismo acerca de la existencia global. Para el sistema (2.6.75), la energía, que se observa mientras las soluciones existen es

$$E(t) = \frac{1}{2}\int_\Omega [\mid \nabla u(x,t) \mid^2 + \mid u_t(x,t) \mid^2] dx - \frac{1}{p+1}\int_\Omega \mid u(x,t) \mid^{p+1} dx \quad (2.6.76)$$

pero, el que esta energía permanezca constante o acotada es perfectamente compatible con la explosión (2.6.71) de las soluciones en tiempo finito. De hecho, en este caso, las soluciones pueden efectivamente explotar en tiempo finito. Para convencerse de este hecho basta ver que existen soluciones de la EDO

$$x'' = \mid x \mid^{p-1} x \quad (2.6.77)$$

que, cuando $p > 1$, explotan en tiempo finito, en un tiempo que tiende a cero cuando el tamaño de los datos iniciales tiende a infinito. El hecho de que las soluciones de la ecuación de ondas dependan exclusivamente de los datos iniciales en la base del cono característico permite entonces construir datos iniciales, independientes de x en una bola de Ω, y de modo que en el interior del cono correspondiente coinciden con la solución de la ODE (2.6.77) y por tanto explotan en tiempo finito.

Esta construcción permite efectivamente probar que, para todo $p > 1$ y todo abierto no vacío Ω de $\mathbb{R}^n$, existen datos iniciales $(u_0, u_1) \in H_0^1(\Omega) \times L^2(\Omega)$ para los que la solución local de (2.6.75) explota en tiempo finito.

Hemos ilustrado el modo en que la Teoría de semigrupos permite resolver la ecuación de ondas y sus variantes. Veamos ahora algunas de las ideas fundamentales de la demostración del Teorema fundamental de esta teoría: El Teorema 7.1 de Hille-Yosida.

La idea central de la demostración de Hille-Yosida, que permite utilizar la propiedad del operador A de ser maximal-disipativo, es introducir y usar la regularización de Yosida del operador A:

$$A_\lambda = -\frac{1}{\lambda}(I - (I - \lambda A)^{-1}). \quad (2.6.78)$$

Es fácil comprobar que, formalmente, A_λ converge a A cuando $\lambda \to 0$. Para ello basta analizar la expresión algebraica de la derecha de (2.6.78) en el caso de números reales:

$$-\frac{1}{\lambda}\Big[1 - \frac{1}{1-\lambda x}\Big] = -\frac{1}{\lambda}\Big[\frac{1-\lambda x - 1}{1-\lambda x}\Big] = \frac{x}{1-\lambda x} \to x,\ \lambda \to 0.$$

Pero para que la definición (2.6.78) tenga rigurosamente sentido es primeramente preciso probar que el operador $I - \lambda A$ es inversible. La hipótesis de maximalidad

sobre A garantiza que esto es así cuando $\lambda = 1$. Veamos que ésto permite probar que $I - \lambda A$ es inversible para todo $\lambda > 1/2$. En efecto, reescribimos la ecuación

$$x - \lambda Ax = y \tag{2.6.79}$$

como

$$x - Ax = \frac{1}{\lambda}y + \Big(1 - \frac{1}{\lambda}\Big)x,$$

o, lo que es lo mismo,

$$x = (I - A)^{-1}\Big[\frac{1}{\lambda}y + \Big(1 - \frac{1}{\lambda}\Big)x\Big]. \tag{2.6.80}$$

Es fácil comprobar que cuando $\mid 1 - 1/\lambda \mid < 1$ el segundo miembro de (2.6.80) admite una única solución para el Teorema de punto fijo de Banach.

Iterando este argumento se puede comprobar que $(I - \lambda A)^{-1}$ está bien definido para todo $\lambda > 0$. Además

$$\| \, (I - \lambda A)^{-1} \, \|_{\mathcal{L}(H,H)} \leqslant 1. \tag{2.6.81}$$

En efecto, como A es disipativo, $\langle Ax,\, x\rangle \leqslant 0$ y por tanto, si $x = (I - \lambda A)^{-1}y$ tenemos

$$\langle (I - \lambda A)^{-1}y,\, y\rangle = \langle x,\, (I - \lambda A)x\rangle = \langle x\, x\rangle - \lambda\langle x\, Ax\rangle \geqslant \langle x,\, x\rangle = \| \, x \, \|_H^2 = \| \, (I - \lambda A)^{-1}y \, \|_H^2,$$

de donde se deduce que, efectivamente,

$$\| \, (I - \lambda A)^{-1}y \, \|_H \leqslant \| \, y \, \|_H, \quad \forall y \in H, \tag{2.6.82}$$

lo cual equivale a (2.6.81).

Deducimos por tanto que A_λ, para cada $\lambda > 0$, es un operador lineal y acotado de H en H.

Esto nos permite resolver la ecuación abstracta

$$\begin{cases} u' = A_\lambda u, \quad t > 0 \\ u(0) = u_0, \end{cases} \tag{2.6.83}$$

Como si se tratase de una EDO.

En efecto, como A_λ es un operador lineal y acotado, $e^{A_\lambda t}$ se puede definir, como en el caso matricial, mediante el desarrollo en serie de potencias de la exponencial

$$e^{A_\lambda t} = \sum_{k=0}^{\infty} \frac{(A_\lambda t)^k}{k!}. \tag{2.6.84}$$

Es fácil comprobar que para cada $\lambda > 0$ y $t > 0$, $e^{A_\lambda t}$ define un operador lineal y acotado de H en H. Además

$$u_\lambda(t) = e^{A_\lambda t} u_0 \in C^\infty([0,\infty);\, H) \tag{2.6.85}$$

y es la única solución de (2.6.83).

La regularización de Yosida genera entonces un semigrupo $S_\lambda(t) = e^{A_\lambda t}$.

Además, para todo $\lambda > 0$, A_λ hereda la propiedad de A de ser disipativo, de modo que

$$\langle A_\lambda x,\, x\rangle \leqslant 0,\, \forall x \in H,\, \forall \lambda > 0. \tag{2.6.86}$$

Entonces

$$\| u_\lambda(t) \|_H \leqslant \| u_0 \|,\, \left|\left| \frac{du_\lambda(t)}{dt} \right|\right|_H = \| A_\lambda u_\lambda(t) \|_H \leqslant \| A_\lambda u_0 \|_H,\, \forall t > 0,\, \forall \lambda > 0. \tag{2.6.87}$$

Para comprobar (2.6.86) basta proceder del modo siguiente

$$\begin{aligned} \langle A_\lambda x,\, x\rangle &= \langle A_\lambda x,\, x - (I - \lambda A)^{-1} x\rangle + \langle A_\lambda x,\, (I - \lambda A)^{-1} x\rangle \\ &= -\lambda \| A_\lambda x \|^2 + \langle A_\lambda x,\, (I - \lambda A)^{-1} x\rangle \\ &= -\lambda \| A_\lambda x \|^2 + \langle A(I - \lambda A)^{-1} x,\, (I - \lambda A)^{-1} x\rangle \leqslant -\lambda \| A_\lambda x \|^2 \leqslant 0. \end{aligned}$$

Como, al menos formalmente, $A_\lambda \to A$ cuando $\lambda \to 0$, en virtud de las cotas uniformes (2.6.87) de las soluciones de las ecuaciones aproximadas (2.6.83) en las que el operador A ha sido sustituido por su regularización de Yosida A_λ, cabe esperar que la solución u de (2.6.22) en el Teorema de Hille-Yosida se obtenga como límite cuando $\lambda \to 0$ de las soluciones aproximadas u_λ.

La clave de la demostración del Teorema de Hille-Yosida consiste en ver que, cuando $u_0 \in D(A)$, $\{u_\lambda(t)\}_{\lambda>0}$ constituye una sucesión de Cauchy cuando $\lambda \to 0$ en $C([0,\infty);\, H)$.

En efecto. tenemos

$$\frac{du_\lambda}{dt} - \frac{du_\mu}{dt} = A_\lambda u_\lambda - A_\mu u_\mu$$

y por tanto

$$\frac{1}{2}\frac{d}{dt} \| u_\lambda - u_\mu \|_H^2 = \langle A_\lambda u_\lambda - A_\mu u_\mu,\, u_\lambda - u_\mu\rangle.$$

Pero

$$\begin{aligned} &\langle A_\lambda u_\lambda - A_\mu u_\mu,\, u_\lambda - u_\mu\rangle = \\ &= \langle A_\lambda u_\lambda - A_\mu u_\mu,\, u_\lambda - (I - \lambda A)^{-1} u_\lambda + (I - \lambda A)^{-1} u_\lambda - (I - \mu A)^{-1} u_\mu + (I - \mu A)^{-1} u_\mu - u_\mu\rangle \\ &= \langle A_\lambda u_\lambda - A_\mu u_\mu,\, -\lambda A_\lambda u_\lambda + \mu A_\mu u_\mu\rangle \\ &\quad + \langle A((I - \lambda A)^{-1} u_\lambda - (I - \mu A)^{-1} u_\mu),\, (I - \lambda A)^{-1} u_\lambda - (I - \mu A)^{-1} u_\mu\rangle \\ &\leqslant \langle A_\lambda u_\lambda - A_\mu u_\mu,\, -\lambda A_\lambda u_\lambda + \mu A_\mu u_\mu\rangle. \end{aligned}$$

Por tanto

$$\frac{1}{2}\frac{d}{dt} \parallel u_\lambda - u_\mu \parallel_H^2 \leqslant 2(\lambda+\mu) \parallel Au_0 \parallel_H^2 . \tag{2.6.88}$$

En este punto hemos utilizado la segunda cota de (2.6.87) junto con $\parallel A_\lambda x \parallel_H \leqslant \parallel Ax \parallel_H$, para todo $x \in H$ y $\lambda > 0$, propiedad esta que se deduce fácilmente de la identidad $A_\lambda = (I - \lambda A)^{-1}A$.

La estimación (2.6.88) proporciona el carácter de Cauchy de la sucesión u_λ en $C([0,\infty); H)$ que habíamos enunciado.

El mismo argumento permite probar que si $u_0 \in D(A^2)$, entonces du_λ/dt es también de Cauchy en $C([0,\infty); H)$. Esto permite pasar al límite en (2.6.83) cuando $\lambda \to 0$ y obtener la solución del problema abstracto (2.6.22) que el Teorema de Hille-Yosida enuncia cuando $u_0 \in D(A)$. Como $D(A^2)$ es denso en $D(A)$, un argumento de densidad permite concluir la existencia de solución para datos $u_0 \in D(A)$, tal y como se enuncia en el Teorema 2.6.1.

El lector interesado en una demostración completa del Teorema de Hille-Yosida puede consultar el capítulo VII del libro de H. Brezis [2]. En el libro de T. Cazenave y A. Haraux [3] se da también una extensión de este resultado a espacios de Banach y diversas aplicaciones a ecuaciones de evolución semilineales entre las que se incluyen la ecuación del calor, de ondas y de Schrödinger.

2.7. La ecuación de ondas con coeficientes variables

En esta sección analizamos brevemente la ecuación de ondas $1-d$ con coeficientes variables

$$u_{tt} - (\gamma(x)u_x)_x = 0,\ x \in \mathbb{R},\ t > 0. \tag{2.7.1}$$

Consideramos en primer lugar el problema de Cauchy en el que ya tendremos que hacer frente a las principales diferencias con respecto al caso de coeficientes constantes.

Desde el punto de vista de la modelización, el hecho que la constante $\gamma = \gamma(x)$ (de rigidez en el caso de un medio elástico) dependa de x indica que las ondas se propagan en un medio heterogéneo compuesto de diferentes materiales.

En el caso en que la densidad del medio también es variable, la ecuación de ondas correspondiente es

$$\rho(x)u_{tt} - (\gamma(x)u_x)_x = 0,\ x \in \mathbb{R}, t > 0. \tag{2.7.2}$$

Supondremos que los coeficientes γ y ρ son medibles y que existen constantes positivas ρ_j, γ_j, $j = 0, 1$ tales que

$$0 < \rho_0 \leqslant \rho(x) \leqslant \rho_1 < \infty,\ 0 < \gamma_0 \leqslant \gamma(x) \leqslant \gamma_1 < \infty \ \text{ p.c.t.Ê } x \in \mathbb{R}. \tag{2.7.3}$$

En estas condiciones ambos sistemas están bien puestos en $H^1(\mathbb{R}) \times L^2(\mathbb{R})$ y la energía de las soluciones se conserva. Las energías de los sistemas (2.7.1) y (2.7.2) es

$$E(t) = \frac{1}{2}\int_{\mathbb{R}} [\mid u_t(x,t) \mid^2 + \gamma(x) \mid u_x(x,t) \mid^2] dx, \tag{2.7.4}$$

y

$$E_\rho(t) = \frac{1}{2}\int_{\mathbb{R}} [\rho(x) \mid u_t(x,t) \mid^2 + \gamma(x) \mid u_x(x,t) \mid^2] dx, \tag{2.7.5}$$

respectivamente.

En virtud de las hipótesis (2.7.3) estas energías son equivalentes a la energía habitual de la ecuación de ondas.

Consideremos ahora la ecuación (2.7.1) y veamos cual es la aproximación semi-dicreta más natural. Proponemos el siguiente esquema

$$u_j'' - \frac{1}{h}\Big[\gamma_{j+1/2}\frac{u_{j+1} - u_j}{h} - \gamma_{j-1/2}\frac{u_j - u_{j-1}}{h}\Big] = 0,\ j \in \mathbb{Z},\ t > 0. \tag{2.7.6}$$

En (2.7.6) el coeficiente γ se evalúa en puntos intermedios del mallado. Así, por ejemplo,

$$\gamma_{j+1/2} = \gamma(x_{j+1/2}),\ x_{j+1/2} = x_j + \frac{h}{2} = \Big(j + \frac{1}{2}\Big)h. \tag{2.7.7}$$

Obviamente, la definición (2.7.7) de $\gamma_{j+1/2}$ es válida cuando γ es continuo. Si no lo fuese, como es habitual, lo más natural sería definir $\gamma_{j+1/2}$ a través de una media:

$$\gamma_{j+1/2} = \frac{1}{h}\int_{x_j}^{x_{j+1}} \gamma(s) ds \tag{2.7.8}$$

Cuando el coeficiente γ es constante (i.e. $\gamma(x) \equiv \gamma$), el esquema (2.7.6) coincide con el esquema semi-discreto centrado de orden dos para la aproximación de la ecuación de ondas:

$$u_j'' + \frac{\gamma}{h^2}[2u_j - u_{j+1} - u_{j-1}] = 0,\ j \in \mathbb{Z},\ t > 0. \tag{2.7.9}$$

El sistema (2.7.6) es conservativo. En efecto, multiplicando en (2.7.6) por u_j' y sumando en $j \in \mathbb{Z}$ obtenemos que

$$\sum_{j\in\mathbb{Z}} u_j'' u_j' - \frac{1}{h}\sum_{j\in\mathbb{Z}}\Big[\gamma_{j+1/2}\frac{u_{j+1} - u_j}{h} - \gamma_{j-1/2}\frac{u_j - u_{j-1}}{h}\Big]u_j' = 0.$$

Por otra parte

$$\sum_{j\in\mathbb{Z}} u_j'' u_j' = \frac{1}{2}\frac{d}{dt}\sum_{j\in\mathbb{Z}} |\, u_j' \,|^2$$

y

$$\begin{aligned}
&-\frac{1}{h}\sum_{j\in\mathbb{Z}}\Big[\gamma_{j+1/2}\frac{u_{j+1}-u_j}{h} - \gamma_{j-1/2}\frac{u_j-u_{j-1}}{h}\Big]u_j' \\
&= -\frac{1}{h}\sum_{j\in\mathbb{Z}}\gamma_{j+1/2}\frac{u_{j+1}-u_j}{h}u_j' + \frac{1}{h}\sum_{j\in\mathbb{Z}}\gamma_{j-1/2}\frac{u_j-u_{j-1}}{h}u_j' \\
&= -\frac{1}{h}\sum_{j\in\mathbb{Z}}\gamma_{j+1/2}\frac{u_{j+1}-u_j}{h}u_j' + \frac{1}{h}\sum_{j\in\mathbb{Z}}\gamma_{j+1/2}\frac{u_{j+1}-u_j}{h}u_{j+1}' \\
&= \sum_{j\in\mathbb{Z}}\gamma_{j+1/2}\frac{u_{j+1}-u_j}{h}\frac{u_{j+1}'-u_j'}{h} = \frac{1}{2}\frac{d}{dt}\sum_{j\in\mathbb{Z}}\gamma_{j+1/2}\Big|\frac{u_{j+1}-u_j}{h}\Big|^2.
\end{aligned}$$

Deducimos por tanto que la siguiente energía discreta se conserva para las soluciones de (2.7.6):

$$E_h(t) = \frac{1}{2}\sum_{j\in\mathbb{Z}}\Big[\,|\,u_j'\,|^2 + \gamma_{j+1/2}\Big|\frac{u_{j+1}-u_j}{h}\Big|^2\Big]. \tag{2.7.10}$$

Es obvio que E_h es una aproximación discreta de la energía (2.7.4) del problema continuo (2.7.1).

El hecho que la energía E_h se conserva garantiza la estabilidad del esquema numérico. Se trata por otra parte de un esquema consistente de orden dos, al menos cuando γ es suficientemente regular. En la medida en que el esquema que consideraremos es semi-discreto y que la variable temporal no ha sido discretizada, basta analizar la consistencia de la discretización espacial. Tenemos

$$\begin{aligned}
&-\frac{1}{h}\Big[\gamma(x_{j+1/2})\frac{u(x_{j+1})-u(x_j)}{h} - \gamma(x_{j-1/2})\frac{u(x_j)-u(x_{j-1})}{h}\Big] \\
= \;& -\frac{1}{h}\Big[\gamma(x_j)\frac{u(x_{j+1})-u(x_j)}{h} - \gamma(x_j)\frac{u(x_j)-u(x_{j-1})}{h}\Big] \\
&-\frac{1}{h}\Big[\frac{h}{2}\gamma'(x_j)\frac{u(x_{j+1})-u(x_j)}{h} + \frac{h}{2}\gamma'(x_j)\frac{u(x_j)-u(x_{j-1})}{h}\Big] \\
&-\frac{1}{h}\Big[\frac{h^2}{8}\gamma''(x_j)\frac{u(x_{j+1})-u(x_j)}{h} - \frac{h^2}{8}\gamma''(x_j)\frac{u(x_j)-u(x_{j-1})}{h}\Big] \\
&+O(h^2)\Big[\Big|\frac{u(x_{j+1})-u(x_j)}{h}\Big| + \Big|\frac{u(x_j)-u(x_{j-1})}{h}\Big|\Big]
\end{aligned}$$

La estabilidad, junto a su consistencia de orden dos garantiza que se trata de un método convergente de orden dos, cuando γ es suficientemente regular.

En el caso de la ecuación (2.7.2) con densidad variable es fácil modificar el esquema (2.7.6). Basta en este caso considerar

$$\rho(x_j)u''-\frac{1}{h}\Big[\gamma_{j+1/2}\frac{u_{j+1}-u_j}{h}-\gamma_{j-1/2}\frac{u_j-u_{j-1}}{h}\Big]=0,\quad j\in\mathbb{Z},\,t>0. \tag{2.7.11}$$

Analicemos ahora uno de los aspectos más importantes en los que las ecuaciones de ondas con coeficientes irregulares más se distinguen de los de coeficientes regulares: *la reflexión y transmisión de energía en las singularidades de los coeficientes*.

Ya hemos visto en el caso de la ecuación con coeficientes constantes, a través de la fórmula de d'Alembert, que las soluciones de la ecuación de ondas son una mera superposición de ondas de transporte que viajan a izquierda y derecha en el espacio a velocidad constante unidad. Esto no ocurre en el caso de ecuaciones con coeficientes variables. Si estos son regulares, las soluciones se propagan a lo largo de curvas características que no son rectilíneas, mientras que en el caso de coeficientes irregulares las ondas pueden incluso llegar a rebotar parcialmente en los puntos de discontinuidad de los coeficientes.

Por ejemplo, si $\gamma=\gamma(x)$ es una función de clase C^1, una ecuación de ondas de la forma

$$\partial_t^2u-\gamma(x)\partial_x^2u-\frac{\gamma'(x)}{2}\partial_x=0, \tag{2.7.12}$$

puede factorizarse como

$$\Big(\partial_t+\sqrt{\gamma(x)}\partial_x\Big)\Big(\partial_t-\sqrt{\gamma(x)}\partial_x\Big)u=0. \tag{2.7.13}$$

Las soluciones pueden entonces escribirse como superposición de las soluciones de ecuaciones de transporte de la forma

$$\Big[\partial_t\pm\sqrt{\gamma(x)}\partial_x\Big]u=0. \tag{2.7.14}$$

Para estas últimas, las soluciones son constantes a lo largo de curvas características que son las curvas parametrizadas $x=x(t)$ en las que

$$x'(t)=\pm\sqrt{\gamma(x(t))}. \tag{2.7.15}$$

Las curvas características están definidas de manera única cuando el coeficiente $\gamma^{1/2}$ tiene continuidad Lipschitz.

Pero cuando el coeficiente $\gamma^{1/2}=\gamma^{1/2}(x)$ deja de ser regular, tanto la definición de características como el hecho de que transporten la información de las soluciones deja de ser válida. Para analizar este hecho consideramos el caso de una ecuación de ondas con coeficientes constantes a trozos en un medio heterogéneo con dos caras:

$$\rho(x)u_{tt}-(\gamma(x)u_x)_x=0. \tag{2.7.16}$$

Suponemos entonces que

$$(\rho(x),\, \gamma(x)) = \begin{cases} (\rho_1,\, \gamma_1),\, x < 0 \\ (\rho_2,\, \gamma_2),\, x > 0. \end{cases} \tag{2.7.17}$$

La velocidad de propagación de las ondas es entonces $c_1 = \sqrt{\gamma_1/\rho_1}$ y $c_2 = \sqrt{\gamma_2/\rho_2}$ en el medio $x > 0$ y $x < 0$ respectivamente.

En este caso es fácil comprobar que, si bien en cada uno de los medios $x > 0$ y $x < 0$ las ondas se transportan sin deformación a velocidad constante c_2 y c_1 respectivamente, al alcanzar la interfase $x = 0$ parte de la onda se transmite mientras que la otra parte rebota. Este fenómeno puede establecerse con claridad a través del estudio de los *coeficientes de reflexión y transmisión*: R y T.

Para introducirlos, consideremos una onda plana que en el medio $x < 0$ se desplaza hacia la derecha a velocidad constante c_1 hasta alcanzar la interfase $x = 0$. Al alcanzarla, parte de la solución rebota produciendo una onda de transporte que en el medio $x < 0$ se propagará en sentido opuesto, siempre a velocidad c_1 y otra parte se transmite al medio $x > 0$ dando lugar a una onda que se transporta hacia la derecha a velocidad c_2.

El conjunto de esta solución que combina los fenómenos descritos puede escribirse en la forma

$$u(x,t) = 1_{(-\infty,0)}(x)\Big[e^{i(\omega t - k_1 x)} + \mathrm{Re}\, {}^{i(\omega t + k_1 x)}\Big] + T 1_{(0,\infty)}(x) e^{i(\omega t - k_2 x)}, \tag{2.7.18}$$

donde $1_{(-\infty,0)}$ y $1_{(0,\infty)}$ denotan respectivamente las funciones características de las semirectas $(-\infty, 0)$ y $(0, \infty)$ y k_1 y k_2 denotan las relaciones de dispersión en cada uno de los medios

$$k_1 = \omega/c_1;\; k_2 = \omega/c_2. \tag{2.7.19}$$

En (2.7.18), R y T denotan las constantes de reflexión y transmisión que precisamente deseamos calcular.

Es fácil comprobar que u en (2.7.18) constituye una solución de (2.7.16) tanto en el semiplano de la izquierda $x < 0$ como en el de la derecha $x > 0$.

Sin embargo el que la función u definida a trozos en (2.7.18) satisfaga (2.7.16) depende también de las condiciones de transmisión que se han de cumplir en el punto de interfase. En este caso son:

$$u^+(0,t) = u^-(0,t);\; \gamma_2 u_x^+(0,t) = \gamma_1 u_x^-(0,t), \quad t > 0. \tag{2.7.20}$$

Para que las condiciones (2.7.20) se verifiquen es preciso que:

$$1 + R = T, \quad k_1\gamma_1 R + k_2\gamma_2 T = k_1\gamma_1. \tag{2.7.21}$$

La solución de este sistema arroja los siguientes valores para los coeficientes R y T:

$$R = \frac{\sigma_1 - \sigma_2}{\sigma_1 + \sigma_2}, \tag{2.7.22}$$

$$T = \frac{2\sigma_1}{\sigma_1 + \sigma_2} \tag{2.7.23}$$

donde

$$\sigma_j = \sqrt{\gamma_j \rho_j},\ j = 1, 2 \tag{2.7.24}$$

es la *impedancia acústica* en cada uno de los medios.

De estas expresiones se deduce que si la impedancia acústica de ambos medios es la misma (i.e. $\sigma_1 = \sigma_2$), entonces toda la onda se transmite mientras que la parte de la onda reflejada se anula.

Veamos ahora lo que ocurre en una aproximación numérica de estas ecuaciones. Consideremos para ello una semi-discretización en diferencias finitas de paso h en cada uno de los medios. En el medio $x < 0$ la semi-discretización adopta la forma

$$\rho_1 u_j'' + \gamma_1 \frac{2u_j - u_{j+1} - u_{j-1}}{h^2} = 0,\ j \leqslant -1,\ t > 0 \tag{2.7.25}$$

mientras que en el medio $x > 0$ la aproximación correspondiente es

$$\rho_2 u_j'' + \gamma_2 \frac{2u_j - u_{j+1} - u_{j-1}}{h^2} = 0,\ j \geqslant 1,\ t > 0. \tag{2.7.26}$$

En el nodo $j = 0$ correspondiente a la interfase $x = 0$ la aproximación correspondiente más natural es

$$\frac{\rho_1 + \rho_2}{2} u_0'' + \frac{1}{h}\Big(\gamma_1 \frac{u_0 - u_{-1}}{h} - \gamma_2 \frac{u_1 - u_0}{h}\Big) = 0,\ t > 0. \tag{2.7.27}$$

Las relaciones de dispersión asociadas a los esquemas (2.7.25) y (2.7.26) son, en cada uno de los casos,

$$k_{1,h} = \frac{2}{h} \operatorname{arc\,sen}\Big(\frac{\omega h}{2c_1}\Big),\ x < 0 \tag{2.7.28}$$

$$k_{2,h} = \frac{2}{h} \operatorname{arc\,sen}\Big(\frac{\omega h}{2c_2}\Big),\ x > 0. \tag{2.7.29}$$

Escribimos entonces la solución numérica, inspirándonos en el caso continuo, del modo siguiente

$$u_j(t) \begin{cases} e^{(\omega t - j k_1 h)} + R_h e^{i(\omega t + j k_1 h)},\ j \leqslant 0 \\ T_h e^{i(\omega t - j k_2 h)},\ j \geqslant 0. \end{cases} \tag{2.7.30}$$

Tenemos nuevamente

$$1 + R_h = T_h. \tag{2.7.31}$$

La ecuación (2.7.27) en el punto $x = 0$ de transmisión proporciona la relación adicional

$$-(\rho_1+\rho_2)\frac{T_h\omega^2}{2} + \frac{1}{h}\Big(\frac{\gamma_1}{h}\Big(1 + R_h - e^{ik_1h} - R_he^{-ik_1h}\Big) - \frac{\gamma_2}{h}T_h\Big(e^{-ik_2h} - 1\Big)\Big) = 0. \tag{2.7.32}$$

De estas ecuaciones obtenemos

$$T_h = \frac{-2i\gamma_1 \operatorname{sen}(k_1h)}{\gamma_2e^{-ik_2h} - \gamma_1 - \gamma_2 + \gamma_1e^{-ik_1h} + (\rho_1+\rho_2)\frac{\omega^2h^2}{2}}. \tag{2.7.33}$$

Mediante un desarrollo de Taylor observamos que

$$T_h = \frac{2\sigma_1}{\sigma_1+\sigma_2} + \frac{\gamma_1\rho_2 - \gamma_2\rho_1}{4c_1c_2(\sigma_1+\sigma_2)}\omega^2h^2 + O(h^3) \tag{2.7.34}$$

de donde se deduce que el coeficiente de transmisión del método numérico es una aproximación de orden dos del caso continuo (2.7.23).

Conviene sin embargo señalar que no siempre los esquemas numéricos proporcionan aproximaciones de los coeficientes de reflexión y transmisión del mismo orden que el que caracteriza al método numérico.

Hemos estudiado ecuaciones de ondas con coeficientes variables dependientes de la variable espacial. Los mismos problemas se plantean con ecuaciones cuyos coeficientes dependen también de la variable tiempo. Consideremos por ejemplo la siguiente ecuación de ondas con densidad variable dependiente de x y t:

$$\rho(x,t)u_{tt} - u_{xx} = 0,\ x \in \mathbb{R},\ t > 0. \tag{2.7.35}$$

Es natural suponer que la densidad es una función medible, acotada superior e inferiormente por constantes positivas ρ_0 y ρ_1:

$$0 < \rho_0 \leqslant \rho(x,t) \leqslant \rho_1 < \infty,\ p.c.t.\, x \in \mathbb{R},\ t > 0. \tag{2.7.36}$$

Cabe entonces plantearse si la ecuación (2.7.35) está bien planteada bajo estas hipótesis. Para entender esta cuestión es conveniente considerar la energía de las soluciones

$$E(t) = \frac{1}{2}\int_{\mathbb{R}} \Big[\rho(x,t) \mid u_t(x,t) \mid^2 + \mid u_x(x,t) \mid^2 \Big]dx. \tag{2.7.37}$$

Sin embargo, la energía verifica, formalmente, la identidad

$$E'(t) = \frac{1}{2}\int_{\mathbb{R}} \rho_t(x,t) \mid u_t(x,t) \mid^2 dx. \tag{2.7.38}$$

En virtud de (2.7.38) es fácil comprobar que la ecuación de ondas (2.7.35) no está bien puesta en el espacio de la energía bajo la mera hipótesis (2.7.36). En efecto para que la ecuación esté bien puesta es necesaria alguna hipótesis sobre $\rho_t = \partial\rho/\partial_t$. Supongamos por ejemplo que

$$| \rho_t(x,t) | \leqslant k, \, \forall x \in \mathbb{R}, \, t > 0. \tag{2.7.39}$$

En este caso, de la identidad de energía (2.7.38) se deduce que

$$| E'(t) | \leqslant \frac{k}{2\rho_0} \int_{\mathbb{R}} \rho(x,t) \mid u_t(x,t) \mid^2 dx \leqslant \frac{k}{\rho_0} E(t), \tag{2.7.40}$$

de donde, por la desigualdad de Gronwall, se obtiene que

$$E(t) \leqslant e^{kt/\rho_0} E(0), \, \forall t \geqslant 0. \tag{2.7.41}$$

Bajo las hipótesis (2.7.36) y (2.7.39) sobre el coeficiente variable de densidad $\rho = \rho(x,t)$ se puede entonces probar que la ecuación (2.7.35) está bien puesta en $H^1(\mathbb{R}) \times L^2(\mathbb{R})$ de modo que para cada par de datos iniciales $(u_0, u_1) \in H^1(\mathbb{R}) \times L^2(\mathbb{R})$ existe una única solución $u \in C([0,\infty); H^1(\mathbb{R})) \cap C^1([0,\infty); L^2(\mathbb{R}))$ que toma este dato inicial y cuya energía $E(t)$ satisface la estimación (2.7.41).

La hipótesis (2.7.39) no es meramente técnica. En efecto, de manera general, la ecuación (2.7.35) no está bien puesta bajo la mera hipótesis (2.7.36). En el caso en que la densidad es en cada instante de tiempo independiente de x y depende del tiempo de modo que sea constante a trozos, la solución de la ecuación de ondas correspondiente puede calcularse explícitamente mediante la fórmula de d'Alembert, prestando especial atención al cambio de los perfiles de la solución en los instantes de tiempo en los que la densidad presenta la discontinuidad. En efecto, consideremos el caso particular en que

$$\rho = \begin{cases} \rho_1, \, 0 \leqslant t \leqslant 1, \, x \in \mathbb{R} \\ \rho_2, \, t > 1, \, x \in \mathbb{R} \end{cases} \tag{2.7.42}$$

siendo ρ_1 y ρ_2 dos constantes positivas distintas.

En el intervalo temporal $0 \leqslant t \leqslant 1$ en el que la densidad es la constante ρ_1 la solución de la ecuación de ondas es de la forma

$$u = f\Big(x - t/\sqrt{\rho_1}\Big) + g\Big(x + t/\sqrt{\rho_1}\Big). \tag{2.7.43}$$

A partir de ese instante, i.e. para $t > 1$, la solución es sin embargo de la forma

$$u = \widetilde{f}\Big(x - t/\sqrt{\rho_2}\Big) + \widetilde{g}\Big(x + t/\sqrt{\rho_2}\Big). \tag{2.7.44}$$

Al imponer la condición de continuidad sobre u y u_t en $t = 1$ obtenemos las ecuaciones

$$\begin{cases} f\Big(x-1/\sqrt{\rho_1}\Big)+g\Big(x+1/\sqrt{\rho_1}\Big)=\widetilde{f}\Big(x-1/\sqrt{\rho_2}\Big)+\widetilde{g}\Big(x+1/\sqrt{\rho_2}\Big) \\ -\dfrac{1}{\sqrt{\rho_1}}\Big[f'\Big(x-1/\sqrt{\rho_1}\Big)-g'\Big(x+1/\sqrt{\rho_1}\Big)\Big]=-\dfrac{1}{\sqrt{\rho_2}}\Big[\widetilde{f}'\Big(x-1/\sqrt{\rho_2}\Big)-\widetilde{g}'\Big(x+1/\sqrt{\rho_2}\Big)\Big] \end{cases} \tag{2.7.45}$$

que permiten calcular los perfiles $\widetilde{f}$ y $\widetilde{g}$ del intervalo temporal $t > 1$ a partir de los perfiles f y g del intervalo $0 < t < 1$.

Es sin embargo obvio que este tipo de procedimiento resulta sumamente costoso y difícil de adaptar a casos más generales de densidades variables.

Los mismos fenómenos que acabamos de describir se presentan también para las aproximaciones numéricas. Consideremos por ejemplo la semidiscretización más natural de (2.7.35)

$$\rho_j(t)u_j'' + \frac{[2u_j - u_{j+1} - u_{j-1}]}{h^2} = 0,\ j \in \mathbb{Z},\ t > 0$$

donde

$$\rho_j(t) = \rho(x_j, t).$$

En este caso la energía viene dada por la expresión

$$E_h(t) = \frac{h}{2}\sum_{j\in\mathbb{Z}}\Big[\rho_j(t) \mid u_j'(t) \mid^2 + \Big|\frac{u_{j+1}(t) - u_j(t)}{h}\Big|^2\Big].$$

Es fácil comprobar que en este caso la energía evoluciona según la ley

$$E_h'(t) = \frac{h}{2}\sum_{j\in\mathbb{Z}}\rho_j'(t) \mid u_j'(t) \mid^2$$

de modo que no es posible obtener estimaciones sobre su evolución temporal, independientes del paso del mallado h, sin imponer condiciones sobre la derivada temporal de la densidad.

Conviene pues abordar el análisis de ecuaciones de ondas y de sus aproximaciones discretas, en presencia de coeficientes dependientes del tiempo, con prudencia, pues, como hemos visto, es habitual que esto exija hipótesis adicionales sobre la regularidad de los coeficientes en su evolución temporal, inesperadas en primera instancia.

2.8. Semi-discretización de la ecuación de ondas semilineal

En la sección 2.6 hemos estudiado la ecuación de ondas semilineal en el contexto de la Teoría de semigrupos. Hemos visto que, bajo condiciones adecuadas sobre el crecimiento de la no-linealidad, (que garantizan que la nolinealidad envía H^1 en L^2 de manera Lipschitz en acotados) la ecuación de ondas semi-lineal está bien puesta, localmente en tiempo. Veíamos posteriormente que a través de una estimación de energía, cuando la no-linealidad tenía una propiedad de "buen signo" la solución podía prolongarse y definirse globalmente en tiempo.

En esta sección discutimos brevemente algunos aspectos de la aproximación numérica de estas ecuaciones.

Consideremos la ecuación de ondas semilineal $1-d$:

$$\begin{cases} u_{tt} - u_{xx} + u^3 = 0, & x \in \mathbb{R}, \quad t > 0 \\ u(x,0) = u_0(x),\ u_t(x,0) = u_1(x), & x \in \mathbb{R}. \end{cases} \tag{2.8.1}$$

Gracias a la inclusión de Sobolev $H^1(\mathbb{R}) \hookrightarrow L^p(\mathbb{R})$, para todo $2 \leqslant p \leqslant \infty$, la ecuación (2.8.1) está bien puesta en el espacio de la energía. Además, la energía

$$E(t) = \frac{1}{2}\int_{\mathbb{R}} \Big[\mid u_t(x,t) \mid^2 + \mid u_x(x,t) \mid^2 \Big] dx + \frac{1}{4}\int_{\mathbb{R}} u^4(x,t) dx \tag{2.8.2}$$

se conserva de modo que las soluciones están globalmente definidas en tiempo.

En este caso la aproximación semi-discreta más natural viene dada por las siguientes ecuaciones

$$\begin{cases} u_j'' + \dfrac{[2u_j - u_{j+1} - u_{j-1}]}{h^2} + u_j^3 = 0, & j \in \mathbb{Z}, \quad t > 0 \\ u_j(0) = u_{0,j},\ u_j'(0) = u_{1,j}, & j \in \mathbb{Z}, \end{cases} \tag{2.8.3}$$

donde $\Big(u_{0,j}\Big)_{j\in\mathbb{Z}}$, $\Big(u_{1,j}\Big)_{j\in\mathbb{Z}}$ son aproximaciones adecuadas de los datos iniciales continuos de (2.8.1).

El sistema (2.8.3) es un conjunto de una infinidad de ecuaciones diferenciales nolineales acopladas. Se puede probar que, para cada paso de mallado $h > 0$, (2.8.3) admite una única solución local en tiempo. Para ello basta aplicar la fórmula de variación de las constantes y utilizar las propiedades ya conocidas sobre el sistema lineal subyacente.

Pero, nuevamente, para probar la existencia global de soluciones, necesitamos de una identidad de energía. En este caso tenemos que la energía del sistema

semi-discreto es

$$E_h(t) = \frac{h}{2}\sum_{j\in\mathbb{Z}}\Big[\mid u_j'(t)\mid^2 + \Big|\frac{u_{j+1}(t)-u_j(t)}{h}\Big|^2\Big] + \frac{h}{4}\sum_{j\in\mathbb{Z}} u_j^4(t). \tag{2.8.4}$$

Nuevamente, esta energía se conserva en tiempo y este hecho permite probar que las soluciones de (2.8.3) están globalmente definidas.

Más adelante analizaremos el modo en que las soluciones de (2.8.3) convergen a las de (2.8.1) cuando $h \to 0$.

Conviene de todos modos analizar con un poco más de cuidado el argumento que permite deducir la existencia global de soluciones a partir de la conservación de las energías (2.8.2) o (2.8.4).

En efecto, el resultado de existencia y unicidad de soluciones de (2.8.1) en $H^1(\mathbb{R})\times L^2(\mathbb{R})$, localmente en tiempo y el posterior argumento de prolongación al intervalo maximal de existencia $[0,\, T_{\max})$ permite establecer la alternativa siguiente: O bien $T_{\max}=\infty$ (*existencia global*); o bien $T_{\max}<\infty$ en cuyo caso se produce la *explosión en tiempo finito* de las soluciones, i.e.

$$\lim_{t\nearrow T_{\max}} \parallel u(t)\parallel_{H^1(\mathbb{R})} + \parallel u_t(t)\parallel_{L^2(\mathbb{R})} = \infty. \tag{2.8.5}$$

El que la energía $E(t)$ de (2.8.2) se conserve junto con que todos los términos que en ella intervienen tengan signo positivo permite garantizar que tanto $\parallel u_x(t)\parallel_{L^2(\mathbb{R})}$ como $\parallel u_t(t)\parallel_{L^2(\mathbb{R})}$ permanecen acotadas en intervalos finitos de tiempo. Sin embargo, en la medida que estamos trabajando en $\mathbb{R}$ y no disponemos de la desigualdad de Poincaré, para probar que (2.8.5) no es posible tenemos también que establecer una cota sobre la norma de la solución en $L^2(\mathbb{R})$. Para ello introducimos la energía perturbada

$$F(t) = \frac{1}{2}\int_{\mathbb{R}}\Big[u_t^2 + u_x^2 + u^2\Big]dx + \frac{1}{4}\int_{\mathbb{R}} u^4 dx.$$

Para esta nueva energía tenemos la identidad

$$\frac{dF(t)}{dt} = \int_{\mathbb{R}} u\, u_t dx \leqslant \frac{1}{2}\int_{\mathbb{R}}\Big[u^2 + u_t^2\Big]dx \leqslant F(t)$$

de modo que, por la desigualdad de Gronwall,

$$F(t) \leqslant F(0)e^t.$$

En particular

$$\int u^2(t)dx \leqslant Ce^t$$

de modo que la explosión de la norma L^2 de la solución en tiempo finito queda también excluida.

El mismo tipo de argumento, basado en la perturbación de la energía natural del sistema, permite también probar que las soluciones de (2.8.3) están globalmente definidas en tiempo.

Se trata de una aplicación más del denominado *método de la energía* que consiste en construir cantidades de interpretación física más o menos directa, para las que se puede establecer alguna desigualdad diferencial que proporcione cotas sobre dicha cantidad permitiendo a su vez obtener estimaciones sobre las soluciones.

Capítulo 3

La ecuación de transporte lineal

Las ecuaciones que modelizan fenómenos de propagación de ondas y vibraciones son típicamente Ecuaciones en Derivadas Parciales (EDP) de orden dos. Sin embargo en todas ellas subyacen las ecuaciones de transporte de orden uno que analizamos en esta sección.

El modelo más sencillo es

$$u_t + u_x = 0. \tag{3.0.1}$$

Es fácil comprobar que $u = u(x,t)$ es solución de esta ecuación si y sólo si es constante a lo largo de las *líneas características*

$$x + t = cte. \tag{3.0.2}$$

De este modo deducimos que las soluciones de (3.0.1) son de la forma

$$u = f(x - t), \tag{3.0.3}$$

donde f es el perfil de la solución en el instante inicial $t = 0$, i.e.

$$u(x, 0) = f(x). \tag{3.0.4}$$

La solución (3.0.3) es entonces una simple onda de transporte pura en la que el perfil f se transporta (avanza) en el eje real a velocidad constante uno[1] .

[1]Si bien en este caso la ecuación puede resolverse explícitamente, el problema (3.0.1) entra también el marco de la Teoría de Semigrupos. En efecto, basta considerar el espacio

Al invertir el sentido del tiempo (i.e. haciendo el cambio de variable $t \to -t$) la ecuación (3.0.1) se transforma en

$$u_t - u_x = 0 \tag{3.0.5}$$

cuyas soluciones son ahora de la forma

$$u = g(x + t), \tag{3.0.6}$$

tratándose de ondas viajeras que se propagan en dirección opuesta a velocidad uno.

Vemos por tanto que las soluciones de la ecuación de transporte pueden calcularse de manera explícita y que en ellas se observa un sencillo fenómeno de transporte lineal sin deformación.

Esta ecuación es por tanto un excelente laboratorio para experimentar algunas de las ideas más sencillas del análisis numérico.

Consideremos pues un paso de discretización $h > 0$ en la variable espacial e introduzcamos el mallado $\{x_j\}_{j\in\mathbf{Z}}$, $x_j = jh$.

Buscamos una semi-discretización (continua en tiempo y discreta en espacio) que reduzca la EDP (3.0.1) a un sistema de ecuaciones diferenciales cuya solución proporcione una aproximación $u_j(t)$ de la solución $u = u(x,t)$ de (3.0.1) en el punto $x = x_j$.

La manera más sencilla de construir esta semi-discretización es utilizar el desarrollo de Taylor para introducir una aproximación de la derivación parcial en la variable espacial.Son varias las posibilidades:

$$u_x(x_j,t) \sim \frac{u(x_{j+1},t) - u(x_j,t)}{h} \sim \frac{u_{j+1}(t) - u_j(t)}{h}, \tag{3.0.7}$$

$$u_x(x_j,t) \sim \frac{u(x_j,t) - u(x_{j-1},t)}{h} \sim \frac{u_j(t) - u_{j-1}(t)}{h} \tag{3.0.8}$$

$$u_x(x_j,t) \sim \frac{u(x_{j+1},t) - u(x_{j-1},t)}{2h} \sim \frac{u_{j+1}(t) - u_{j-1}(t)}{2h}. \tag{3.0.9}$$

Cada una de estas elecciones corresponde a un determinado sentido de avance a lo largo del eje x. En efecto (3.0.7) y (3.0.8) y (3.0.9) corresponden a diferencias progresivas, regresivas y centradas respectivamente.

---

de Hilbert $H = L^2(\mathbf{R})$ y el operador $A = -\partial_x$ con dominio $D(A) = H^1(\mathbf{R})$ para que el problema (3.0.1) entre en el marco abstracto del Teorema de Hille-Yosida. En efecto, el operador A así definido es maximal disipativo. Para ver que es disipativo basta con observar que $< Au, u >_{L^2(\mathbf{R})} = -\int_{\mathbf{R}} \partial_x u u dx = -\frac{1}{2}\int_{\mathbf{R}} \partial_x(u^2)dx = 0$. Además A es maximal. En efecto, dado $f \in L^2(\mathbf{R})$, existe una única solución $u \in H^1(\mathbf{R})$ de $u + \partial_x u = f$. Esta solución puede calcularse explícitamente y se obtiene: $u(x) = \int_{-\infty}^{x} f(s)e^{s-x}ds = \int_{-\infty}^{0} f(z+x)e^z dz$. Tomando normas en $L^2(\mathbf{R})$ y aplicando la desigualdad de Minkowski se deduce fácilmente que $||u||_{L^2(\mathbf{R})} \leq \int_{-\infty}^{0} ||f||_{L^2(\mathbf{R})} e^z dz = ||f||_{L^2(\mathbf{R})}$. Como $u_x = f - u$ vemos inmediatamente que, efectivamente, u pertenece a $H^1(\mathbf{R})$.

Cada una de estas elecciones proporciona un sistema semi-discreto diferente de aproximación de la EDP (3.0.1) en diferencias finitas:

- *Esquema progresivo:*

$$u_j'(t) + \frac{u_{j+1}(t) - u_j(t)}{h} = 0,\ j \in \mathbf{Z},\ t > 0, \tag{3.0.10}$$

- *Esquema regresivo:*

$$u_j'(t) + \frac{u_j(t) - u_{j-1}(t)}{h} = 0,\ j \in \mathbf{Z},\ t > 0, \tag{3.0.11}$$

- *Esquema centrado:*

$$u_j'(t) + \frac{u_{j+1}(t) - u_{j-1}(t)}{2h} = 0,\ j \in \mathbf{Z},\ t > 0. \tag{3.0.12}$$

Estos sistemas constituyen un conjunto numerable de ecuaciones diferenciales de orden uno lineales acopladas.

Al tratarse de sistema infinitos su resolución no entra en el marco de la teoría clásica de EDO. Sin embargo, es fácil verificar que su solución existe y es única sin necesidad de utilizar la Teoría de Semigrupos desarrollada en la sección anterior. Para ello basta considerar el espacio de Hilbert $H = \ell^2$ de las sucesiones de cuadrado sumables. La solución de cualquiera de estas ecuaciones semi-discretas puede entonces verse como un elemento de este espacio: $\vec{u} = \{u_j\}_{j\in\mathbf{Z}} \in \ell^2$. Estos sistemas pueden escribirse entonces en forma abstracta

$$\frac{d}{dt}\vec{u} = A_h\vec{u}. \tag{3.0.13}$$

Es fácil comprobar que en cada uno de los casos anteriores el operador A_h involucrado puede representarse a través de una matriz infinita, tridiagonal y acotada con norma $1/h$. Se trata pues de ecuaciones de evolución en espacios de Hilbert de dimensión infinita pero en las que el generador A_h está acotado. Esto nos permite calcular el semigrupo $e^{A_h t}$ mediante la representación en desarrollo de serie de potencias de la exponencial. Obtenemos así que estas ecuaciones generan semigrupos en $H = \ell^2$. De este modo deducimos que para cada dato inicial dado en ℓ^2 cada una de estas ecuaciones admite una única solución $C^\infty(\mathbf{R}, \ell^2)$ que toma ese dato en el instante $t = 0$. Las soluciones dependen en realidad de manera analítica con respecto a la variable temporal.

Todos estos esquemas son consistentes con la ecuación de transporte. Es decir, al llevar a estos esquemas una solución regular de la ecuación de transporte continua vemos que se produce un error que tiende a cero a medida que $h \to 0$.

La mejor manera de analizar la estabilidad es a través del método de von Neumann. Así, introduciendo

$$\check{u}(\theta,t)=\sum_{j\in\mathbf{Z}}u_j(t)e^{i\theta j} \tag{3.0.14}$$

obtenemos que $\check{u}$, en cada uno de los casos, satisface

$$\check{u}'(\theta,t)+\left(\frac{e^{-i\theta}-1}{h}\right)\check{u}(\theta,t)=0,\ t>0, \tag{3.0.15}$$

$$\check{u}'(\theta,t)+\left(\frac{1-e^{i\theta}}{h}\right)\check{u}(\theta,t)=0,\ t>0, \tag{3.0.16}$$

$$\check{u}'(\theta,t)+\left(\frac{e^{-i\theta}-e^{i\theta}}{2h}\right)\check{u}(\theta,t)=0,\ t>0. \tag{3.0.17}$$

La transformada discreta de Fourier no sólo tiene la virtud de transformar los sistemas de ecuaciones semi-discretas (3.0.10)-(3.0.12) en ecuaciones diferenciales con paramétro θ (3.0.15)-(3.0.17) que son inmediatas de resolver, sino que define también una isométrica de ℓ^2 a valores en $L^2(0,2\pi)$. En efecto, la fórmula (3.0.14) puede invertirse fácilmente. de hecho tenemos

$$u_j(t)=\frac{1}{2\pi}\int_0^{2\pi}\check{u}(\theta,t)e^{-ij\theta}d\theta. \tag{3.0.18}$$

Además

$$\frac{1}{2\pi}\int_0^{2\pi}|\check{u}(\theta,t)|^2d\theta=\sum_{j\in\mathbf{Z}}|u_j(t)|^2. \tag{3.0.19}$$

Obtenemos por tanto

$$\check{u}(\theta,t)=e^{a_h(\theta)t}\check{u}(\theta,0) \tag{3.0.20}$$

donde $a_h(\theta)$ varía de un caso a otro. De manera más precisa se tiene

$$a(\theta)=\begin{cases}\dfrac{1-e^{-i\theta}}{h}, & \text{(esquema progresivo)}\\ \dfrac{e^{i\theta}-1}{h}, & \text{(esquema regresivo)}\\ \dfrac{e^{i\theta}-e^{-i\theta}}{2h}, & \text{(esquema centrado).}\end{cases} \tag{3.0.21}$$

Como es bien sabido, la convergencia de un método numérico exige su estabilidad[2] y ésta pasa por que $Re\, a_h(\theta)$ permanezca acotada superiormente cuando $h\to 0$ uniformemente en $\theta\in[0,2\pi)$. Verifiquemos si esta propiedad se cumple en cada uno de los casos:

[2]En este punto estamos haciendo uso del clásico Teorema de Lax que dice que la convergencia de un esquema es equivalente a su estabilidad más consistencia. En el caso más sencillo de

- *Esquema progresivo:* Tenemos

$$a_h(\theta) = \frac{1 - e^{i\theta}}{h} = \frac{1 - \cos(\theta)}{h} - \frac{i \operatorname{sen}(\theta)}{h}. \tag{3.0.22}$$

Por tanto

$$Re\, a_h(\theta) = \frac{1 - \cos(\theta)}{h}.$$

Obviamente,

$$Re\, a_h(\theta) \nearrow \infty,\ h \to 0,\ \forall 0 < \theta < 2\pi, \tag{3.0.23}$$

lo cual demuestra la falta de estabilidad y por tanto de convergencia de este esquema.

- *Esquema regresivo:* En este caso

$$a_h(\theta) = \frac{e^{-i\theta} - 1}{h} = \frac{\cos(\theta) - 1}{h} - \frac{i \operatorname{sen}\theta}{h} \tag{3.0.24}$$

de modo que

$$Re\, a_h(\theta) = \frac{\cos(\theta) - 1}{h} \leq 0,\ \forall \theta \in [0, 2\pi). \tag{3.0.25}$$

La estabilidad del esquema está por tanto garantizada. Esto demuestra que el esquema es también convergente, propiedad que analizaremos más adelante.

- *Esquema centrado:* En este caso

$$a_h(\theta) = \frac{e^{-i\theta} - e^{i\theta}}{2h} = -\frac{i \operatorname{sen}\theta}{h}. \tag{3.0.26}$$

Obviamente,

$$Re\, a_h(\theta) = 0 \tag{3.0.27}$$

por lo que este esquema es también estable y convergente.

---

la resolución de un sistema lineal $Ax = b$, podemos interpretar este resultado del siguiente modo. Aproximemos este problema por otro de características semejantes $A_\varepsilon x_\varepsilon = b_\varepsilon$. Suponemos que $b_\varepsilon \to b$ cuando $\varepsilon \to 0$. Deseamos probar que $x_\varepsilon \to x$. Para ello hacemos las dos siguientes hipótesis: a) $A_\varepsilon y \to Ay$ para todo y (*consistencia*) y b) Las matrices inversas $(A_\varepsilon)^{-1}$ están uniformemente acotadas (*estabilidad*). Deducimos entonces la convergencia de las soluciones: $x_\varepsilon \to x$ cuando $\varepsilon \to 0$. En efecto, tenemos $A_\varepsilon(x_\varepsilon - x) = b_\varepsilon - b + (A - A_\varepsilon)x = r_\varepsilon$. Por las hipótesis realizadas sobre la aproximación deducimos que $r_\varepsilon \to 0$. La hipótesis de estabilidad garantiza entonces que $x_\varepsilon - x \to 0$. El Teorema de Lax generaliza este resultado al caso de las EDP y sus aproximaciones numéricas. La ecuación $Ax = b$ del ejemplo anterior juega el papel de la EDP, la ecuación cuya solución deseamos aproximar. La ecuación aproximada $A_\varepsilon x_\varepsilon = b_\varepsilon$ juega el papel de la aproximación numérica, y ε es el parámetro destinado a tender a cero, lo mismo que hace h en las aproximaciones numéricas.

En realidad bastaría verificar las propiedades geométricas más elementales asociadas a la evolución temporal que la ecuación continua y semi-discreta generan para ver que el esquema progresivo no puede de ningún modo ser convergente y que, sin embargo, los otros dos esquemas pueden perfectamente serlo.

En efecto, en virtud de la expresión explícita (3.0.3) de la solución de la ecuación de transporte (3.0.1), observamos que el dominio de dependencia de la solución en el punto (x,t) se reduce al punto $x-t$ en el instante inicial. Veamos ahora cuáles son los dominio de dependencia en los esquemas discretos.

En el esquema progresivo, fijado un punto $x = x_j$, vemos que la ecuación que gobierna la dinámica de $u_j(t)$ depende de $u_{j+1}(t)$, la aproximación de la solución en el nodo x_{j+1} inmediatamente a la derecha de x_j, que a su vez depende de $u_{j+1}(t)$, etc. Vemos pues que, en este caso, el sistema semi-discreto depende del valor del dato inicial a la derecha de x_j mientras que el único valor relevante para la solución real es el punto $x-t$ que está al lado opuesto, a la izquierda de x.

Por lo tanto el esquema semi-discreto progresivo viola la condición indispensable para la convergencia de un esquema numérico según la cual *el dominio de dependencia del esquema numérico ha de contener el dominio de dependencia de la ecuación original*[3].

Sin embargo, los otros dos esquemas si que verifican esta propiedad geométrica, lo cual garantiza su convergencia.

El esquema progresivo para la ecuación de transporte que consideramos suele normalmente denominarse "upwind", que viene a significar algo asi como "a favor de la correinte". Con este término se pone de manifiesto que en los problemas en los que está presente el fenómeno de transporte, el sentido y orientación del mismo ha de ser tenido en cuenta a la hora de diseñar métodos numéricos convergentes.

El análisis que acabamos de realizar indica que:

* Las ondas continuas se propagan en el espacio-tiempo con una velocidad y dirección determinadas.

* Los esquemas numéricos, a pesar de estar basados en un mecanismo aparentemente coherentes de discretización, pueden generar ondas que se pro-

[3]Se trata efectivamente de una condición necesaria para la convergencia de un método numérico. Cuando no se cumple, hay puntos del dominio de dependencia del problema continuo que no pertenecen al del problema discreto. En estas circunstancias, modificando los datos iniciales en esos puntos, podemos conseguir alterar la solución del problema continuo sin que la del problema discreto sufra ningún cambio. Esto excluye cualquier posibilidad de convergencia del método numérico.

pagan con velocidades y direcciones distintas y no converger a medida que el paso del mallado tiende a cero.

Los tres esquemas que hemos analizado son en principio coherentes. En realidad en la terminología del Análisis Numérico se dice que son esquemas consistentes. De manera más precisa, mientras que el esquema progresivo y regresivo son consistentes de orden 1, el esquema centrado es consistente de orden 2. En efecto, supongamos que u es una solución suficientemente regular de la ecuación de transporte (3.0.1) (basta con que u tenga una derivada continua en la variable tiempo y tres en la variable espacial).

Sea entonces

$$\underline{u}_j(t) = u(x_j, t), \tag{3.0.28}$$

la restricción de (3.0.1) a los puntos del mallado.

Para analizar la consistencia de los esquemas numéricos introducidos consideramos $\left(\underline{u}_j\right)_{j\in\mathbf{Z}}$ como una solución aproximada de dicho esquema[4].

Tenemos entonces, en el caso de esquema progresivo

$$\begin{aligned} \underline{u}'_j + \frac{\underline{u}_{j+1} - \underline{u}_j}{h} &= u_t(x_j,t) + \frac{u(x_{j+1},t) - u(x_j,t)}{h} \\ &= u_t(x_j,t) + \frac{u(x_j,t) + h\,u_x(x_j,t) + O\left(h^2\right) - u(x_j,t)}{h} \\ &= u_t(x_j,t) + u_x(x_j,t) + O(h) = O(h), \end{aligned} \tag{3.0.29}$$

lo cual indica que se trata efectivamente de un esquema consistente de orden 1.

Por último, el esquema centrado es consistente de orden 2:

$$\begin{aligned} \underline{u}'_j + \frac{\underline{u}_{j+1} - \underline{u}_{j-1}}{2h} &= u_t(x_j,t) + \frac{u(x_{j+1},t) - u(x_{j-1},t)}{2h} \\ &= u_t(x_j,t) + \Bigg[u(x_j,t) + hu_x(x_j,t) + \frac{h^2}{2}u_{xx}(x_j,t) + O(h^3) \\ &\qquad -u(x_j,t) + hu_x(x_j,t) - \frac{h^2}{2}u_{xx}(x_j,t) + O\left(h^3\right)\Bigg] \Big/ 2h, \\ &= u_t(x_j,t) + u_x(x_j,t) + O\left(h^2\right) = O\left(h^2\right). \end{aligned} \tag{3.0.30}$$

En virtud del Teorema de equivalencia de P. Lax que garantiza que la convergencia equivale a la consistencia más la estabilidad cabe entonces esperar que el

[4] Conviene subrayar que, a la hora de comprobar la consistencia de un método numérico, lo que comunmente se hace es considerar la solución del problema continuo como una solución aproximada del esquema discreto y no al revés, como podría esperarse en la medida en que el esquema numérico tiene como objeto aproximar la ecuación continua. Así, el error de truncatura es el resto resultante de considerar la solución del problema continuo como una solución aproximada del problema discreto. Cuando el error de truncatura τ es del orden de $O(h^p)$ se dice que el método es consistente de orden p.

esquema regresivo sea convergente de orden 1 y que el centrado sea convergente de orden 2.

Comprobémoslo. Consideremos en primer lugar el esquema regresivo y analicemos el error

$$\varepsilon_j(t) = \underline{u}_j(t) - u_j(t) = u(x_j,t) - u_j(t), \tag{3.0.31}$$

es decir la diferencia entre la solución real y la numérica sobre los puntos del mallado. Para simplificar la presentación suponemos que el dato inicial es continuo[5], lo cual permite tomar datos iniciales exactos en el esquema semi-discreto:

$$u_j(0) = f(x_j),\ j \in \mathbf{Z}. \tag{3.0.32}$$

En virtud del análisis de consistencia anterior, sustrayendo la ecuación verificada por $\underline{u}_j$ y u_j deducimos que

$$\begin{cases} \varepsilon_j' + \frac{\varepsilon_j - \varepsilon_{j-1}}{h} = O_j(h),\ j \in \mathbf{Z},\ t > 0 \\ \varepsilon_j(0) = 0,\ j \in \mathbf{Z}. \end{cases} \tag{3.0.33}$$

Multiplicando en (3.0.33) por ε_j y sumando en $j \in \mathbf{Z}$ obtenemos

$$\frac{1}{2}\frac{d}{dt}\left[\sum_{j\in\mathbf{Z}} |\varepsilon_j(t)|^2\right] + \frac{1}{h}\sum_{j\in\mathbf{Z}} \left(\varepsilon_j^2 - \varepsilon_{j-1}\varepsilon_j\right) = \sum_{j\in\mathbf{Z}} O_j(h)\varepsilon_j.$$

En este punto conviene observar que

$$\begin{aligned} \sum_{j\in\mathbf{Z}} \left(\varepsilon_j^2 - \varepsilon_{j-1}\varepsilon_j\right) &= \frac{1}{2}\sum_{j\in\mathbf{Z}} \left(\varepsilon_j^2 + \varepsilon_{j-1}^2 - 2\varepsilon_{j-1}\varepsilon_j\right) \\ &= \frac{1}{2}\sum_{j\in\mathbf{Z}} \left(\varepsilon_j - \varepsilon_{j-1}\right)^2 \geq 0. \end{aligned}$$

Por tanto la identidad de energía anterior puede reescribirse del siguiente modo

$$\frac{1}{2}\frac{d}{dt}\left[\sum_{j\in\mathbf{Z}} |\varepsilon_j(t)|^2\right] + \frac{1}{2h}\sum_{j\in\mathbf{Z}} \left(\varepsilon_j(t) - \varepsilon_{j-1}(t)\right)^2 = \sum_{j\in\mathbf{Z}} O_j(h)\varepsilon_j(t). \tag{3.0.34}$$

En este punto introducimos la norma en ℓ^2, el espacio de las sucesiones de cuadrado sumable a escala h:

$$||(\varepsilon_j)||_h = \left[h\sum_{j\in\mathbf{Z}} |\varepsilon_j|^2\right]^{1/2}. \tag{3.0.35}$$

[5]Si el dato inicial no fuese continuo sino solamente localmente integrable, por ejemplo, tomaríamos como dato inicial para el problema discreto una media del dato inicial $f = f(x)$ en torno a los puntos del mallado. Por ejemplo, $u_j(0) = \frac{1}{h}\int_{x_j-h/2}^{x_j+h/2} f(s)ds$.

En lo sucesivo utilizaremos la notación vectorial $\vec{\varepsilon}$ para denotar el vector infinito numerable de componentes $(\varepsilon_j)_{j\in\mathbf{Z}}$.

Conviene observar que (3.0.35) es una aproximación discreta de la norma continua en $L^2(\mathbf{R})$ en el mallado de paso h.

Con esta notación, y denotando mediante $\tau_{-1}\vec{\varepsilon}$ la sucesión trasladada una unidad con componentes $(\varepsilon_{j-1})_{j\in\mathbf{Z}}$, la identidad (3.0.34) puede reescribirse del modo siguiente:

$$\frac{1}{2}\frac{d}{dt}\left|\vec{\varepsilon}(t)\right|_h^2+\frac{1}{2h}\left|\vec{\varepsilon}(t)-\tau_{-1}\vec{\varepsilon}(t)\right|_h^2=h\sum_{j\in\mathbf{Z}}O_j(h)\varepsilon_j(t)\leq\left|\vec{O}(h)\right|_h\left|\varepsilon_j(t)\right|_h. \tag{3.0.36}$$

De esta desigualdad se deduce que

$$\frac{d}{dt}\left|\vec{\varepsilon}(t)\right|_h\leq\left|\vec{O}(h)\right|_h$$

de donde se sigue que

$$\left|\vec{\varepsilon}(t)\right|_h\leq\int_0^t\left|\vec{O}(h)\right|_h ds \tag{3.0.37}$$

puesto que $\vec{\varepsilon}(0)=0$.

En este punto tenemos que analizar el error de truncatura $\vec{O}(h)$. En vista del análisis de la consistencia realizado previamente se observa que cada componente $O_j(h)$ del error es de la forma

$$O_j(h)=\frac{h}{2}u_{xx}\left(\xi_j,t\right)$$

donde ξ_j es un punto en el intervalo $[x_{j-1},x_j]$.

Con el objeto de concluir la prueba de la convergencia suponemos que el dato inicial $f=f(x)$ es de clase C^2 y de soporte compacto: $f\in C_c^2(\mathbf{R})$. Entonces, la solución u, cuya forma explícita fue derivada en (3.0.3), tiene la misma propiedad para todo $t>0$ y además:

$$\max_{x\in\mathbf{R},\,t\geq 0}\left|u_{xx}(x,t)\right|=C<\infty,$$

de donde, habida cuenta que el soporte de u_{xx} está contenido en una traslación del soporte compacto de f, se sigue que

$$\left|\vec{O}(h)\right|_h\leq Ch,\ \forall t\geq 0,\ \forall h>0. \tag{3.0.38}$$

Combinando (3.0.37)-(3.0.38) se concluye que

$$\left|\vec{\varepsilon}(t)\right|_h\leq Cth,\ \forall t\geq 0,\ \forall h>0, \tag{3.0.39}$$

lo cual concluye la demostración de que el método semi-discreto de diferencias finitas regresivas es convergente de orden uno.

El método empleado en la prueba de la convergencia es el denominado método de la energía y está basado en la siguiente ley de energía que las soluciones del problema semi-discreto verifican

$$\frac{1}{2}\frac{d}{dt}\sum_{j\in\mathbf{Z}}|u_j(t)|^2 + h\sum_{j\in\mathbf{Z}}\left\{\frac{u_j(t)-u_{j-1}(t)}{h}\right\}^2 = 0$$

y que, con las notaciones anteriores, puede reescribirse como

$$\frac{1}{2}\frac{d}{dt}\,|\vec{u}(t)|_h^2 + h\,|\vec{u}(t)|_{1,h}^2 = 0. \tag{3.0.40}$$

Aquí y en lo sucesivo $\|\cdot\|_{1,h}$ denota la versión discreta de la semi-norma $\left(\int_{\mathbf{R}} u_x^2 dx\right)^{1/2}$, i.e.

$$|\vec{u}|_{1,h} = \left[h\sum_{j\in\mathbf{Z}}\left|\frac{u_j-u_{j-1}}{h}\right|^2\right]^{1/2}. \tag{3.0.41}$$

Conviene comparar (3.0.40) con la ley de conservación de la energía para la ecuación de transporte continua (3.0.1) donde, multiplicando por u e integrando con respecto a x se deduce que

$$\frac{d}{dt}\,\|\,u(t)\,\|_{L^2(\mathbf{R})}^2 = 0. \tag{3.0.42}$$

Obviamente, la ley de conservación de energía (3.0.42) para el problema continuo (3.0.1) es perfectamente coherente con la forma explícita (3.0.3) de la solución (3.0.1).

Sin embargo, es de señalar que, en contraste con la ley de conservación de energía (3.0.42) de la ecuación continua (3.0.1), la identidad (3.0.40) establece el carácter disipativo del esquema semi-discreto regresivo. Este carácter disipativo no está reñido con la convergencia del esquema cuando $h \to 0$, esencialmente por dos razones:

- La tasa de disipación del esquema semi-discreto decrece a medida que $h \to 0$, tal y como se observa con claridad en (3.0.40).
- El carácter disipativo del equema numérico contribuye a su estabilidad.

Verifiquemos la ley de energía de los otros dos esquemas considerados.

En el esquema progresivo tenemos

$$\frac{1}{2}\frac{d}{dt}\sum_{j\in\mathbf{Z}}|u_j(t)|^2+\sum_{j\in\mathbf{Z}}\frac{(u_{j+1}-u_j)}{h}u_j$$
$$=\frac{1}{2}\frac{d}{dt}\sum_{j\in\mathbf{Z}}|u_j(t)|^2-\frac{1}{2h}\sum_{j\in\mathbf{Z}}|u_{j+1}-u_j|^2=0. \tag{3.0.43}$$

En esta identidad queda claramente de manifiesto el carácter anti-disipativo del método progresivo, causante de su inestabilidad.

En el caso del esquema centrado tenemos sin embargo

$$\frac{d}{dt}\sum_{j\in\mathbf{Z}}|u_j(t)|^2=0, \tag{3.0.44}$$

identidad que garantiza su carácter puramente conservativo y su estabilidad.

Las propiedades disipativas, anti-disipativas y conservativas de los esquemas regresivo, progresivo y centrado pueden interpretarse fácilmente de la siguiente manera.

Consideremos por ejemplo el esquema regresivo en el que hemos adoptado la siguiente aproximación de la derivada espacial

$$u_x(x,t)\sim\frac{u(x,t)-u(x-h,t)}{h}.$$

Un análisis más cuidadoso indica que, en realidad,

$$\frac{u(x,t)-u(x-h,t)}{h}=u_x(x,t)-\frac{h}{2}u_{xx}(x,t)+O(h^2).$$

Por lo tanto, el esquema regresivo es en realidad una aproximación de orden dos de la ecuación de transporte perturbada

$$u_t+u_x-\frac{h}{2}u_{xx}=0. \tag{3.0.45}$$

La ecuación (3.0.45) es una aproximación parabólica o viscosa de la ecuación de transporte puro [6] (3.0.1). Multiplicando en (3.0.45) por u e integrando en x

[6] No es difícil comprobar que la ecuación (3.0.45) genera un semigrupo de contracciones en $L^2(\mathbf{R})$ para cada $h>0$ y que, dado un dato inicial $f\in L^2(\mathbf{R})$, la solución $u_h=u_h(x,t)$ de (3.0.45) converge a la solución de la ecuación de transporte puro $u(x,t)=f(x-t)$, cuando $h\to 0$ en $L^2(\mathbf{R})$ para cada $t>0$. Para ello basta observar que $v_h(x,t)=u_h(x+t,t)$ es solución de la ecuación del calor $v_t-hv_{xx}=0$ que, tras el cambio de variables $w_h(x,t)=v_h(x,t/h)$, se convierte en una solución de la ecuación del calor $w_t-w_{xx}=0$. Así, vemos que $v_h(x,t)=[G_h(t)*f](x)$ siendo G_h el núcleo del calor reescalado: $G_h(x,t)=(4\pi ht)^{-1/2}\exp(-x^2/4ht)$. De esta expresión se deduce fácilmente que $v_h(x,t)\to f(x)$ en $L^2(\mathbf{R})$, para cada $t>0$, o, lo que es lo mismo, $u_h(x,t)\to f(x-t)$.

deducimos que

$$\frac{1}{2}\frac{d}{dt}\int_{\mathbf{R}} u^2(x,t)dx + \frac{h}{2}\int_{\mathbf{R}} u_x^2(x,t)dx = 0, \tag{3.0.46}$$

lo cual refleja el carácter disipativo del término de regularización $-hu_{xx}/2$ añadido en la ecuación (3.0.45) y supone, claramente, la versión continua de la ley de disipación de energía (3.0.40) del esquema regresivo.

El mismo argumento permite detectar el carácter inestable de la aproximación progresiva puesto que

$$\frac{u(x+h,t)-u(x,t)}{h} = u_x(x,t) + \frac{h}{2}u_{xx}(x,t) + O(h^2). \tag{3.0.47}$$

En este caso, el esquema progresivo resulta ser una aproximación de orden dos de la EDP de segundo orden

$$u_t + u_x + \frac{h}{2}u_{xx} = 0. \tag{3.0.48}$$

En esta ocasión (3.0.48) es una ecuación parabólica retrógrada de carácter inestable[7]

tal y como queda de manifiesto en la ley de amplificación de la energía que las soluciones de (3.0.48) satisfacen

$$\frac{1}{2}\frac{d}{dt}\int_{\mathbf{R}} u^2(x,t)dx = \frac{h}{2}\int_{\mathbf{R}} u_x^2(x,t)dx. \tag{3.0.49}$$

Sin embargo, este argumento permite confirmar el carácter puramente conservativo de la aproximación centrada. En efecto:

$$\frac{u(x+h,t)-u(x-h,t)}{2h} = u_x(x,t) + \frac{h^2}{3!}\partial_x^3 u(x,t) + \cdots + \frac{h^{2\ell}}{(2\ell+1)!}\partial_x^{2\ell+1}u(x,t) + \cdots . \tag{3.0.50}$$

Es fácil comprobar, en efecto, que cualquiera de las aproximaciones de la ecuación de transporte (3.0.1) obtenidas truncando el desarrollo en serie de potencias (3.0.50) de la forma

$$u_t + \sum_{\ell=0}^{L} \frac{h^{2\ell}}{(2\ell+1)!}\partial_x^{2\ell+1} = 0 \tag{3.0.51}$$

[7] La inestabilidad de esta ecuación a medida que $h \to 0$ se pone claramente de manifiesto a través del cambio de variable $v_h(x,t) = u_h(x+t,t)$. En este caso, se trata de una solución de la ecuación del calor retrógrada $v_t + hv_{xx} = 0$ que, tras el cambio de variables $w_h(x,t) = v_h(x,t/h)$, se convierte en una solución de la ecuación del calor retrógrada normalizada $w_t + w_{xx} = 0$. Así, vemos que $v_h(x,t) = [G_h(\tau - t) * v_h(\cdot,\tau)](x)$, para cada par de instantes de tiempo $0 < t < \tau$, siendo G_h el núcleo del calor reescalado: $G_h(x,t) = (4\pi ht)^{-1/2}\exp(-x^2/4ht)$. De esta expresión, aplicada con $t = 0$ de modo que $v_h(x,t) = f(x)$, se deduce fácilmente que $v_h(x,t)$ no está acotada en $L^\infty(0,T;L^2(\mathbf{R}))$, para ningún $T > 0$.

Tiene un carácter puramente conservativo.

Las ecuaciones (3.0.51) tienen sin embargo un carácter dispersivo que analizaremos más adelante.

En relación a la ecuación de transporte (3.0.5) en la que el sentido de progresión de las ondas ha sido invertido, como es de esperar, se tiene que el esquema regresivo es inestable y no converge mientras que el progresivo y centrado son convergentes de orden 1 y 2, respectivamente.

Para concluir esta sección consideremos el siguiente esquema completamente discreto para la aproximación de (3.0.1):

$$\frac{u_j^{k+1} - u_j^{k-1}}{2\Delta t} + \frac{u_{j+1}^k - u_{j-1}^k}{2\Delta x} = 0. \tag{3.0.52}$$

Aquí y en lo sucesivo utilizamos las notaciones habituales de modo que Δt y Δx denotan los pasos del mallado en la dirección temporal y espacial respectivamente. Por otra parte, u_j^k denota la aproximación de la solución continua $u = u(x,t)$ de (3.0.1) en el punto $(x,t) = (x_j, t_k) = (j\Delta x, k\Delta t)$.

El esquema (3.0.52) está perfectamente centrado tanto en la variable espacial como temporal y se denomina esquema "leap-frog".

Se trata de un esquema consistente de orden 2 y puede ser escrito en la forma

$$u_j^{k+1} = u_j^{k-1} + \mu \left[u_{j-1}^k - u_{j+1}^k \right] \tag{3.0.53}$$

donde μ es el número de Courant:

$$\mu = \Delta t / \Delta x. \tag{3.0.54}$$

El método de von Neumann permite analizar fácilmente la estabilidad del esquema. En este caso, la transformada de Fourier $\check{u}^k(\theta)$ de la solución de (3.0.53) satisface

$$\check{u}^{k+1}(\theta) = \check{u}^{k-1}(\theta) + \mu \left[e^{-i\theta} - e^{i\theta} \right] \check{u}^k(\theta) = \check{u}^{k-1}(\theta) - 2i\mu \operatorname{sen}(\theta) \check{u}^k(\theta),$$

es decir,

$$\check{u}^{k+1}(\theta) + 2i\mu \operatorname{sen}(\theta) \check{u}^k(\theta) - \check{u}^{k-1}(\theta) = 0. \tag{3.0.55}$$

En (3.0.55) vemos que cada componente de Fourier $\check{u}^k(\theta)$ satisface un esquema de evolución discreto de dos pasos cuyos coeficientes dependen de $\theta \in [0, 2\pi)$. Basta por tanto verificar si se satisface el criterio de la raíz. En este caso los ceros del polinomio característico del esquema (3.0.55) son

$$\lambda_\pm(\theta) = -i\mu \operatorname{sen}(\theta) \pm \sqrt{-\mu^2 \operatorname{sen}^2(\theta) + 1}. \tag{3.0.56}$$

Conviene entonces distinguir los tres siguientes casos:

- *Caso 1:* $\mu < 1$.

  En este caso

  $$|\lambda_\pm|^2 = \left[\mu^2 \operatorname{sen}^2\theta + 1 - \mu^2 \operatorname{sen}^2\theta\right] = 1$$

  con lo cual la estabilidad queda garantizada al ser las raices $\lambda_\pm$ simples.

- *Caso 2:* $\mu = 1$.

  En este caso límite se observa que cuando $\theta = \pi/2$ o $\theta = 3\pi/2$ el discriminante se anula. Tenemos entonces raíces dobles de módulo unidad, lo cual produce la inestabilidad del esquema.

- *Caso 3:* $\mu > 1$.

  En este caso, cuando $\theta \sim \pi/2$ tenemos que

  $$-4\mu^2 \operatorname{sen}^2(\theta) + 4 < 0$$

  y por lo tanto los ceros son de la forma

  $$\lambda_\pm(\theta) = -i\left[\mu \operatorname{sen}\theta \mp \sqrt{\mu^2 \operatorname{sen}^2\theta - 1}\right].$$

  La raíz de mayor módulo es la que corresponde al signo negativo. En este caso tenemos

  $$|\lambda_-(\theta)| = \mu \operatorname{sen}\theta + \sqrt{\mu^2 \operatorname{sen}^2\theta - 1} > 1$$

  puesto que $\mu \operatorname{sen}\theta > 1$.

  El método es por tanto inestable en este caso.

De este análisis deducimos que el método completamente discreto de leap-frog es convergente de orden dos si y sólo si $\mu < 1$.

Es fácil comprobar también que $\mu \leq 1$ es precisamente la condición que garantiza que el dominio de dependencia del esquema discreto contiene el de la ecuación continua.

Señalemos por último que el método consistente en sustituir el esquema numérico por una aproximación semejante escrita en términos de EDP puede también aplicarse en este caso. Obtendríamos ahora aproximaciones conservativas pero dispersivas de la ecuación de transporte de la forma

$$\sum_{m=0}^{M} \frac{(\Delta t)^{2m}}{(2m+1)!}\partial_t^{2m+1} + \sum_{\ell=0}^{L} \frac{(\Delta x)^{2\ell}}{(2\ell+1)!}\partial_x^{2\ell+1} = 0 \tag{3.0.57}$$

Tomando por ejemplo $L = M = 1$ obtenemos la ecuación:

$$\partial_t u + \partial_x u + \frac{(\Delta t)^2}{6}\partial_t^3 u + \frac{(\Delta x)^2}{6}\partial_x^3 u = 0.$$

Ahora bien, la ecuación de transporte indica que $\partial_t u = -\partial_x u$ y por tanto $\partial_t^2 = \partial_x^2$, de modo que la ecuación anterior puede escribirse del modo siguiente:

$$\partial_t[u + \frac{(\Delta t)^2}{6}\partial_x^2 u] + \partial_x[u + \frac{(\Delta x)^2}{6}\partial_x^2 u] = 0.$$

En esta última expresión es fácil comprobar el carácter conservativo de estas aproximaciones. En efecto, multiplicando en la ecuación por u e integrando en $\mathbb{R}$ obtenemos:

$$\int_{\mathbb{R}} \partial_t\Big(u + \frac{(\Delta t)^2}{6}\partial_x^2 u\Big)u dx + \int_{\mathbb{R}} \partial_x\Big(u + \frac{(\Delta x)^2}{6}\partial_x^2 u\Big)u dx + 0.$$

Ahora bien, tenemos,

$$\int_{\mathbb{R}} \partial_t u u dx = \frac{1}{2}\frac{d}{dt}\int_{\mathbb{R}} u^2 dx; \quad \int_{\mathbb{R}} \partial_x^3 u\, u dx = -\int_{\mathbb{R}} \partial_x^2 u\, \partial_x u\, dx = 0.$$

$$\int_{\mathbb{R}} \partial_t \partial_x^2 u\, u dx = -\int_{\mathbb{R}} \partial_t \partial_x u \partial_x u dx = -\frac{1}{2}\frac{d}{dt}\int_{\mathbb{R}} |\, \partial_x u\,|^2\, dx; \quad \int_{\mathbb{R}} \partial_x u\, u dx = 0.$$

Obtenemos así la ley de conservación de la energía:

$$\frac{d}{dt}\Big[\frac{1}{2}\int_{\mathbb{R}}\Big[u^2 - \frac{(\Delta t)^2}{6}\,|\,u_x\,|^2\,\Big]dx\Big] = 0$$

Vemos sin embargo que el efecto dispersivo introducido por el esquema numérico hace que no sea la norma de u en $L^2(\mathbb{R})$ la que se conserve en tiempo sino la cantidad:

$$\int_{\mathbb{R}}\Big(u^2 - \frac{(\Delta t)^2}{6}\,|\,u_x\,|^2\,\Big)dx$$

que, incluso puede ser negativa si la función u oscila rápidamente. Conviene sin embargo no olvidar que en las soluciones numéricas su máxima oscilación está limitada por el paso del mallado, por lo que esta cantidad nunca se puede hacer negativa en ellas.

Son muchos los esquemas completamente discretos que surgen de manera natural en la aproximación de la ecuación de transporte (3.0.1), además del esquema "leap-frog" ya estudiado. La mayoría de ellos aparecen al realizar una aproximación discreta en tiempo de un esquema semidiscreto, pero no siempre es así. Obviamente, en caso de proceder a la obtención del esquema completamente discreto mediante la discretización temporal de un esquema semi-discreto,

elegiremos uno que sea convergente puesto que si el esquema semi-discreto de partida fuese divergente, el esquema completamente discreto obtenido tampoco convergería.

En vista de este hecho, conviene excluir inmediatamente los esquemas completamente discretos derivados del esquema semi-discreto progresivo (3.0.10) puesto que ya vimos que es inestable y por tanto divergente.

Sin embargo, como el esquema regresivo (3.0.11) es convergente parece natural introducir el esquema de Euler regresivo

$$\frac{u_j^{k+1}-u_j^k}{\Delta t}+\frac{u_j^k-u_{j-1}^k}{\Delta x}=0 \tag{3.0.58}$$

o su versión implícita

$$\frac{u_j^{k+1}-u_j^k}{\Delta t}+\frac{u_j^{k+1}-u_{j-1}^{k+1}}{\Delta x}=0 \tag{3.0.59}$$

Nos referiremos a estos esquemas con ER (Euler regresivo explícito) y ERI (Euler regresivo implícito), respectivamente.

Ambos esquemas son de un paso temporal y consistentes de orden uno. Comprobemos pues su estabilidad.

El esquema ER puede reescribirse como

$$u_j^{k+1}=u_j^k+\mu(u_{j-1}^k-u_j^k). \tag{3.0.60}$$

El análisis de von Neumann conduce al esquema discreto

$$\breve{u}^{k+1}(\theta)=\breve{u}^k(\theta)+\mu(e^{i\theta}\breve{u}^k(\theta)-\breve{u}^k(\theta))=\Big[1+\mu(e^{i\theta}-1)\Big]\breve{u}^k(\theta) \tag{3.0.61}$$

Para su estabilidad basta entonces comprobar si $\mid 1+\mu(e^{i\theta}-1)\mid\leqslant 1$. Como

$$1+\mu(e^{i\theta}-1) \;=\; 1+\mu[\cos\theta+i\,\mathrm{sen}\,\theta-1]=1+\mu(\cos\theta-1)+i\mu\,\mathrm{sen}\,\theta$$

tenemos que

$$\begin{aligned}\mid 1+\mu(e^{i\theta}-1)\mid^2 &= (1+\mu(\cos\theta-1))^2+\mu^2\,\mathrm{sen}^2\,\theta\\ &= 1+\mu^2(\cos\theta-1)^2+2\mu(\cos\theta-1)+\mu^2\,\mathrm{sen}^2\,\theta\\ &= 1+\mu^2(\cos^2\theta+1-2\cos\theta)+2\mu(\cos\theta-1)+\mu^2\,\mathrm{sen}^2\,\theta=1-2\mu\end{aligned}$$

de donde deducimos que es estable, y por tanto convergente de orden uno, si y sólo si

$$\mid 1-2\mu\mid\leqslant 1\Leftrightarrow\mu\leqslant 1. \tag{3.0.62}$$

Es fácil comprobar que esta condición de estabilidad es precisamente la que se obtiene al imponer que el dominio de dependencia del esquema discreto contenga al de la ecuación de transporte continua.

En el caso del ERI tenemos

$$u_j^{k+1} = u_j^k - \mu(u_j^{k+1} - u_{j-1}^{k+1})$$

que al aplicar la transformada de Fourier, se convierte en

$$\check{u}^{k+1}(\theta) = \check{u}^k(\theta) - \mu(\check{u}^{k+1}(\theta) - e^{i\theta}\check{u}^{k+1}(\theta)), \tag{3.0.63}$$

es decir,

$$\Big[1 + \mu(1 - e^{i\theta})\Big]\check{u}^{k+1}(\theta) = \check{u}^k(\theta),$$

o

$$\check{u}^{k+1}(\theta) = [1 + \mu(1 - e^{i\theta})]^{-1}\check{u}^k(\theta). \tag{3.0.64}$$

La condición de estabilidad es entonces en este caso $\mid 1 + \mu(1 - e^{i\theta}) \mid \geqslant 1$. Como

$$1 + \mu(1 - e^{i\theta}) = 1 + \mu(1 - \cos\theta) - i\mu \operatorname{sen}\theta$$

tenemos que

$$\begin{aligned} \mid 1 + \mu(1 - e^{i\theta}) \mid^2 &= (1 + \mu(1 - \cos\theta))^2 + \mu^2 \operatorname{sen}^2\theta \\ &= 1 + \mu^2(1 + \cos^2\theta - 2\cos\theta) + 2\mu(1 - \cos\theta) + \mu^2 \operatorname{sen}^2\theta \\ &= 1 + 2\mu(1 - \cos\theta) + 2\mu^2(1 - \cos\theta) \geqslant 1 \end{aligned}$$

y por tanto el método es incondicionalmente estable.

A primera vista puede resultar sorprendente que el método ERI sea convergente para cualquier valor del número de Courant pues cabría preguntarse si la condición de inclusión de los dominios de dependencia se cumple con independencia del valor de μ. Esto es efectivamente así puesto que en el esquema discreto (3.0.59) el cálculo de u_j^{k+1} involucra a u_{j-1}^{k+1}, cuyo valor a su vez involucra a $u_{j-2}^{k+1}, \cdots$. Vemos pues que el dominio de dependencia de ERI es el conjunto de todos los nodos del mallado, con independencia del valor de μ.

De hecho cabe preguntarse sobre cual es el modo de resolver el sistema (3.0.59). Este sistema, con la notación vectorial habitual puede escribirse en la forma

$$B_\mu \overrightarrow{u}^{k+1} = \overrightarrow{u}^k$$

donde B_μ es una matriz infinita con valores $1 + \mu$ en la subdiagonal. Se trata por tanto de una matriz infinita "triangular inferior" que define un operador acotado de ℓ^2 en ℓ^2. Pero, >se puede invertir el operador B_μ?

Para comprobar que esto es efectivamente así, conviene utilizar el análisis de von Neumann. En efecto, el sistema equivalente (3.0.63) se resuelve inmediatamente y tiene como solución (3.0.64). Además, tal y como hemos visto en el análisis de la estabilidad del esquema

$$| \breve{u}^{k+1}(\theta) | \leqslant | \breve{u}^{k}(\theta) |, \quad \forall \theta \in [0, 2\pi).$$

Deducimos por tanto que

$$\begin{aligned} \| (u_j^{k+1})_{j\in\mathbb{Z}} \|_{\ell^2}^2 &= \sum_{j\in\mathbb{Z}} | u_j^{k+1} |^2 = \frac{1}{2\pi} \int_0^{2\pi} | \breve{u}^{k+1}(\theta) |^2 \, d\theta \\ &\leqslant \tfrac{1}{2\pi} \int_0^{2\pi} | \breve{u}^{k}(\theta) |^2 \, d\theta = \sum_{j\in\mathbb{Z}} | u_j^{k} |^2 = \| (u_j^{k})_{j\in\mathbb{Z}} \|_{\ell^2}^2 \end{aligned}$$

de modo que B_μ^{-1} está bien definido y es un operador acotado de ℓ^2 en ℓ^2 con norma no superior a uno. Vemos pues que la transformada discreta de Fourier permite probar la resolubilidad del sistema algebráico (3.0.59) que el método ERI plantea.

Evidentemente hay muchos otros esquemas que pueden considerarse. Por ejemplo, el esquema de Crank-Nicolson (CN) inspirado en la regla del trapecio para la resolución de ecuaciones diferenciales y en la diferencia finita centrada para la aproximación de la derivada espacial:

$$\frac{u_j^{k+1} - u_j^k}{\Delta t} = -\frac{1}{2}\Big[\frac{u_{j+1}^k - u_{j-1}^k}{2\Delta x} + \frac{u_{j+1}^{k+1} - u_{j-1}^{k+1}}{2\Delta x}\Big]. \tag{3.0.65}$$

El método CN es convergente de orden dos para cualquier valor del parámetro de Courant μ. Vemos pues que CN preserva la propiedad que el método ERI de converger para todo valor de μ, pero tiene además la propiedad de ser de orden dos. El orden dos proviene de la combinación de los dos hechos siguientes: a) La utilización de diferencias centradas en la aproximación de la derivada espacial, lo cual da, efectivamente, una aproximación de orden dos de la derivada espacial; b) La utilización del método del trapecio en la aproximación de la derivada temporal, que es también un método de orden dos, aunque esta vez en tiempo.

Nuevamente (3.0.65) es un sistema implícito. Pero se puede ver que es resoluble utilizando el argumento que hemos usado para el método ERI, mediante la transformada discreta de Fourier.

Existen otros muchos métodos que proporcionan aproximaciones convergentes de la ecuación de transporte. Tenemos por ejemplo de "leap-frog" de orden cuatro (LF4):

$$\frac{u_j^{k+1} - u_j^{k-1}}{2\Delta t} = -\Big[\frac{4}{3}\Big[\frac{u_{j+1}^k - u_{j-1}^k}{2\Delta x}\Big] - \frac{1}{3}\Big[\frac{u_{j+2}^k - u_{j-2}^k}{4\Delta x}\Big]\Big] \tag{3.0.66}$$

que es de orden dos en tiempo y orden cuatro en espacio.

En todos los métodos descritos hasta ahora puede observarse que han sido derivados en dos pasos, discretizando primero la variable espacial y después la temporal. Sin ir más lejos es obvio que (3.0.66) proviene de una semi-discretización de la forma

$$u_j'(t) = -\Big[\frac{4}{3}\Big[\frac{u_{j+1}(t) - u_{j-1}(t)}{2\Delta x}\Big] - \frac{1}{3}\Big[\frac{u_{j+2}(t) - u_{j-2}(t)}{4\Delta x}\Big]\Big] \tag{3.0.67}$$

que es efectivamente consistente con la ecuación de transporte. El paso de (3.0.67) a (3.0.66) es claro. Basta con utilizar el esquema de dos pasos

$$\frac{y^{k+1} - y^{k-1}}{2\Delta t} = f(y^k)$$

para la resolución de la ecuación diferencial

$$y'(t) = f(y(t)).$$

Pero, como decíamos, no todos los métodos discretos provienen de discretizar en tiempo una semi-discretización. Por ejemplo el método de la derivada oblicua es de la forma

$$u_j^{k+2} = (1 - 2\mu)\Big(u_j^{k+1} - u_{j-1}^{k+1}\Big) + u_{j-1}^k. \tag{3.0.68}$$

Se trata de un método de orden dos. Para ver que, efectivamente, es un método consistente con la ecuación de transporte lo escribimos como

$$\frac{u_j^{k+2} - u_{j-1}^k - u_j^{k+1} + u_{j-1}^{k+1}}{\Delta t} = -2\frac{(u_j^{k+1} - u_{j-1}^{k+1})}{\Delta x},$$

o, de manera más clara aún,

$$\frac{1}{2}\Big[\frac{u_j^{k+2} - u_j^{k+1}}{\Delta t} + \frac{u_{j-1}^{k+1} - u_{j-1}^k}{\Delta t}\Big] + \frac{u_j^{k+1} - u_{j-1}^{k+1}}{\Delta x} = 0, \tag{3.0.69}$$

expresión en la que queda claramente de manifiesto la analogía del esquema discreto con la ecuación de transporte continua.

Citemos por último los esquemas de Lax-Wendroff

$$u_j^{k+1} = \frac{1}{2}\mu(1+\mu)u_{j-1}^k + (1-\mu^2)u_j^k - \frac{1}{2}\mu(1-\mu)u_{j+1}^k \tag{3.0.70}$$

y el esquema de Lax-Friedrichs

$$u_j^{k+1} = \frac{1}{2}(1-\mu)u_{j-1}^k + \frac{1}{2}(1+\mu)u_{j+1}^k. \tag{3.0.71}$$

El esquema de Lax-Friedrichs puede escribirse en la forma

$$\frac{u_j^{k+1} - \frac{1}{2}(u_{j-1}^k + u_{j+1}^k)}{\Delta t} = -\frac{u_{j+1}^k - u_{j-1}^k}{2\Delta x} \tag{3.0.72}$$

en la que queda claramente de manifiesto la consistencia con la ecuación de transporte. En esta expresión se observa que este esquema se obtiene a partir de la semi-discretización centrada realizando una discretización explícita de la derivada temporal en la que el valor u_j^k de la discretización más típica de u_t, (i.e. $(u_j^{k+1} - u_j^k)/\Delta t$) ha sido sustituido por la media de los valores u_{j-1}^k y u_{j+1}^k.

Es fácil comprobar que (3.0.72) es una aproximación difusiva de la ecuación de transporte. En efecto, basta aplicar formalmente en (3.0.72) el desarrollo de Taylor para observar que (3.0.72) da lugar en realidad a la siguiente corrección difusiva de la ecuación de transporte:

$$u_t - \frac{(\Delta x)^2}{2\Delta t} u_{xx} + u_x = 0,$$

que es análoga a la que obtuvimos para el esquema semi-discreto regresivo.

Obviamente, se trata de un esquema discreto. Es de orden uno y tiene la propiedad de conservar la masa de la solución. En efecto, definimos la masa de la solución discreta como

$$m^k = \sum_{j \in \mathbb{Z}} u_j^k.$$

Basta entonces sumar con respecto a $j \in \mathbb{Z}$ en (3.0.71) para obtener que las soluciones del esquema de Lax-Friedrichs satisfacen $m^{k+1} = m^k$. Esto es una versión discreta de la propiedad de conservación de la masa que las soluciones $u(x,t) = f(x-t)$ de (3.0.1) satisfacen, i.e.

$$\int_{\mathbb{R}} u(x,t)dx = \int_{\mathbb{R}} f(x)dx, \quad \forall t \in \mathbb{R}.$$

Esta propiedad de conservación de la masa juega un papel relevante en la aproximación numérica de las ecuaciones de transporte no-lineales, como la ecuación de Burgers, y los esquemas que la verifican se dicen *conservativos*.

Es fácil comprobar que el esquema de Lax-Friedrichs es estable (y, por tanto, convergente de orden uno) si y sólo

$$\mu = \Delta t/\Delta x \leqslant 1.$$

El esquema de Lax-Wendroff es consistente de orden dos y es fácil comprobar que es también un esquema conservativo y es convergente bajo la misma condición $\mu \leqslant 1$.

3.1. Dispersión numérica y velocidad de grupo

En el apartado anterior hemos estudiado la convergencia de diversos esquemas semi-discretos y completamente discretos de aproximación de la ecuación

continua (3.0.1). Hemos comprobado que esquemas numéricos convergentes pueden introducir efectos disipativos o anti-disipativos que pueden ser respectivamente la causa de su estabilidad o inestabilidad o, por el contrario, ser puramente conservativos.

Sin embargo con independencia de su carácter convergente o divergente la mayoría de los esquemas numéricos tienen un carácter dispersivo. Por dispersión entendemos la propiedad de un sistema dinámico continuo o discreto en tiempo de propagar a diferentes velocidades las diversas componentes de la solución.

La ecuación de transporte (3.0.1) es precisamente un ejemplo claro de sistema no dispersivo pues, como vimos en la sección 3, todas sus soluciones son ondas de transporte puras que se propagan en el espacio a velocidad uno. Este hecho, obvio de la expresión explícita de la solución (3.0.3), puede también comprobarse a través del análisis de Fourier.

En efecto, consideremos soluciones u de (3.0.1) de la forma

$$u = e^{i\omega t} e^{i\xi x}, \tag{3.1.1}$$

es decir soluciones sinusoidales en variables separadas de frecuencia temporal ω y longitud de onda espacial $2\pi/\xi$, i.e. número de onda ξ.

Es fácil comprobar que u de la forma (3.1.1) es solución de (3.0.1) si y sólo si

$$\omega = -\xi. \tag{3.1.2}$$

En este caso la solución (3.1.1) adquiere la forma

$$u(x,t) = e^{i\omega(t-x)} \tag{3.1.3}$$

y se confirma lo observado en (3.0.3) en el sentido que las soluciones de (3.0.1) son meras ondas de transporte progresivas con velocidad uno.

La relación (3.1.2) es la que se denomina *relación de dispersión* para la ecuación de transporte (3.0.1).

Analicemos ahora, por ejemplo, el esquema semi-discreto regresivo que, como vimos en la sección anterior, es convergente de orden uno:

$$u_j' + \frac{u_j - u_{j-1}}{h} = 0. \tag{3.1.4}$$

Buscamos ahora soluciones de la forma

$$u_j(t) = e^{i\omega t} e^{i\xi x_j}. \tag{3.1.5}$$

Llevando la expresión (3.1.5) a la ecuación (3.1.4) obtenemos la ecuación

$$i\omega + \frac{1 - e^{-i\xi h}}{h} = 0,$$

es decir

$$\omega = \frac{i}{h}\left[1 - e^{-i\xi h}\right]. \tag{3.1.6}$$

La ecuación (3.1.6) es la relación de dispersión para el esquema semi-discreto (3.1.4).

Un simple desarrollo de Taylor permite comprobar que, en una primera aproximación, (3.1.6) coincide con la relación de dispersión (3.1.2) de la ecuación de transporte continua. En efecto,

$$\frac{i}{h}\left[1 - e^{-i\xi h}\right] = \frac{i}{h}\left[1 - \left[1 - i\xi h - \frac{\xi^2 h^2}{2} + O(h^3)\right]\right] = -\xi + \frac{i\xi^2 h}{2} + O(h^2). \tag{3.1.7}$$

En virtud de (3.1.6) la solución (3.1.5) de (3.1.4) puede escribirse en la forma

$$u_j(t) = e^{i\xi(x_j + \frac{\omega}{\xi}t)},$$

de donde vemos que la solución semi-discreta es una onda de transporte progresiva que avanza a una velocidad

$$c_h(\xi) = -\frac{\omega_h(\xi)}{\xi} = -\frac{i}{h\xi}(1 - e^{-i\xi h}) = 1 - \frac{i\xi h}{2} + O(h^2), \tag{3.1.8}$$

denominada *velocidad de fase*.

En la expresión (3.1.8) queda claramente de manifiesto el carácter dispersivo de la ecuación semi-discreta, en la medida en que la velocidad de propagación de la onda depende de la longitud de la misma.

Pero, cabría argumentar que la expresión (3.1.8) es un número complejo, por lo que no representa realmente una velocidad de transporte en el espacio físico. Esto es debido al efecto disipativo que el esquema (3.1.4) introduce y que quedó claramente de manifiesto en su análogo continuo (3.0.45).

Consideremos ahora el esquema centrado

$$u_j' + \frac{u_{j+1} - u_{j-1}}{2h} = 0, \tag{3.1.9}$$

que, como vimos en la sección 3, es convergente de orden 2 y puramente conservativo.

En este caso se obtiene la relación de dispersión

$$i\omega + \frac{e^{i\xi h} - e^{-i\xi h}}{2h} = 0.$$

Es decir,

$$i\omega + \frac{i\,\mathrm{sen}(\xi h)}{h} = 0$$

o, equivalentemente,

$$\omega = -\frac{\operatorname{sen}(h\xi)}{h}. \tag{3.1.10}$$

Nuevamente observamos que (3.1.10) es una aproximación de la relación de dispersión (3.1.2) de la ecuación de transporte continua. De (3.1.10) se deduce que la velocidad de propagación de las ondas semi-discretas es en este caso

$$c_h(\xi) = \frac{\operatorname{sen}(\xi h)}{\xi h} = 1 - \frac{\xi^2 h^2}{3!} + O(h^4). \tag{3.1.11}$$

Comprobamos por lo tanto que las ondas en el medio semi-discreto se propagan más lentamente que en el medio continuo si bien, fijada la longitud de onda espacial, la velocidad de propagación $c_h(\xi)$, cuando $h \to 0$, converge a la velocidad de propagación en el caso continuo $c \equiv 1$. Obviamente, la convergencia de las velocidades de propagación está motivada por el hecho que el esquema sea convergente. En efecto, el esquema numérico no podría ser convergente si para algunas longitudes de onda las velocidades de propagación no convergiesen cuando el paso del mallado tiende a cero.

Consideremos por último el esquema "leap-frog" completamente discreto (3.0.52). En este caso buscamos ondas discretas de la forma

$$u_j^k = e^{i\omega\Delta t_k} e^{i\xi x_j} = e^{i\omega k\Delta t} e^{i\xi j\Delta x}. \tag{3.1.12}$$

Obtenemos entonces la relación de dispersión:

$$\frac{e^{i\omega\Delta t} - e^{-i\omega\Delta t}}{2\Delta t} + \frac{e^{i\xi\Delta x} - e^{-i\xi\Delta x}}{2\Delta x} = 0$$

que, en función del número de Courant $\mu = \Delta t/\Delta x$, puede reeescribirse como

$$\operatorname{sen}(\omega\Delta t) = -\mu \operatorname{sen}(\xi\Delta x),$$

o, de otro modo,

$$\omega = -\frac{1}{\Delta t}\operatorname{arcsen}\left[\mu \operatorname{sen}(\xi\Delta x)\right]. \tag{3.1.13}$$

Nuevamente es evidente que a medida que $\Delta x \to 0$, $\Delta t \to 0$ la relación de dispersión (3.1.13) se aproxima a la de la ecuación de transporte continua.

El caso

$$\mu = 1 \tag{3.1.14}$$

es particularmente interesante puesto que la relación de dispersión (3.1.13) se reduce a

$$\omega = -\xi \tag{3.1.15}$$

que es precisamente la correspondiente a la ecuación de transporte continua. En este caso las ondas discretas se propagan a velocidad constante idénticamente igual a uno, como lo hacen en el caso continuo.

Con el objeto de entender esta coincidencia de las velocidades de propagación continua y discretas conviene reescribir el esquema discreto con $\mu = 1$. Se obtiene en este caso

$$u_j^{k+1} - u_j^{k-1} + u_{j+1}^k - u_{j-1}^k = 0.$$

Es decir

$$u_j^{k+1} + u_{j+1}^k = u_j^{k-1} + u_{j-1}^k. \tag{3.1.16}$$

Habida cuenta que las soluciones de la ecuación de transporte continua son de la forma $u = f(x - t)$, se comprueba que, en este caso, son también soluciones exactas del esquema discreto (3.1.16) con $\mu = 1$. El esquema numérico es por tanto en este caso de orden infinito y reproduce de manera exacta las soluciones de la ecuación de transporte continua sobre los puntos del mallado.

Pero esto ocurre sólo cuando $\mu = 1$. Cuando $0 < \mu < 1$ el esquema es convergente pero también dispersivo. En efecto, en este caso la velocidad de propagación es

$$c_h(\xi) = \frac{1}{\Delta t \xi} \text{arcsen}\left[\mu \,\text{sen}(\xi \Delta x)\right]. \tag{3.1.17}$$

Nuevamente observamos que, para cualquier valor del número de Courant $0 < \mu < 1$:

$$\bullet\ c_h(\xi) \to 1,\ h \to 0,\ \forall \xi;$$

$$\bullet\ \mid c_h(\xi) \mid < 1,\ \forall h > 0,\ \forall \xi.$$

La velocidad $c_h(\xi)$ describe de manera adecuada la propagación de las ondas semi-discretas o discretas que involucran un solo modo de Fourier. Son las que llamaremos ondas monocromáticas. Pero, tal y como vimos en la sección 2 en el marco del análisis del movimiento armónico simple, cuando se superponen dos ondas con velocidades de propagación semejantes pero no idénticas surgen paquetes de ondas que pueden propagarse a velocidades distintas. Con el objeto de entender este fenómeno es conveniente introducir la noción de *velocidad de grupo.*

Para introducir esta noción consideremos cualquiera de los anteriores esquemas semi-discretos que admite soluciones de la forma

$$u_j(t) = e^{i\omega_h(\xi)t} e^{i\xi x_j}. \tag{3.1.18}$$

Superponiendo dos soluciones de esta forma con longitudes de ondas ξ y $\xi + \Delta\xi$ respectivamente obtenemos una nueva solución

$$u_{\Delta\xi,j}(t) = \frac{e^{i\omega_h(\xi)t}e^{i\xi x_j} - e^{i\omega_h(\xi+\Delta\xi)t}e^{i(\xi+\Delta\xi)x_j}}{\Delta\xi}$$

cuyo límite, cuando $\Delta\xi \to 0$, viene dado por

$$w_j(t) = -i[\omega_h'(\xi)t + x_j]e^{i\omega_h(\xi)t}e^{i\xi x_j}.$$

El resultado es un nuevo tipo de onda, producto de la solución (3.1.18) que se propaga a la velocidad de fase habitual $c_h(\xi)$ con la onda $g(x,t) = -i[\omega_h'(\xi)t + x_j]$ que se propaga a velocidad $\omega_h'(\xi)$ que se denomina *velocidad de grupo*.

La velocidad de grupo es la que determina la propagación de paquetes de ondas conteniendo varias ondas de números de onda semejantes. Para comprobar este hecho basta con considerar la solución que se obtendría a partir de un dato inicial $f = f(x)$ con transformada de Fourier $F(\xi)$. La solución tendría entonces la expresión [8]:

$$u(x,t) = \int_{-\infty}^{+\infty} F(\xi)e^{i(\omega_h(\xi)t+\xi x)}d\xi = \int_{-\infty}^{+\infty} F(\xi)e^{it(\omega_h(\xi)+\xi x/t)}d\xi. \qquad (3.1.19)$$

Supongamos ahora que fijamos el valor de x/t, lo cual corresponde a mover el origen de referencia a velocidad $x/t = cte$. Evidentemente, cuando $t \to \infty$ la exponencial del integrando oscila más y más con respecto a la variable ξ y tiende a cero en un sentido débil haciendo que la integral tienda a anularse. Esta cancelación ocurre efectivamente para todos los valores de ξ salvo para aquéllos en los que

$$\frac{d}{d\xi}(\omega_h(\xi) + \xi x/t) = 0. \qquad (3.1.20)$$

Este hecho puede probarse de manera rigurosa mediante el Teorema de la Fase Estacionaria (TFE) (véase [8]).

La ecuación (3.1.20) puede también escribirse del modo siguiente:

$$\omega_h'(\xi) = -x/t. \qquad (3.1.21)$$

Esta relación indica que, a medida que nos trasladamos en el espacio a velocidad x/t, sólo podemos ver las componentes cuyo número de onda ξ satisfaga la relación (3.1.20), o, dicho de otro modo, la energía asociada al número de onda ξ se propaga a una *velocidad de grupo*

$$C_h(\xi) = -\omega_h'(\xi). \qquad (3.1.22)$$

[8] Evitamos aquí las constantes multiplicativas de la transformada y antitransformada de Fourier que en nada afectan al fenómeno cualitativo que pretendemos ilustrar.

Conviene en este punto señalar que la velocidad de fase $c_h(\xi)$ y la velocidad de grupo $C_h(\xi)$, en general, no coinciden. Analicemos este hecho en los ejemplos que hemos introducido más arriba.

En el caso de la ecuación de transporte continua teníamos que $w(\xi) = -\xi$ para todo ξ. En este caso, obviamente $c_h(\xi) \equiv C_h(\xi) \equiv 1$, lo cual indica que todas las ondas se propagan a velociad uno en este modelo.

Sin embargo en el esquema semi-discreto regresivo teníamos que

$$\omega = \frac{i}{h}\left[1 - e^{-i\xi h}\right]; \quad c_h(\xi) = -\frac{\omega_h(\xi)}{\xi} = 1 - \frac{i\xi h}{2} + O(h^2), \tag{3.1.23}$$

mientras la velocidad de grupo viene dada por la expresión

$$C_h(\xi) = -\omega_h'(\xi) = e^{-i\xi h} = 1 - i\xi h + O(h^2), \tag{3.1.24}$$

Se observa efectivamente una sutil diferencia entre las expresiones obtenidas en (3.1.23) y (3.1.24).

Consideramos ahora el esquema centrado en el que, como veíamos anteriormente,

$$\omega = -\frac{\text{sen}(h\xi)}{h}; \quad c_h(\xi) = \frac{\text{sen}(\xi h)}{\xi h} = 1 - \frac{\xi^2 h^2}{6} + O(h^4). \tag{3.1.25}$$

En este caso la velocidad de grupo es

$$C_h(\xi) = \cos(\xi h) = 1 - \frac{\xi^2 h^2}{2} + O(h^4). \tag{3.1.26}$$

Nuevamente se observa una ligera diferencia en las expresiones de velocidad de fase y de grupo.

Consideremos por último el esquema completamente discreto de “leap-frog". En aquél caso veíamos que

$$\omega_h(\xi) = \frac{-1}{\Delta t}\text{arcsen}\left[\mu\,\text{sen}(\xi\Delta x)\right]; \quad c_h(\xi) = \frac{1}{\Delta t\xi}\text{arcsen}\left[\mu\,\text{sen}(\xi\Delta x)\right]. \tag{3.1.27}$$

Sin embargo, la velocidad de grupo viene dada por la expresión:

$$C_h(\xi) = \frac{\Delta x\mu\cos(\xi\Delta x)}{\Delta t\sqrt{1-\mu^2\text{sen}^2(\xi\Delta x)}}, \tag{3.1.28}$$

que, nuevamente, difiere de la velocidad de fase, salvo en el caso $\mu = 1$ en el que $c_h(\xi) \equiv C_h(\xi) \equiv 1$.

Estas, aparentemente, pequeñas diferencias entre la velocidad de fase y de grupo pueden sin embargo ser la causa de comportamientos inesperados de las soluciones de los esquemas numéricos.

Hemos antes mencionado que el Teorema de la Fase Estacionaria (TFE) juega un papel fundamental a la hora de entender el fenómeno de la velocidad de grupo. El lector interesado en una presentación básica pero completa de este Teorema puede consultar la sección 4.5.3 del libro de Evans [8].

El TFE parte de la observación siguiente: Si $a \in C_c^\infty(\mathbb{R}^n)$, entonces

$$\int_{\mathbb{R}^n} e^{\frac{p\cdot x}{\varepsilon}} a(x)dx = O(\varepsilon^m),\ \varepsilon \to 0 \tag{3.1.29}$$

para todo $p \in \mathbb{R}^n$, $p \neq 0$ y $m \geqslant 1$.

En (3.1.29) se pone de manifiesto el hecho de que la integral se anula a un orden arbitrariamente grande cuando $\varepsilon \to 0$. Esto, evidentemente, no es así porque el integrando tienda a cero en módulo. Todo lo contrario, tenemos $\mid e^{ip\cdot x/\varepsilon} a(x) \mid = \mid a(x) \mid$ para todo $\varepsilon > 0$. Sin embargo es el carácter rápidamente oscilante de la exponencial compleja del integrando lo que hace que la integral se anule a cualquier orden. La prueba de (3.1.29) es muy sencilla. Basta integrar por partes teniendo en cuenta que

$$e^{ip\cdot x/\varepsilon} = \frac{\varepsilon}{p_k}\frac{\partial}{\partial x_k}(e^{ip\cdot x/\varepsilon}).$$

El número de integrales por partes que podemos realizar es ilimitado pues la función $a = a(x)$ es de clase C^∞. Además, al ser su soporte compacto, las integrales por partes no aportan términos de frontera.

En el caso en que en la exponencial aparecen términos cuadráticos obtenemos la expresión

$$\frac{1}{(2\pi\varepsilon)^{n/2}} \int_{\mathbb{R}^n} e^{\frac{1}{2\varepsilon} y\cdot Ay} a(y)dy = \frac{e^{i\frac{\pi}{4}\mathrm{sgn}\,(A)}}{\mid \det A \mid^{1/2}}(a(0) + O(\varepsilon))$$

para todo $a \in C_c^\infty(\mathbb{R}^n)$, A matriz real, simétrica, no singular, siendo $\det A$ el determinante de A y $\mathrm{sgn}\,(A)$ su signatura, i.e. el número de autovalores positivos de A menos el número de autovalores negativos. Este resultado se prueba primero en el caso diagonal para después abordar el caso general utilizando una rotación que permita representar A como una matriz diagonal en la base de sus autovectores.

Estos resultados, junto con el desarrollo de Taylor, permiten describir el comportamiento de la integral

$$I_\varepsilon = \int_{\mathbb{R}^n} e^{\frac{i\phi(y)}{\varepsilon}} a(y)dy$$

para una función real, regular ϕ general.

En efecto, suponiendo que $\nabla\phi$ sólo se anula en un número finito de puntos $y_1 \cdots, y_N$ del soporte de la función a y que las matrices $D^2\phi(y_k)$ no son singulares, $k = 1, \cdots, N$; obtenemos que

$$I_\varepsilon = (2\pi\varepsilon)^{n/2} \sum_{k=1}^{N} \frac{e^{\frac{i\phi(y_k)}{\varepsilon}}}{\mid \det D^2\phi(y_k) \mid^{1/2}} e^{i\frac{\pi}{4}\mathrm{sgn}\,(D^2\phi(y_k))}(a(y_k) + O(\varepsilon)).$$

Las pruebas detalladas de estos resultados se encuentran en 4.5.4. del libro de Evans [8].

3.2. Transformada discreta de Fourier a escala h

En la sección 3 hemos introducido y utilizado el método de von Neumann para el análisis de la estabilidad de un esquema numérico que está basado en la utilización de una transformada discreta de Fourier que permite:

- Definir una isometría entre ℓ^2 y $L^2(0, 2\pi)$;
- Transformar un esquema en diferencias en una ecuación diferencial dependiente de un parámetro $\theta \in [0,\, 2\pi)$.

La transformada de Fourier que introducimos en su momento, sin embargo, no tiene en cuenta el paso h del mallado puesto que se aplica meramente sobre sucesiones en ℓ^2, sin tener en cuenta el mallado al que están asociadas. Con el objeto de analizar el comportamiento de las soluciones cuando $h \to 0$ es conveniente introducir una *transformada de Fourier a escala h,* cuyo límite cuando $h \to 0$ sea la clásica transformada de Fourier, de modo que recuperemos en el límite la ecuación en derivadas parciales.

Recordemos en primer lugar la definición clásica de la transformada de Fourier continua

$$\widehat{f}(\xi) = \int_{\mathbb{R}} f(x)e^{-i\xi x}dx = \mathcal{F}(f). \tag{3.2.1}$$

Es bien sabido que la transformada de Fourier define una isometría de $L^2(\mathbb{R})$ en sí mismo. La transformada inversa de Fourier viene dada por

$$\mathcal{F}^{-1}(g)(x) = \frac{1}{2\pi}\int_{\mathbb{R}} g(\xi)e^{i\xi x}d\xi. \tag{3.2.2}$$

Una de las mayores utilidades de la transformada continua de Fourier es su posible utilización para la resolución de EDP con coeficientes constantes. En

esto la siguiente propiedad juega un papel fundamental[9]

$$\widehat{\partial_x f}(\xi) = i\xi \widehat{f}(\xi). \tag{3.2.3}$$

Por ejemplo, gracias a la propiedad (3.2.3), la ecuación de transporte

$$u_t + u_x = 0, \tag{3.2.4}$$

mediante la aplicación de la transformada de Fourier en la variable x, se convierte en

$$\widehat{u}_t + i\xi \widehat{u} = 0 \tag{3.2.5}$$

de donde deducimos que

$$\widehat{u}(\xi, t) = e^{-i\xi t} \widehat{f}(\xi). \tag{3.2.6}$$

La expresión (3.2.6) ya nos confirma el carácter conservativo de la ecuación de transporte (3.2.4) puesto que proporciona la identidad

$$| \widehat{u}(\xi, t) | = | \widehat{f}(\xi) |, \quad \forall \xi \in \mathbb{R}, \quad \forall t > 0 \tag{3.2.7}$$

que, tras integración en $\xi \in \mathbb{R}$, asegura que

$$\| \widehat{u}(t) \|_{L^2(\mathbb{R})} = \| \widehat{f} \|_{L^2(\mathbb{R})}, \, \forall t > 0 \tag{3.2.8}$$

lo cual, a su vez, por el carácter isométrico de la transformada de Fourier, garantiza que

$$\| u(t) \|_{L^2(\mathbb{R})} = \| f \|_{L^2(\mathbb{R})}, \, \forall t > 0. \tag{3.2.9}$$

Introduzcamos pues ahora la transformada discreta de Fourier a escala h, una de cuyas propiedades más relevantes será que, en el límite cuando $h \to 0$, recuperaremos la transformada continua de Fourier que acabamos de definir.

Dada la sucesión $(f_j)_{j \in \mathbb{Z}}$ proveniente de un mallado espacial de paso h (i.e. de modo que $f_j \sim f(x_j)$ con $x_j = jh$), definimos la transformada discreta de Fourier a escala h como

$$\overset{\sqcap}{f}(\xi) = h \sum_{j \in \mathbb{Z}} f_j e^{-i\xi hj}, \quad -\frac{\pi}{h} \leqslant \xi \leqslant \frac{\pi}{h}. \tag{3.2.10}$$

Denotamos la transformada discreta de Fourier mediante el símbolo $\overset{\sqcap}{\cdot}$ para distinguirla de la transformada continua. A pesar de que la transformada (3.2.10) depende del parámetro h, no lo expresamos explícitamente en la notación para aligerarla.

[9] El lector interesado en un estudio de las propiedades básicas de la Transformada de Fourier y su aplicación a las EDP puede consultar los textos de F. John [15] y J. Rauch [24].

Vemos que la imagen mediante la transformada discreta de Fourier de una sucesión de paso h es una función continua con soporte en el intervalo $[-\pi/h,\, \pi/h]$. Obviamente, a medida que $h \to 0$, este soporte converge a toda la recta real. Este hecho refleja una de las propiedades fundamentales de la transformada de Fourier, a medida que el carácter oscilante de la función en el espacio físico aumenta, su transformada de Fourier se amplifica para las altas frecuencias. La sucesión discreta de paso h puede verse como una función que oscila a escala h (basta para ello extender la sucesión discreta de valores a una función constante o lineal a trozos definida en toda la recta real). El soporte de su transformada de Fourier aumenta, consecuentemente.

La transformada de Fourier discreta puede invertirse con facilidad. Tenemos

$$f_j = \frac{1}{2\pi} \int_{-\pi/h}^{\pi/h} \overset{\sqcap}{f}(\xi) e^{i\xi h j} d\xi \tag{3.2.11}$$

La analogía entre las fórmulas (3.2.1) y (3.2.2) de la transformada continua de Fourier y (3.2.10)-(3.2.11) de la transforma discreta a escala h son evidentes. Mientras que (3.2.10) se asemeja a una suma de Riemann de la integral (3.2.1) que define la transformada continua de Fourier sobre la partición $x_j = jh$, $j \in \mathbb{Z}$, la transformada discreta inversa (3.2.11) es simplemente una versión truncada de la integral (3.2.2) que define la transformada inversa de Fourier.

Es fácil también comprobar que la transformada discreta define una isometría:

$$\| \vec{f} \|_h^2 = h \sum_{j \in \mathbb{Z}} |f_j|^2 = \frac{1}{2\pi} \int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} | \overset{\sqcap}{f}(\xi) |^2 \, d\xi = \frac{1}{2\pi} \left|\left| \overset{\sqcap}{f} \right|\right|^2_{L^2(-\frac{\pi}{h}, \frac{\pi}{h})}. \tag{3.2.12}$$

Esto, evidentemente, no es más que la versión discreta de la identidad de Parseval para la transformada de Fourier

$$\left|\left| f \right|\right|^2_{L^2(\mathbb{R})} = \frac{1}{2\pi} \int_{\mathbb{R}} | \widehat{f}(\xi) |^2 \, d\xi = \frac{1}{2\pi} \left|\left| \widehat{f} \right|\right|^2_{L^2(\mathbb{R})}. \tag{3.2.13}$$

La relación entre transformada continua y discreta se hace más clara aún si utilizamos la *función cardinal* (también denominada función *cardinal de Whittaker* of *función de Shannon*, por su papel relevante en teoría de la comunicación):

$$\psi_0(x) = \frac{\operatorname{sen}(\pi x/h)}{\pi x/h}. \tag{3.2.14}$$

Denotamos mediante ψ_j su trasladada al punto $x_j = jh$, i.e.

$$\psi_j(x) = \frac{\operatorname{sen}(\pi(x - x_j)/h)}{\pi(x - x_j)/h}. \tag{3.2.15}$$

Dada una función discreta $(f_j)_{j\in\mathbb{Z}}$ de paso h definimos entonces la función continua

$$f^*(x) = \sum_{j\in\mathbb{Z}} f_j \psi_j(x). \tag{3.2.16}$$

Es fácil comprobar que la función continua f^* interpola la sucesión $(f_j)_{j\in\mathbb{Z}}$. En efecto,

$$f^*(x_j) = f_j,\ \forall j \in \mathbb{Z}. \tag{3.2.17}$$

Esto es simplemente debido a que

$$\psi_j(x_k) = \delta_{jk},\ \forall j,k \in \mathbb{Z}. \tag{3.2.18}$$

La función cardinal ψ_0 tiene además la interesante propiedad que[10]

$$\widehat{\psi_0}(\xi) = h1_{(-\pi/h,\,\pi/h)}(\xi). \tag{3.2.19}$$

Su transformada de Fourier es por tanto, módulo un factor multiplicativo h, la función característica del intervalo $\Big(-\dfrac{\pi}{h}, \dfrac{\pi}{h}\Big)$.

Es fácil probar que, como ψ_j se obtiene de ψ_0 mediante una nueva traslación, entonces

$$\widehat{\psi_j}(\xi) = e^{-i\xi jh}\widehat{\psi_0}(\xi). \tag{3.2.20}$$

Por otra parte, utilizando la identidad de Plancherel obtenemos que

$$\int_{\mathbb{R}} \psi_j(x)\psi_k(x)dx = \frac{1}{2\pi}\int_{\mathbb{R}} \widehat{\psi_j}(\xi)\overline{\widehat{\psi_k}}(\xi)d\xi = \frac{h^2}{2\pi}\int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} e^{i\xi h(k-j)}d\xi = h\delta_{jk}. \tag{3.2.21}$$

Vemos por tanto que las funciones $\{\psi_j(x)\}_{j\in\mathbb{Z}}$ son ortogonales. De esta propiedad de ortogonalidad deducimos fácilmente que

$$\Big|\Big|f^*\Big|\Big|^2_{L^2(\mathbb{R})} = h\sum_{j\in\mathbb{Z}} |\ f_j\ |^2\ . \tag{3.2.22}$$

Por tanto, la extensión continua f^* de sucesiones de paso h define en realidad una isometría de ℓ^2 en un subespacio de $L^2(\mathbb{R})$.

Por otra parte, la transformada continua de Fourier de la función f^* está íntimamente ligada a la transformada discreta de la sucesión $(f_j)_{j\in\mathbb{Z}}$. En efecto,

$$\widehat{f^*}(\xi) = \sum_{j\in\mathbb{Z}} f_j\widehat{\psi_j}(\xi) = \sum_{j\in\mathbb{Z}} f_j e^{-i\xi jh}\widehat{\psi_0}(\xi) = h\sum_{j\in\mathbb{Z}} f_j e^{-i\xi jh}1_{(-\pi/h,\,\pi/h)} = \overset{\sqcap}{f}(\xi).$$

[10]Para comprobarlo observamos que $\partial_\xi 1_{[-A,A]} = \delta_A - \delta_{-A}$. Además $\mathcal{F}^{-1}(\partial_\xi 1_{[-A,A]}) = -i\xi\mathcal{F}^{-1}(1_{[-A,A]})$ y, por otra parte, $\mathcal{F}^{-1}(\delta_A - \delta_{-A}) = \frac{1}{2\pi}(e^{ixA} - e^{-ixA}) = \frac{i}{\pi}\,\mathrm{sen}(xA)$. De estas identidades deducimos (3.2.19) fácilmente, utilizando el hecho que la transformada de Fourier es un isomorfismo.

Vemos pues que *la transformada discreta de Fourier no es más que la transformada continua aplicada a la interpolación de la sucesión mediante la función cardinal.*

Estos resultados, probados por Whitakker en 1915 y utilizados en 1949 por Shannon, contribuyendo de manera decisiva a la teoría de la comunicación, indican que una función de *banda limitada* (cuya transformada de Fourier se anula fuera del intervalo $\xi \in [-B, B]$), puede ser reconstruida a través de la interpolación mediante la función cardinal a partir del muestreo de sus valores en los puntos $x_j = jh$, siempre que $h \leqslant \pi/B$. Retomaremos esta cuestión en la siguiente sección.

3.3. Revisión de la ecuación de transporte y sus aproximaciones a través de la transformada discreta de Fourier

Consideremos el problema de Cauchy para la ecuación de transporte

$$u_t + u_x = 0,\ x \in \mathbb{R},\ t > 0;\ u(x,0) = f(x),\ x \in \mathbb{R}. \tag{3.3.1}$$

Sabemos que la solución es una onda de transporte pura

$$u(x,t) = f(x-t). \tag{3.3.2}$$

Esta expresión puede también obtenerse mediante la transformación de Fourier. En efecto, como veíamos en (3.2.5),

$$\widehat{u}_t + i\xi\widehat{u} = 0,\ \xi \in \mathbb{R},\ t > 0, \widehat{u}(\xi,0) = \widehat{f}(\xi),\ \xi \in \mathbb{R}, \tag{3.3.3}$$

de donde deducimos que

$$\widehat{u}(\xi,t) = e^{-i\xi t}\widehat{f}(\xi). \tag{3.3.4}$$

Aplicando la transformada inversa obtenemos

$$u(x,t) = \mathcal{F}^{-1}(e^{-i\xi t}\widehat{f}(\xi)) = f(x-t). \tag{3.3.5}$$

En este último punto hemos usado el hecho de que la transformada de Fourier de la masa de Dirac es la constante unidad ($\widehat{\delta_0} \equiv 1$) o, equivalentemente, $\widehat{\delta_{x_0}} \equiv e^{-i\xi x_0}$.

Retomemos ahora el problema de la aproximación numérica de la solución.

Suponiendo que el dato inicial f es continuo, es natural tomar los datos discretos

$$f_j = f(x_j) = f(jh),\ j \in \mathbb{Z}, \tag{3.3.6}$$

lo cual supone realizar un muestreo de la función f.

Gracias a la fórmula de sumación de Poisson[11] es fácil comprobar que:

$$\overset{\sqcap}{f}(\xi) = \sum_{k\in\mathbb{Z}} \widehat{f}(\xi + k\omega_0),\ \forall -\pi/h \leqslant \xi \leqslant \pi/h, \tag{3.3.7}$$

donde

$$\omega_0 = 2\pi/h. \tag{3.3.8}$$

Si el dato inicial f es de banda limitada o, más precisamente, si $\widehat{f}(\xi) = 0$ para todo ξ tal que $\mid \xi \mid > \pi/h$, tenemos entonces

$$\overset{\sqcap}{f}(\xi) = \widehat{f}(\xi) \tag{3.3.9}$$

y por tanto el muestreo del dato inicial sobre los puntos del mallado no introduce error alguno.

Sin embargo, cuando f no es de banda limitada, en virtud de (3.3.7), las componentes de $\widehat{f}$ de altas frecuencias se superponen con las de la banda principal $[-\pi/h,\, \pi/h]$ dando lugar a lo que se conoce como fenómeno de *aliasing*. En este caso, la tranformada discreta de Fourier de la sucesión obtenida al muestrar f a lo largo de la sucesión $x_j = jh$ no permite recuperar la tranformada de Fourier de f y por tanto no permite codificar todas las características de la función f.

De este análisis deducimos que una función f es de banda limitada si y sólo si se obtiene como una función de la forma f^* a través de las funciones de Shannon a partir de su muestreo a lo largo de la sucesión $x_j = jh$.

En la práctica es por tanto recomendable aproximar en primer lugar la función $f(x)$ para una familia de funciones de banda limitada

$$f_h(x) = \mathcal{F}^{-1}\Big(\widehat{f}(\xi) 1_{(-\pi/h,\,\pi/h)}(\xi)\Big) \tag{3.3.10}$$

que tienen la virtud de converger a f en $L^2(\mathbb{R})$ cuando $h \to 0$ y de forma que su muestreo no introduzca ningún error.[12]

[11] La fórmula de sumación de Poisson asegura que $\sum_{j\in\mathbf{Z}} f(j) = \sum_{k\in\mathbf{Z}} \widehat{f}(2\pi k)$. Para comprobar esta fórmula basta considerar la función $g(x) = \sum_{j\in\mathbf{Z}} f(x+j)$, observar que es periódica de período uno y aplicar su desarrollo en series de Fourier. Los términos del sumando de la derecha son precisamente sus coeficientes de Fourier en la base $e^{i2\pi kx}$. Al aplicar este desarrollo en $x = 0$ obtenemos esta fórmula de sumación de Poisson. Al aplicar esta identidad a escala h a la función $f(x)e^{-i\xi x}$ obtenemos la identidad (3.3.7).

[12] La prueba de la convergencia de f_h a f en $L^2(\mathbb{R})$ se realiza combinando el Teorema de la Convergencia Dominada con el hecho de que $\mathcal{F}$ sea una isometría en $L^2(\mathbb{R})$.

Pero, dejando de lado los errores introducidos por la aproximación de los datos iniciales, consideremos el generado por los esquemas numéricos. Consideramos por tanto el esquema semi-discreto regresivo y progresivo (3.0.10) y (3.0.11) que, como vimos, son convergentes y divergentes respectivamente.

Revisión del esquema semi-discreto regresivo.

Consideremos en primer lugar el esquema

$$u_j'(t) + \frac{u_j(t) - u_{j-1}(t)}{h} = 0,\ j \in \mathbb{Z},\ t > 0. \tag{3.3.11}$$

Aplicando la transformada discreta de Fourier a escala h obtenemos

$$\frac{d}{dt}\overset{\sqcap}{u}(\xi,t) + \frac{1}{h}(1 - e^{-i\xi h})\overset{\sqcap}{u}(\xi,t) = 0,\ t > 0,\ \xi \in [-\pi/h,\ \pi/h]. \tag{3.3.12}$$

En este punto hemos utilizado la siguiente propiedad fundamental de la transformada discreta de Fourier:

$$(\tau_{-1}f)^{\sqcap}(\xi) = h\sum_{j\in\mathbb{Z}} f_{j-1}e^{-i\xi jh} = e^{-i\xi h}h\sum_{j\in\mathbb{Z}} f_j e^{-i\xi jh} = e^{-i\xi h}\overset{\sqcap}{f}(\xi). \tag{3.3.13}$$

La proximidad entre la ecuación de transporte continua (3.3.1) y la aproximación semi-discreta regresiva (3.3.11) es evidente. El coeficiente

$$\omega_h(\xi) = \frac{1}{h}(1 - e^{-i\xi h}) \tag{3.3.14}$$

que interviene en la ecuación diferencial (3.3.12) converge, cuando $h \to 0$, de manera evidente, al coeficiente

$$\omega(\xi) = i\xi \tag{3.3.15}$$

correspondiente a la ecuación de transporte continua.

De hecho, mediante el desarrollo de Taylor se observa que

$$\omega_h(\xi) = i\xi + \frac{\xi^2 h}{2} + \cdots. \tag{3.3.16}$$

En la expresión (3.3.16) se observa que, efectivamente, para cada $\xi \in \mathbb{R}$,

$$\omega_h(\xi) \to \omega(\xi) \text{ cuando } h \to 0. \tag{3.3.17}$$

Además en (3.3.16) volvemos a constatar el carácter difusivo de la aproximación regresiva. En efecto, esto queda de manifiesto en que el primer término corrector en (3.3.16) ($\xi^2 h/2$) sea real y positivo.

Conviene sin embargo observar que la convergencia (3.3.17) es sólamente uniforme en conjuntos R_h en los que

$$\max_{\omega \in R_h} \xi^2 h \to 0,\ h \to 0. \tag{3.3.18}$$

En otras palabras, la convergencia de los *símbolos* (3.3.17) sólo se produce en regiones en las que

$$|\ \xi\ | = O(h^{-1/2}),\ h \to 0. \tag{3.3.19}$$

Sin embargo, conviene señalar que la convergencia (3.3.17) interesa para cualquier $\xi \in \mathbb{R}$. En efecto, en el límite cuando $h \to 0$, la banda de frecuencias del dato inicial continuo $f = f(x)$ de la ecuación de transporte es toda la recta real. Por otra parte, a medida que $h \to 0$, la banda de frecuencias de los datos iniciales del problema discreto $[-\pi/h,\ \pi/h]$ aumenta hasta cubrir toda la recta real.

En virtud de que la convergencia (3.3.17) es uniforme en conjuntos de la forma (3.3.19) es fácil ver que las soluciones del problema discreto convergen a las del continuo para datos con una banda de frecuencias limitada, independiente de h. La estabilidad del esquema permite después extender esta convergencia a un dato inicial cualquiera $f \in L^2(\mathbb{R})$.

En efecto, en virtud de (3.3.12) tenemos

$$\overset{\sqcap}{u}(\xi, t) = e^{-\omega_h(\xi)t}\overset{\sqcap}{f}(\xi) = e^{-\frac{1}{h}(1-e^{-i\xi h})t}\overset{\sqcap}{f}(\xi) = e^{-\frac{1}{h}(1-\cos(\xi h))t}e^{-i\,\mathrm{sen}(\xi h)t/h}\overset{\sqcap}{f}(\xi).$$

Aplicando la anti-transformada discreta de Fourier tenemos

$$u_j(t) = \frac{1}{2\pi}\int_{-\pi/h}^{\pi/h} e^{-\frac{1}{h}(1-\cos(\xi h))t}e^{i\xi(jh-\mathrm{sen}(\xi h)t/\xi h)}\overset{\sqcap}{f}(\xi)d\xi. \tag{3.3.20}$$

En virtud de (3.3.9) sabemos que, si f es de banda acotada, $\overset{\sqcap}{f} \equiv \widehat{f}$, para h suficientemente pequeño. Bajo estas hipótesis es por tanto evidente que la solución del problema numérico puede reescribirse como

$$u_j(t) = \frac{1}{2\pi}\int_{-B}^{B} e^{-\frac{1}{h}(1-\cos(\xi h)t)}e^{i\xi t(jh-\mathrm{sen}(\xi h)/\xi h)t}\widehat{f}(\xi)d\xi, \tag{3.3.21}$$

donde $B > 0$ es tal que $\mathrm{sop}(\widehat{f}) \subset [-B,\ B]$.

Elegimos ahora $j \in \mathbb{Z}$ de modo que $jh = x_0$, siendo $x_0 \in \mathbb{R}$ un punto fijado. Evidentemente, esto supone elegir $j = x_0/h$ que, para que $j \in \mathbb{Z}$, exige a su vez tomar una sucesión determinada de $h \to 0$. Bajo esta condición ($x_0 = jh$), deberíamos ser capaces de ver que la expresión (3.3.21) converge, cuando $h \to 0$,

al valor de la solución continua $u(x_0,t) = f(x_0 - t)$. Veámos que esto es efectivamente así. Cuando $h \to 0$, el integrando de (3.3.21) converge a $e^{i\xi(x_0-t)}\widehat{f}(\xi)$. La aplicación del Teorema de la convergencia dominada permite entonces ver que el límite de (3.3.21) es

$$u(x_0,t) = \frac{1}{2\pi}\int_{-B}^{B} e^{i\xi(x_0-t)}\widehat{f}(\xi)d\xi = \frac{1}{2\pi}\int_{-B}^{B} e^{i\xi x_0}e^{-i\xi t}\widehat{f}(\xi)d\xi = f(x_0-t), \tag{3.3.22}$$

que coincide con la solución del problema continuo, gracias a la hipótesis de que f sea de banda limitada.

En realidad, bajo estas hipótesis, se puede probar que la convergencia de la solución discreta a la continua tiene lugar en la norma L^2. Se puede ver esto de dos maneras. Tomando normas discretas de diferencias en ℓ^2 o bien tomando normas continuas en $L^2(\mathbb{R})$ observando que la expresión (3.3.21) de la solución del esquema numérico puede extenderse a una función continua con respecto a la variable espacial x, dependiente del parámetro h:

$$u_h(x,t) = \frac{1}{2\pi}\int_{-B}^{B} e^{-\frac{1}{h}(1-\cos(\xi h))t}e^{i\xi(x-\operatorname{sen}(\xi h)t/\xi h)}\widehat{f}(\xi)d\xi. \tag{3.3.23}$$

Combinando (3.3.22) y (3.3.23) vemos que, tanto la solución continua como la discreta pueden ser escritas de un modo semejante mediante la transformada inversa de Fourier

$$u_h(x,t) = \mathcal{F}^{-1}\left[e^{-\frac{t}{h}(1-\cos(\xi h))}e^{-i\operatorname{sen}(\xi h)t/h}\widehat{f}\right](x) \tag{3.3.24}$$

y

$$u(x,t) = \mathcal{F}^{-1}\left[e^{-i\xi t}\widehat{f}\right](x). \tag{3.3.25}$$

Para comprobar que $u_h(t) \to u(t)$ en $L^2(\mathbb{R})$ cuando $h \to 0$ basta entonces ver que

$$e^{-\frac{t}{h}(1-\cos(\xi h))}e^{i\operatorname{sen}(\xi h)t/h}\widehat{f}(\xi) \to e^{-i\xi t}\widehat{f}(\xi) \text{ en } L^2(\mathbb{R})$$

cuando $h \to 0$. Teniendo en cuenta que f es, por hipótesis, de banda acotada, vemos que esto es equivalente a que

$$\int_{-B}^{B}\left|e^{-\frac{t}{h}(1-\cos(\xi h))}e^{-i\operatorname{sen}(\xi h)t/h} - e^{-i\xi t}\right|^2 \mid \widehat{f}(\xi)\mid^2 d\xi \to 0$$

y esto, efectivamente, ocurre en virtud del Teorema de la convergencia dominada.

Esto confirma la convergencia del esquema regresivo para datos iniciales con banda acotada. Para considerar el caso general, dado $f \in L^2(\mathbb{R})$ basta introducir su aproximación

$$f_B(x) = \mathcal{F}^{-1}\Big(\widehat{f}(\xi)1_{(-B,B)}(\xi)\Big)$$

que es, por definición, una función de banda acotada tal que

$$f_B \to f \text{ en } L^2(\mathbb{R}) \text{ cuando } B \to \infty.$$

Denotamos mediante u y u_B la solución de la ecuación de transporte con datos f y f_B respectivamente. Como

$$\| u(t) - u_B(t) \|_{L^2(\mathbb{R})} = \| f - f_B \|_{L^2(\mathbb{R})}$$

vemos que

$$u_B \to u \text{ en } L^\infty(0,\infty; L^2(\mathbb{R})),\ B \to \infty.$$

Por tanto, dado $\varepsilon > 0$ existe $B_0 > 0$ suficientemente grande tal que

$$\| u - u_{B_0} \|_{L^\infty(0,\infty;L^2(\mathbb{R}))} \leqslant \varepsilon/2.$$

Fijado este valor de B_0 y resolviendo la ecuación semi-discreta con dato inicial f_{B_0} muestreado sobre el mallado tenemos que

$$u_{h,B_0}(t) \to u_{B_0}(t),\ h \to 0, \text{ en } L^2(\mathbb{R})$$

para cada $t > 0$. Por tanto, para h suficientemente pequeño

$$\| u(t) - u_{h,B_0}(t) \|_{L^2(\mathbb{R})} \leqslant \| u(t) - u_{B_0}(t) \|_{L^2(\mathbb{R})} + \| u_{B_0}(t) - u_{h,B_0}(t) \|_{L^2(\mathbb{R})} \leqslant \varepsilon/2 + \varepsilon/2 = \varepsilon.$$

Esto demuestra la convergencia para datos iniciales obtenidos muestreando una aproximación del dato inicial de banda acotada. Evidentemente, gracias a la estabilidad del esquema numérico, se puede obtener el mismo resultado de convergencia para cualquier elección de la aproximación de los datos iniciales que converja al dato inicial del problema continuo.

Revisión del esquema semi-discreto progresivo.

Consideramos ahora el caso progresivo que, como vimos, es inestable y divergente. Analicémoslo pues con la herramienta que las transformadas de Fourier proporcionan.

En este caso el esquema es de la forma

$$u_j'(t) + \frac{u_{j+1} - u_j(t)}{h} = 0,\ j \in \mathbb{Z},\ t > 0. \tag{3.3.26}$$

Aplicando la transformada discreta de Fourier obtenemos

$$\frac{d}{dt}\overset{\sqcap}{u}(\xi,t)+\frac{1}{h}(e^{i\xi h}-1)\overset{\sqcap}{u}(\xi,t)=0,\ t>0,\ \xi\in[-\pi/h,\ \pi/h], \tag{3.3.27}$$

de modo que

$$\overset{\sqcap}{u}(\xi,t)=e^{-\frac{1}{h}(e^{i\xi h}-1)t}\overset{\sqcap}{f}(\xi)=e^{\frac{1}{h}(1-\cos(\xi h))t}e^{-i\,\mathrm{sen}(\xi h)t/h}\overset{\sqcap}{f}(\xi). \tag{3.3.28}$$

Pretendemos ahora ilustrar de manera aún más explícita la ausencia de convergencia de este método. Para ello tomamos un dato inicial f de banda acotada de modo que $\overset{\sqcap}{f}\equiv\widehat{f}$ para h suficientemente pequeño.

Obtenemos así

$$\overset{\sqcap}{u}(\xi,t)=e^{\frac{1}{h}(1-\cos(\xi h))t}e^{-i\,\mathrm{sen}(\xi h)t/h}\widehat{f}(\xi), \tag{3.3.29}$$

de modo que

$$|\ \overset{\sqcap}{u}(\xi,t)\ |=e^{\frac{1}{h}(1-\cos(\xi h))t}\ |\ \widehat{f}(\xi)\ |, \tag{3.3.30}$$

y entonces

$$\|\ \vec{u}_h\ \|_h^2=\frac{1}{2\pi}\int_{-B}^{B}e^{\frac{2}{h}(1-\cos(\xi h))t}\ |\ \widehat{f}(\xi)\ |^2\ d\xi. \tag{3.3.31}$$

Habida cuenta que $\xi\in[-B,\ B]$, tenemos que

$$1-\cos(\xi h)\geqslant c\xi^2h^2 \tag{3.3.32}$$

con $c>0$ para todo $\xi\in[-B\,B]$, a condición que h sea suficientemente pequeño.

Combinando (3.3.31) y (3.3.32) vemos que

$$\|\ \vec{u}_h(t)\ \|_h^2\geqslant\frac{1}{2\pi}\int_{-B}^{B}e^{ch\xi^2t}\ |\ \widehat{f}(\xi)\ |^2\ d\xi. \tag{3.3.33}$$

Pero esta estimación es claramente insuficiente para concluir la divergencia del método puesto que la integral a la derecha de (3.3.33) permanece acotada cuando $h\to 0$.

Para ilustrar la divergencia hemos considerar datos iniciales de banda más ancha. Dado $f\in L^2(\mathbb{R})$ tal que el soporte de su transformada de Fourier sea toda la recta real (por ejemplo la Gaussiana[13]), introducimos el dato inicial del esquema discreto $f_h(x)$ truncando la transformada de Fourier de f a la banda admisible $[-\pi/h,\ \pi/h]$, i.e.

$$f_h(x)=\mathcal{F}^{-1}(\widehat{f}1_{(-\pi/h,\,\pi/h)}(\xi)). \tag{3.3.34}$$

[13]Es bien sabido que la transformada de Fourier de la función $f(x)=e^{-x^2/2}$ es la Gaussiana $\widehat{f}(\xi)=\sqrt{2\pi}e^{-\xi^2/2}$.

En este caso la norma de la aproximación discreta viene dada por

$$\| \vec{u}_h(t) \|_h^2 = \frac{1}{2\pi} \int_{-\pi/h}^{\pi/h} e^{\frac{2}{h}(1-\cos(\xi h))t} \mid \widehat{f}(\xi) \mid^2 d\xi. \tag{3.3.35}$$

Utilizando ahora el hecho que

$$1 - \cos(\eta) \geqslant c\eta^2, \quad \forall \eta \in [-\pi, \pi], \tag{3.3.36}$$

vemos que

$$\| \vec{u}_h(t) \|_h^2 \geqslant \frac{1}{2\pi} \int_{-\pi/h}^{\pi/h} e^{ch\xi^2 t} \mid \widehat{f}(\xi) \mid^2 d\xi. \tag{3.3.37}$$

En esta ocasión la integral a la derecha de (3.3.37) puede diverger puesto que en la banda $\xi \in [-\pi/h, \pi/h]$ hay zonas donde $h\xi^2 \to \infty$. Para comprobar la divergencia de esta integral con más detalle consideremos el dato inicial f de modo que

$$\widehat{f}(\xi) = \sum_{k \in \mathbb{Z}} \alpha_k 1_{I_k}(\xi) \tag{3.3.38}$$

donde $(I_k)_{k\in\mathbb{Z}}$ son intervalos disjuntos de $\mathbb{R}$. Para que $f = \mathcal{F}^{-1}(\widehat{f}) \in L^2(\mathbb{R})$ basta entonces con que

$$\sum_{k \in \mathbb{Z}} \alpha_k^2 \mid I_k \mid < \infty, \tag{3.3.39}$$

donde $\mid I_k \mid$ denota la longitud del intervalo I_k.

En este caso la integral a la derecha (3.3.37) puede reescribirse como

$$\frac{1}{2\pi} \int_{-\pi/h}^{\pi/h} e^{ch\xi^2 t} \mid \widehat{f}(\xi) \mid^2 d\xi = \frac{1}{2\pi} \sum_{k \in \mathbb{Z}} \alpha_k^2 \int_{I_k \cap [-\pi/h,\, \pi/h]} e^{ch\xi^2 t} d\xi. \tag{3.3.40}$$

Si elegimos los intervalos $I_k = (k, k+1)$ observamos que el último sumatorio puede acotarse inferiormente por

$$\frac{1}{2\pi} \alpha_{k_0}^2 e^{ch(\frac{\pi}{h}-1)^2 t} \tag{3.3.41}$$

con $k_0 = \frac{\pi}{h} - 1$.

En este caso, además, la condición (3.3.39) puede simplemente reescribirse como

$$\sum_{k \in \mathbb{Z}} \alpha_k^2 < \infty. \tag{3.3.42}$$

Es evidente que es perfectamente posible elegir una sucesión α_k de la forma $\alpha_k = 1/k$, de modo que (3.3.42) se cumpla y que, sin embargo, la cota inferior (3.3.41) de la norma ℓ^2 de la solución discreta correspondiente diverja cuando $h \to 0$ con un orden $e^{ct/h}$.

Vemos por tanto que la utilización de la transformada de Fourier a escala h permite ilustrar de manera mucho más cuantitativa la divergencia del método semi-discreto progresivo que habíamos predicho mediante el análisis de von Neumann. En el caso en que el esquema es convergente puede también aplicarse para ilustrar de un modo más claro la convergencia hacia la solución del problema continuo.

A pesar de que en esta sección sólo hemos analizado las aproximaciones semidiscretas progresiva y regresiva las ideas que hemos desarrollado son completamente generales y pueden ser aplicadas al estudio de cualquier otro esquema, en particular para los completamente discretos.

Revisión del comportamiento de las velocidades de fase y grupo.

Ahora que sabemos que el rango de frecuencias relevantes para una aproximación numérica es $-\pi/h \leq \xi \leq \pi/h$, conviene revisar los conceptos de velocidades de fase y grupo. Consideremos en primer lugar el esquema centrado semi-discreto. En este caso la velocidad de fase viene dada por

$$c_h(\xi) = \frac{\text{sen}(\xi h)}{\xi h}. \tag{3.3.43}$$

Vemos entonces que la velocidad de fase se anula cuando $\xi h = \pm\pi$. Se trata evidentemente de un fenómeno nuevo con respecto a la ecuación de transporte continua donde todas las componentes de Fourier de las soluciones se transportan a velocidad constante uno. En virtud de este hecho, para cada $h > 0$ fijo, existen soluciones del problema numérico que apenas se tranportan. Esto no es incompatible con la convergencia de orden dos del esquema numérico centrado que ya comprobamos. En efecto, en el problema clásico de la convergencia, el dato inicial se supone fijo, lo cual, en la práctica, gracias a la propiedad de estabilidad del esquema, permite filtrar las altas frecuencias del dato inicial y considerar únicamente datos cuya transformada de Fourier tiene soporte compacto. El hecho de que la velocidad de propagación se anule cuando $|\xi| \sim \pi/h$, no tiene entonces efectos a nivel de la convergencia. Pero, insistimos, si lo que nos interesa es la dinámica de las soluciones para h pequeño pero fijo, este hecho tiene un gran impacto puesto que surgen soluciones que nada tienen que ver con el comportamiento de la ecuación de transporte continua. Se trata del mismo fenómeno que surge al estudiar la estabilidad absoluta de los sistemas stiff de ecuaciones diferenciales ordinarias (véase por ejemplo [14]). En el caso del sistema centrado este hecho especialmente grave puesto que el esquema es puramente conservativo y por tanto estas soluciones a altas frecuencias en absoluto se disipan. Diremos que se trata de *soluciones espúreas*, en el sentido que

son ficticias puesto que no corresponden a la ecuación de transporte continua y sólo surgen como soluciones del esquema numérico. Por otra parte, la velocidad de grupo en este caso toma el valor

$$C_h(\xi) = \cos(\xi h). \tag{3.3.44}$$

Vemos que la situación es aún peor puesto que se anula cuando $\xi h = \pi/2$ y tiene signo negativo para todo $\xi h \in (\pi/2, \pi]$. En este caso por tanto tendremos incluso soluciones que se transportan en la dirección opuesta a la de la ecuación de transporte continua. Se trata de un fenómeno de soluciones numéricas espúreas que no es incompatible con la convergencia del esquema numérico.

Consideremos ahora el esquema numérico regresivo que, como vimos, tiene un carácter disipativo. En este caso la velocidad de fase viene dada por:

$$c_h(\xi) = \frac{i(e^{-i\xi h} - 1)}{\xi h} = \frac{\text{sen}(\xi h)}{\xi h} + i\frac{\cos(\xi h) - 1}{\xi h}. \tag{3.3.45}$$

Vemos que la parte real de la velocidad de fase se comporta como en el caso de la aproximación centrada de modo que ésta se anula cuando $\xi h = \pm\pi$. Sin embargo, vemos también que para estos valores de frecuencias la parte imaginaria de la velocidad de grupo es estrictamente negativa, lo cual asegura que estas componentes de Fourier de la solución decaen exponencialmente en tiempo. Vemos pues que la aproximación regresiva, a pesar de introducir soluciones numéricas espúreas, las disipa. Es a causa de este hecho que la soluciones del esquema regresivo para $h > 0$ pequeño y fijo se comportan de manera mucho más semejante a las de la ecuación de transporte continua que las del esquema centrado. Lo mismo ocurre con la velocidad de grupo.

De este análisis concluimos que más allá de las propiedades de convergencia clásicas de un esquema numérico, con el objeto de garantizar que para $h > 0$ pequeño la dinámica del esquema discreto se asemeja a la del continuo es preciso tener en cuenta el comportamiento de las velocidades de fase y de grupo en frecuencias $|\xi|$ del orden de c/h con $0 < c < \pi$.

Capítulo 4

Ecuaciones de convección-difusión

4.1. Introducción

En estas notas recogemos un resumen del curso de doctorado impartido en la UAM en el segundo cuatrimestre del curso 04-05.

El curso estará esencialmente orientado a analizar algunos métodos numéricos relevantes en la simulación y aproximación numérica de las ecuaciones de la Mecánica de Fluidos.

Como es bien sabido las ecuaciones de Euler y de Navier-Stokes son, respectivamente, las que describen el movimiento de los fluidos perfectos y viscosos incomprensibles. Se trata sin duda de dos de los modelos más relevantes de las Ciencias y Tecnologías pues intervienen de manera decisiva en la modelización y diseño de fenómenos muy diversos como las corrientes marinas, el movimiento del aire en meteorología y aeronáutica y la circulación sanguínea.

La gran complejidad de estas ecuaciones, sobre todo en tres dimensiones espaciales, ha hecho que durante el siglo XX hayan sido un tema de investigación permanente. Buena parte de los desarrollos del Análisis Matemático y de las ecuaciones diferenciales han estado orientados a la comprensión y resolución de estas ecuaciones. A pesar de ello, algunas de las cuestiones más elementales, como la regularidad y unicidad de las soluciones en tres dimensiones esté aún abierto.

La necesidad de obtener aproximaciones numéricas efectivas y la relativa incapacidad de los métodos analíticos para describir de manera más precisa el

comportamiento de estas ecuaciones y sus soluciones han hecho que estas hayan constituido también uno de los motores principales del Análisis Numérico.

En este curso abordaremos algunos de los métodos principales que en la actualidad se emplean en la resolución de estas ecuaciones. El curso está fuertemente inspirado en el excelente libro recopilatorio de R. Glowinski [9]. Sin embargo, dada la extensión de aquella monografía, en este curso sólo cubriremos los aspectos más básicos puesto que los más avanzados necesitarían de mucho más tiempo y de unos conocimientos sobre la Teoría de las Ecuaciones en Derivadas Parciales que no podemos presuponer en los cursos de doctorado.

Con el objeto de evitar algunas de las dificultades propias de las ecuaciones de Euler y de Navier-Stokes y poder ilustrar con más facilidad las ideas fundamentales de los aspectos más numéricos, nos ocuparemos casi exclusivamente de la ecuación de Burgers. Se trata de un modelo relativamente simple que, en una dimensión espacial, reproduce algunos de los aspectos más importantes de las ecuaciones de Navier-Stokes como son la difusión y la convección no-lineal.

Como complemento a estas notas sugerimos las siguientes lecturas:

- Las notas de J.L. Vázquez [28] sobre Mecánica de Fluidos que recoge los aspectos más importantes de la modelización y un estudio cualitativo de los modelos más importantes.

- Las notas introductorias al Análisis Numérico de Ecuaciones en Derivadas Parciales [36]. Para el lector interesado en una introducción más completa, recogiendo también lo más relevante del Análisis Numérico de Ecuaciones Diferenciales Ordinarias, recomendamos el libro de A. Iserles [14].

- Las notas [33], [34] donde recogemos el contenido de los cursos de doctorado de los años 02-03 y 03-04. En ellos abordamos los métodos de diferencias finitas, la descomposición de dominios, los métodos de descenso y los elementos finitos.

4.2. La ecuación de Burgers y la transformación de Hopf-Cole

La ecuación de Burgers viscosa es la siguiente

$$u_t - \nu u_{xx} + (u^2)_x = 0. \tag{4.2.1}$$

En (4.2.1) $\nu > 0$ es el parámetro de viscosidad.

En el caso $\nu = 0$ la ecuación (4.2.1) se convierte en la siguiente ecuación hiperbólica no-lineal de orden uno:

$$u_t + (u^2)_x = 0. \tag{4.2.2}$$

Se trata en ambos casos del análogo $1-d$ de las ecuaciones de Navier-Stokes y de Euler para el movimiento de los fluidos.

Las ecuaciones de Navier-Stokes para un fluido incompresible y homogéneo pueden escribirse como

$$\begin{cases} u_t - \nu\Delta u + u \cdot \nabla u = \nabla p \\ \operatorname{div} u = 0. \end{cases} \tag{4.2.3}$$

En (4.2.3) $u = u(x, t)$ es el vector velocidad del fluido que puede por tanto tener tres componentes o sólo dos en casos simplificados de fluidos bidimensionales. El parámetro $\nu > 0$ es el de viscosidad del fluido. La función $p = p(x, t)$ es escalar y denota la presión.

La primera ecuación de (4.2.3) es en realidad un sistema de tres (resp. dos en dimensión dos) ecuaciones en derivadas parciales con tres (resp. dos) incógnitas. La segunda ecuación es la que refleja la incompresibilidad del fluido.

En el caso de ausencia de viscosidad, i.e. cuando $\nu = 0$, obtenemos las ecuaciones de Euler para un fluido perfecto o ideal:

$$\begin{cases} u_t + u \cdot \nabla u = \nabla p \\ \operatorname{div} u = 0. \end{cases} \tag{4.2.4}$$

Se trata efectivamente de una idealización puesto que todo fluido tiene un cierto grado de viscosidad. Pero (4.2.4) es un modelo útil para describir el movimiento de fluidos poco viscosos.

Las ecuaciones (4.2.1) y (4.2.2) son modelos reducidos de las ecuaciones de Navier-Stokes (4.2.3) y de Euler (4.2.4) respectivamente.

La ecuación (4.2.1), también denominada ecuación del calor no-lineal, es la más sencilla en la que se combinan los efectos de la viscosidad y de la convección no-lineal cuadrática. Esta ecuación, mediante un cambio de variables introducido independientemente por Hopf y Cole en los años 50, puede reducirse a la ecuación del calor lineal y, por tanto, ser resuelta explícitamente. En la siguiente sección describiremos este cambio de variables. La solución así obtenida refleja adecuadamente la presencia de los dos términos y efectos principales presentes en (4.2.1): la viscosidad y la convección no-lineal. La viscosidad hace que la solución tenga una escala Gaussiana, mientras que la convección no-lineal hace

que la solución desarrolle una asimetría, producto de un fenómeno de transporte a una velocidad no homogénea.

Las soluciones de (4.2.2) son el límite cuando $\nu \to 0$ de las de (4.2.1). En ellas el efecto de la viscosidad desaparece y eso hace que las soluciones puedan dejar de ser regulares en tiempo finito desarrollando choques.

Con el objeto de introducir esta transformación consideraremos soluciones $u = u(x,\, t)$ de (4.2.1) tales que $\mid u(x,\, t) \mid + \mid u_x(x,\, t) \mid \to 0$ cuando $\mid x \mid \to \infty$.

Si $u = u(x,\, t)$ es una solución de (4.2.1) de este tipo, la función

$$v = v(x,\, t) = \int_{-\infty}^{x} u(s,\, t)ds \tag{4.2.5}$$

satisface

$$v_t - \nu v_{xx} + \mid v_x \mid^2 = 0. \tag{4.2.6}$$

Definimos entonces

$$w = v(x,\, t/\nu)$$

que verifica

$$w_t - w_{xx} + \frac{1}{\nu} \mid w_x \mid^2 = 0. \tag{4.2.7}$$

Por otra parte

$$z = 2/\nu \tag{4.2.8}$$

verifica entonces

$$z_t - z_{xx} + \mid z_x \mid^2 = 0. \tag{4.2.9}$$

Introducimos por último

$$\eta(x,\, t) = e^{-z} \tag{4.2.10}$$

que satisface entonces la ecuación del calor

$$\eta_t - \eta_{xx} = 0. \tag{4.2.11}$$

Deshaciendo este cambio de variables vemos que

$$\begin{aligned} &u = v_x \\ &v(\cdot,\, t/\nu) = w(\cdot,\, t) = \nu z(\cdot,\, t) = -\nu \log(\eta). \end{aligned}$$

Por tanto

$$u(x,\, t) = -\nu \frac{\eta_x(x,\, \nu t)}{\eta(x,\, \nu t)}. \tag{4.2.12}$$

La solución η de la ecuación del calor se obtiene por convolución con el núcleo de Gauss:

$$G(x,\, t) = (4\pi t)^{-1/2} \exp\Big(- \mid x \mid^2 \Big/ 4t\Big), \tag{4.2.13}$$

de modo que

$$\eta(x,\,t) = \Big[G(\cdot,\,t) * \eta_0(\cdot)\Big](x), \tag{4.2.14}$$

donde η_0 es el dato inicial de η.

Por otra parte

$$G_x(x,\,t) = -\frac{x}{4\sqrt{\pi}t^{3/2}} \exp\Big(-\mid x \mid^2 \Big/ 4t\Big). \tag{4.2.15}$$

Obtenemos así

$$u(x,\,t) = \frac{\displaystyle\int_{\mathbb{R}} (x-y)e^{-|x-y|^2/4\nu t}\eta_0(y)dy}{\displaystyle 2t\int_{\mathbb{R}} e^{-|x-y|^2/4\nu t}\eta_0(t)dy}. \tag{4.2.16}$$

Ahora bien

$$\eta_0(x) = e^{-\int_{-\infty}^{x} u_0(\sigma)d\sigma/\nu}. \tag{4.2.17}$$

Por tanto

$$u_\nu(x,\,t) = \frac{\displaystyle\int_{\mathbb{R}} (x-y)e^{-H(x,\,y,\,t)/\nu}dy}{\displaystyle 2t\int_{\mathbb{R}} e^{-H(x,\,y,\,t)/\nu}dy} \tag{4.2.18}$$

donde

$$H(x,\,y,\,t) = \frac{\mid x-y \mid^2}{4t} + \int_{-\infty}^{y} u_0(\sigma)d\sigma. \tag{4.2.19}$$

De la expresión de esta solución observamos que:

• La función solución $u = u_\nu(x,\,t)$ es regular para todo $x \in \mathbb{R}$ y $t > 0$ cuando u_0 pertenece, por ejemplo, a $L^1(\mathbb{R})$.

Basta para ello utilizar el efecto regularizante de la convolución con el núcleo de Gauss que se deduce de la desigualdad de Young:

$$\| f * g \|_\infty \leq \| f \|_1 \| g \|_\infty . \tag{4.2.20}$$

Basta entonces aplicar esta desigualdad utilizando que $G(\cdot,\,t)$ y sus derivadas de orden arbitrario son, para todo $t > 0$, funciones regulares e integrables y que el dato inicial $e^{-\nu\int_{-\infty}^{x} u_0(\sigma)d\sigma}$ está acotado por estarlo $\displaystyle\int_{-\infty}^{x} u_0(\sigma)d\sigma$, lo cual es a su vez consecuencia de que el dato inicial u_0 sea integrable.

• Cuando el dato inicial u_0 es par, la solución deja de serlo.

Esto es consecuencia del efecto de la convolución no-lineal. Esto no ocurre para las soluciones de la ecuación del calor lineal. En efecto, las soluciones de

$$\begin{cases} u_t - u_{xx} = 0, & x \in \mathbb{R}, \quad t > 0 \\ u(x,\,0) = u_0(x), & x \in \mathbb{R} \end{cases}$$

son de la forma

$$u(x,\, t) = \Big[G(\cdot,\, t) * u_0(\cdot)\Big](x).$$

Es por tanto fácil ver que u es par (resp. impar) con respecto a x, para todos $t > 0$, en función de que el dato inicial u_0 lo sea.

Esto, sin embargo, no es así para las soluciones de la ecuación de Burgers que vienen dadas por (4.2.8). Esto es por una parte debido a que, en (4.2.1) interviene el dato inicial $e^{-\nu \int_{-\infty}^{x} u_0(\sigma) d\sigma}$, que no preserva la paridad del dato inicial, y a que en el numerador de dicha expresión interviene no sólo el núcleo G sino también G_x.

Lo dicho hasta ahora es válido cuando $\nu > 0$. La expresión (4.2.18) presenta una singularidad para $\nu = 0$. Tal y como veremos más adelante, el límite cuando $\nu \to 0$ de la solución $u_\nu = u_\nu(x,\, t)$ de (4.2.1) es la solución de la ecuación de Burgers sin viscosidad (4.2.2).

Analicemos ahora (4.2.2). En este caso la transformación de Hopf-Cole no se aplica. De las ideas utilizadas en el caso viscoso la única que puede ser empleada es la de realizar el cambio de variables

$$v(x,\, t) = \int_{-\infty}^{x} u(\sigma,\, t) d\sigma$$

que reduce (4.2.2) a la ecuación de Hamilton-Jacobi

$$v_t + \mid v_x \mid^2 = 0.$$

Pero esta ecuación no se puede linealizar de modo que es mejor resolver directamente (4.2.2) mediante el método de las características.

Escribimos (4.2.2) en la forma

$$u_t + 2uu_x = 0. \tag{4.2.21}$$

Esto permite comprobar que las soluciones de (4.2.20), mientras son de clase C^1, son constantes a lo largo de las características, i.e.

$$u(x(t),\, t) = C,$$

donde $x = x(t)$ está caracterizada por la ecuación

$$x'(t) = 2u(x(t),\, t). \tag{4.2.22}$$

Estas curvas son fáciles de calcular. En efecto, como u es constante a lo largo de características, $u(x(t),\, t)$ ha de coincidir con su valor en $t = 0$, de modo que

$$u(x(t),\, t) = u_0(x_0), \tag{4.2.23}$$

donde x_0 es el punto de partida de la característica. La ecuación de la recta característica es entonces

$$x(t) = 2u_0(x_0)t + x_0 \tag{4.2.24}$$

y por tanto,

$$u(x,\, t) = u_0(x_0), \tag{4.2.25}$$

donde $(x,\, t)$ y x_0 están relacionadas a través de la identidad (4.2.24).

En virtud de este análisis se comprueba que u es constante a lo largo de líneas características de pendiente $1/2u_0$ en el plano $(x,\, t)$. De este análisis se deduce que si el dato inicial u_0 es decreciente, u ha de generar una discontinuidad en tiempo finito. En efecto, en la medida en que existen dos puntos x_0, x_1 tales que $x_0 < x_1$ y $u(x_0) > u(x_1)$, las características que arrancan de x_0 y x_1 se cruzarán en un tiempo finito t^* en un punto x. La solución habrá de ser discontinua en $(x^*,\, t^*)$ puesto que los dos valores $u_0(x_0)$ y $u_0(x_1)$ son incompatibles.

Un análisis más cuidadoso permite calcular el tiempo t^* en el que se produce la discontinuidad o choque. En efecto, las características que, según la ecuación (4.2.23), arrancan de x_0 y x_1, se encuentran en tiempo t^* si

$$2u_0(x_0)t + x_0 = 2u_0(x_1)t + x_1.$$

Esto se produce en tiempo

$$t^* = \frac{x_1 - x_0}{2(u_0(x_0) - u_0(x_1))} = -\frac{x_0 - x_1}{2(u_0(x_0) - u_0(x_1))}. \tag{4.2.26}$$

Cuando $x_0 \to x_1$ el tiempo t^* tiene como límite

$$t^* = -\frac{1}{2u_0'(x_0)}. \tag{4.2.27}$$

De esta expresión se deduce que el tiempo mínimo en el que se produce el choque es

$$t^* = \frac{1}{2\,\text{máx}_{x_0\in\mathbb{R}}\Big(-u_0'(x_0)\Big)}. \tag{4.2.28}$$

Vemos por tanto que el comportamiento de la ecuación de Burgers viscosa (4.2.1) y la no viscosa (4.2.2) es muy dispar en la medida que, mientras las soluciones de (4.2.1) son regulares para cualquier $\nu > 0$, las soluciones de (4.2.2) son discontinuas cuando el dato inicial es decreciente. Basta de hecho que el dato inicial sea decreciente en un intervalo para que los choques se produzcan en tiempo finito. A pesar de ello, como veremos en la próxima sección, las soluciones en ausencia de viscosidad, son el límite de las de la ecuación de Burgers viscosa cuando $\nu \to 0^+$.

4.3. Viscosidad evanescente

En esta sección analizamos el comportamiento límite cuando $\nu \to 0$ de las soluciones (4.2.18) de la ecuación de Burgers (4.2.1).

Al pasar al límite en (4.2.18) cuando $\nu \to 0^+$, las integrales que intervienen se concentran en torno a los puntos en los que H alcanza su mínimo. Calculamos por tanto los puntos críticos de H:

$$H_y = -\frac{x-\xi}{2t} + u_0(y) = 0 \Leftrightarrow \xi = x - 2tu_0(y), \tag{4.3.1}$$

en los que

$$H = -tu_0^2(\xi) + \int_{-\infty}^{\xi} u_0(\sigma)d\sigma. \tag{4.3.2}$$

La contribución de una integral

$$\int_{\mathbb{R}} f(y)e^{-H/\nu}dy \tag{4.3.3}$$

en un entorno de un punto de mínimo $y = \xi$ es

$$f(\xi)\sqrt{\frac{2\pi\nu}{H''(\xi)}}e^{-H(\xi)/\nu}. \tag{4.3.4}$$

En nuestro caso

$$H''(\xi) = \frac{1}{2t}. \tag{4.3.5}$$

Aplicando estas fórmulas en las integrales que intervien en (4.2.18) obtenemos

$$\int_{\mathbb{R}} (x-y)e^{-H/\nu}dy \sim (x-\xi)\sqrt{\frac{\pi\nu}{t}}e^{-[tu_0^2(\xi)+\int_{-\infty}^{\xi} u_0(\sigma)d\sigma]\nu}, \tag{4.3.6}$$

$$\int_{\mathbb{R}} e^{-H/\nu}dy \sim \sqrt{\frac{\pi\nu}{t}}e^{-[tu_0^2(\xi)+\int_{-\infty}^{\xi} u_0(\sigma)d\sigma]}. \tag{4.3.7}$$

Por tanto,

$$u_\nu(x,\, t) \sim \frac{(x-\xi)}{2t} \tag{4.3.8}$$

donde ξ viene caracterizado por la ecuación

$$\xi = x - 2tu_0(\xi) \tag{4.3.9}$$

que es exactamente la expresión obtenida en la sección anterior para la ecuación de Burgers sin viscosidad (4.2.2), puesto que $(x-\xi)/2t = u_0(\xi)$.

Esta expresión es válida cuando la función H sólo tiene un mínimo. En caso en que H tiene varios puntos de mínimo $\xi_1, \ldots, \xi_N$, cada uno de ellos

contribuye de manera semejante a las integrales que aparecen en (4.2.18). Sin embargo, debido a la presencia el factor exponencial, en la determinación de la forma asintótica de u_ν, sólo intervienen los mínimos absolutos de H. En caso, por ejemplo, de que H posea dos mínimos absolutos ξ_1, ξ_2, la forma asintótica de u_ν sería:

$$u_\nu(x,\, t) \sim u_0(\xi_1) + u_0(\xi_2). \tag{4.3.10}$$

Más adelante justificaremos rigurosamente estas asíntotas. Pero, por el momento analicemos su significado en lo que se refiere a la ecuación de Burgers (4.2.2).

Distinguimos dos casos:

Caso 4.3.1 *Dato inicial u_0 creciente.*

En este caso, tal y como vimos en la sección anterior, el método de las características no predice ningún choque y esperamos que la ecuación de Burgers admita una solución regular siempre y cuando el dato inicial u_0 sea regular.

El análisis asintótico que acabamos de realizar confirma este hecho. En efecto,

$$u_\nu(x,\, t) \to u_0(\xi),\ \nu \to 0, \tag{4.3.11}$$

donde $\xi \in \mathbb{R}$ viene caracterizado por la ecuación

$$\xi + 2tu_0(\xi) = x. \tag{4.3.12}$$

Para $t > 0$, si $u_0(\cdot)$ es creciente, la función

$$\xi \to \xi + 2tu_0(\xi), \tag{4.3.13}$$

también lo es, y por tanto (4.3.12) admite una única solución. Esto nos confirma sin ambigüedad alguna que el límite de las soluciones u_ν de la ecuación de Burgers viscosa, cuando la viscosidad $\nu \to 0$, es la solución de Burgers sin viscosidad obtenida por el método de las características.

Consideramos ahora el caso particular de un dato discontinuo, constante a trozos

$$u_0(x) = \begin{cases} 0, & x < 0 \\ 1, & x > 0. \end{cases} \tag{4.3.14}$$

Resolvemos la ecuación de Burgers sin viscosidad (4.2.2) con este dato. Es lo que se conoce como problema de Riemann, elemento básico del conocido método de Riemann para la aproximación del dato inicial mediante datos iniciales constantes a trozos.

Si aplicamos el método de las características con el dato inicial u_0 de (4.3.14) obtenemos que la solución u de (4.2.2) es de la forma

$$u = \begin{cases} 0, & x < 0,\, t > 0 \\ 1, & x \geq t. \end{cases} \tag{4.3.15}$$

Sin embargo, el método de las características no proporciona ningún valor para la solución u en la región $0 < x < t$.

El método de la viscosidad evanescente es en realidad un modo de obtener una solución globalmente definida en este caso. En efecto, consideramos la ecuación (4.3.12) en este caso particular. Este sistema se escribe

$$\begin{cases} \xi = x, & \text{si} \quad \xi < 0 \\ \xi + 2t\xi = x, & \text{si} \quad \xi > 0. \end{cases} \tag{4.3.16}$$

Cuando $x < 0$, esto nos da como solución $\xi = x$ y por consiguiente, según (4.3.11), el valor límite $u = u_0(\xi) = 0$. Esto coincide con lo obtenido en (4.3.15). Cuando $x > 0$ obtenemos $\xi = x/(1+2t)$. En el límite cuando $\nu \to 0$, por (4.3.12) obtenemos entonces

$$u(x,\, t) = \frac{x - \xi}{t} = \frac{x}{t} \tag{4.3.17}$$

que está ahora globalmente definida para todo $x \in \mathbb{R}$ y $t > 0$.

Evidentemente el frente $u = x/t$ conecta adecuadamente el valor constante $u = 0$ a izquierda y el valor $x = 1$ a derecha. Se trata de una *onda de rarefacción*.

Acabamos de ver que el método de la viscosidad evanescente proporciona en el límite una solución de la ecuación de Burgers sin viscosidad (4.2.2). Es lo que se denomina *la solución de entropía*.

Tal y como vamos a ver, la ecuación de Burgers puede incluso tener varias soluciones débiles. En este caso, la solución de entropía es la que tiene significado físico puesto que el modelo sin difusión ha de entenderse como una idealización del caso en que la viscosidad es pequeña y tiende a cero.

En la situación presente existen también otras soluciones débiles aunque, como decíamos, la única con significado físico es la denominada de entropía que acabamos de obtener. En vista de que la solución u se anula para $x < 0$ y toma el valor $u = 1$ para $x > t$ es natural considerar también soluciones de la forma

$$u = \begin{cases} 0, & x < \alpha t \\ 1, & x > \alpha t \end{cases} \tag{4.3.18}$$

donde la recta $x = \alpha t$ es a determinar. Veámos cual es el valor de α que hemos de elegir para que la función u dada por (4.3.14) puede ser una solución débil

de la ecuación de Burgers. Recordemos que por ser solución de la ecuación de Burgers

$$u_t + (u^2)_x = 0 \tag{4.3.19}$$

es preciso que

$$\int_0^\infty \int_{\mathbb{R}} \left(u\varphi_t + u^2\varphi_x\right) dxdt = 0 \tag{4.3.20}$$

para toda función test $\varphi \in C_0^\infty(\mathbb{R} \times (0, \infty))$.

En el caso de una función de la forma (4.3.14), (4.3.16) se reduce a

$$\int_0^\infty \int_{\alpha t}^\infty (\varphi_t + \varphi_x)\, dxdt = 0 \tag{4.3.21}$$

lo cual ocurre sí y sólo sí $\alpha = 1$.

Vemos por tanto que la ecuación de Burgers admite también como solución débil aquella que propaga el choque a velocidad 1. Ahora bien esta última solución de choque no es una solución de entropía. Tal y como hemos visto anteriormente la única solución de entropía es aquella que desarrolla la onda de rarefacción.

Un resultado clásico e importante de Kruzkov (véase la sección 11.4.3 de [8]) asegura que la solución de entropía es única. Esta puede caracterizarse no sólo como la solución obtenida por el método de viscosidad evanescente. Otra manera de caracterizarla es, por ejemplo, la siguiente cota unilateral sobre la derivada de la solución

$$u_x \leq 1/2t. \tag{4.3.22}$$

Esta cota se obtiene del siguiente modo. Formalmente, si u resuelve (4.3.19), entonces $v = u_x$ satisface

$$v_t + (2uv)_x = v_t + 2v^2 + 2uv_x = 0. \tag{4.3.23}$$

Aplicando el principio del máximo, deducimos que

$$v \leq w \tag{4.3.24}$$

donde $w = w(t)$ es la solución de

$$w_t + 2w^2 = 0 \tag{4.3.25}$$

con dato inicial $w(0) = \infty$. Esta solución puede calcularse explícitamente: $w(t) = 1/2t$.

Ahora bien, >podemos justificar la aplicación del principio del máximo para ecuaciones de la forma (4.3.23) para obtener la comparación (4.3.23)? Esta justificación puede en efecto hacerse para las soluciones de entropía. En efecto, si u

es solución de entropía, es el límite cuando $\nu \to 0$ de soluciones u_ν con viscosidad $\nu > 0$. Derivando sus ecuaciones vemos que $v_\nu = u_{\nu,\,x}$ satisface entonces

$$v_{\nu,\,t} - \nu v_{v,\,xx} + 2v_\nu^2 + 2u_\nu v_{\nu,\,x} = 0. \tag{4.3.26}$$

En esta ecuación parabólica podemos aplicar el principio del máximo y deducir que

$$v_\nu \leq 1/2t \tag{4.3.27}$$

para todo $\nu > 0$. Pasando al límite cuando $\nu \to 0$ obtenemos la cota (4.3.22) para $u_x = v$.

Caso 4.3.2 *Dato inicial u_0 decreciente.*

Según el análisis previo debemos de analizar los valores de ξ para los que

$$\xi + 2tu_0(\xi) = x. \tag{4.3.28}$$

Ahora bien, como u_0 es decreciente, no está garantizado que (4.3.28) admita una única solución para cada $x \in \mathbb{R}$ y $t > 0$. Más bien al contrario, para $t > 0$ suficientemente grande, el término $2tu_0(\cdot)$ de la izquierda de (4.3.28) dominará y destruirá el carácter monótono creciente de la función $\xi + 2tu_0(\xi)$. De hecho, el análisis de las características ya predecía la aparición de choques en este caso. Ambos hechos son reflejo del mismo fenómeno.

Consideramos nuevamente como caso modelo el problema de Riemann con dato inicial

$$u_0(x) = \begin{cases} 1, & x < 0 \\ 0, & x > 0. \end{cases} \tag{4.3.29}$$

En este caso, el cambio de variables (4.3.28) no puede aplicarse puesto que u_0 no es integrable en $-\infty$. Para evitar esta dificultad definimos

$$\tilde{u}_\nu(x,\,t) = -u_\nu(-x,\,t) \tag{4.3.30}$$

de modo que

$$u_\nu(x,\,t) = -\tilde{u}_\nu(-x,\,t). \tag{4.3.31}$$

La función $\tilde{u}$ es solución de la misma ecuación de Burgers viscosa pero con dato inicial

$$\tilde{u}_0 = \begin{cases} 0, & x < 0 \\ -1, & x > 0. \end{cases} \tag{4.3.32}$$

Tenemos ahora que estudiar las raíces de la ecuación

$$\xi + 2t\tilde{u}_0(\xi) = x \tag{4.3.33}$$

que puede escribirse en la forma

$$\begin{cases} \xi = x, & \text{si} \quad \xi < 0 \\ \xi - 2t = x, & \text{si} \quad \xi > 0. \end{cases} \tag{4.3.34}$$

De (4.3.34) deducimos que para $-2t < x < 0$ se obtienen dos soluciones distantes $\xi_1 = x$ y $\xi_2 = x + 2t$. Para el resto de valores de $(x,\, t)$ se tiene una única solución.

Tenemos ahora que analizar en cuál de estos puntos el valor crítico ξ el valor de H es mínimo. Según (4.3.2), en cada punto crítico ξ el valor de H es

$$H(\xi) = t\tilde{u}_0(\xi) + \int_{-\infty}^{\xi} \tilde{u}_0(s)ds. \tag{4.3.35}$$

Por tanto, cuando $\xi < 0$,

$$H(\xi) = 0 \tag{4.3.36}$$

mientras que, cuando $\xi > 0$, por ejemplo, si $\xi = \xi_2$,

$$H(\xi_2) = t - \int_0^{\xi_2} ds = t - \xi_2 = t - x - 2t = -(t + x). \tag{4.3.37}$$

Por tanto

$$\begin{cases} H(\xi_1) < H(\xi_2) & \text{si} \quad t + x < 0 \\ H(\xi_2) < H(\xi_1) & \text{si} \quad t + x > 0. \end{cases} \tag{4.3.38}$$

Deducimos así que:

$$u_\nu(x,\, t) \sim u(x,\, t) \tag{4.3.39}$$

donde

$$u(x,\, t) = \begin{cases} 1, & x < t \\ 0, & x > t. \end{cases} \tag{4.3.40}$$

Nuevamente en el límite (4.3.40) obtenemos una solución de entropía de la ecuación de Burgers sin viscosidad (4.2.2). Se trata de una onda de choque que se propaga a una velocidad 1. Esto es coherente con la condición de Rankine-Hugoniot que se precisa para que una función discontinua pueda ser solución débil de la ecuación de Burgers. Suponiendo que $u^{\pm}$ son los valores a izquierda y derecha del choque la velocidad de propagación es precisamente

$$s = \frac{(u^+)^2 - (u^-)^2}{u^+ - u^-}.$$

En nuestro caso, con valores $u^- = 1$ y $u^+ = 0$, esto nos da una velocidad de propagación unidad.

Hemos por tanto visto que el método de la viscosidad evanescente permite obtener la solución de la ecuación de Burgers de entropía. Aunque no hayamos probado aquí la unicidad, esta solución es la única que tiene sentido físico.

Mientras la solución es regular coincide con la obtenida por características. Cuando no lo es puede ser de dos tipos: O bien una onda de rarefacción, o una onda de choque, dependiendo del signo del choque, i.e. de los valores relativos de la solución a cada lado del choque. Cuando se tiene choque, éste se propaga con una velocidad que es la indicada por la condición de Rankine-Hugoniot.

Con el objeto de completar esta sección comprobamos por último que el valor asintótico de la integral

$$\int_{\mathbb{R}} f(y)e^{-H/\nu}dy$$

cuando $\nu \to 0$ es del orden de

$$f(\xi)\sqrt{\frac{2\pi\nu}{H''(\xi)}}e^{-H(\xi)/\nu},$$

siendo ξ el punto mínimo de H. De manera rigurosa, lo que esto significa es que

$$\lim_{\nu\to\infty} \frac{\displaystyle\int_{\mathbb{R}} f(y)e^{-H(y)/\nu}dy}{f(\xi)\sqrt{\frac{2\pi\nu}{H''(xi)}}e^{-H(\xi)/\nu}} = 1.$$

Para comprobarlo, basta ver que

$$J_\nu(y) = \left(\frac{2\pi\nu}{H''(\xi)}\right)^{-1/2} e^{-(H(y)-H(\xi))/\nu}$$

es una aproximación de la identidad cuando $\nu \to \infty$. Es obvio que, cuando $\nu \to \infty$, $J_\nu(y) \to 0$ exponencialmente salvo en el punto de mínimo $y = \xi$.

Por otra parte,

$$\int_{\mathbb{R}} J_\nu(y)dy \to 1,\ \nu \to \infty.$$

En efecto, habida cuenta de que

$$H(y) - H(\xi) = \frac{H''(\xi)}{2}(y\xi)^2 + O\left(\mid y - \xi \mid^3\right)$$

y haciendo el cambio de variables

$$\sigma = \sqrt{\frac{H''(\xi)}{2\nu}}(y - \xi)$$

el problema se reduce a probar que

$$\frac{1}{\sqrt{\pi}}\int_{\mathbb{R}} e^{-\eta^2} e^{-0(\nu^{1/2}\eta^3)} d\eta \to 1$$

lo cual puede probarse mediante el Teorema de la Convergencia dominada.

El lector interesado en un estudio más detallado de estas cuestiones, para leyes de conservación hiperbólicas más generales, podrá consultar el libro de C. Evans [8].

4.4. Aplicación del "splitting" a la ecuación de Burgers

En esta sección vamos a aplicar el θ-método de splitting descrito en la sección anterior a la ecuación de Burgers viscosa:

$$\begin{cases} u_t - \nu u_{xx} + (u^2)_x = f, & 0 < x < 1, \quad t > 0 \\ u = 0, & x = 0,\, 1, \quad t > 0 \\ u(x,\, 0) = u_0(x), & 0 < x < 1. \end{cases} \tag{4.4.1}$$

Estamos considerando el problema en un abierto acotado con condiciones de contorno de Dirichlet. Obviamente, al abordar el problema de Cauchy en toda la recta real esto exige un primer paso en la aproximación consistente en sustituir la recta real $\mathbb{R}$ por un intervalo acotado $[-k,\, k]$ con k suficientemente grande. El error cometido en dicha aproximación puede estimarse mediante un método de energía. De este modo acabamos habiendo de abordar el problema de Dirichlet en un intervalo acotado, problema que nos ocupa en esta sección.

Aplicando el θ-método obtenemos la siguiente secuencia de ecuaciones:

$$\begin{cases} \dfrac{u^{k+\theta} - u^k}{\theta \Delta t} - \alpha\nu u_{xx}^{k+\theta} = f^{k+\theta} = \beta\nu u_{xx}^k - \left(\left(u^k\right)^2\right)_x, & 0 < x < 1 \\ u^{k+\theta} = 0,\, x = 0,\, 1, \end{cases} \tag{4.4.2}$$

$$\begin{cases} \dfrac{u^{k+1-\theta} - u^{k+\theta}}{(1-2\theta)\Delta t} - \beta\nu u_{xx}^{k+1-\theta} + \left(\left(u^{k+1-\theta}\right)^2\right)_x = f^{k+\theta} + \alpha\nu u_{xx}^{k+\theta}, & 0 < x < 1 \\ u^{k+1-\theta} = 0,\, x = 0,\, 1, \end{cases} \tag{4.4.3}$$

$$\begin{cases} \dfrac{u^{k+1} - u^{k+1-\theta}}{\Delta t} - \alpha\nu u_{xx}^{k+1} = f^{k+1} + \beta\nu u_{xx}^{k+1-\theta} - \left(\left(u^{k+1-\theta}\right)^2\right)_x, & 0 < x < 1 \\ u^{k+1} = 0,\, x = 0,\, 1. \end{cases} \tag{4.4.4}$$

Conviene señalar que (4.4.2) y (4.4.4) se reducen a resolver un *problema de Dirichlet lineal*, que puede ser resuelto utilizando un método de elementos finitos, lo cual proporciona un esquema completamente discreto.

El problema (4.4.3) es no-lineal aunque los experimentos numéricos indican que se obtienen esencialmente los mismos resultados si la no-linealidad $\left(\left(u^{k+1-\theta}\right)^2\right)_x$ se sustituye por $2\left(u^{k+\theta}\left(u^{k+1-\theta}\right)_x\right)$, en cuyo caso el sistema reducido que se obtiene tiene la virtud de ser lineal.

En el contexto de las ecuaciones de Navier-Stokes, el θ-método tiene la virtud de permitir desacoplar la no-linealidad de la condición de incompresibilidad. Obtenemos así

$$\begin{cases} \dfrac{u^{k+\theta}-u^k}{\theta\Delta t} - \alpha\nu\Delta u^{k+\theta} + \nabla p^{k+\theta} = f^{k+\theta} + \beta\nu\Delta u^k - \left(u^k\cdot\nabla\right)u^k \text{ en } \Omega, \\ \nabla\cdot u^{k+\theta} = 0 \text{ en } \Omega, \\ u^{k+\theta} = 0 \text{ en } \partial\Omega, \end{cases}$$

$$\begin{cases} \dfrac{u^{k+1-\theta}-u^{k+\theta}}{(1-2\theta)\Delta t} - \beta\nu\Delta u^{k+1-\theta} + \left(u^{k+1-\theta}\cdot\nabla\right)u^{k+1-\theta} = f^{k+\theta} + \alpha\nu\Delta u^{k+\theta} - \nabla p^{k+\theta} \text{ en } \Omega \\ u^{k+1-\theta} = 0 \text{ en } \Omega, \end{cases}$$

$$\begin{cases} \dfrac{u^{k+1}-u^{k+1-\theta}}{\theta\Delta t} - \alpha\nu\Delta u^{k+1} + \nabla p^{k+1} = f^{k+1} + \beta\nu\Delta u^{k+1-\theta} - \left(u^{k+1-\theta}\cdot\nabla\right)u^{k+1-\theta} \text{ en } \Omega \\ \nabla\cdot u^{k+1} = 0 \text{ en } \Omega, \\ u^{k+1} = 0 \text{ en } \partial\Omega. \end{cases}$$

4.5. Ecuaciones elípticas de convección-difusión

Como hemos visto en la sección anterior, al aplicar el método de splitting a la ecuación de Burgers viscosa obtenemos una familia de problemas elípticos con convección cuadrática de la forma

$$\begin{cases} -\nu u_{xx} + (u^2)_x = f, \quad 0 < x < 1 \\ u = 0,\ x = 0,\ 1. \end{cases} \tag{4.5.1}$$

En esta sección analizamos brevemente la existencia y unicidad de soluciones para esta ecuación.

Comenzamos recordando los resultados elementales para el problema de Dirichlet lineal en ausencia de convección:

$$\begin{cases} -\nu u_{xx} = f, \qquad 0 < x < 1 \\ u = 0,\ x = 0,\ 1. \end{cases} \tag{4.5.2}$$

En este caso es bien sabido que para cada $f \in H^{-1}(0,\,1)$ existe una única solución débil $u \in H_0^1(0,\,1)$ que satisface

$$\begin{cases} \nu \displaystyle\int_0^1 u_x \varphi_x dx = \langle f,\, \varphi \rangle,\ \forall \varphi \in H_0^1(0,\,1) \\ u \in H_0^1(0,\,1). \end{cases} \tag{4.5.3}$$

En (4.5.3) $\langle \cdot,\, \cdot \rangle$ denota el producto de dualidad entre $H^{-1}(0,\,1)$ y $H_0^1(0,\,1)$. Además, cuando $f \in L^2(0,\,1)$ la solución pertence a $H^2(0,\,1)$.

La solución débil puede obtenerse mediante la aplicación directa del Lema de Lax-Milgram o bien mediante el Método Directo del Cálculo de Variaciones (MDCV), minimizando el funcional

$$\begin{cases} J:\, H_0^1(0,\,1) \longrightarrow \mathbb{R}, \\ J(v) = \dfrac{1}{2} \displaystyle\int_0^1 \mid v_x \mid^2 dx - \langle f,\, v \rangle. \end{cases} \tag{4.5.4}$$

El modo más simple de abordar el problema no-lineal (4.5.1) es mediante una técnica de punto fijo. Para cada $v \in H_0^1(0,\,1)$ podemos considerar el problema

$$\begin{cases} -\nu u_{xx} = f - \left(v^2\right)_x,\ 0 < x < 1 \\ u(0) = u(1) = 0. \end{cases} \tag{4.5.5}$$

Como $f - \left(v^2\right)_x \in H^{-1}(0,\,1)$, el problema (4.5.5) admite una única solución $u \in H_0^1(0,\,1)$. Esto nos permite definir una aplicación no-lineal $\mathcal{N}:\, H_0^1(0,\,1) \to H_0^1(0,\,1)$ que a $v \in H_0^1(0,\,1)$ asocia $\mathcal{N}v = u$. No es difícil de comprobar que esta aplicación es compacta. Basta para ello constatar que si v varía en un conjunto acotado de $H_0^1(0,\,1)$, entonces $\left(v^2\right)_x = 2vv_x$ varía en un conjunto acotado de $L^2(0,\,1)$ y por tanto en un conjunto compacto de $H^{-1}(0,\,1)$. Es pues natural aplicar el Teorema de Schauder. Pero esto no es directamente posible. En efecto

$$\begin{aligned} \nu \parallel \mathcal{N}(v) \parallel^2_{H_0^1(0,\,1)} &= \parallel f - \left(v^2\right)_x \parallel_{H^{-1}(0,\,1)} \leq \parallel f \parallel_{H^{-1}(0,\,1)} + \parallel v^2 \parallel_{L^2(0,\,1)} \\ &\leq \parallel f \parallel_{H^{-1}(0,\,1)} + C \parallel v \parallel^2_{H_0^1(0,\,1)}. \end{aligned} \tag{4.5.6}$$

Por tanto,

$$\parallel u \parallel_{H_0^1(0,\,1)} \leq \frac{\parallel f \parallel_{H^{-1}(0,\,1)}}{\nu} + \frac{C}{\nu} \parallel v \parallel^2_{H_0^1(0,\,1)}. \tag{4.5.7}$$

Con el objeto de aplicar el Teorema de Schauder lo más simple es buscar una bola B_R de $H_0^1(0,\,1)$ en la que la aplicación de $\mathcal{N}$ sea invariante. En virtud de la estimación (4.5.7) esto exige que

$$\frac{\parallel f \parallel_{H^{-1}(0,\,1)}}{\nu} + \frac{C}{\nu} R^2 \leq R, \tag{4.5.8}$$

lo cual es posible para cualquier $f \in H^{-1}(0,\,1)$ si $\nu > 0$ es suficientemente grande o bien para $\nu > 0$ arbitrario, siempre y cuando $\| f \|_{H^{-1}(0,\,1)}$ sea suficientemente pequeño con respecto a ν.

Pero este punto de vista no permite resolver el problema no-lineal (4.5.1) en toda su generalidad. Con el objeto de hacerlo introducimos una aproximación acotada de la no-linealidad cuadrática que interviene en (4.5.1), i.e.

$$\phi_k(s) = \text{mín}(s^2,\, k). \tag{4.5.9}$$

Consideramos ahora, en lugar de (4.5.1), el problema con no-linealidad truncada

$$\begin{cases} -\nu u_{xx} + (\phi_k(u))_x = f, \quad 0 < x < 1 \\ u(0) = u(1) = 0. \end{cases} \tag{4.5.10}$$

En esta ocasión el argumento anterior permite concluir la existencia de una solución $u_k \in H_0^1(0,\,1)$. En efecto, en el presente caso, la condición sobre el radio R de la bola B_R que se precisa para aplicar el Teorema del punto fijo de Schauder es simplemente

$$\frac{\| f \|_{H^{-1}(0,\,1)}}{\nu} + \frac{Ck^2}{\nu} \leq R, \tag{4.5.11}$$

que, evidentemente, se cumple si $R > 0$ es suficientemente grande.

Por otra parte, no es difícil obtener una cota uniforme sobre la sucesión de soluciones aproximadas $\{u_k\}_{k\geq 0}$. En efecto, multiplicando en (4.5.10) por u_k e integrando por partes o, más bien, utilizando la propia función u_k como función test en la formulación débil de (4.5.10) obtenemos

$$\nu \int_0^1 |u_{k,\,x}|^2\, dx - \int_0^1 \phi_k(u_k) u_{k,\,x} dx = \langle f,\, u_k \rangle. \tag{4.5.12}$$

Ahora bien, como

$$\phi_k(u_k) u_{k,\,x} = \frac{\partial}{\partial x} \left(\psi_k \left(u_k \right) \right)$$

donde

$$\psi_k(s) = \int_\sigma^s \phi_k(\sigma) d\sigma,$$

se tiene

$$\int_0^1 \phi_k(u_k) u_{k,\,x} dx = \int_0^1 \frac{\partial}{\partial x} \left(\psi_k \left(u_k \right) \right) dx = 0. \tag{4.5.13}$$

Por tanto

$$\nu \int_0^1 |u_{k,\,x}|^2\, dx = \langle f,\, u_k \rangle \tag{4.5.14}$$

lo cual implica la cota

$$\| u_k \|_{H_0^1(0,\,1)} \leq \frac{1}{\nu} \| f \|_{H^{-1}(0,\,1)},\ \forall k \geq 1. \tag{4.5.15}$$

Esto permite pasar al límite en las soluciones aproximadas. En efecto, como $\{u_k\}$ está acotada en $H_0^1(0,\,1)$ se puede extraer una subsucesión, que seguimos denotando mediante $\{u_k\}$, tal que

$$u_k \rightharpoonup u \text{ débilmente en } H_0^1(0,\,1). \tag{4.5.16}$$

Esta sucesión puede además extraerse de modo que

$$u_k \longrightarrow u \text{ fuertemente en } L^2(0,\,1) \tag{4.5.17}$$

y

$$u_k \longrightarrow u \,\text{p.c.t. } x \in (0,\,1). \tag{4.5.18}$$

Esto permite pasar al límite en la formulación débil de (4.5.10):

$$\int_0^1 u_{k,\,x}\varphi_x dx - \int_0^1 \phi_k(u_k)\varphi_x dx = \langle f,\,\varphi\rangle,\ \forall \varphi \in H_0^1(0,\,1). \tag{4.5.19}$$

En efecto, pasando al límite en (4.5.19) obtenemos que el límite $u \in H_0^1(0,\,1)$ es solución débil de (4.5.1) puesto que satisface

$$\int_0^1 u_x\varphi_x dx - \int_0^1 u^2\varphi_x dx = \langle f,\,\varphi\rangle,\ \forall \varphi \in H_0^1(0,\,1). \tag{4.5.20}$$

En virtud de la convergencia débil en $H_0^1(0,\,1)$ de $\{u_k\}$, la única dificultad para obtener (4.5.20) de (4.5.19) es el paso al límite en el término no-lineal. Esto puede hacerse comprobando que

$$\phi_k(u_k) \rightharpoonup u^2 \text{ débilmente en } L^2(0,\,1).$$

Esto es así puesto que, en virtud de (4.5.18),

$$\phi_k(u_k) \longrightarrow u^2,\ \text{p.c.t. } x \in (0,\,1)$$

y, por otra parte,

$$||\phi_k(u_k)||_{L^\infty(0,\,1)} \leq \left|\left|u_k^2\right|\right|_{L^\infty(0,\,1)} \leq \| u_k \|^2_{L^\infty(0,\,1)} \leq C||u\|^2_{H_0^1(0,\,1)} \leq C. \tag{4.5.21}$$

En este punto hemos usado el siguiente lema clásico de convergencia que se demuestra gracias al Teorema de Egorov.

Lemma 4.5.1 *Sea Ω un dominio acotado de $\mathbb{R}^n$. Sea $h_k : \Omega \longrightarrow \mathbb{R}$ una sucesión de funciones medibles tales que*

$$\| h_k \|_{L^p(\Omega)} \leq C, \forall k \geq 1$$

con $p > 1$ y

$$h_k \longrightarrow h, \text{ p.c.t. } x \in \Omega.$$

Entonces

$$h_k \longrightarrow h \text{ en } L^q(\Omega), \forall 1 \leq q < p$$

y

$$h_k \longrightarrow h \text{ débilmente en } L^p(\Omega),$$

si $p < \infty$ o débil-$*$ en L^∞ si $p = \infty$.

Gracias a este argumento que combina la aproximación por truncatura y el paso al límite hemos probado la existencia de al menos una solución de (4.5.1) para cada $f \in H^{-1}(0, 1)$ y cada $\nu > 0$.

Veamos ahora que esta solución es única.

Como es habitual en estos casos suponemos que existen dos soluciones u_1, u_2, definimos $v = u_1 - u_2$ e intentamos probar que $v \equiv 0$. La función v satisface

$$\begin{cases} -\nu v_{xx} + \left(u_1^2 - u_2^2\right)_x = 0, \quad 0 < x < 1, \\ v(0) = v(1) = 0, \end{cases} \tag{4.5.22}$$

Si bien el resultado de unicidad se cumple para todo $\nu > 0$, en estas notas lo probaremos únicamente para $\nu > 0$ suficientemente grande.

Multiplicando en (4.5.22) por v e integrando por partes obtenemos

$$\begin{aligned} \nu \int_0^1 v_x^2 dx &= \int_0^1 (u_1 + u_x)\, v v_x dx \\ &\leq \| u_1 + u_2 \|_{L^\infty(0,1)} \| v \|_{L^2(0,1)} \| v_x \|_{L^2(0,1)} \end{aligned}$$

de donde deducimos que

$$\nu \| v \|_{H_0^1(0,1)} \leq C \| u_1 + u_2 \|_{L^\infty(0,1)} \| v \|_{H_0^1(0,1)}$$

con $C > 0$, independiente de ν. Es decir

$$\nu \leq C \| u_1 + u_2 \|_{L^\infty(0,1)} . \tag{4.5.23}$$

Ahora bien, cualquier solución de (4.5.1) satisface

$$\nu \int_0^1 u_x^2 dx = \langle f, u \rangle.$$

Esto es así puesto que

$$\int_0^1 u^2 u_x dx = 0.$$

Por tanto

$$\nu \parallel u_x \parallel_{H_0^1(0,1)} \leq \parallel f \parallel_{H^{-1}(0,1)},$$

y por consiguiente,

$$\nu \parallel u \parallel_{L^\infty(0,1)} \leq C \parallel f \parallel_{H^{-1}(0,1)} . \tag{4.5.24}$$

Combinando (4.5.23) y (4.5.24) se deduce que

$$\nu \leq \frac{C}{\nu} \parallel f \parallel_{H^{-1}(0,1)},$$

lo cual es imposible si $\nu > 0$ es suficientemente grande.

Esto prueba la unicidad si $\nu > 0$ es suficientemente grande.

Los mismos argumentos permiten probar la existencia y unicidad (si $\nu > 0$ es grande) para las soluciones del problema de Navier-Stokes estacionario

$$\begin{cases} -v\Delta u + (u \cdot \nabla)u + \nabla p = f & \text{en} \quad \Omega \\ \nabla \cdot u = 0 & \text{en} \quad \Omega \\ u = 0 & \text{en} \quad \partial\Omega. \end{cases}$$

Esto justifica la existencia y unicidad de las soluciones aproximadas obtenidas mediante la aplicación del θ-método de splitting de la sección anterior.

4.6. Sistemas de leyes de conservación y soluciones de entropía

En este capítulo recogemos brevemente algunos aspectos teóricos sobre la existencia y unicidad de soluciones de ecuaciones escalares hiperbólicas.

Se trata de modelos de gran importancia en numerosos campos y, en particular, en Mecánica de Fluidos. Precisamente a causa de su importancia han sido objeto de un intensísimo estudio tanto en el ámbito teórico como en el diseño de métodos numéricos eficaces.

En este capítulo nos centraremos en los aspectos más básicos analizando sólo el caso de ecuaciones escalares aunque muchas de las ideas que desarrollaremos son aplicables en un contexto mucho más amplio y, son hoy en día utilizadas en aplicaciones muy importantes, en particular en el ámbito de la Aeronaútica.

En estas notas nos guiaremos por el libro de E. Godlewski y P. A. Raviart [11].

Aunque, como decíamos, nos centraremos en el caso de ecuaciones escalares, empezaremos presentando el sistema vectorial tipo:

$$\frac{\partial u}{\partial t} + \sum_{j=1}^{d} \frac{\partial}{\partial x_j}\big(f_j(u)\big) = 0,\ x = (x_1, \cdots, x_d) \in \mathbb{R}^d,\ t > 0 \tag{4.6.1}$$

donde la incógnita $u = u(x,\, t)$ es un vector de p componentes

$$u = \begin{pmatrix} u_1 \\ \vdots \\ u_p \end{pmatrix}.$$

Las no-linealidades o funciones de flujo f_j son también funciones vectoriales

$$f_j = \begin{pmatrix} f_{1j} \\ \vdots \\ f_{pj} \end{pmatrix},$$

que supondremos de clase C^1.

Se trata pues de un sistema de p ecuaciones en $\mathbb{R}^d_x \times \mathbb{R}_t$.

Frecuentemente escribiremos el sistema de manera más compacta del siguiente modo:

$$\frac{\partial u}{\partial t} + \operatorname{div}\big(f(u)\big) = 0, \tag{4.6.2}$$

donde div denota el operador de divergencia en las variables espaciales.

El sistema (4.6.1) representa una ley de conservación. En efecto, integrando las ecuaciones en un recinto D de $\mathbb{R}^d$ obtenemos que

$$\frac{d}{dt}\int_D u dx + \int_{\partial D} f(u)\cdot n\, d\sigma = 0$$

donde n denota el vector exterior unitario a D y $\cdot$ el producto escalar en $\mathbb{R}^d$.

Consideraremos la matriz Jacobiana del flujo

$$A_j(u) = \left(\frac{\partial f_{ij}(u)}{\partial u_k}\right)_{1\le i,\, k\le p}.$$

Algunos de los ejemplos más relevantes de este tipo de sistemas son:

• **La ecuación de Burgers**

Las ecuaciones de Burgers en su versión viscosa y no viscosa respectivamente son:

$$u_t - \nu u_{xx} + u\frac{\partial u}{\partial x} = 0,$$

y

$$u_t + \frac{\partial}{\partial x}\left(\frac{u^2}{2}\right) = 0.$$

Esta última se trata de un caso particular de ecuaciones escalares de la forma

$$u_t + \frac{\partial}{\partial x}\big(f(u)\big) = 0,$$

en las que el flujo f es convexo.

• **La ecuación de Buckley - Leverett**

Se trata de un modelo uni-dimensional para un flujo bi-fásico de fluidos inmiscibles. Un ejemplo típico de este tipo de modelos con porosidad constante $(= 1)$ y en el que se ignora los efectos de capilaridad y gravedad es:

$$\frac{\partial s}{\partial t} + \frac{\partial}{\partial x}\big(f(s)\big) = 0,$$

con

$$f(s) = \frac{s^2}{s^2 + (1-s^2)\frac{\mu_w}{\mu_o}},$$

donde μ denota la viscosidad (que suponemos constante) para los dos medios (el subíndice w se refiere al agua (water) y el o al petróleo (oil)).

La incógnita s representa la saturación. La no-linealidad en este caso es una función convexa y regular de [0, 1] en [0, 1].

• **El p-sistema**

Se trata de un sistema de dos ecuaciones para la dinámica de gases isentrópicos uni-dimensionales en coordenadas Lagrangianas:

$$\begin{cases} \dfrac{\partial v}{\partial t} - \dfrac{\partial u}{\partial x} = 0 \\ \dfrac{\partial u}{\partial t} + \dfrac{\partial}{\partial x}\big(p(v)\big) = 0. \end{cases}$$

En este sistema v es el volumen específico, u la velocidad y $p = p(v)$ una función de presión dada.

Conviene observar que toda ecuación de ondas de la forma

$$w_{tt} - \frac{\partial}{\partial x}\left(\sigma\left(\frac{\partial w}{\partial x}\right)\right) = 0$$

puede escribirse en esta forma en la variable

$$u = \frac{\partial w}{\partial t},\ v = \frac{\partial w}{\partial x}$$

con $p(v) = -\sigma(v)$.

• **El sistema de la dinámica de gases en coordenadas Eulerianas**

Bajo condiciones de simetría adecuadas el sistema es de la forma:

$$\begin{aligned}
&\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x}(\rho u) = 0,\\
&\frac{\partial (\rho u)}{\partial t} + \frac{\partial}{\partial x}\big(\rho u^2 + p\big) = 0,\\
&\frac{\partial}{\partial t}(\rho e) + \frac{\partial}{\partial x}\big((\rho e + p)u\big) = 0.
\end{aligned}$$

En este caso ρ representa la densidad, u la velocidad, p la presión y $e = \varepsilon + \mid u \mid^2 /2$ la energía específica total, siendo ε la energía interna específica. Este sistema puede escribirse como un sistema de la forma general (4.6.1) con $d = 1$ y $p = 3$ en las incógnitas ρ, $m = \rho\, u$ y $E = \rho e$.

Una de las propiedades más relevantes de estos sistemas es que, contrariamente a lo que ocurre típicamente en los sistemas lineales, incluso cuando los datos del problema son regulares, las soluciones no lo son y, más concretamente, desarrollan discontinuidades en tiempo finito.

Para convencerse de ello basta con considerar la ecuación escalar en una dimensión espacial:

$$\begin{cases} u_t + \partial_x\big(f(u)\big), & x \in \mathbb{R}, \quad t > 0 \\ u(x,\, 0) = u_0(x), & x \in \mathbb{R}. \end{cases} \tag{4.6.3}$$

Sea $a(u) = f'(u)$.

Es fácil entonces comprobar que las soluciones, mientras son regulares, son constantes a lo largo de las características que son rectas de la forma

$$x = x_0 + t\, a\big(u_0(x_0)\big).$$

Supongamos ahora que el dato inicial u_0 y la no-linealidad f son tales que existen dos puntos $x_1 < x_2$ tales que

$$m_1 = 1/a_0\big(u_0(x_1)\big) < m_2 = 1/a\big(u_0(x_2)\big).$$

Entonces, las características que arrancan de los puntos x_1 y x_2, que tienen pendientes m_1 y m_2 respectivamente necesariamente se cruzan en un punto P en tiempo finito. La solución no puede ser continua en este punto puesto que los dos valores $u_0(x_1)$ y $u_0(x_2)$ son incompatibles. El tiempo en el que se produce esta discontinuidad o choque es

$$t = \big(x_2 - x_1\big)/\big(a\big(u_0(x_1)\big) - a\big(u_0(x_2)\big)\big).$$

Esto demuestra que, salvo que la función $x \to a\big(u_0(x)\big)$ sea monótona creciente, el choque necesariamente se producirá en un cierto tiempo finito $t > 0$. El tiempo mínimo de explosión resulta ser:

$$T^* = \min_{y \in \mathbb{R}} \frac{d\big(a\big(u_0(y)\big)\big)}{dy}. \tag{4.6.4}$$

Este método, denominado de características, permite construir soluciones regulares, constantes a lo largo de las mismas, hasta el tiempo de explosión en el que se genera un choque o discontinuidad. Esto hace que sea necesario que elaboremos un concepto de solución débil, posiblemente discontinua.

Las soluciones débiles pueden caracterizarse del modo usual mediante funciones test. Más concretamente han de satisfacer

$$\begin{cases} 0 = \displaystyle\int_0^\infty \int_{\mathbb{R}^d} \left(u \cdot \frac{\partial \varphi}{\partial t} + \sum_{j=1}^d f_j(u) \cdot \frac{\partial \varphi}{\partial x_j} \right) dxdt + \int_{\mathbb{R}^d} u_0(x) \cdot \varphi(x,\, 0) dx \\ \quad \forall \varphi \in C_0^1\big(\mathbb{R}^d \times [0,\, \infty)\big), \end{cases} \tag{4.6.5}$$

siendo C_0^1 el espacio de las funciones de clase C^1 y de soporte compacto.

Este concepto de solución en el sentido de (4.6.5) tiene perfecto sentido en el marco de las funciones $u \in L^\infty_{loc}\big(\mathbb{R}^d \times (0,\, \infty)\big)$.

Esta formulación débil puede ser interpretada de manera muy gráfica y geométrica para funciones u de clase C^1 y que son discontinuas a lo largo de una hipersuperficie Σ. Supongamos que $n = \big(n_1, \cdots,\, n_d,\, n_t\big)$ es el vector normal a esta superficie y que u^+ y u^- son los valores de la solución a ambos lados de ella. Se tiene entonces que u es solución débil de (4.6.1) sí y sólo si se verifican las dos siguientes condiciones:

- u es una solución clásica de la ecuación a cada lado de la superficie Σ en la que es C^1;
- u satisface la siguiente condición de salto sobre Σ:

$$\big(u_+ - u_-\big) n_t + \sum_{j=1}^d \big(f_j\big(u_+\big) - f_j\big(u_-\big)\big) n_{x_j} = 0. \tag{4.6.6}$$

La condición (4.6.6) se denomina *condición de Rankine-Hugoniat (RH)*. Utilizando la notación $[\cdot]$ para el salto puede escribirse de manera más compacta del siguiente modo:

$$n_t[u] + \sum_{j=1}^d n_{x_j}[f_j(u)] = 0. \tag{4.6.7}$$

Esta condición relaciona la velocidad de propagación del salto con su amplitud. Por ejemplo, en una dimensión espacial ($d = 1$), suponiendo que Σ es una curva regular de parametrización $\big(t,\, \xi(t)\big)$, el vector normal es de la forma $n = (1,\, -s)$, $s = \xi'(t)$, y por tanto la condición de RH se escribe

$$s[u] = [f(u)]. \tag{4.6.8}$$

En el caso particular de la ecuación de Burgers en que $f(z) = z^2/2$ obtenemos

$$s = \big(u^+ + u^-\big)/2,$$

lo cual indica que el salto se propaga con una velocidad que coincide con la media de los valores de la solución a cada lado del mismo.

Consideremos algunos ejemplos.

Resolvemos la ecuación de Burgers

$$u_t + \partial_x \left(\frac{u^2}{2}\right) = 0 \tag{4.6.9}$$

con dato inicial

$$u(x,\, 0) = u_0(x) = \begin{cases} 1, & x \le 0 \\ 1 - x, & 0 \le x \le 1 \\ 0, & x > 1. \end{cases} \tag{4.6.10}$$

Las características son entonces de la forma

$$x(x_0,\, t) = \begin{cases} x_0 + t, & x_0 \le 0 \\ x_0 + t(1 - x_0), & 0 \le x_0 \le 1, \\ x_0, & x_0 \le 1. \end{cases}$$

Para $t < 1$ las características no se cruzan. Eso permite obtener la solución por el método de las características que preserva la misma regularidad que el dato inicial: Se trata de una función lineal a trozos continua y, por tanto, es Lipschitz.

Obtenemos así:

$$u = \begin{cases} 1, & x \le t \\ (1 - x)/(1 - t), & t \le x \le 1 \\ 0, & x \ge 1, \end{cases} \tag{4.6.11}$$

en el intervalo temporal $0 < t < 1$.

Cuando $t = 1$ se produce un choque en el punto $x = 1$ en el que entran en conflicto los valores 0 y 1 de $u\big(u^+$ y $u^-\big)$. En este caso la condición de RH nos

indica que el choque se habrá de propagar con velocidad $s = 1/2$. Definimos por tanto para $t \geq 1$ la siguiente función constante a trozos:

$$u(x,\, t) = \begin{cases} 1, & x < (t+1)/2, \\ 0, & x > (t+1)/2. \end{cases} \tag{4.6.12}$$

Las expresiones (4.6.11) y (4.6.12) proporcionan la solución de (4.6.9)-(4.6.10) buscada.

Consideramos ahora la ecuación de Burgers pero con dato inicial discontinuo:

$$u_0 = \begin{cases} u_\ell, & x < 0 \\ u_r, & x > 0. \end{cases}$$

Este problema de Cauchy se denomina *el problema de Riemann*.

La condición de RH nos proporciona una solución constante a trozos con una discontinuidad que se propaga con velocidad $s = \big(u_\ell + u_r\big)/2$. Obtenemos así

$$u(x,\, t) = \begin{cases} u_\ell, & x < \big(u_\ell + u_r\big)/2 \\ u_r, & x > \big(u_\ell + u_r\big)/2. \end{cases}$$

Sin embargo esta no es la única solución débil posible. De hecho pueden encontrarse muchas otras. En efecto, para cada $a \geq \max\big(u_\ell,\, -u_r\big)$ la función definida por

$$u = \begin{cases} u_\ell, & x < s_1 t \\ -a, & s_1 t < x < 0, \\ a, & 0 < x < s_2 t, \\ u_r, & x > s_2 t, \end{cases}$$

es también una solución débil si

$$s_1 = \big(u_1 - a\big)/2,\; s_2 = \big(u_r + a\big)/2,$$

de forma que la condición de RH se satisface a lo largo de cada choque. Obtenemos así una familia uniparamétrica de soluciones débiles discontinuas.

Por otra parte, cuando $u_\ell \leq u_r$, podemos también encontrar una solución continua pues las características no se cortan. De hecho, el método de las características permite determinar la solución (que toma valores u_ℓ y u_r) salvo en la región $u_\ell \leq x/t \leq u_r$. Este espacio puede rellenarse gracias a la función $v = x/t$ que es una solución de la ecuación de Burgers. Obtenemos así la solución continua

$$u(x,\, t) = \begin{cases} u_\ell, & x \leq u_\ell t \\ x/t, & u_\ell t \leq x \leq u_r t \\ u_r, & x \geq u_r t. \end{cases}$$

Es lo que se denomina una *onda de rarefacción.*

Estos ejemplos demuestran que, a pesar de que la noción de solución débil regular a trozos y con posibles choques a lo largo de los cuales se satisface la condición de RH, permite incorporar las soluciones discontinuas que las características necesariamente generan, no es un marco funcional suficiente para garantizar la unicidad. Para ello es necesario introducir un criterio de entropía que seleccione la solución "física" o "buena" en la clase de soluciones débiles existentes.

Son varias las maneras de introducir el criterio de entropía. En estas notas nos limitaremos al estudio de ecuaciones escalares de la forma

$$u_t + \operatorname{div}\big(f(u)\big) = 0,\ x \in \mathbb{R}^d,\ t > 0, \tag{4.6.13}$$

en las que la incógnita $u = u(x,\, t)$ es una función escalar.

La ecuación (4.6.13) es un modelo simplificado que se asemeja a las ecuaciones de Euler para un fluido perfecto o ideal, en ausencia de viscosidad. Pero estos no son más que una idealización o aproximación de los fluidos con viscosidad pequeña. Es por tanto natural considerar (4.6.13) como una simplificación del modelo viscoso

$$u_t - \varepsilon\Delta u + \operatorname{div}\big(f(u)\big) = 0,\ x \in \mathbb{R}^d,\ t > 0. \tag{4.6.14}$$

Consideraremos que una solución débil de (4.6.13) es una solución de entropía cuando sea el límite cuando $\varepsilon \to 0$ de soluciones u_ε del problema viscoso.

En el caso particular de la ecuación de Burgers

$$u_t + \left(\frac{u^2}{2}\right)_x = 0, \tag{4.6.15}$$

se trata por tanto de definir sus soluciones de entropía como aquéllas que son límite cuando $\varepsilon \to 0$ de las soluciones de Burgers viscosas:

$$u_t - \varepsilon u_{xx} + \left(\frac{u^2}{2}\right)_x = 0. \tag{4.6.16}$$

En este caso particular, la transformación de Hopf-Cole permite transformar la ecuación viscosa (4.6.16) en la ecuación del calor

$$v_t - \varepsilon v_{xx} = 0, \tag{4.6.17}$$

cuya solución puede calcularse explícitamente por convolución con el núcleo de Gauss.

De este modo la solución de entropía de (4.6.15) puede calcularse explícitamente. Se observa entonces que, en particular, en el caso del problema de Riemann, la solución de entropía es como sigue:

- Cuando $u_\ell > u_r$, obtenemos la solución discontinua con valores u_ℓ y u_r a izquierda y derecha del choque que se propaga con velocidad $s = \big(u_\ell + u_r\big)/2$.
- Cuando $u_\ell < u_r$, obtenemos la onda de rarefacción.

Vemos por tanto que, en este caso particular, el criterio de entropía obtenido por la viscosidad evanescente permite identificar de manera única la solución de entropía.

La solución de entropía en este caso puede también caracterizarse mediante la utilización de principio del máximo. Indiquemos brevemente cómo ésto puede hacerse.

El problema (4.6.16) es parabólico y por tanto sus soluciones son regulares. Derivando (4.6.16) con respecto a x obtenemos que $w = u_x$ satisface

$$w_t - \varepsilon w_{xx} + (uw)_x = 0,$$

que puede también reescribirse como

$$w_t - \varepsilon w_{xx} + w^2 + uw_x = 0. \tag{4.6.18}$$

La ecuación (4.6.18) satisface el principio del máximo y admite la solución explícita

$$w^* = 1/t. \tag{4.6.19}$$

Como su dato inicial es $w(0) = \infty$, toda solución de (4.6.18) está por debajo de ella. Obtenemos por tanto que

$$u_x = w \leq 1/t \tag{4.6.20}$$

para todo $\varepsilon > 0$.

Como la solución de entropía de (4.6.15) es el límite de soluciones de (4.6.16) cuando $\varepsilon \to 0$, la desigualdad (4.6.20) se preserva en el límite. La condición (4.6.20) caracteriza la solución de entropía.

Es fácil ver que (4.6.20) selecciona las soluciones de entropía obtenidas mediante el método de la viscosidad evanescente. En efecto, cuando $u_\ell > u_r$, hemos elegido la solución discontinua con la condición de propagación del choque $s = \big(u_\ell + u_r\big)/2$. Esta solución satisface (4.6.20) puesto que, en este caso,

$$u_x = (u_r - u_\ell)\delta_{x(t)}$$

siendo $x(t)$ el punto donde se ubica el choque.

Cuando $u_\ell < u_r$, hemos obtenido la onda de rarefacción. En este caso (4.6.20) se verifica con igualdad puesto que la derivada espacial de $v = x/t$ es precisamente $1/t$.

La condición (4.6.20) es conocida como la condición de Oleinick.

La condición de entropía puede aún describirse de otro modo alternativo. Se trata de la manera más conveniente para probar la unicidad de la misma, siguiendo el trabajo pionero de Kruzkov.

Consideremos la siguiente familia uniparamétrica de funciones de entropía

$$U(u) =\mid u - k \mid,\ k \in \mathbb{R}. \tag{4.6.21}$$

Multiplicando por (4.6.14) por $\operatorname{sgn}(u-k)$ obtenemos

$$u_t \operatorname{sgn}(u-k) - \varepsilon \Delta u \operatorname{sgn}(u-k) + \operatorname{div}\big(f(u)\big) \operatorname{sgn}(u-k) = 0. \tag{4.6.22}$$

La función $-\varepsilon \Delta u \operatorname{sgn}(u-k)$ es no-negativa en el sentido de las distribuciones puesto que $\operatorname{sgn}(u-k)$ es monótona creciente. Por otra parte, el término

$$\operatorname{div}\big(f(u)\big) \operatorname{sgn}(u-k)$$

se puede describir como

$$\operatorname{div}\big(F(u)\big)$$

donde $F(u)$ es la función del flujo de entropía

$$F(u) = \big(f(u) - f(k)\big) \operatorname{sgn}(u-k). \tag{4.6.23}$$

De este modo vemos que

$$\frac{\partial U(u)}{\partial t} + \operatorname{div}\big(F(u)\big) \leq 0, \tag{4.6.24}$$

para toda solución del problema viscoso (4.6.14), para todo $\varepsilon > 0$, y para todo par de entropía-flujo de entropía $(U,\, F)$ con $k \in \mathbb{R}$ arbitrario.

Nuevamente, es natural imponer a la solución de entropía de la ecuación hiperbólica (4.6.13) que la condición (4.6.24) se cumpla como resultado de paso a límite $\varepsilon \to 0$.

Al adoptar este punto de vista se plantean dos problemas:

- *Existencia:* Para probar la existencia de la solución de entropía que satisface (4.6.24) es preciso demostrar que la solución débil de (4.6.13) es límite cuando $\varepsilon \to 0$ de soluciones de (4.6.14) en un sentido suficientemente fuerte como para poder pasar al límite en el sentido de las distribuciones en los términos no-lineales $U(u)$ y $F(u)$ de (4.6.24).

- *Unicidad:* Una vez que la solución que satisface (4.6.24) ha sido construida es preciso probar su unicidad.

En lo sucesivo vamos a abordar estas dos cuestiones con el objeto de concluir un resultado de existencia y unicidad de soluciones de entropía para la ley de conservación.

Para ello consideramos en primer lugar el problema viscoso:

$$\begin{cases} u_t - \varepsilon \Delta u + \operatorname{div}\big(f(u)\big) = 0, & x \in \mathbb{R}^d, \quad t > 0 \\ u(x,\, 0) = u_0, & x \in \mathbb{R}^d. \end{cases} \tag{4.6.25}$$

Se tiene el siguiente resultado de existencia y unicidad de soluciones:

Theorem 4.6.1 *Supongamos que f es de clase C^m, $m \geq 1$ y que el dato inicial u_0 pertenece a $H^m(\mathbb{R}^d) \cap L^\infty(\mathbb{R}^d)$. Entonces, para cada $\varepsilon > 0$, (4.6.25) admite una única solución tal que*

$$u \in L^2\big(0,\, T;\, H^{m+1}(\mathbb{R}^d 0\big) \cap L^\infty\big(0,\, T;\, H^m(\mathbb{R}^d)\big) \tag{4.6.26}$$

para todo $T > 0$ y

$$\frac{\partial^k u}{\partial_t^k} \in \begin{cases} L^2\big(0,\, T;\, H^{m+1-2k}(\mathbb{R}^d)\big) \cap L^\infty\big(0,\, T;\, H^{m-2k}(\mathbb{R}^d)\big), & m > 2k \\ L^2\big(0,\, T;\, L^2(\mathbb{R}^d)\big), & m = 2k-1. \end{cases} \tag{4.6.27}$$

Demostración del Teorema 4.6.1. Para una demostración detallada sugerimos la sección II. 2 de [11]. En estas notas nos limitaremos a indicar las ideas principales de la prueba.

• **Paso 1.** Suponemos en primer lugar que f es globalmente Lipschitz y que el dato inicial u_0 pertenece a $L^2(\mathbb{R}^d)$. Es entonces fácil probar mediante un argumento de punto fijo que (4.6.25) admite una única solución en la clase

$$u \in W(0,\, T) \tag{4.6.28}$$

donde $W(0,\, T)$ es el espacio

$$W(0,\, T) = L^2\big(0,\, T;\, H^1(\mathbb{R}^d)\big) \cap H^1\big(0,\, T;\, H^{-1}(\mathbb{R}^d)\big). \tag{4.6.29}$$

Este espacio está incluido con continuidad en $BC\big([0,\, T];\, L^2(\mathbb{R}^d)\big)$.

El argumento de punto fijo puede aplicarse tanto en la ecuación integral asociada a (4.6.25) como en su formulación variacional.

De este modo se obtiene la existencia local de soluciones. Es decir, la existencia de un tiempo $T > 0$ (que puede depender de $\varepsilon > 0$) para el que existe una única solución en la clase (4.6.28).

La solución puede ser prolongada a una solución global definida para todo $t \geq 0$. Para ello es preciso obtener una estimación a priori que excluya la posibilidad de explosión en tiempo finito. Esto se obtiene con facilidad en este caso. En efecto, multiplicando en (4.6.25) por u e integrando en $\mathbb{R}^d$ obtenemos

$$\frac{1}{2}\frac{d}{dt}\int_{\mathbb{R}^d} u^2 dx + \varepsilon \int_{\mathbb{R}^d} |\nabla u|^2 \, dx = -\int_{\mathbb{R}^d} \operatorname{div}\big(f(u)\big)u \, dx.$$

Ahora bien,

$$\int_{\mathbb{R}^d} \operatorname{div}\big(f(u)\big)u \, dx = \int_{\mathbb{R}^d} f'(u)\cdot\nabla u \, u \, dx = \int_{\mathbb{R}^d} \operatorname{div}\big(\mathcal{G}(u)\big)dx = 0$$

donde

$$\mathcal{G}(z) = \int_0^z f'(s)s \, ds.$$

Deducimos por tanto que

$$\frac{1}{2}\frac{d}{dt}\int_{\mathbb{R}^d} u^2 dx + \varepsilon \int_{\mathbb{R}^d} |\nabla u|^2 \, dx = 0.$$

Integrando esta ecuación en tiempo obtenemos

$$\| u(t) \|^2_{L^2(\mathbb{R}^d)} + 2\varepsilon \int_0^t \| \nabla u(s) \|^2_{L^2(\mathbb{R}^d)} \, ds = \| u_0 \|^2_{L^2(\mathbb{R}^d)} . \tag{4.6.30}$$

De esta identidad se deduce que no se puede producir explosión en tiempo finito. La solución se prolonga por tanto a una solución global $u \in BC\big([0,\infty);\, L^2(\mathbb{R}^d)\big)$ tal que $\nabla u \in L^2\big(\mathbb{R}^d \times (0,\infty)\big)$.

Obtenemos así el resultado de regularidad (4.6.26) con $m = 0$.

• **Paso 2.** Supongamos ahora que $u_0 \in L^\infty(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$. Vamos a probar que la solución definida en el paso anterior verifica la cota

$$\| u \|_\infty \leq \| u_0 \|_\infty . \tag{4.6.31}$$

Esto permite extender el resultado del paso anterior a no-linealidades $f \in C^1$, sin necesidad de suponer que sean globalmente Lipschitz.

Con el objeto de probar (4.6.31) basta con demostrar que $u \leq k$ siendo $k = \max(u_0)$. Del mismo modo se prueba que $u \geq -\| u_0 \|_\infty$. Para ello multiplicamos la ecuación por $\operatorname{sgn}(u-k)^+$ e integramos en $\mathbb{R}^d$. Obtenemos que

$$\frac{d}{dt}\int_{\mathbb{R}^d} (u-k)^+ dx \leq 0, \tag{4.6.32}$$

puesto que:

$$\int_{\mathbb{R}^d} \operatorname{div}\big(f(u)\big)\,\operatorname{sgn}(u-k)^+ dx = 0 \tag{4.6.33}$$

$$-\int_{\mathbb{R}^d} \Delta u\,\operatorname{sgn}(u-k)^+ dx \le 0. \tag{4.6.34}$$

El hecho de que (4.6.33) se cumpla es consecuencia, nuevamente de que

$$\operatorname{div}\big(f(u)\big)\,\operatorname{sgn}(u-k)^+ = \operatorname{div}\big(\mathcal{G}_k(u)\big)$$

para una función $\mathcal{G}_k$ adecuada.

El que (4.6.34) se satisfaga es producto del hecho que $\operatorname{sgn}(u-k)^+$ es una función monótona creciente. En realidad la prueba de (4.6.34) pasa por utilizar como función test funciones de la forma $\varphi_\varepsilon(u-k)$, siendo φ_ε una regularización de la función sgn^+.

• **Paso 3.** En el marco de las hipótesis generales del Teorema basta probar la regularidad añadida de la solución (4.6.26)-(4.6.27) con $m \ge 1$ arbitrario. Esto puede hacerse tomando derivadas sucesivas de la ecuación.

En primer lugar observamos que la ecuación que u satisface puede escribirse en la forma

$$u_t - \varepsilon\Delta u = -f'(u)\cdot\nabla u.$$

Como $u \in L^2\big(0,\,T;\,H^1(\mathbb{R}^d)\big) \cap L^\infty\big(\mathbb{R}^d\times(0,\,T)\big)$, deducimos que

$$f'(u)\cdot\nabla u \in L^2\big(\mathbb{R}^d\times(0,\,T)\big).$$

Suponiendo que $u_0 \in H^1(\mathbb{R}^d)$, resultados clásicos de regularidad para la ecuación del calor lineal garantizan que

$$\begin{cases} u \in BC\big([0,\,T];\,H^1(\mathbb{R}^d)\big) \cap L^2\big(0,\,T;\,H^2(\mathbb{R}^d)\big); \\ u_t \in L^2\big(\mathbb{R}^d\times(0,\,T)\big). \end{cases}$$

Esto proporciona el resultado del Teorema 4.6.1 cuando $m = 1$.

Con el objeto de obtener regularidad adicional consideramos las sucesivas derivadas de u. Tomando la deriva con respecto a t de la ecuación que u satisface deducimos que $u = v_t$ verifica la ecuación

$$v_t - \varepsilon\Delta v = -\operatorname{div}\big(f'(u)u_t\big).$$

De los resultados anteriores deducimos que el segundo miembro pertenece a $L^2\big(0,\,T;\,H^{-1}(\mathbb{R}^d)\big)$ de donde deducimos que $v \in BC\big([0,\,T];\,L^2(\mathbb{R}^d)\big)\cap L^2\big(0,\,T;\,H^1(\mathbb{R}^d)\big)$ siempre y cuando

$$v(0) = u_t(0) = \varepsilon\Delta u(0) - \operatorname{div}\big(f\big(u(0)\big)\big) = \varepsilon\Delta u_0 - \operatorname{div}\big(f(u_0)\big) \;\in L^2(\mathbb{R}^d),$$

cosa que está plenamente garantizada bajo las hipótesis del Teorema con $m = 2$.

Iterando este proceso el resultado puede demostrarse para $m \geq 1$ arbitrario.■

Para probar la existencia de una solución de entropía hemos de ser capaces de pasar al límite cuando $\varepsilon \to 0$ en las soluciones obtenidas en el Teorema 4.6.1. La dificultad mayor reside en la obtención de la compacidad suficiente para pasar al límite en el término no-lineal. En particular, la estimación de energía (4.6.30) es insuficiente a tal fin.

Tenemos el siguiente resultado, que proporciona estimaciones uniformes sobre ∇u en $L^1(\mathbb{R}^d)$, uniformemente en $t \geq 0$ y en $\varepsilon > 0$.

Lemma 4.6.1 *Supongamos que* $u_0 \in L^1(\mathbb{R}^d) \cap L^\infty(\mathbb{R}^d) \cap BV(\mathbb{R}^d)$.

Entonces la solución u_ε *de* (4.6.25) *satisface*

$$\| u_\varepsilon \|_{L^1(\mathbb{R}^d)} \leq \| u_0 \|_{L^1(\mathbb{R}^d)}, \tag{4.6.35}$$

$$\int_{\mathbb{R}^d} u_\varepsilon(x, t)dx = \int_{\mathbb{R}^d} u_0(x)dx, \tag{4.6.36}$$

$$\| u_\varepsilon - v_\varepsilon \|_{L^1(\mathbb{R}^d)} \leq \| u_0 - v_0 \|_{L^1(\mathbb{R}^d)} \tag{4.6.37}$$

$$\| \nabla u_\varepsilon \|_{L^1(\mathbb{R}^d)} \leq TV(u_0), \tag{4.6.38}$$

para todo $t > 0$ *y* $\varepsilon > 0$, u_0, $v_0 \in L^1(\mathbb{R}^d)$, *siendo* u_ε *y* v_ε *las soluciones correspondientes de* (4.6.25).

Demostración. Para obtener (4.6.36) basta integrar la ecuación que u satisface en $\mathbb{R}^d$ y usar que

$$\int_{\mathbb{R}^d} \Delta u dx = \int_{\mathbb{R}^d} \operatorname{div}\big(f(u)\big)dx = 0.$$

Para obtener (4.6.35) multiplicamos la ecuación por $\operatorname{sgn}(u)$. Utilizando que

$$-\int_{\mathbb{R}^d} \Delta u \operatorname{sgn}(u)dx \geq 0$$

y que

$$\int_{\mathbb{R}^d} \operatorname{div}\big(f(u)\big) \operatorname{sgn}(u)dx = 0,$$

deducimos que

$$\frac{d}{dt}\int_{\mathbb{R}^d} | u | dx \leq 0$$

de donde se deduce a su vez (4.6.35).

Para obtener (4.6.37) consideramos las soluciones u y v de (4.6.25) con datos iniciales u_0 y v_0. Entonces $w = u - v$ satisface

$$w_t - \varepsilon\Delta w = -\operatorname{div}\big(f(u) - f(v)\big).$$

Multiplicando en esta ecuación por $\operatorname{sgn}(u-v)=\operatorname{sgn}(w)$ obtenemos que

$$\frac{d}{dt}\int_{\mathbb{R}^d} \mid u-v \mid dx \leq 0.$$

Esto es así puesto que $-\displaystyle\int_{\mathbb{R}^d} \Delta w \operatorname{sgn}(w) dx \geq 0$ y que

$$\int_{\mathbb{R}^d} \operatorname{div}\big(f(u)-f(v)\big)\operatorname{sgn}(u-v)dx = 0.$$

La propiedad de contracción en $L^1(\mathbb{R}^d)$ (4.6.37) permite deducir la estimación (4.6.38) en $BV(\mathbb{R}^d)$. En efecto, sean u y u_h las soluciones de (4.6.25) con datos u_0 y $u_{0,\,h}$ respectivamente, siendo $u_{0,\,h}$ una traslación de paso h de u_0, i.e.

$$u_{0,\,h}(x) = u_0\big(x + h\, e_\alpha\big)$$

siendo e_α, $\alpha = 1, \cdots, d$, uno de los vectores unitarios de la base canónica de $\mathbb{R}^d$.

Por invarianza (4.6.25) con respecto a traslaciones y por la unicidad de la solución, deducimos que

$$u_h(x,\,t) = u\big(x + h\, e_{\alpha,\,t}\big).$$

Por la propiedad de contracción en $L^1(\mathbb{R}^d)$ se concluye que

$$\frac{\parallel u - u_h \parallel_{L^1(\mathbb{R}^d)}}{h} \leq \frac{\parallel u_0 - u_{0,\,h} \parallel_{L^1(\mathbb{R}^d)}}{h}.$$

Pasando al límite cuando $h \to 0$, deducimos (4.6.38).

Con el objeto de probar la compacidad de las soluciones u_ε conviene probar una estimación sobre las derivadas temporales $\partial_t u_\varepsilon$:

Lemma 4.6.2 *Supongamos que los datos iniciales de* (4.6.25) *son tales que, además de estar uniformemente acotados en* $L^1(\mathbb{R}^d) \cap L^\infty(\mathbb{R}^d) \cap w^{1,\,1}(\mathbb{R}^d)$, *satisfacen la cota uniforme*

$$\varepsilon \parallel \Delta u_{0,\,\varepsilon} \parallel_{L^1(\mathbb{R}^d)} \leq C. \tag{4.6.39}$$

Entonces

$$\parallel \partial_t u_\varepsilon \parallel_{L^1(\mathbb{R}^d)} \leq C\, TV(u_0), \tag{4.6.40}$$

para todo $t > 0$ *y* $\varepsilon > 0$.

Demostración. Consideramos en este caso traslaciones temporales $u_{\varepsilon,\,\tau}$ de las soluciones u_ε de (4.6.25). Procediendo como en la prueba del Lema anterior deducimos que

$$\| u_\varepsilon(t) - u_{\varepsilon,\,\tau}(t) \|_{L^1(\mathbb{R}^d)} \leq \| u_{\varepsilon,\,0} - u_{\varepsilon,\,\tau}(0) \|_{L^1(\mathbb{R}^d)} .$$

Pasando al límite cuando $\tau \to 0^+$ deducimos que

$$\begin{array}{rcl} \| \partial_t u_\varepsilon \|_{L^1(\mathbb{R}^d)} & \leq & \| \partial_t u_\varepsilon(0) \|_{L^1(\mathbb{R}^d)} \\ & \leq & \varepsilon \| \Delta u_{0,\,\varepsilon} \|_{L^1(\mathbb{R}^d)} + \| \operatorname{div}\big(f(u_{\varepsilon,\,0})\big) \|_{L^1(\mathbb{R}^d)} \end{array} \tag{4.6.41}$$

la hipótesis (4.6.39) permite acotar el primer término del miembro de la derecha de (4.6.41). El segundo miembro puede acotarse del modo siguiente:

$$\| \operatorname{div}\big(f(u_{\varepsilon,\,0})\big) \|_{L^1(\mathbb{R}^d)} \leq M_\varepsilon \| \nabla u_{\varepsilon,\,0} \|_{L^1(\mathbb{R}^d)}$$

siendo

$$M_\varepsilon = \max_{|s| \leq \|u_{0,\,\varepsilon}\|_\infty} | f'(s) | .$$

De este modo se concluye la prueba de (4.6.40).■

Estamos ahora en condiciones de probar la existencia de una solución de entropía para (4.6.13).

Theorem 4.6.2 *Supongamos que el dato inicial u_0 pertenece al espacio $L^1(\mathbb{R}^d) \cap L^\infty(\mathbb{R}^d) \cap BV(\mathbb{R}^d)$ y que la no-linealidad f es de clase C^1. Entonces el problema*

$$\left\{ \begin{array}{lll} u_t + \operatorname{div}\big(f(u)\big) = 0 & \text{en} & \mathbb{R}^d, \quad t > 0 \\ u(x,\,0) = u_0(x) & \text{en} & \mathbb{R}^d \end{array} \right. \tag{4.6.42}$$

tiene una solución de entropía que verifica (4.6.24).

Además, la solución pertenece a la clase

$$u \in L^\infty\big(\mathbb{R}^d \times (0,\,\infty)\big) \cap BC\big([0,\,\infty);\, L^1(\mathbb{R}^d)\big) \tag{4.6.43}$$

y satisface las estimaciones

$$\| u \|_\infty \leq \| u_0 \|_\infty, \tag{4.6.44}$$

$$TV\big(u(t)\big) \leq TV(u_0),\ \forall t \geq 0, \tag{4.6.45}$$

$$\| u(t_2) - u(t_1) \| \leq C\, TV(u_0)\, | t_2 - t_1 |,\ \forall t_1,\, t_2 \geq 0. \tag{4.6.46}$$

Demostración. Utilizamos el método de aproximación y paso al límite basado en las regularizaciones viscosas (4.6.25).

En primer lugar hemos de construir los datos iniciales para las ecuaciones regularizadas (4.6.25). Esto puede hacerse convolucionando el dato inicial u_0 con una aproximación de la identidad. De este modo no sólo se preservan, uniformemente en $\varepsilon > 0$, las estimaciones que las hipótesis sobre u_0 proporcionan, sino que podemos además garantizar que se satisface (4.6.39).

Aplicando los resultados anteriores obtenemos una familia de soluciones u_ε de (4.6.25) que satisfacen las siguientes cotas, uniformemente en $\varepsilon > 0$:

$$\| u_\varepsilon \|_\infty \leq \| u_0 \|_\infty, \tag{4.6.47}$$

$$\| u_\varepsilon \|_{L^\infty\left(0,\,T;\,L^1,\,L^2(\mathbb{R}^d)\right)} \leq C, \tag{4.6.48}$$

$$\| \nabla u_\varepsilon \|_{L^\infty\left(0,\,T;\,L^1(\mathbb{R}^d)\right)} \leq C, \tag{4.6.49}$$

$$\| \partial_t u_\varepsilon \|_{L^\infty\left(0,\,T;\,L^1(\mathbb{R}^d)\right)} \leq C. \tag{4.6.50}$$

Dado un dominio acotado Ω de $\mathbb{R}^d$, como la inclusión $W^{1,\,1}(\Omega) \hookrightarrow L^1(\Omega)$ es compacta, deducimos que para todo $0 \leq t \leq T$, $u_\varepsilon(t)$ permanece en subconjunto compacto de $L^1(\Omega)$. Por otra parte, de la cota (4.6.50) se deduce que u_ε es uniformemente equicontinua en $[0,\,T]$ a valores en $L^1(\mathbb{R}^d)$.

Por el Teorema de Ascoli-Arzela deducimos por tanto la existencia de una subsucesión (que seguimos denotando por el subíndice ε) tal que

$$u_\varepsilon \to u \text{ en } C\big([0,\,T];\,L^1(\Omega)\big).$$

Utilizando una familia de conjuntos compactos que cubran todo $\mathbb{R}^d$ y un procedimiento de extracción diagonal deducimos que

$$u_\varepsilon \to u \text{ en } C\big([0,\,T];\,L^1_{loc}(\mathbb{R}^d)\big).$$

De hecho, el límite $u \in C\big([0,\,T];\,L^1(\mathbb{R}^d)\big)$. En efecto, de la cota (4.6.48) se deduce que

$$\| u(t) \|_{L^1(\mathbb{R}^d)} \leq C,\ \forall 0 \leq t \leq T.$$

Además de (4.6.50) se deduce que

$$\| u(t_2) - u(t_1) \|_{L^1(\mathbb{R}^d)} \leq C \mid t_2 - t_1 \mid,$$

de donde se concluye la continuidad en tiempo a valores en $L^1(\mathbb{R}^d)$.

Por otra parte, de (4.6.47) se deduce que

$$\| u \|_\infty \leq \| u_0 \|_\infty .$$

De estas estimaciones se concluye que $u \in C\big([0,\,T];\,L^p(\mathbb{R}^d)\big)$ para todo p finito
$1 \le p \le \infty$.

La estimación (4.6.45) se obtiene también como consecuencia inmediata del paso al límite.

De las convergencias anteriores es fácil comprobar que el límite u construido de este modo es una solución débil de entropía de (4.6.25). La única dificultad a la hora de comprobar este hecho es el paso al límite en los términos no-lineales pero ésto es posible como combinación de la cota L^∞ uniforme (que hace irrelevante el crecimiento de la no-linealidad en el infinito) y de la convergencia fuerte en $C\big([0,\,T];\,L^p(\mathbb{R}^d)\big)$.

En el Teorema anterior hemos probado la existencia de una solución de entropía. Queda probar su unicidad. Se trata de un resultado de una importancia histórica en este campo debido a Kruzhov.

Recordemos que una solución débil de (4.6.25) se dice solución de entropía si verifica que

$$\int_0^\infty \int_{\mathbb{R}^d} \big[\, |\, u-k\,|_t\, \partial\varphi + \operatorname{sgn}(u-k)\big(f(u)-f(k)\big)\cdot\nabla u\big]\,dxdt \ge 0$$

para toda función test no-negativa $\varphi \in C_0^\infty\big(\mathbb{R}^d \times (0,\,T)\big)$, $\varphi \ge 0$ y todo $k \in \mathbb{R}$.

Tenemos el siguiente resultado:

Theorem 4.6.3 *Sean u y v dos soluciones de entropía de* (4.6.25) *asociados a datos iniciales $u_0,\, v_0 \in L^\infty(\mathbb{R}^d)$, tales que $u,\, v \in L^\infty\big(\mathbb{R}^d \times (0,\,\infty)\big) \cap BC\big([0,\,T];\,L^1_{loc}(\mathbb{R}^d)\big)$, para todo $T > 0$.*

Entonces, siendo

$$M = \text{máx}\big\{\, |\, f'(\xi)\,| : |\,\xi\,| \le \text{máx}(\|\,u\,\|_\infty,\, \|\,v\,\|_\infty)\big\},$$

tenemos que

$$\int_{|x|\le R} |\,u(x,\,t) - v(x,\,t)\,|\,dx \le \int_{|x|\le R+Mt} |\,u_0(x) - v_0(x)\,|\,dx,\ \forall R > 0,\ \text{p.c.t.}\, t \ge 0.$$

Omitiremos la prueba de este resultado. El lector interesado podrá encontrar una presentación de la misma en el libro [11]

4.7. Esquemas numéricos de aproximación de leyes de conservación escalares

En las secciones anteriores hemos desarrollado una teoría que permite concluir la existencia de unicidad de soluciones de entropía para leyes de conser-

vación escalares en una y en varias dimensiones espaciales. El método de construcción empleado ha sido el de la viscosidad evanescente, que es un método de aproximación. El objetivo de esta sección es obtener resultados de aproximación mediante esquemas numéricos. La tarea es en este caso más compleja que en el marco de las ecuaciones en derivadas parciales lineales. En efecto, tal y como hemos visto, las soluciones de entropía de las leyes de conservación escalares en algunos casos desarrollan ondas de choque y en otros ondas de rarefacción. Es por tanto indispensable que los métodos que desarrollemos sean capaces de capturar y reproducir estas soluciones discerniendo los choques admisibles de los que no lo son.

Comenzamos con el caso uni-dimensional

$$u_t + \partial_x\Big(f(u)\Big) = 0,\ x \in \mathbb{R},\ t > 0, \tag{4.7.1}$$

$$u(x,\,0) = u_0(x),\ x \in \mathbb{R}. \tag{4.7.2}$$

Supondremos que $f \in C^2$ y usaremos la notación

$$a(s) = f'(s). \tag{4.7.3}$$

Consideramos ahora un paso espacial y temporal Δx y Δt e introducimos el ratio

$$\lambda = \Delta t/\Delta x. \tag{4.7.4}$$

Con el objeto de aproximar las soluciones de (4.7.1)-(4.7.2) utilizaremos esquemas de la forma

$$u_j^{n+1} = H\Big(u_{j-k}^n, \ldots, u_{j+k}^n\Big),\ \forall n \geq 0,\ j \in \mathbb{Z}. \tag{4.7.5}$$

La función discreta u_j^n representa una aproximación de la solución continua u en el punto $(x_j = j\Delta x,\ t_k = k\Delta t)$ y $H : \mathbb{R}^{2k+1} \to \mathbb{R}$ es una función continua.

Con el objeto de simplificar la notación denotaremos mediante H_Δ para la función que envía la sucesión $(v_j)_{j\in\mathbb{Z}}$ en la sucesión imagen $H_\Delta(v) = \Big(\big(H_\Delta(v)\big)_j\Big)_{j\in\mathbb{Z}}$ tal que

$$\big(H_\Delta(v)\big)_j = H\big(v_{j-k}, \ldots,\, v_{j+k}\big).$$

De este modo el esquema numérico (4.7.5) puede reescribirse como

$$v^{n+1} = H_\Delta(v^n). \tag{4.7.6}$$

Diremos que este esquema puede ponerse en *forma conservativa* si existe una función continua $g : \mathbb{R}^{2k} \to \mathbb{R}$ tal que

$$H\big(v_{-k}, \ldots,\, v_k\big) = v_0 - \lambda\Big[g\big(v_{-k+1}, \ldots,\, v_k\big) - g\big(v_{-k}, \ldots,\, v_{k-1}\big)\Big]. \tag{4.7.7}$$

La función g se denomina *flujo numérico*.

El esquema numérico tiene entonces la forma

$$u_j^{n+1} = u_j^n - \lambda\Big[g\big(u_{j-k+1}^n, \dots, u_{j+k}^n\big) - g\big(u_{j-k}^n, \dots, u_{j+k-1}^n\big)\Big]. \tag{4.7.8}$$

Denotando

$$g_{j+1/2}^n = g\big(u_{j-k+1}^n, \dots, u_{j+k}^n\big) \tag{4.7.9}$$

el esquema puede escribirse como

$$u_j^{n+1} = u_j^n - \lambda\big(g_{j+1/2}^n - g_{j-1/2}^n\big). \tag{4.7.10}$$

Entonces (4.7.5) puede ponerse en forma conservativa sí y sólo sí

$$\sum_{j\in\mathbb{Z}} H\big(v_{j-k}, \dots, v_{j+k}\big) = \sum_{j\in\mathbb{Z}} v_j. \tag{4.7.11}$$

Veamos cómo podemos calcular el flujo g para un esquema conservativo. Observamos que

$$\mathcal{G}\big(v_{-k}, \dots, v_k\big) = -\frac{1}{\lambda}\Big(H(v_{-k}, \dots, v_k) - v_0\Big)$$

satisface

$$\sum_{j\in\mathbb{Z}} \mathcal{G}\big(v_{j-k}, \dots, v_{j+k}\big) = 0. \tag{4.7.12}$$

Se puede entonces comprobar la existencia de una función continua g tal que

$$\mathcal{G}\big(v_{-k}, \dots, v_k\big) = g\big(v_{-k+1}, \dots, v_k\big) - g\big(v_{-k}, \dots, v_{k-1}\big). \tag{4.7.13}$$

Todo esquema conservativo, además de preservar la integral discreta (4.7.11), envía $L^1(\mathbb{Z})$ en sí mismo.

Por otra parte, con el objeto de garantizar la consistencia del esquema (4.7.10) con la ecuación (4.7.1) necesitamos que $\Big(g\big(u_{k+1}, \dots, u_k\big) - g\big(u_{-k}, \dots, u_{k-1}\big)\Big)/\Delta x$ sea una aproximación de $\partial_x\big(f(u)\big)$. Diremos por tanto que el esquema (4.7.5), (4.7.7) es *consistente* si

$$g(v, \dots, v) = f(v). \tag{4.7.14}$$

Con el objeto de analizar la convergencia del método introducimos la función constante a trozos

$$u_\Delta(x, t) = u_j^n,\ x_{j-1/2} < x < x_{j+1/2},\ t_n \le t \le t_{n+1}, \tag{4.7.15}$$

donde $x_{j+1/2} = \big(x_j + x_{j+1}\big)/2$. Analizamos entonces la convergencia de la función constante a trozos u_Δ hacia u. Para ello debemos introducir una aproximación del dato inicial. Son varias las posibilidades. Entre ellas

$$u_j^0 = \frac{1}{\Delta x}\int_{x_j-1/2}^{x_j+1/2} u_0(x)dx. \tag{4.7.16}$$

El primer resultado en esta dirección, debido a Lax-Wendroff, asegura que si la sucesión u_Δ está acotada en $L^\infty\big(\mathbb{R}\times(0,\,\infty)\big)$ y converge en $L^1_{loc}\big(\mathbb{R}\times(0,\,\infty)\big)$ y en casi todo punto a una función u, ésta es una solución débil de (4.7.1)-(4.7.2).

La prueba de este resultado consiste en pasar al límite en las formulaciones débiles de ambos problemas, discretos y continuos. Obviamente se presenta la dificultad adicional del paso de una formulación discreta a la continua. El modo de proceder consiste en tomar una función test para el problema continuo y, de ella, evaluándola en el mallado discreto, obtener una función test ara el problema discreto. De manera más precisa, dada $\varphi\in C^1_0\big(\mathbb{R}\times(0,\,\infty)\big)$ consideramos la función test discreta

$$\varphi^n_j=\varphi(x_j,\,t_n),\ j\in\mathbb{Z},\ n\geq 0. \tag{4.7.17}$$

Utilizando esta función test en (4.7.10) deducimos que

$$\Delta x\sum_{j,\,n}\big(u^{n+1}_j-u^n_j\big)\varphi^n_j+\Delta t\sum_{j,\,n}\big(g^n_{j+1/2}-g^n_{j-1/2}\big)\varphi^n_j=0. \tag{4.7.18}$$

Sumando por partes deducimos que

$$\Delta x\sum_n\sum_j u^{n+1}_j\big(\varphi^{n+1}_j-\varphi^n_j\big)+\Delta t\sum_n\sum_j g^n_{j+1/2}\big(\varphi^n_{j+1}-\varphi^n_j\big)+\Delta x\sum_j y^0_j\varphi^0_j=0. \tag{4.7.19}$$

Con el objeto de comparar los términos no-lineales del esquema discreto y de la ecuación continua introducimos la siguiente extensión constante a trozos de g que denotamos como g_Δ:

$$g_\Delta(x,\,t)=g^n_{j+1/2}=g\big(v^n_{j-k+1},\ldots,\,v^n_{j+k}\big),\ x_j<x<x_{j+1},\ t_n\leq t\leq t_{n+1}. \tag{4.7.20}$$

Esto nos permite escribir la expresión discreta anterior en forma integral

$$\begin{aligned}&\int_{\mathbb{R}\times(0,\,\infty)}u_\Delta(x,\,t)\Big(\varphi_\Delta(x,\,t)-\varphi_\Delta(x,\,t-\Delta t)\Big)/\Delta t\,dxdt \qquad (4.7.21)\\ +\ &\int_{\mathbb{R}\times(0,\,\infty)}g_\Delta(x,\,t)\Big(\varphi_\Delta(x+\Delta x/2,\,t)-\varphi_\Delta(x-\Delta x/2,\,t)\Big)/\Delta x\,dxdt\\ +\ &\int_{\mathbb{R}}u_0(x)\varphi_\Delta(x,\,0)dx=0.\end{aligned}$$

Aplicando la convergencia uniforme de φ_Δ a φ, es fácil comprobar que las integrales correspondientes a los datos iniciales convergen, i.e.

$$\int_{\mathbb{R}}u_0(x)\varphi_\Delta(x,\,0)dx\longrightarrow\int_{\mathbb{R}}u_0(x)\varphi(x,\,0)dx, \tag{4.7.22}$$

cuando $\Delta x,\, \Delta t \to 0$.

De forma análoga se puede comprobar que

$$\int_{\mathbb{R}\times(0,\,\infty)} u_\Delta(x,\,t)\left(\frac{\varphi_\Delta(x,\,t)-\varphi_\Delta(x,\,t-\Delta t)}{\Delta t}-\partial_t\varphi(x,\,t)\right)dxdt\to 0. \quad (4.7.23)$$

Además, de la convergencia de u_Δ a u en L^1_{loc} deducimos que

$$\int_{\mathbb{R}\times(0,\,\infty)} u_\Delta(x,\,t)\partial_t\varphi(x,\,t)\,dxdt \longrightarrow \int_{\mathbb{R}\times(0,\,\infty)} u(x,\,t)\partial_t\varphi(x,\,t)\,dxdt. \quad (4.7.24)$$

Se concluye por tanto que

$$\int_{\mathbb{R}\times(0,\,\infty)} u_\Delta(x,\,t)\left(\frac{\varphi_\Delta(x,\,t)-\varphi_\Delta(x,\,t-\Delta t)}{\Delta t}\right)dxdt \longrightarrow \int_{\mathbb{R}\times(0,\,\infty)} u(x,\,t)\partial_t\varphi(x,\,t)\,dxdt. \quad (4.7.25)$$

Por último hemos de considerar el término no-lineal. El mismo argumento anterior reduce el problema al estudio del límite de

$$\int_{\mathbb{R}\times(0,\,\infty)} g_\Delta(x,\,t)\partial_t\varphi(x,\,t)dxdt.$$

Observamos que

$$g_\Delta(x,\,t)=g\left(u_\Delta(x-(k+1/2)\Delta x,\,t),\ldots,\,u_\Delta\big(x+(k-1/2)\Delta x,\,t\big)\right), \quad (4.7.26)$$

y, consecuentemente, introducimos la notación

$$w^j_\Delta(x,\,t)=u_\Delta\big(x+\big(j-1/2\big)\Delta x,\,t\Big),\ -k\le j\le k. \quad (4.7.27)$$

Si $\mathcal{K}$ es un subconjunto compacto de $\mathbb{R}\times(0,\,\infty)$ y $\mathcal{K}_1=\mathcal{K}+\big(j-\frac{1}{2}\big)\Delta x$, entonces

$$\begin{aligned}\int_{\mathcal{K}}\big|w^j_\Delta(x,\,t)-u(x,\,t)\big|dxdt \;&\le\; \int_{\mathcal{K}_1}\big|u_\Delta(x,\,t)-u(x,\,t)\big|dxdt \quad (4.7.28)\\ &+\; \int_{\mathcal{K}}\big|u\big(x+(j-1/2)\Delta x,\,t\big)-u(x,\,t)\big|dxdt.\end{aligned}$$

Se deduce entonces que w^j_Δ converge a u en $L^1(k)$ cuando $\Delta\to 0$, para cualquier $|\ j\ |\le k$, gracias a la convergencia de u_Δ. Podemos entonces extraer subsucesiones de modo que w^j_Δ converja para casi todo $(x,\,t)\in\mathbb{R}\times(0,\,\infty)$ y todo $j:|\ j\ |\le k$. Utilizando la continuidad de g deducimos que

$$\begin{aligned}g_\Delta(x,\,t)=g\big(w^{-k+1}_\Delta(x,\,t),\ldots,w^k_\Delta(x,\,t)\big)\to g\big(u(x,\,t),\ldots,\,(x,\,t)\big) \quad (4.7.29)\\ p.c.t.\,(x,\,t)\in\mathbb{R}\times(0,\,\infty).\end{aligned}$$

De la condición de consistencia sobre g se deduce que $g(u,\, u,\, \ldots,\, u) = f(u)$. Por tanto, aplicando el Teorema de convergencia dominada de Lebesgue, deducimos que

$$\int_{\mathbb{R}\times(0,\,\infty)} g_\Delta(x,\, t)\partial_t\varphi(x,\, t)\, dxdt \longrightarrow \int_{\mathbb{R}\times(0,\,\infty)} f(u)\partial_t\varphi(x,\, t)\, dxdt. \tag{4.7.30}$$

De ésto se desprende que el límite u es una solución débil de (4.7.1)-(4.7.2) en el sentido que

$$\int_{\mathbb{R}\times(0,\,\infty)} \Big(u(x,\, t)\partial_t\varphi(x,\, t) + f\big(u(x,\, t)\big)\partial_x\varphi(x,\, t)\Big)\, dxdt + \int_{\mathbb{R}} u_0(x)\varphi(x,\, 0)dx = 0. \tag{4.7.31}$$

Este resultado muestra que el límite de soluciones numéricas obtenidas por un esquema conservativo necesariamente es una solución débil del problema continuo, de donde se deduce en particular que se verifican las condiciones de Rankine-Hugoniat. Queda por comprobar que este límite es la solución de entropía además de probar la compacidad de la sucesión u_Δ de soluciones aproximadas.

Pero analicemos primero el *orden de consistencia o de precisión*. Diremos que el esquema es de orden $p \geq 1$ si para toda solución regular u de (4.7.1)-(4.7.2) y fijando $\lambda = \Delta x/\Delta t$, se tiene

$$u(x,\, t + \Delta t) - H\big(u(x - k\Delta x,\, t), \ldots,\, u(x + k\Delta x,\, t)\big) = O\big(\Delta t^{p+1}\big), \tag{4.7.32}$$

cuando $\Delta t \to 0$. El término que se estima a la izquierda de (4.7.32) es lo que se denomina el *error de truncación* y se estima típicamente realizando un desarrollo de Taylor de u y de H, usando la ecuación que u satisface y la estructura de H.

La siguiente Proposición resulta útil para estimar el error de truncación.

Proposition 4.7.1 *Consideremos el esquema en diferencias* (4.7.5) *que suponemos puede escribirse en la forma conservativa* (4.7.7) *y que es consistente con la ecuación* (4.7.1). *Supongamos que* $H \in C^3$. *Entonces, para toda solución* u *de* (4.7.1)-(4.7.2) *suficientemente regular y con* $\lambda = \Delta x/\Delta t$ *constante, el error de truncación admite la siguiente expresión*

$$\begin{aligned} &u(x,\, t + \Delta t) - H\big(u(x - k\Delta x,\, t), \ldots,\, u(x + k\Delta x,\, t)\big) \\ &= -\Delta t^2\partial_x\big(\beta(u,\, \lambda)\partial_x u(x,\, t)\big) + O(\Delta t^3) \end{aligned} \tag{4.7.33}$$

con

$$\beta(u,\, \lambda) = \left(\sum_{j=-k}^{k} j^2 \frac{\partial H}{\partial \nu}(u,\, u, \ldots,\, u)\right) \Big/ 2\lambda^2 - a(u)^2/2. \tag{4.7.34}$$

Demostración. En primer lugar observamos por (4.7.7) que

$$H(u,\, u, \ldots,\, u) = u. \tag{4.7.35}$$

Con el objeto de simplificar la notación introducimos la siguiente:

$$\bar{v} = (v_{-k}, \ldots,\, v_{k-1});\; T\bar{v} = (v_{-k+1}, \ldots,\, v_k). \tag{4.7.36}$$

Entonces

$$H(v_{-k}, \ldots,\, v_k) = v_0 - \lambda\big(g(T\bar{v}) - g(\bar{v})\big) \tag{4.7.37}$$

y

$$\frac{\partial H}{\partial v_j} = \delta_j^0 - \lambda\left(\frac{\partial g}{\partial v_{j-1}}(T\bar{v}) - \frac{\partial g}{\partial v_j}(\bar{v})\right),\; -k \le j \le k, \tag{4.7.38}$$

siempre y cuando usemos la convención

$$\frac{\partial g}{\partial v_{-k-1}} = \frac{\partial g}{\partial v_k} = 0. \tag{4.7.39}$$

Por tanto

$$\begin{aligned}\sum_{j=-k}^{k} j\frac{\partial H}{\partial v_j}(u, \ldots,\, u) &= -\lambda \sum_{j=-k}^{k} j\left(\frac{\partial g}{\partial v_{j-1}}(u, \ldots,\, u) - \frac{\partial g}{\partial v_j}(u, \ldots, u)\right) \\ &= -\lambda \sum_{j=-k}^{k} \frac{\partial g}{\partial v_j}(u, \ldots,\,).\end{aligned} \tag{4.7.40}$$

La condición de consistencia (4.7.14) asegura que

$$\sum_{j=-k}^{k} \frac{\partial g}{\partial v_j}(u, \ldots,\, u) = a(u). \tag{4.7.41}$$

Obtenemos así

$$\sum_{j=-k}^{k} j\frac{\partial H}{\partial v_j}(u, \ldots,\, u) = -\lambda a(u). \tag{4.7.42}$$

Diferenciando de nuevo en (4.7.38) deducimos que

$$\frac{\partial^2 H}{\partial v_i \partial v_j} = \lambda\left\{\frac{\partial^2 g}{\partial v_{j-i}\partial v_{j-i}}(T\bar{v}) - \frac{\partial^2 g}{\partial v_i \partial v_j}(\bar{v})\right\}$$

y

$$\begin{aligned}\sum_{i,\,j=-k}^{k} (i-j)^2 \frac{\partial^2 H}{\partial v_i \partial v_j}(u, \ldots,\, u) &= -\lambda \sum_{i,\,j=-k}^{+k} (i-j)^2 \\ \left\{\frac{\partial^2 g}{\partial v_{j_i}\partial v_{j-i}}(u, \ldots,\, u) - \frac{\partial^2 g}{\partial v_i \partial v_j}(u, \ldots,\, u)\right\} &= 0.\end{aligned} \tag{4.7.43}$$

Utilizamos ahora el desarrollo de Taylor de H en el punto $(u, \ldots, u)$:

$$\begin{aligned} H(u_{-k}, \ldots, u_k) &= u + \sum_{j=-k}^{k} \frac{\partial H}{\partial v_j}(u, \ldots, u)(u_j - u) \\ &\quad + \frac{1}{2} \sum_{i,\, j=-k}^{k} \frac{\partial^2 H}{\partial v_i \partial v_j}(u, \ldots, u)(u_j - u)(u_i - u) + O(\Delta x^3) \end{aligned}$$

en donde hemos usado la notación $u = u(x, t)$ y $u_j = u(x + j\Delta x, t)$. Por el desarrollo de Taylor de u tenemos que

$$u_j - u = j\Delta x \partial_x u + \frac{(j\Delta x)^2}{2} \partial_x^2 u + O(\Delta x^3)$$

y

$$(u_i - u)(u_j - u) = ij(\Delta x)^2 (\partial_x u)^2 + O(\Delta x^3).$$

Por tanto

$$\begin{aligned} H(u_{-k}, \ldots, u_k) &= u + \Delta x \partial_x u \sum_{j=-k}^{k} j \frac{\partial H}{\partial v_j}(u, \ldots, u) \\ &\quad + \frac{(\Delta x)^2 (\partial_x u)^2}{2} \sum_{i,\, j=-k}^{k} ij \frac{\partial^2 H}{\partial v_i \partial v_j}(u, \ldots, u) \\ &\quad + \frac{(\Delta x)^2}{2} \partial_x^2 u \sum_{i,\, j=-k}^{k} j^2 \frac{\partial H}{\partial v_j}(u, \ldots, u) + O(\Delta x^3). \end{aligned}$$

Por otra parte

$$\begin{aligned} &\sum_{j=-k}^{k} j^2 \frac{\partial H}{\partial v_j}(u, \ldots, u) \partial_x^2 u + \sum_{i,\, j=-k}^{k} ij \frac{\partial^2 H}{\partial v_i \partial v_j}(u, \ldots, u)(\partial_x u)^2 \qquad (4.7.44) \\ &= \frac{\partial}{\partial x} \left(\sum_{j=-k}^{k} j^2 \frac{\partial H}{\partial v_j}(u, \ldots, u) \partial_x u \right) + \sum_{i,\, j=-k}^{k} (ij - j^2) \frac{\partial^2 H}{\partial v_i \partial v_j}(u, \ldots, u)(\partial_x u)^2 \\ &= \partial_x \left(\sum_{j=-k}^{k} j^2 \frac{\partial H}{\partial v_j}(u, \ldots, u) \partial_x u \right). \end{aligned}$$

Esto es así puesto que

$$\sum_{i,\, j=-k}^{k} (ij - j^2) \frac{\partial H}{\partial v_i \partial v_j}(u, \ldots, u) = 0, \qquad (4.7.45)$$

lo cual es consecuencia de (4.7.43) y de la simetría de la matriz $\partial^2 H/\partial v_j \partial v_i$ puesto que

$$\sum_{i,\, k=-k}^{k} (ij - j^2) \frac{\partial^2 H}{\partial v_i \partial v_j}(u, \ldots, u) = -\frac{1}{2} \sum_{i,\, j=-k}^{k} (i-j)^2 \frac{\partial H}{\partial v_j \partial v_i}(u, \ldots, u).$$

Deducimos de este modo que

$$H(u_{-k}, \ldots, u_k) = u - \Delta t\, a(u)\partial_x u + \frac{(\Delta x)^2}{2} \partial_x \left(\sum_{j=-k}^{k} j^2 \frac{\partial H}{\partial v_j} \partial_x u \right) + O(\Delta x^3). \tag{4.7.46}$$

Por otra parte, el desarrollo de Taylor de $u(x,\, t + \Delta t)$ garantiza que

$$u(x,\, t + \Delta t) = u + \Delta t \partial_t u + \frac{(\Delta t)^2}{2} \partial_t^2 u + O(\Delta t^3).$$

Utilizando que

$$\partial_t u = -\partial_x \big(f(u) \big) = -a(u) \partial_x u$$

y

$$\partial_t^2 u = -\partial_u \Big(\partial_t \big(f(u) \big) = -\partial_x \big(a(u) \partial_t u \Big) = \partial_x \big(a^2(u) \partial_x u \big),$$

deducimos

$$u(x,\, t + \Delta t) = u - \Delta t\, a(u)\partial_x u + \frac{(\Delta t)^2}{2} \partial_x \left(\big(a(u) \big)^2 \partial_x u \right) + O(\Delta t^3). \tag{4.7.47}$$

Combinando (4.7.46) y (4.7.47) deducimos la identidad (4.7.33) deseada.

■

Observación: Supongamos que el esquema numérico es consistente de orden 1 en cuyo caso $\beta(u,\, \lambda) \neq 0$. Entonces, el esquema proporciona una aproximación de orden dos de la ecuación viscosa

$$v_t + \partial_x \big(f(v) \big) - \lambda \Delta x \partial_x \big(\beta(v,\, \lambda) \partial_x v \big) = 0.$$

Una manera heurística de estudiar el comportamiento del esquema numérico es precisamente estudiar el de esta aproximación parabólica o viscosa.■

Con el objeto de analizar la estabilidad de los esquemas numéricos es preciso considerar en primer lugar la ecuación lineal

$$u_t + a \partial_x u = 0$$

en la que no-linealidad $f(s)$ se reduce a la ecuación lineal $f(s) = a\, s$ en la que a es una constante.

Supongamos que el esquema es de la forma

$$v_j^{n+1} = \sum_{\ell=-k}^{+k} c_\ell v_{j+\ell}^n,\ n \geq 0,\ j \in \mathbb{Z}$$

con coeficientes constantes c_ℓ, $-k \leq \ell \leq k$, que dependen sólo de a y de λ.

Consideramos la norma L^2-discreta:

$$\| v \|_{L^2(\Delta)} = \left(\Delta x \sum_{j \in \mathbb{Z}} v_j^2 \right)^{1/2}.$$

El esquema se dice estable si existe una constante $c > 0$ independiente de $\Delta t > 0$ tal que

$$\| v^n \|_{L^2(\Delta)} \leq C \| v^0 \|_{L^2(\Delta)},\ \forall n \geq 0.$$

Es conveniente extender el esquema a funciones definidas para todo x:

$$v^{n+1}(x) = \sum_{\ell=-k}^{+k} c_\ell v^n(x + \ell \Delta_x).$$

La propiedad de estabilidad se reescribe entonces como

$$\| v^n \|_{L^2(\mathbb{R})} \leq C \| v^0 \|_{L^2(\mathbb{R})},\ \forall n \geq 0.$$

Utilizando la transformada de Fourier

$$\hat{\varphi}(\xi) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-ix\xi} \varphi(x) dx,$$

el esquema se traduce en

$$\hat{v}^{n+1}(\xi) = h(\xi) \hat{v}^n(\xi)$$

donde

$$h(\xi) = \sum_{\ell=-k}^{+k} c_\ell e^{i\ell \Delta_x \xi}$$

se denomina el factor de amplificación.

Es fácil entonces comprobar que el esquema es estable sí y sólo sí se verifica la condición de Von-Neumann

$$| h(\xi) | \leq 1,\ \forall \xi \in \mathbb{R}.$$

Consideremos por ejemplo el esquema de tres puntos

$$v_j^{n+1} = c_{-1} v_{j-1}^n + c_0 v_j^n + c_1 v_{j+1}^n$$

que puede ponerse en forma conservariva si

$$c_{-1} + c_0 + c_1 = 1,$$

siendo su flujo numérico

$$g(u,\, v) = \big(c_{-1}u - c_1 v\big)/\lambda.$$

La condición de consistencia impone que

$$c_{-1} - c_1 = \lambda a.$$

Obtenemos así una familia uniparamétrica de esquemas dependiente de

$$q = c_{-1} + c_1 = 1 - c_0$$

que pueden ser escritos en la siguiente forma viscosa

$$v_j^{n+1} = v_j^n - \lambda a\big(v_{j+1}^n - v_{j-1}^n\big)/2 + q\big(v_{j+1}^n - 2v_j^n + v_{j-1}^n\big)/2.$$

Es fácil comprobar que el esquema es estable si el número de Courant $v = \lambda a$ satisface

$$(\lambda a)^2 \leq q \leq 1.$$

Para que el esquema sea de orden 2 es preciso que

$$\beta(u,\, \lambda) = \frac{q}{2\lambda^2} - \frac{a^2}{2} = 0,$$

es decir,

$$q = \lambda^2 a^2.$$

Existe por tanto un sólo esquema de tres puntos de orden 2. Es el esquema denominado de Lax-Wendroff

$$v_j^{n+1} = v_j^n - \lambda a \frac{(v_{j+1}^n - v_{j-1}^n)}{2} + \frac{\lambda^2 a^2}{2}\big(v_{j+1}^n - 2v_j^n + v_{j-1}^n\big),$$

que es estable bajo la condición

$$\lambda \mid a \mid \leq 1.$$

Se trata precisamente de la condición de Courant-Friedrichs-Lewy (CFL) que garantiza que el dominio de dependencia del esquema numérico contiene al de la EDP.

Este análisis lineal de la estabilidad de los esquemas nos permite derivar algunos criterios heurísticos para los esquemas de ecuaciones no-lineales por linearización. La ecuación linealizada es de la forma

$$w_t + a(u)\partial_x w = 0$$

siendo u una solución de la ecuación no-lineal.

Lo mismo puede hacerse para el esquema numérico obteniéndose el siguiente esquema linealizado:

$$w_j^{n+1} = \sum_{\ell=-k}^{+k} \frac{\partial H}{\partial v_\ell}\big(v_j^n, \ldots, v_j^n\big) w_{j+\ell}^n.$$

Aunque la estabilidad del esquema linealizado no es una condición suficiente de estabilidad del esquema permite obtener algunas condiciones heurísticas. De este modo, para el esquema de Lax-Wendroff obtenemos la condición

$$\max_{n,\, j\in\mathbb{Z}} \lambda \big|a\big(v_j^n\big)\big| \leq 1.$$

Mencionemos ahora algunos ejemplos relevantes de esquemas de tres puntos:

El esquema de Lax-Friedrichs:

Consideremos primero el esquema centrado

$$v_j^{n+1} = v_j^n - \lambda \left(\frac{f\big(v_{j+1}^n\big) - f\big(v_{j-1}^n\big)}{2} \right)$$

que es linealmente inestable.

Sustituyendo la aproximación de la derivada

$$u_t \sim \frac{\big[v_j^{n+1} - \big(v_{j+1}^n + v_{j-1}^n\big)/2\big]}{\Delta t}$$

obtenemos el esquema de Lax-Friedrichs

$$v_j^{n+1} = \frac{v_{j+1}^n + v_{j-1}^n}{2} - \lambda \frac{\big(f\big(v_{j+1}^n\big) - f\big(v_{j-1}^n\big)\big)}{2}.$$

Puede ponerse en forma conservativa con el flujo numérico

$$g^{LF}(u,\, v) = \frac{f(u) + f(v)}{2} - \frac{(v-u)}{2\lambda}$$

y es de orden uno.

En el caso lineal este esquema corresponde a un coeficiente de viscosidad $q = 1$ de modo que es estable si

$$\lambda \mid a \mid \leq 1.$$

En el caso lineal el esquema proporciona la interpolación entre los valores de la solución en $\big(x_{j-1},\, t_n\big)$ y $\big(x_{j+1},\, t_n\big)$.

El esquema upwind.

En el caso lineal teniendo en cuenta la orientación de las características en función del signo de a obtenemos

$$v_j^{n+1} = \begin{cases} v_j^n - \lambda a\big(v_j^n - v_{j-1}^n\big), & \text{si} \quad a > 0 \\ v_j^n - \lambda a\big(v_{j+1}^n - v_j^n\big), & \text{si} \quad a < 0. \end{cases}$$

Cuando la no-linealidad f es monótona el esquema es extiende con facilidad:

$$v_j^{n+1} = \begin{cases} v_j^n - \lambda\Big(f\big(v_j^n\big) - f\big(v_{j-1}^n\big)\Big), & \text{si} \quad f' > 0 \\ v_j^n - \lambda\Big(f\big(v_{j+1}^n\big) - f\big(v_j^n\big)\Big), & \text{si} \quad f' < 0. \end{cases}$$

El esquema de Godunov.

El esquema de Godunov está basado en la resolución de problemas de Riemann locales de la forma:

$$u_t + \partial_x\big(f(u)\big) = 0,$$

$$u(x,\, t) = \begin{cases} u_\ell, & \text{si} \quad x < 0 \\ u_r, & \text{si} \quad x > 0, \end{cases}$$

cuya solución es autosemejante de la forma

$$u(x,\, t) = w_R\Big(\frac{x}{t};\, u_\ell,\, u_r\Big).$$

El esquema de Godunov genera v^{n+1} a partir de v^n del siguiente modo.

Paso 1 Resolvemos exactamente el problema

$$\begin{cases} w_t + \partial_x\big(f(w)\big) = 0 \\ w(x,\, t_n) = v_\Delta(x,\, t_n) \end{cases}$$

donde v_Δ está definida del siguiente modo:

$$v_\Delta(x,\, t_n) = v_j^n,\; x_{j-1/2} \leq x \leq x_{j+1/2},\; j \in \mathbb{Z}.$$

Como el dato inicial es constante a trozos presenta una discontinuidad en cada punto de la forma $x_{k+1/2}$. Esto da lugar a un problema de Riemann en cada uno de ellos. Estos no interactuan antes de un tiempo Δt si

$$\lambda \,\text{máx}\big(\mid a(v)\mid :\, v \in \big[v_{j-1}^n,\, v_j\big]\big) \leq 1/2,\, j \in \mathbb{Z}.$$

Paso 2 v_j^{n+1} se define como la media de la solución precedente en el intervalo $\big[x_{j-1/2},\, x_{j+1/2}\big]$.

Con el objeto de obtener una expresión sencilla para v_j^{n+1} integramos la ecuación que w satisface en $\big(x_{j-1/2},\, x_{j+1/2}\big) \times (0,\, \Delta t)$, cosa que puede hacerse por tratarse de una solución débil regular a trozos y que satisface las condiciones de Rankine-Hugoniot. Obtenemos de este modo

$$v_j^{n+1} = v_j^n - \lambda\big[f\big(w_R\big(0;\, v_j^n,\, v_{j+1}^n\big)\big) - f\big(w_R\big(0;\, v_{j-1}^n,\, v_j^n\big)\big)\big].$$

Se trata de un esquema en forma conservativa con flujo

$$g_{\mathcal{G}}(u,\, v) = f\big(w_R(0;\, u,\, v)\big).$$

Es fácil comprobar que cuando f es lineal o, más generalmente, f es convexa en las regiones en que es monótona, el esquema de Godunov coincide con el upwind.

En el caso general el flujo puede también representarse de manera sencilla del siguiente modo. Como hemos visto antes, este esquema es linealmente estable pero es no-linealmente estable cerca de los puntos de estagnación en los que $a(u) = 0$. Esto es debido a que el esquema pierde su carácter disipativo en esos puntos.

Hasta ahora hemos visto algunos ejemplos de esquemas numéricos conservativos y hemos analizado su estabilidad lineal. Pero no hemos estudiado el problema de la convergencia a las soluciones de entropía de (4.7.1)-(4.7.2) cuando $\Delta x \to 0$ y $\Delta t \to 0$. Tal y como hemos, por el Teorema de Lax-Wendroff, si el esquema es conservativo y las soluciones numéricas están acotadas en L^∞ y convergen en L^1_{loc} y para casi todo punto, su límite es una solución débil.

Con el objeto de completar el análisis de la convergencia de los esquemas debemos:

• Analizar para cuáles de los esquemas introducidos las soluciones numéricas tienen las propiedades de acotación y de compacidad requeridas;

• Estudiar cuando la solución débil obtenida en el límite es también una solución de entropía.

Para analizar estas cuestiones vamos a considerar la clase de esquemas monótonos y T.V.D.

El esquema en diferencias (4.7.6) es monótono si H_Δ es monótona creciente con respecto a cada una de las variables involucradas.

Es por ejemplo fácil de comprobar que el esquema de Lax-Friedrichs es monótono si se cumple la condición de CFL

$$\lambda \max \mid f'(v) \mid \leq 1$$

y que el esquema de Engquist-Osher lo es bajo la misma condición. Es fácil comprobar que el esquema de Godunov es también monótono.

La monotonía de un esquema conservativo puede también caracterizarse a través de la función de flujo. En efecto, el esquema conservativo asociado a la función de flujo $g(u, v)$ es monótono si g es creciente en la primera variable y decreciente en la segunda.

Una de las limitaciones de los esquemas monótonos es que no pueden ser de orden mayor que uno.

Con el objeto de analizar la convergencia de las soluciones discretas introducimos las siguientes normas:

$$\begin{aligned} \| v \|_{L^1(\Delta)} &= \Delta x \sum_{j\in\mathbb{Z}} \mid v_j \mid \\ \| v \|_{L^\infty(\Delta)} &= \max_{j\in\mathbb{Z}} \mid v_j \mid \\ TV(v) &= \sum_{j\in\mathbb{Z}} \mid v_{j+1} - v_j \mid . \end{aligned}$$

Se trata de los tres casos de las versiones discretas de las normas canónicas de $L^1(\mathbb{R})$, $L^\infty(\mathbb{R})$ y $BV(\mathbb{R})$.

Diremos que un esquema es TVD (*total variation diminishing*), es decir, que hace decrecer la variación total si

$$TV\big(H_\Delta(v)\big) \leq TV(v). \tag{4.7.48}$$

Nuevamente los esquemas TVD pueden ser a lo sumo de orden 1.

Por otra parte, diremos que un esquema es L^∞-estable si existe una constante independiente de n y de $\Delta t > 0$ tal que

$$\| v^n \|_{L^\infty(\Delta)} \leq C, \ \forall n \geq 0. \tag{4.7.49}$$

Tenemos el siguiente resultado que asocia la monotonía de un esquema a su estabilidad L^∞ y su carácter TVD.

Sin embargo el esquema de Murman-Roe presenta soluciones discontinuas estacionarias con $u_\ell < u_r$ que violan las condiciones de entropía.

El esquema de Engquist-Osher.

Este esquema no tiene el inconveniente del de Roe de admitir soluciones que no verifican la condición de entropía. El esquema se escribe como

$$v_j^{n+1} = v_j^n - \frac{\lambda}{2}\left[f\big(v_{j+1}^n\big) - f\big(v_{j-1}^n\big)\right] + \frac{\lambda}{2}\left[\int_{v_j}^{v_{j+1}} \mid a(\xi) \mid d\xi - \int_{v_{j-1}}^{v_j} \mid a(\xi) \mid d\xi\right].$$

El esquema es conservativo con flujo

$$g^{\xi 0}(u,\, v) = \frac{1}{2}\left[f(u) + f(v) - \int_u^v \mid a(\xi) \mid d\xi\right].$$

Conviene observar que en los intervalos en que f' tiene signo constante coincide con el esquema upwind.

El esquema de Lax-Wendroff.

Se trata de derivar un esquema de orden 2 mediante la expansión de Taylor de una solución regular.

Tenemos

$$u(x,\, t+\Delta t) = u(x,\, t) - \Delta t \partial_x\big(f(u)\big) + \frac{(\Delta t)^2}{2}\partial_x\big(a(u)\partial_x f(u)\big) + O\big((\Delta t)^3\big).$$

Escribimos entonces

$$\partial_x\big(f(u)\big) = \big(f(u(x+\Delta x,\, t)) - f\big(u(x-\Delta x,\, t)\big)\big) + O\big((\Delta x)^2\big)$$

y para cualquier $\theta \in [0,\, 1]$:

$$\begin{aligned}\partial_x\big(a(u)\partial_x\big(f(u)\big)\big) = \big[&a(u(x+\theta\Delta x,\, t)(f(u(x+\Delta x,\, t) - f(u(x,\, t))) \\ &-a\big(u(x-(1-\theta)\Delta x,\, t)\big)\big(f(u(x,\, t)) - f(u(x-\Delta x,\, t))\big)\big] + O(\Delta x).\end{aligned}$$

Obtenemos así un esquema de segundo orden tomando

$$\begin{aligned}v_j^{n+1} = v_j^n - \frac{\lambda}{2}\big[f\big(v_{j+1}^n\big) - f\big(v_{j-1}^n\big)\big] + \frac{\lambda^2}{2}\big[&a_{j+1/2}^n\big(f\big(v_{j+1}^n\big) - f\big(v_{j-1}^n\big)\big) \\ &-a_{j-1/2}^n\big(f\big(v_j^n\big) - f\big(v_{j-1}^n\big)\big)\big]\end{aligned}$$

siendo $a_{j+1/2}^n$ el valor de f' evaluado en un punto intermedio entre v_j^n y v_{j+1}^n.

$$g^{\mathcal{G}}(u,\, v) = \begin{cases} \min\limits_{w\in[u,\, v]} f(w), & \text{si} \quad u \le v \\ \min\limits_{w\in[v,\, u]} f(w), & \text{si} \quad v \le u. \end{cases}$$

El esquema de Murman-Roe.

Se trata de evitar el mayor inconveniente del esquema de Godunov que exige resolver un problema de Riemann. El esquema de Murman-Roe utiliza un resolvedor aproximado muy sencillo del problema de Riemann. Para introducirlo consideramos la función

$$a(u,\, v) = \begin{cases} \dfrac{f(u) - f(v)}{u - v} & \text{si} \quad u \neq v \\ f'(u) & \text{si} \quad u = v. \end{cases}$$

Procedemos como en el esquema de Godunov pero sustituyendo la resolución del problema de Riemann por la del problema lineal

$$w_t + a\big(v_j^n,\, v_{j+1}^n\big) w_x = 0$$

$$w(x,\, 0) = \begin{cases} w_j^n, & x < x_{j+1/2} \\ v_{j+1}^n, & x > x_{j+1/2}. \end{cases}$$

El resultado es una onda discontinua propagándose a velocidad $a\big(v_j^n,\, v_{j+1}^n\big)$. La solución que se obtiene es

$$x(x,\, t) = w_R^{Roe}\big(\xi,\, v_j^n,\, v_{j+1}^n\big) = \begin{cases} v_j, & \xi < a\big(v_j^n,\, v_{j+1}^n\big) \\ v_{j+1}, & \xi > a\big(v_j^n,\, v_{j+1}^n\big). \end{cases}$$

Sustituyendo este resolvedor en el esquema de Godunov obtenemos

$$v_j^{n+1} = v_j^m - \lambda \left[f\left(w_R^{Roe}\big(0;\, v_j^n,\, v_{j+1}^n\big)\right) - f\left(w_R^{Roe}\big(0;\, v_{j-1}^n,\, v_j^n\big)\right)\right].$$

El flujo numérico asociado es entonces

$$g^R(u,\, v) = f\left(w_R^{Roe}(0;\, u,\, v)\right) = \begin{cases} f(u), & \text{si} \quad a(u,\, v) > 0 \\ f(v), & \text{si} \quad a(u,\, v) < 0, \end{cases}$$

que puede también escribirse como

$$g^R(u,\, v) = \frac{1}{2}\Big(f(u) + f(v) - \mid a(u,\, v) \mid (v - u)\Big).$$

Proposition 4.7.2 *Sea* (4.7.6) *un esquema de diferencias monótono y conservativo. Entonces es T.V.D. y L^∞-estable. De hecho,*

$$\| \, v^n \, \|_{L^\infty(\Delta)} \leq \| \, v^0 \, \|_{L^\infty(\Delta)} \, . \tag{4.7.50}$$

Además para cualquier par de sucesiones u y v tenemos

$$\| \, H_\Delta(u) - H_\Delta(v) \, \|_{L^1(\Delta)} \leq \| \, u - v \, \|_{L^1(\Delta)} \, . \tag{4.7.51}$$

Demostración.

En primer lugar vemos que por la monotonía del esquema se verifica el principio del máximo. Es decir

$$\min_{j-k\le\ell\le j+k} v_\ell \le \big(H_\Delta(v)\big)_j \le \max_{j-k\le\ell\le j+k} v_\ell. \tag{4.7.52}$$

Para ello, dada la sucesión acotada v_j introducimos la sucesión constante w_j:

$$w_j = \max_{\ell\in\mathbb{Z}} v_\ell = c. \tag{4.7.53}$$

Como el esquema es conservativo tenemos

$$\big(H_\Delta(v)\big)_j = c \tag{4.7.54}$$

y por la monotonía del esquema, como $v \le w$, tenemos

$$H_\Delta(v) \le H_\Delta(w). \tag{4.7.55}$$

Como H_Δ depende sólo de $2k+1$ variables, de (4.7.55) se deduce (4.7.52). De (4.7.52) se deduce (4.7.50).

Conviene señalar que en este punto hemos usado un argumento clásico en EDP que consiste en deducir propiedades de estabilidad en L^∞ a partir del principio del máximo.

Observamos ahora que la propiedad T.V.D. es consecuencia inmediata de (4.7.51), que es una propiedad de contracción en $L^1(\Delta)$ de la aplicación no-lineal H_Δ.

Para verlo constatamos que H_Δ preserva la integral, gracias a que es un esquema conservativo, que es monótono, y que está acotado en $L^1(\Delta)$:

$$\| H_\Delta(v) \|_{L^1(\Delta)} = \Delta x \sum_{j\in\mathbb{Z}} | H_\Delta(v)_j | \le (2k+1)\Delta x \sum_j | v_j | \tag{4.7.56}$$

lo cual es consecuencia de (4.7.52). Esto es así como consecuencia del Lema de Crandall-Tartar siguiente:

Lema (Crandall-Tartar). *Sea C un subconjunto de $L^1(\Omega)$ tal que*

$$f \vee_{\mathrm{v}} g = \sup(f,\, g) \in C,\ \forall f,\, g \in C.$$

Sea T una aplicación de C en $L^1(\Omega)$ que verifica

$$\int_\Omega T(f) = \int_\Omega f.$$

Entonces las siguientes propiedades son equivalentes:

(a) $f,\, g \in C,\, f \leq g \quad p.c.t. \quad x \in \Omega \Rightarrow T(f) \leq T(g) \quad p.c.t. \quad x \in \Omega;$

(b) $\displaystyle\int_\Omega \mid T(f) - T(g) \mid \leq \int_\Omega \mid f - g \mid,\, \forall f,\, g \in C;$

(c) $\displaystyle\int_\Omega \big(T(f) - T(g)\big)^+ \leq \int_\Omega (f-g)^+,\, \forall f,\, g \in C.$

Como consecuencia de la proposición tenemos que:

Corolario. Si el esquema es conservativo y monótono entonces satisface las siguientes estimaciones

$$\| v_\Delta(\cdot,\, t) \|_{L^\infty(\mathbb{R})} \leq \| v_\Delta(\cdot,\, 0) \|_{L^\infty(\Delta)}, \tag{4.7.57}$$

$$\| v_\Delta(\cdot,\, t) \|_{L^1(\mathbb{R})} \leq \| v_\Delta(\cdot,\, 0) \|_{L^1(\Delta)}, \tag{4.7.58}$$

$$Tv\big(v_\Delta(\cdot,\, t)\big) \leq TV\big(v_\Delta(\cdot,\, 0)\big). \tag{4.7.59}$$

Hemos probado que todo esquema conservativo monótono es T.V.D. El recíproco también es cierto. No demostraremos sin embargo este resultado con el objeto de concluir la prueba de la convergencia.

En virtud de las estimaciones (4.7.57)-(4.7.59) y aplicando los mismos argumentos empleados en el estudio del comportamiento de las soluciones de las ecuaciones continuas en el límite de la viscosidad evanescente, no es difícil de concluir los resultados de acotación y compacidad necesarios sobre las soluciones discretas. De este hecho y del resultado de convergencia general de Lax-Wendroff, al tratarse de esquemas conservativos, podemos concluir que el límite es una solución débil de (4.7.1)-(4.7.2).

Queda por ver que se trata de la solución de entropía. Para ello es preciso que el esquema numérico verifique la condición adicional de ser consistente con la condición de entropía.

Para ello introducimos una versión discreta de la condición de entropía. Recordemos que para la ecuación (4.7.1) la condición de entropía es de la forma

$$\partial_t U(u) + 0_x\big(F(u)\big) \leq 0, \tag{4.7.60}$$

para todo par $(U,\, F)$ de entropía + flujo de entropía relacionados a través de la condición

$$F'(s) = U'(s) f'(s).$$

De manera análoga diremos que el esquema es entrópico si

$$\begin{cases} \forall (U,\, F),\, \exists\, \mathcal{G} : \mathcal{G}(s,\, s, \ldots,\, s) = F(s) \quad t.q. \\ v_j^{n+1} \leq v_j^n - \lambda\big(\mathcal{G}_{j+1/2}^n - \mathcal{G}_{j-1/2}^n\big). \end{cases} \tag{4.7.61}$$

Lo mismo que ocurre en el caso de la EDP podemos limitarnos al caso en que

$$U(s) =\mid s-k\mid,\ F(s) = \operatorname{sgn}(s-k)\big(f(s)-f(k)\big),$$

con $k \in \mathbb{R}$.

Se puede comprobar que todo esquema monótono y consistente es entrópico. Por otra parte, el mismo tipo de argumentos utilizados en la prueba del Teorema de Lax-Wendroff según el cual los límites de soluciones discretas de ecuaciones conservativas son soluciones débiles de (4.7.1), en el caso de esquemas entrópicos, permite probar que se trata de soluciones de entropía.

El último resultado que precisamos es el que garantiza que todo esquema monótono y consistente es entrópico.

De este modo obtenemos el siguiente resultado de convergencia:

Theorem 4.7.1 *Supongamos que el esquema es conservativo, consistente y monótono y que la función de flujo numérico es localmente Lipschitz. Entonces, si* $u_0 \in L^\infty(\mathbb{R}) \cap L^1(\mathbb{R}) \cap BV(\mathbb{R})$ *la solución del esquema numérico converge a la solución de entropía de* (4.7.1)-(4.7.2) *de modo que*

$$u_\Delta \to u \text{ en } L^\infty\big(0,\, T;\, L^1_{loc}(\mathbb{R})\big),\ \forall T > 0.$$

4.8. Ejercicios

1 Comprueba que la superposición de dos movimientos armónicos de la misma frecuencia de la forma

$$x_j(t) = A_j \cos(\omega_0 t + \phi_j),\ j = 1, 2$$

puede escribirse en la forma

$$x(t) = Ae^{i(\omega_0 t+\phi)}$$

con

$$\begin{aligned} A &= \sqrt{A_1^2 + A_2^2 + A_1 A_2 \cos(\phi_1 - \phi_2)} \\ \operatorname{tg}\phi &= \frac{A_1 \operatorname{sen}\phi_1 + A_2 \operatorname{sen}\phi_2}{A_1\cos\phi_1 + A_2\cos\phi_2}. \end{aligned}$$

2 a) Comprueba mediante un cambio de variables que la ecuación de ondas

$$\rho v_{tt} - \sigma v_{xx} = 0$$

con ρ y σ constantes positivas, puede reducirse al caso particular en que $\rho = \sigma = 1$.

b) Deduce la fórmula de d'Alembert para esta ecuación y analiza la velocidad de propagación, dominio de dependencia y región de influencia.

c) Comprueba que, mediante un cambio de variables adecuado, la ecuación

$$\rho(x)v_{tt} - (\sigma(x)v_x)_x = 0$$

con coeficientes regulares positivos puede reducirse a una ecuación de ondas de la forma

$$v_{tt} - v_{xx} + a(x)v = 0$$

en la que la parte principal es el operador de d'Alembert, que se ve perturbada por un potencial $a = a(x)$ dependiente de los coeficientes $\rho(x)$ y $\sigma(x)$.

3 a) Comprueba que no existe ninguna función discreta $\varepsilon(N)$ tal que $\varepsilon(N) \to 0$ cuando $N \to \infty$ y que satisfaga

$$\| \vec{a} - \vec{a}_N \|_{\ell^2} \leqslant \varepsilon(N) \| \vec{a} \|_{\ell^2}, \quad \forall \vec{a} \in \ell^2$$

donde $\vec{a}_N$ denota la sucesión truncada en el N-ésimo elemento.

b) Comprueba que existe dicha función si nos limitamos a una versión más débil de esta desigualdad:

$$\| \vec{a} - \vec{a}_N \|_{\ell^2} \leqslant \varepsilon(N) \| \vec{a} \|_{h^1}, \quad \forall \vec{a} \in h^1,$$

donde

$$h^1 = \left\{ \vec{a} = (a_j)_{j \geqslant 1} : \sum_{j=1}^{\infty} j^2 a j^2 < \infty \right\}$$

y

$$\| \vec{a} \|_{h^1} = \Big[\sum_{j=1}^{\infty} j^2 a j^2 \Big]^{1/2}.$$

Prueba primeramente que h^1 es denso en ℓ^2.

c) Demuestra que una desigualdad semejante se verifica si sustituimos h^1 por h_ρ:

$$h_\rho = \left\{ \vec{a} = (a_j)_{j \geqslant 1} : \sum_{j=1}^{\infty} \rho_j a_j^2 < \infty \right\}$$

siendo $\rho = (\rho_j)_{j \geqslant 1}$ una función discreta tal que $\rho_j \to \infty$ cuando $j \to \infty$.

Calcula $\varepsilon = \varepsilon(N)$ en función de $(\rho_j)_{j \geqslant 1}$.

4 Demuestra que existe $C > 0$ tal que

$$\| f \|^2_{H^2(0,\pi)} \leqslant C \| f'' \|^2_{L^2(0,\pi)}$$

para toda función $f \in H^2(0,\pi) \cap H^1_0(0,\pi)$.

5 Consideramos la ecuación de ondas disipativa

$$\begin{cases} \varepsilon u_{tt} - \Delta u + u_t = 0 & \text{en} \quad \Omega \times (0,\infty) \\ u = 0 & \text{en} \quad \partial\Omega \times (0,\infty) \\ u(0) = u_0,\ u_t(0) = u_1 & \text{en} \quad \Omega. \end{cases}$$

a) Desarrolla las soluciones en serie de Fourier en la tasa de las autofunciones del Laplaciano.

b) Calcula la base exponencial de convergencia de la energía cuando $t \to \infty$.

c) Comprueba que a medida que ε tiende a cero el comportamiento de las soluciones se asemeja cada vez más al de las soluciones de la ecuación del calor

$$\begin{cases} u_t - \Delta u = 0 & \text{en} \quad \Omega \times (0,\infty) \\ u = 0 & \text{en} \quad \partial\Omega \times (0,\infty) \\ u(0) = u_0 & \text{en} \quad \Omega. \end{cases}$$

d) Dibuja el espectro de la ecuación de ondas disipativa en el plano complejo, describe su evolución a medida que $\varepsilon \to 0$ y comprueba que en el límite se recupera el espectro de la ecuación del calor.

e) >Cómo se refleja a nivel del espectro el hecho de que la ecuación de ondas sea una ecuación de orden dos en tiempo y, sin embargo, su límite singular sea simplemente una ecuación de ondas uno en tiempo?

6 Consideremos la ecuación de ondas disipativa

$$\begin{cases} u_{tt} - \Delta u + \alpha u + \beta u_t = 0 & \text{en} \quad \Omega \times (0,\infty) \\ u = 0 & \text{en} \quad \partial\Omega \times (0,\infty) \end{cases}$$

en un dominio acotado y regular Ω de $\mathbb{R}^n$.

Demuestra que eligiendo $\alpha, \beta > 0$ adecuados se puede conseguir que la tasa de decaimiento exponencial de las soluciones sea arbitrariamente grande.

7 Escribe la ecuación de ondas

$$\begin{cases} u_{tt} - \Delta u = 0 & \text{en} \quad \Omega \times (0, \infty) \\ u = 0 & \text{en} \quad \partial\Omega \times (0, \infty) \end{cases}$$

en la forma abstracta de una ecuación de evolución de orden uno en el espacio de la energía. Utilizando la descomposición espectral del Laplaciano calcula la forma explícita de la aproximación de Yosida del generador del semigrupo de ondas.

Comprueba que se trata de un operador anti-adjunto acotado. Discute en qué modo se aproxima al generador de la ecuación de ondas.

8 Utilizando una base ortonormal del Laplaciano $\{\phi_j\}_{j\geqslant 1}$ asociado a sus autofunciones con condiciones de contorno de Dirichlet y definiendo el producto escalar en $H^{-1}(\Omega)$ del siguiente modo

$$(p, q)_{H^{-1}} = \sum_{j\geqslant 1} \frac{p_j q_j}{\lambda_j}$$

donde

$$p(x) = \sum_{j\geqslant 1} p_j \phi_j(x);\ q(x) = \sum_{j\geqslant 1} q_j \phi_j(x),$$

comprueba que este producto escalar coincide con el que se obtiene mediante la definición

$$(p, q)_{H^{-1}(\Omega)} = \Big((\Delta)^{-1} p,\ (-\Delta)^{-1} q\Big)_{H_0^1(\Omega)}.$$

9 Consideramos la ecuación de ondas

$$\begin{cases} u_{tt} - \Delta u = 0 & \text{en} \quad \Omega \times (0, \infty) \\ u = 0 & \text{en} \quad \partial\Omega \times (0, \infty) \\ u(0) = u_0,\ u_t(0) = u_1 & \text{en} \quad \Omega \end{cases}$$

en un dominio acotado Ω de $\mathbb{R}^n$ con datos iniciales $u_0 \in L^2(\Omega)$ y $u_1 \in H^{-1}(\Omega)$.

a) Comprueba que si definimos

$$v(x, t) = \int_0^t u(x, s) ds + \chi(x)$$

con una función χ adecuadamente elegida, entonces v es solución de la ecuación de ondas con datos iniciales en $H_0^1(\Omega) \times L^2(\Omega)$.

b) Utilizando el resultado de existencia y unicidad de las soluciones de energía finita con datos iniciales en $H_0^1(\Omega) \times L^2(\Omega)$, deduce que con datos iniciales en $L^2(\Omega) \times H^{-1}(\Omega)$ existe una única solución en el espacio

$$u \in C\Big([0,\infty);\, L^2(\Omega)\Big) \cap C^1\Big([0,\infty); H^{-1}(\Omega)\Big).$$

10 Consideramos la ecuación de transporte

$$\begin{cases} u_t + u_x = 0, & x \in \mathbb{R}, \quad t > 0 \\ u(x,0) = f(x), & x \in \mathbb{R}. \end{cases}$$

Demuestra que si $f \in C^3(\mathbb{R})$ es de soporte compacto, las soluciones del esquema semi-discreto centrado

$$\begin{cases} u_j' + \dfrac{u_{j+1} - u_{j-1}}{2h} = 0, & j \in \mathbb{Z}, \quad t > 0 \\ u_j(0) = f(x_j), & j \in \mathbb{Z} \end{cases}$$

convergen a la solución de la ecuación de transporte con un orden dos de convergencia.

Obtén una estimación explícita del error.

11 Consideramos un operador no acotado diagonal en ℓ^2, con dominio

$$D(A) = \left\{ \vec{u} \in \ell^2 : \sum_{j\in\mathbb{Z}} \lambda_j^2 u_j^2 < \infty \right\}.$$

definido como $A\vec{u} = (\lambda_j u_j)_{j\in\mathbb{Z}}$ sobre los elementos $\vec{u} = (u_j)_{j\in\mathbb{Z}}$ del dominio.

a) Demuestra que A es un operador acotado y que $D(A) = \ell^2$ si y sólo si

$$\sup_{j\in\mathbb{Z}} \mid \lambda_j \mid < \infty.$$

b) Demuestra que A es un operador compacto (envía conjuntos acotados de ℓ^2 en conjuntos relativamente compactos) si y sólo si

$$\lim_{|j|\to\infty} \mid \lambda_j \mid = 0.$$

Comprueba que en este caso A puede ser aproximado en $\mathcal{L}(\ell^2, \ell^2)$ para operadores de rango finito.

12 Sea $\{e^{At}\}_{t\geqslant 0}$ un semigrupo de contracciones en un espacio de Hilbert H. Prueba que las dos siguientes condiciones son equivalentes:

- $\exists T > 0 : \quad \| e^{AT} \|_{\mathcal{L}(H,H)} < 1;$
- $\exists C, \omega > 0 : \quad \| e^{At} \|_{\mathcal{L}(H,H)} \leqslant Ce^{-\omega t}.$

13 Probar que el recíproco de la desigualdad de Poincaré no es cierto en ningún abierto no vacío de $\mathbb{R}^n$. Es decir, probar que si Ω es un abierto no vacío de $\mathbb{R}^n$,

$$\sup_{u\in H_0^1(\Omega)} \frac{\int_\Omega |\nabla u|^2 \, dx}{\int_\Omega u^2 dx} = \infty.$$

Indicación: Considérese en primer lugar el caso de una variable espacial con $u(x) = \varphi(x/\varepsilon)$ siendo $\varphi \in C_c^\infty(\mathbb{R})$ y $\varepsilon \to 0$. Abórdese después el caso multi-dimensional por separación de variables.

14 a) Resuelve mediante el método de las características la ecuación de transporte

$$u_t + a(x) \cdot \nabla u = 0$$

donde $a = a(x)$ es una función regular.

b) Resuelve posteriormente la ecuación perturbada

$$u_t + a(x) \cdot \nabla u + b(x)u = 0.$$

15 Utiliza el método de descomposición de Fourier y la fórmula de variación de las constantes para obtener una ecuación integral de las soluciones de la ecuación de ondas semi-lineal

$$\begin{cases} u_{tt} - u_{xx} + u^3 = 0, & 0 < x < \pi, \quad t > 0 \\ u(0,t) = u(\pi,t) = 0, & t > 0 \\ u(x,0) = \varphi(x),\ u_t(x,0) = \psi(x), & 0 < x < \pi. \end{cases}$$

Comprueba que se trata de un sistema de infinitas ecuaciones con infinitas incógnitas acopladas. Analiza la naturaleza de este acoplamiento utilizando el valor explícito de las integrales

$$\int_0^\pi \operatorname{sin}(k_1 x) \operatorname{sen}(k_2 x) \operatorname{sen}(k_3 x) \operatorname{sen}(k_4 x) dx.$$

16 Supongamos que f es una función continua para la que existe $g \in L^1(\mathbb{R})$ tal que

$$| f(x) | \leqslant g(x), \text{ p.c.t. } x \in \mathbb{R}$$

de modo que g sea decreciente para $x > 0$ y creciente para $x < 0$.

Demuestra que

$$h \sum_{j \in \mathbb{Z}} f(jh) \to \int_{\mathbb{R}} f(x)dx,\ h \to 0.$$

17 Consideramos el siguiente operador lineal acotado de ℓ^2 en ℓ^2:

$$T\Big[(u_j)_{j\in\mathbb{Z}}\Big] = \Big(\alpha u_j + \beta u_{j-1}\Big)_{j\in\mathbb{Z}} \tag{1}$$

donde

$$0 < \beta < \alpha < \infty. \tag{2}$$

Pretendemos probar que, bajo la condición (2), este operador es inversible y que su inverso T^{-1} es también un operador acotado de ℓ^2 en ℓ^2.

Procedemos de dos modos distintos.

a) Utilizando la transformación de von Neumann obtén una expresión explícita de T^{-1} y una cota de su norma como operador lineal acotado de ℓ^2 en ℓ^2.

b) Aproximamos la ecuación (1) por sistemas de dimensión finita
(3)
$T_N(u_{-N}, \ldots, u_{-1}, u_0, u_1, \ldots, u_N) = A_N(u_{-N}, \ldots, u_{-1}, u_0, u_1, \ldots, u_N)$

donde A_N es la matriz $(2N+1) \times (2N+1)$ de la forma

$$A_N = \begin{pmatrix} \alpha\, 0 \,.....,0 \\ \beta\, \alpha\, 0\, ..,0 \\ ..... \\ 0\, ...,0\, \alpha \end{pmatrix}. \tag{4}$$

Comprobad que A_N^{-1} es inversible para cada N y verificar que, para cada $(f_j)_{j\in\mathbb{Z}}$, $(A_N^{-1})(f_j)_{j\in\mathbb{Z}}$ es una sucesión acotada en ℓ^2. (En este punto abusamos un poco de la notación. En efecto, A_N^{-1} no se aplica a la sucesión $(f_j)_{j\in\mathbb{Z}}$ sino al vector de dimensión $2N+1$ truncado $(f_{-N}, \ldots, f_{-1}, f_0, f_1, \ldots, f_N)$. Análogamente, $(A_N^{-1})(f_j)_{j\in\mathbb{Z}}$ no pertenece a ℓ^2 sino que es un vector de $2N+1$ componentes. Se convierte en un elemento de ℓ^2 cuando lo prolongamos mediante el valor cero para todos los índices j con $| j | > N$.

18 a) Escribe la ecuación de ondas con coeficientes constantes y condiciones de contorno de Dirichlet homogéneas en un dominio acotado en forma abstracta

$$U_t = AU.$$

b) Utilizando las autofunciones del Laplaciano Dirichlet, realiza la descomposición espectral del operador A que adopta la forma de un operador diagonal.

c) Calcula la regularización Yosida de A.

d) Verifica que la regularización Yosida A_λ satisface:

$$* \quad \| A_\lambda \| \leqslant 1/\lambda;$$

$$* \quad A_\lambda U \to AU, \quad \forall U \in D(A).$$

e) Comenta la idoneidad de la aproximación Yosida en el sentido de que genera una dinámica infinito-dimensional muy semejante que la de la ecuación de ondas.

19 Aplicar el resultado general del ejercicio #17 para probar que se puede resolver el esquema implícito de Crank-Nicholson siguiente

$$\frac{u_j^{k+1} - u_j^k}{\Delta t} = -\frac{1}{2}\Big[\frac{u_{j+1}^k - u_{j-1}^k}{2\Delta x} + \frac{u_{j+1}^{k+1} - u_{j-1}^{k+1}}{2\Delta x}\Big]$$

para la aproximación de la ecuación de transporte

$$u_t + u_x = 0.$$

Comprobar que se trata de un esquema convergente de orden dos para cualquier valor del parámetro de Courant μ.

20 Comprueba que los espacios $L^2(0,\infty;\, L^3(0,\pi))$ y $L^3(0,\infty;\, L^2(0,\pi))$ no son comparables (ninguno de los dos está contenido en el otro).

21 Demuestra que cualquier función de $L^2(0,\pi)$ puede aproximarse para funciones regulares que en los extremos $x = 0,\ \pi$ toman un valor arbitrario. Comprueba que ésto no es posible en $H_0^1(0,\pi)$.

22 Sea Ω un dominio regular, estríctamente convexo de $\mathbb{R}^N$. Demuestra que la traza de u tiene sentido si $u \in L^2(\Omega)$ y $\partial_1 u \in L^2(\Omega)$.

>A qué espacio pertenece la traza?

23 Sea A un operador no acotado en un espacio de Hilbert H que genera un semigrupo de contracciones.

Comprueba que las dos siguientes condiciones son equivalentes:

(1) $$\exists T > 0 : \| e^{AT} \| < 1;$$

(2) $$\exists C, \omega > 0 : \| e^{At} \| \leqslant Ce^{-\omega t}, \forall t > 0.$$

24 Demuestra de manera rigurosa en el contexto de la ecuación de transporte

$$u_t + u_x = 0$$

que un esquema numérico semi-discreto o completamente discreto que no verifica la condición de CFL sobre los dominios de dependencia no puede ser convergente.

25 Consideremos el esquema de Lax-Friedrichs

(1) $$u_j^{k+1} = \frac{1}{2}(1-\mu)u_{j-1}^k + \frac{1}{2}(1+\mu)u_{j+1}^k$$

para aproximar las soluciones de la ecuación de transporte

(2) $$u_t + u_x.$$

a) Comprueba que el esquema puede escribirse de la forma

(3) $$\frac{u_j^{k+1} - \frac{1}{2}\left(u_{j-1}^k + u_{j+1}^k\right)}{\Delta t} + \frac{u_{j+1}^k - u_{j-1}^k}{2\Delta x} = 0$$

y comenta la analogía entre (3) y (2).

b) Verifica que se trata de un esquema consistente de orden uno.

c) Comprueba que es estable si y sólo si el número de Courant $\mu \leqslant 1$

d) Escribe un modelo de EDP que sea intermedio entre (2) y (3). >Se percibe algún efecto disipativo en el mismo?

Comenta este resultado en relación con lo observado en el análisis de von Neumann del esquema (3).

26 Desarrolla el programa del ejercicio anterior en el caso de la aproximación de Lax-Wendroff:

$$u_j^{k+1} = \frac{1}{2}\mu(1+\mu)u_{j-1}^k + (1-\mu^2)u_j^k - \frac{1}{2}\mu(1-\mu)u_{j+1}^k.$$

27 Como es bien sabido y es fácil de comprobar, la ecuación de transporte

$$u_t + u_x = 0$$

conserva la masa. Es decir,

$$\frac{d}{dt}\Big[\int_{\mathbb{R}} u(x,t)dx\Big] = 0.$$

Esta propiedad puede obtenerse tanto a través de la fórmula explícita de solución ($u(x,t) = f(x,t)$) como integrando con respecto a $x \in \mathbb{R}$ en la ecuación de transporte.

Verifica si los esquemas de aproximación semi-discretos y discretos de la ecuación de transporte introducidos en las notas reproducen esta propiedad.

>Es necesario que esta propiedad de conservación se cumpla para que un esquema sea convergente?>Es necesario que se cumpla en un sentido aproximado cada vez más preciso a medida que los parámetros de discretización tienden a cero?

28 Obtén la expresión explícita de la velocidad de fase y de grupo en los esquemas numéricos introducidos en las notas. Comenta en particular si se percibe la existencia de efectos disipativos en la aproximación numérica. Compara con el comportamiento de la ecuación en derivadas parciales, corrección de la ecuación de transporte puro, que más se asemeja al esquema numérico.

29 Consideramos la función $f \in H^s(\mathbb{R})$ con $s \geqslant 0$. Definimos la aproximación

$$f_h(x) = \mathcal{F}^{-1}(\widehat{f}1_{(-\pi/h,\,\pi/h)}).$$

La sucesión $\{f_h\}_{h>0}$ está constituida por funciones de banda limitada representables en un mallado de paso h.

a) Demuestra que si $s = 0$,

$$f_h \to f \text{ en } L^2(\mathbb{R}),\ h \to 0.$$

b) Cuando $s > 0$, obtén una estimación superior de la tasa de convergencia de f_h a f en $L^2(\mathbb{R})$.

c) Construye un ejemplo explícito de función f donde se observe que el orden de convergencia obtenido en el apartado anterior es óptimo.

d) demuestra que, cuando $s = 0$ el orden de converegncia puede ser arbitrariamente lento.

e) Comprueba que lo dicho en el caso $s > 0$ sirve cuando $s > s'$, si se trata de medir la convergencia en el espacio

30 Consideramos el siguiente esquema centrado

$$u_j'' + \frac{2u_j - u_{j+1} - u_{j-1}}{h^2} = 0,\ j \in \mathbb{Z},\ t > 0$$

para la aproximación de la ecuación de ondas

$$u_{tt} - u_{xx} = 0.$$

a) Prueba que tanto en el problema de Cauchy como en el problema de Dirichlet la energía

$$E_h(t) = \frac{1}{2}\sum_j \Big[\mid u_j'(t) \mid^2 + \Big| \frac{u_{j+1}(t) - u_j(t)}{h} \Big|^2 \Big]$$

se conserva en tiempo.

Deduce la estabilidad del método.

b) Comprueba en el caso del problema de Cauchy esta propiedad de estabilidad mediante el método de von Neumann.

c) Comprueba que el esquema es consistente de orden dos.

d) Enuncia un resultado preciso de la convergencia de orden dos tanto para el problema de Cauchy como para el de Dirichlet, imponiendo condiciones adecuadas sobre los datos iniciales.

e) Escribe una corrección de la ecuación de las ecuaciones de ondas que refleje mejor que ella la dinámica del sistema semi-discreto.

31 Responde a las cuestiones del problema anterior en el marco del esquema semi-discreto

$$u_j'' + \Big[\frac{2u_j - u_{j+1} - u_{j-1}}{h^2}\Big] + h^\alpha \Big[2u_j' - u_{j-1}' - u_{j+1}'\Big] = 0$$

En particular, calcula el orden del método en función del valor del parámetro $\alpha > -2$.

32 Ilustra en Matlab el comportamiento del esquema progresivo, centrado y regresivo para la aproximación de la ecuación de transporte

$$u_t + u_x = 0$$

con dato inicial

$$f(x) = e^{-x^2}.$$

En particular asegurate de que los siguientes hechos quedan claramente reflejados:

a) El esquema progresivo diverge.

b) A pesar de que el esquema centrado da lugar a una aproximación de mejor orden que el regresivo, la dinámica de este último se asemeja más a la de la ecuación de transporte que la del esquema centrado >Por qué?

33 Consideramos el esquema discreto progresivo para la aproximación numérica de la ecuación de transporte

$$u_t + u_x = 0.$$

Sabemos que se trata de un método inestable.

a) Comprueba que para cualquier $s > 0$ existe un dato inicial $f \in H^s(\mathbb{R})$ tal que si tomamos como dato inicial del problema semi-discreto

$$f_h = \mathcal{F}^{-1}\left(\widehat{f}(\xi) 1_{(-\pi/h,\, \pi/h)}(\xi)\right),$$

entonces la solución semi-discreta tiene, para cada $t > 0$, una norma en $L^2(\mathbb{R})$ que diverge cuando $h \to 0$.

b) Impón una condición sobre el dato inicial f que garantice la convergencia del método.

34 Consideramos la ecuación de ondas disipativa

$$\begin{cases} u_{tt} - u_{xx} + u_t = 0, & x \in \mathbb{R}, \quad t > 0 \\ u(x,0) = f(x), u_t(x,0) = g(x), & x \in \mathbb{R}. \end{cases} \tag{4.8.1}$$

A Comprobar formalmente que las soluciones regulares de (4.8.1) que decaen cuando $x \to \infty$ son tales que la energía

$$E(t) + \frac{1}{2}\int_{\mathbb{R}} \left[u_x^2(x,t) + u_t^2(x,t)\right] dx \tag{4.8.2}$$

decrece en tiempo.

Obtener una fórmula explícita para la evolución de la energía.

Justificar el calificativo de "disipativa" de la ecuación.

B Escribir el sistema (4.8.1) como un problema abstracto de Cauchy

$$\begin{cases} \cup_t = A\cup, & t > 0 \\ \cup(0) = \cup_0, \end{cases} \tag{4.8.3}$$

siendo A un operador lineal no acotado en un espacio de Hilbert H.

C Indicar la estructura funcional adecuada del espacio H y del dominio del operador A.

D Demostrar que el operador A así construido es maximal disipativo. Deducir un resultado de existencia y unicidad de soluciones de (4.8.1).

E Comparar el valor explícito de $\langle A\cup, \cup\rangle_H$ con la fórmula de disipación de la energía obtenida en el apartado A.

Consideramos a partir de este momento la aproximación semi-discreta

$$\begin{cases} u_j'' + \dfrac{[2u_j - u_{j+1} - u_{j-1}]}{h^2} + u_j' = 0, & j \in \mathbb{Z}, \quad t > 0 \\ u_j(0) = f_j,\ u_j'(0) = g_j, & j \in \mathbb{Z} \end{cases} \tag{4.8.4}$$

con las notaciones habituales.

F Construye la energía discreta E_h asociada al sistema (4.8.4). Obtén una ley de disipación para esta energía. Compárala con la obtenida en el apartado A para la ecuación continua (4.8.1).

G Escribe (4.8.4) en la forma de un problema de Cauchy abstracto

$$\begin{cases} \cup_t = A_h\cup, & t > 0 \\ \cup(0) = \cup_0. \end{cases}$$

Comprueba que, en este caso, A_h es un operador acotado en $\ell^2 \times \ell^2$ siendo

$$\ell^2 = \left\{ (a_j)_{j\in\mathbb{Z}} : \sum_{j\in\mathbb{Z}} \mid a_j \mid^2 < \infty \right\},$$

dotado de la norma canónica usual.

Deduce la existencia y unicidad de soluciones de (4.8.4).

H Comprueba la consistencia, estabilidad y convergencia del esquema (4.8.4) a la ecuación continua (4.8.1). >Cuál es el orden de convergencia? Enuncia de manera precisa el resultado de convergencia señalando las hipótesis necesarias sobre la solución u del problema continuo (4.8.1) y por tanto de sus datos iniciales $f = f(x)$ y $g = g(x)$ y la norma en la que se tiene la convergencia.

I Utilizando el desarrollo de Taylor escribe una ecuación en derivadas parciales intermedia entre la EDP (4.8.1) y el esquema discreto (4.8.4).

J Integrando la ecuación (4.8.1) con respecto a x comprueba que la cantidad

$$\int_{\mathbb{R}} [u(x,t) + u_t(x,t)]dx$$

se conserva en tiempo para las soluciones de (4.8.1).

Verifica si el esquema semi-discreto satisface una propiedad similar.

K Aplica la transformada discreta de Fourier a escala h en (4.8.4) y deduce una ecuación que gobierne la evolución de la tranformada de Fourier discreta de la solución de (4).

Obtén una expresión de esta tranformada de Fourier discreta como combinación de exponenciales complejos.

Compárala con la transformada continua de Fourier de la solución de (1).

L Define la velocidad de fase. >Es puramente real? En caso de que no lo sea, >cuál es el significado de la parte imaginaria en relación a las propiedades de disipatividad del esquema semi-discreto (4.8.4)?

M Dibuja un diagrama para la parte real de la velocidad de fase y compáralo con el de la ecuación continua.

Comenta las semejanzas y diferencias de ambos casos.

N Suponiendo que los datos iniciales f y g son de banda limitada (su transformada de Fourier tiene soporte en un intervalo acotado) argumenta

que la velocidad de propagación en el esquema semi-discreto se asemeja cada vez más, a medida que $h \to 0$, a la de la ecuación continua.

O Calcula la velocidad de grupo, dibuja su diagrama y coméntalo.

35 Consideramos la ecuación del calor

$$(1) \qquad \begin{cases} u_t - \Delta u = 0 & \text{en } \mathbb{R}^N, \quad t > 0 \\ u(x,0) = f(x), & \text{en } \mathbb{R}^N. \end{cases}$$

a) Comprueba que la solución de (1) puede escribirse como

$$u = G(\cdot, t) * f(\cdot)$$

donde $*$ denota la convolución en la variable espacial y G en el núcleo de Gauss

$$G(x,t) = (4\pi t)^{-N/2} \exp\left(-\frac{|\ x\ |^2}{4t}\right).$$

Para ello, comprueba (directamente o usando la transformada de Fourier) que G es la solución fundamental de la ecuación del calor.

b) Demuestra que si $f \in L^1(\mathbb{R}^N)$ entonces u es de clase C^∞ en $t > 0$.

c) Demuestra que si $f \in L^1(\mathbb{R}^N)$ y $\displaystyle\int_{\mathbb{R}^N} f(x)dx = 0$, entonces $u(t) \to 0$ en $L^1(\mathbb{R}^N)$ cuando $t \to \infty$.

d) Obtén la expresión explícita de la solución de

$$\begin{cases} u_t - \dfrac{h}{2} u_{xx} + u_x = 0, & x \in \mathbb{R}, \quad t > 0 \\ u(x,0) = f(x), & x \in \mathbb{R}, \end{cases}$$

con $h > 0$.

36 Consideramos el problema de Dirichlet

$$(1) \qquad \begin{cases} u_t - \dfrac{h}{2} u_{xx} + u_x = 0, & x \in \mathbb{R}, \quad t > 0 \\ u(0,t) = 0, & t > 0 \\ u(x,0) = f(x), & x \in \mathbb{R}. \end{cases}$$

a) Demuestra que (1) genera un semigrupo de contracciones en $L^2(\mathbb{R})$.

b) Obtén la ley de disipación de la energía en $L^2(\mathbb{R})$.

c) Demuestra que si $f \in L^2(\mathbb{R}) \cap L^\infty(\mathbb{R})$ entonces las normas de u en $L^p(\mathbb{R})$ con $2 \leqslant p \leqslant \infty$ decrecen cuando $t \to \infty$.

37 Consideramos el siguiente problema de transporte en la semi-recta con condiciones de contorno de Dirichlet:

$$(1) \qquad \begin{cases} u_t + u_x = 0, & x > 0, \quad t > 0, \\ u(0,t) = 0, & t > 0 \\ u(x,0) = f(x), & x > 0. \end{cases}$$

a) Demuestra que (1) genera un semigrupo de contracciones en $L^2(\mathbb{R}^+)$.

b) Calcula explícitamente el valor de la solución.

c) Demuestra que el esquema semi-discreto centrado y regresivo habitual junto con la condición

$$u_0(t) = 0, \quad t > 0$$

en el nodo $x_0 = 0$, proporcionan una aproximación convergente de orden dos y uno respectivamente.

A partir de este momento buscamos calcular una aproximación de la solución exclusivamente en el intervalo espacial $0 < x < 1$.

d) Comprueba que la solución del problema continuo para $0 < x < 1$ y $t > 0$ arbitrario depende exclusivamente del valor del dato inicial $f(x)$ en el intervalo $0 < x < 1$.

e) Comprueba que el esquema regresivo puede resolverse considerando exclusivamente los índices $j \in \mathbb{N}$ tales que $x_j = jh \in (0,1)$ proporcionando una aproximación convergente de la solución en el intervalo $(0,1)$.

f) Demuestra que esto no es así en el esquema centrado.

A partir de este momento intentamos localizar el esquema centrado para los índices

$$j = 0, \cdots, N-1$$

donde $N \in \mathbb{N}$ es tal que $Nh = 1$.

Para ello, evidentemente, hemos de proporcionar el valor de u_N.

g) Analiza la convergencia del esquema centrado para los índices $j = 0, \cdots, N-1$, considerando para el nodo x_N cada una de las aproximaciones siguientes:

- $U_N(t) = f(Nh - t)$.

  Obsérvese que esta aproximación es exacta puesto que $f(Nh - t)$ es el valor de la solución continua de (1) durante un cierto intervalo de tiempo.

- $u_N(t) = u_{N-1}(t)$.

  Se trata también de una aproximación natural puesto que, en el caso continuo,

  $$u(x_N, t) = u(x_{N-1}, t) + hu_x(x_{N-1}, t) + O(h^2)$$

  de modo que

  $$u(x_N, t) = u(x_{N-1}, t) + O(h).$$

  Obsérvese sin embargo que esta aproximación es de orden uno y no de orden dos como es habitual en el esquema centrado.

- $u_N(t) = 2u_{N-1}(t) - u_{N-2}(t)$.

  Nuevamente, por el desarrollo de Taylor, se trata de una aproximación de orden dos. En este caso por tanto el esquema debería ser convergente de orden 2.

- $u_N(t) = u_{N-1}(t) - u'_{N-1}(t)$.

  En vista de la ecuación de transporte

  $$u_t = -u_x$$

  que la solución continua verifica y del argumento del caso anterior, se trata de una aproximación de orden dos

- $u_N(t) = 2u_{N-1}(t) - u_{N-2}(t)$.

  Nuevamente, por el desarrollo de Taylor, se trata de una aproximación de orden dos. En este caso por tanto el esquema debería ser convergente de orden 2.

Capítulo 5

El problema de Dirichlet en un dominio acotado

La idea de cómo el método de descomposición de dominios (MDD) se aplica en más de una dimensión espacial es la misma que en $1-D$ tanto en el marco continuo como en el discreto. En este último ámbito podemos considerar métodos de aproximación en diferentes finitas o cualquier otro método eficiente de aproximación numérica de EDP como puede ser el método de elementos finitos.

El estudio de la convergencia del método es técnicamente más elaborado en este caso. Es por ello que previamente recordamos algunos elementos básicos de la teoría variacional de las EDP (para más detalles el lector interesado podrá consultar [2], [8] y [36]).

Consideremos el problema de Dirichlet

$$\begin{cases} -\Delta u = f \quad \text{en} \quad \Omega \\ u\Big|_{\partial\Omega} = 0 \end{cases} \tag{5.0.1}$$

y su formulación variacional

$$\begin{cases} u \in H_0^1(\Omega) \\ \displaystyle\int_\Omega \nabla u \cdot \nabla \varphi dx = \int_\Omega f\varphi dx, \, \forall \varphi \in H_0^1(\Omega). \end{cases} \tag{5.0.2}$$

Suponiendo que Ω es un dominio acotado de $\mathbb{R}^n$ es fácil comprobar que (5.0.1) admite una única solución débil que satisface (5.0.2). Más aún, la solución puede obtenerse minimizando el funcional

$$J(w) = \frac{1}{2}\int_\Omega \mid \nabla w \mid^2 dx - \int_\Omega f w dx \tag{5.0.3}$$

en el espacio $H_0^1(\Omega)$.

El funcional J alcanza efectivamente su mínimo tal y como garantiza el método directo del Cálculo de Variaciones (MDCV)

5.1. Reducción al problema de valores de contorno no homogéneos

Pretendemos ahora reducir el problema al estudio de la ecuación con segundo miembro nulo y condiciones de contorno no homogéneas.

La manera aparentemente más sencilla de proceder es resolver el problema en $\mathbb{R}^n$:

$$-\Delta w = \tilde{f} \text{ en } \mathbb{R}^n \tag{5.1.1}$$

donde $\tilde{f}$ denota la extensión por cero de f fuera de Ω, i.e.

$$\tilde{f} = \begin{cases} f & \text{en} \quad \Omega \\ 0 & \text{en} \quad \Omega^c. \end{cases} \tag{5.1.2}$$

La formulación débil de (5.1.1) es

$$\int_{\mathbb{R}^n} \nabla w \cdot \nabla \varphi dx = \int_\Omega f\varphi dx, \, \forall \in C_c^\infty(\mathbb{R}^n). \tag{5.1.3}$$

Sin embargo la resolución de este problema conlleva dificultades técnicas innecesarias derivadas del hecho que la desigualdad de Poincaré no se cumple en $H^1(\mathbb{R}^n)$.

Es por eso que es más natural considerar el problema perturbado

$$-\Delta w + a(x)w = \tilde{f} \text{ en } \mathbb{R}^n \tag{5.1.4}$$

donde

$$a = 1_{\Omega^c}, \tag{5.1.5}$$

denota la función característica del conjunto Ω^c.

La formulación débil de este problema es:

$$\begin{cases} w \in H^1(\mathbb{R}^n) \\ \displaystyle\int_{\mathbb{R}^n} \nabla w \cdot \nabla \varphi dx + \int_{\Omega^c} w\varphi dx = \int_\Omega f\varphi dx, \, \forall \varphi \in H^1(\mathbb{R}^n). \end{cases} \tag{5.1.6}$$

Es fácil comprobar que este problema admite una única solución que se obtiene por minimización del funcional

$$J(w) = \frac{1}{2}\int_{\mathbb{R}^n} \mid \nabla w \mid^2 dx + \frac{1}{2}\int_{\Omega^c} \mid w \mid^2 dx - \int_\Omega fwdx \tag{5.1.7}$$

en el espacio de Hilbert $H^1(\mathbb{R}^n)$.

Este funcional alcanza su mínimo en un único punto de $H^1(\mathbb{R}^n)$. Para verlo basta aplicar el Método Directo del Cálculo de Variaciones (MDCV). La única dificultad al hacerlo es probar la coercividad del funcional J pero ésto se obtiene fácilmente a partir de la desigualdad de Poincaré:

$$\int_{\mathbb{R}^n} |\ \varphi\ |^2\ dx \leqslant C\left[\int_{\mathbb{R}^n} |\ \nabla\varphi\ |^2\ dx + \int_{\Omega^c} \varphi^2 dx\right]. \tag{5.1.8}$$

La prueba de esta desigualdad puede realizarse de las tres formas más habituales:

- Un argumento de contradicción basado en la compacidad de la inyección de $H^1(\mathbb{R}^n)$ en $L^2_{\text{loc}}(\mathbb{R}^n)$;
- Directamente de la desigualdad de Poincaré en dominios acotados mediante un argumento de truncatura del soporte de φ.
- La prueba directa por integración de la desigualdad de Poincaré en $H^1(\Omega)$, para un dominio acotado Ω o meramente acotado en una dirección.

Consideramos ahora

$$v = u - w \tag{5.1.9}$$

que satisface entonces

$$\begin{cases} -\Delta v = 0 & \text{en} \quad \Omega \\ v\Big|_{\partial\Omega} = w\Big|_{\partial\Omega}. \end{cases} \tag{5.1.10}$$

A partir de ahora trabajaremos sobre el problema (5.1.10).

Conviene observar que por los resultados clásicos de trazas que revisaremos más adelante, como $w \in H^1(\mathbb{R}^n)$ su traza $w\Big|_{\partial\Omega}$ pertenece a $H^{1/2}(\partial\Omega)$.

Cambiando la notación consideramos por tanto el problema

$$\begin{cases} -\Delta u = 0 & \text{en} \quad \Omega \\ u = g & \text{en} \quad \partial\Omega. \end{cases} \tag{5.1.11}$$

Antes de proceder a aplicar el MDD en esta ecuación analizamos en primer lugar algunas de sus propiedades más importantes.

5.2. El problema de contorno no homogéneo

Suponemos en primer lugar que la condición de contorno g en (5.1.11) pertenece a la clase $H^{1/2}(\partial\Omega)$.

La formulación variacional del problema (5.1.11) es la siguiente:

$$\begin{cases} u \in H^1(\Omega), & u\Big|_{\partial\Omega} = g \\ \displaystyle\int_\Omega \nabla u \cdot \nabla \varphi dx = 0, & \forall \varphi \in H_0^1(\Omega). \end{cases} \tag{5.2.1}$$

Teniendo en cuenta que, como $g \in H^{1/2}(\partial\Omega)$ existe $u^* \in H^1(\Omega)$ tal que $u^*\Big|_{\partial\Omega} = g$ el problema puede también reescribirse del modo siguiente

$$\begin{cases} u - u^* \in H_0^1(\Omega) \\ \displaystyle\int_\Omega \nabla(u-u^*) \cdot \nabla \varphi dx = -\int_\Omega \nabla u^* \cdot \nabla \varphi dx, \, \forall \varphi \in H_0^1(\Omega). \end{cases} \tag{5.2.2}$$

Las técnicas variacionales habituales permiten probar la existencia de una única solución $u - u^*$ de (5.2.2) o, lo que es lo mismo, una única solución u de (5.2.1).

Otro método posible para el estudio del problema no homogéneo (5.1.11) es el denominado método de transposición o dualidad (véase por ejemplo [2]). Recordamos brevemente sus principios fundamentales.

Consideramos el problema auxiliar:

$$\begin{cases} -\Delta \varphi = f & \text{en} \quad \Omega \\ \varphi = 0 & \text{en} \quad \partial\Omega. \end{cases} \tag{5.2.3}$$

Multiplicando (por ahora formalmente) en (5.1.11) por φ e integrando por partes en Ω obtenemos:

$$\int_\Omega f u dx = -\int_{\partial\Omega} g \frac{\partial \varphi}{\partial n} d\sigma. \tag{5.2.4}$$

Adoptamos (5.2.4) como definición de solución de (5.1.11) en el sentido de la transposición. Más concretamente, decimos que $u \in L^2(\Omega)$ es solución de (5.1.11) en el sentido de la transposición si (5.2.4) se satisface para todo $f \in L^2(\Omega)$.

Por el Teorema de representación de Riesz la existencia de una única solución en el sentido de la transposición se deduce a partir de la siguiente estimación de las soluciones de (5.2.3),

$$\Big|\Big|\frac{\partial \varphi}{\partial n}\Big|\Big|_{L^2(\partial\Omega)} \leqslant C \parallel f \parallel_{L^2(\Omega)}, \, \forall f \in L^2(\Omega) \tag{5.2.5}$$

en cuanto $g \in L^2(\partial\Omega)$.

En la siguiente sección probaremos la desigualdad (5.2.5). Deducimos por tanto que para cada $g \in L^2(\partial\Omega)$ existe una única solución $u \in L^2(\Omega)$ de (5.1.11) en el sentido de la transposición.

Pero la regularidad de la solución que este resultado proporciona no es óptima. En efecto, los resultados clásicos de regularidad elíptica garantizan que si Ω es un dominio de C^2, $\varphi \in H^2 \cap H_0^1(\Omega)$ y, más concretamente, existe una constante $C > 0$ tal que

$$\| \varphi \|_{H^2 \cap H_0^1(\Omega)} \leqslant C \| f \|_{L^2(\Omega)} \tag{5.2.6}$$

para todo $f \in L^2(\Omega)$.

Los resultados clásicos de trazas garantizan entonces que

$$\Big|\Big| \frac{\partial \varphi}{\partial n} \Big|\Big|_{H^{1/2}(\partial\Omega)} \leqslant C \| f \|_{L^2(\Omega)}, \, \forall f \in L^2(\Omega). \tag{5.2.7}$$

De este resultado se deduce que si $g \in H^{-1/2}(\partial\Omega)$ entonces existe una única solución $u \in L^2(\Omega)$. En este resultado se observa una ganancia de $1/2$ derivada que es óptima, de acuerdo con los resultados clásicos de trazas.

5.3. La desigualdad de Rellich

En esta sección probamos la desigualdad (5.2.5), debida a Rellich y que es válida en un contexto muy amplio de ecuaciones elípticas y de evolución.

Suponemos que el dominio Ω es de clase C^2, de modo que el campo normal $\nu = \nu(x)$ exterior unitario es de clase C^1 en el borde $\partial\Omega$. En estas circunstancias existe una extensión $q : \Omega \to \mathbb{R}^n$ de clase $C^1(\bar{\Omega})$ del campo normal de modo que

$$q\Big|_{\partial\Omega} = \nu. \tag{5.3.1}$$

Multiplicamos la ecuación (5.2.3) por $q \cdot \nabla\varphi$. Obtenemos

$$\left| -\int_\Omega \Delta\varphi q \cdot \nabla\varphi dx \right| = \left| \int_\Omega f \, q \cdot \nabla\varphi dx \right| \leqslant C \| f \|_{L^2(\Omega)} \| \nabla\varphi \|_{L^2(\Omega)} \leqslant C \| f \|^2_{L^2(\Omega)} \, . \tag{5.3.2}$$

Por otra parte,

$$-\int_\Omega \Delta\varphi q \cdot \nabla\varphi dx = -\int_{\partial\Omega} \frac{\partial\varphi}{\partial n} q \cdot \nabla\varphi d\sigma + \int_\Omega \partial_j\varphi \partial_j (q \cdot \nabla\varphi) dx. \tag{5.3.3}$$

Además

$$\begin{aligned} \int_\Omega \partial_j\varphi \partial_j (q \cdot \nabla\varphi) dx &= \int_\Omega \partial_j\varphi \left(q_i \partial^2_{ij}\varphi + \partial_j q_i \partial_i\varphi \right) dx \\ &= \int_\Omega \partial_j\varphi \partial_j q_i \partial_i\varphi dx + \int_\Omega q \cdot \nabla \left[\frac{| \nabla\varphi |^2}{2} \right] dx. \end{aligned} \tag{5.3.4}$$

Obviamente,

$$\left|\int_\Omega \partial_j\varphi\partial_j q_i\partial_i\varphi dx\right| \leqslant C(q)\int_\Omega \mid \nabla\varphi \mid^2 dx \leqslant C(q)\int_\Omega f^2 dx. \tag{5.3.5}$$

Por otra parte,

$$\int_\Omega q\cdot\nabla\left[\frac{\mid \nabla\varphi \mid^2}{2}\right]dx = -\int_\Omega \operatorname{div} q\frac{\mid \nabla\varphi \mid^2}{2}dx \ + \int_{\partial\Omega}\frac{q\cdot\nu}{2}\mid \nabla\varphi \mid^2 d\sigma. \tag{5.3.6}$$

Nuevamente

$$\left|\int_\Omega \operatorname{div} q\frac{\mid \nabla\varphi \mid^2}{2}dx\right| \leqslant C(q)\int_\Omega \mid \nabla\varphi \mid^2 dx \leqslant C(q)\parallel f\parallel^2_{L^2(\Omega)}. \tag{5.3.7}$$

Combinando las identidades y estimaciones anteriores deducimos que

$$\begin{aligned}&\left|\int_{\partial\Omega}\frac{q\cdot\nu}{2}\mid \nabla\varphi \mid^2 d\sigma - \int_{\partial\Omega}\frac{\partial\varphi}{\partial n}q\cdot\nabla\varphi d\sigma\right| \\ &\quad = \int_{\partial\Omega}\frac{q\cdot\nu}{2}\left|\frac{\partial\varphi}{\partial\nu}\right|^2 d\sigma = \frac{1}{2}\int_{\partial\Omega}\left|\frac{\partial\varphi}{\partial\nu}\right|^2 d\sigma \leqslant C\parallel f\parallel^2_{L^2(\Omega)}.\end{aligned} \tag{5.3.8}$$

En estas identidades hemos usado que, como $\varphi = 0$ en $\partial\Omega$,

$$\begin{cases} \nabla\varphi = \dfrac{\partial\varphi}{\partial n}\nu & \text{en} \quad \partial\Omega \\ q\cdot\nu = 1 & \text{en} \quad \partial\Omega. \end{cases} \tag{5.3.9}$$

5.4. Un resultado de trazas

El resultado óptimo de trazas garantiza que si Ω es un dominio de clase C^1, las trazas de funciones de clase $H^1(\Omega)$ pertenecen al espacio $H^{1/2}(\partial\Omega)$.

En esta sección recordamos brevemente los ingredientes principales de la demostración de este resultado.

En primer lugar señalamos que para definir el espacio $H^{1/2}(\partial\Omega)$, usando cartas locales (en este punto usamos que el dominio es de clase C^1), basta hacerlo en el espacio euclídeo $\mathbb{R}^n$ ($\mathbb{R}^{n-1}$ en el caso en que Ω es un dominio de $\mathbb{R}^n$ pues entonces $\partial\Omega$ es una hipersuperficie de dimensión $n-1$).

En el caso del espacio euclídeo, el modo más sencillo de definir el espacio $H^{1/2}(\mathbb{R}^n)$ y, de manera más general, el espacio $H^s(\mathbb{R}^n)$ con $s > 0$ arbitrario (o, incluso, para cualquier $s \in \mathbb{R}$) es mediante la transformada de Fourier. Recordemos que $u \in L^2(\mathbb{R}^n)$, si y sólo si $\hat{u} \in L^2(\mathbb{R}^n)$. Del mismo modo, $\nabla u \in L^2(\mathbb{R})$ si y sólo si $\mid \xi \mid \hat{u}(\xi) \in L^2(\mathbb{R}^n)$.

El espacio $H^s(\mathbb{R}^n)$ con $s > 0$ se define entonces como el subespacio de $L^2(\mathbb{R}^n)$ de las funciones u tales que

$$\int_{\mathbb{R}^n} \left(1+ \mid \xi \mid^{2s}\right) \mid \hat{u}(\xi) \mid^2 d\xi. \tag{5.4.1}$$

La cantidad (5.4.1) constituye en realidad el cuadrado de la norma en $H^s(\mathbb{R}^n)$. Se trata de espacios de Hilbert.

Con esta definición, cuando s recorre el intervalo $[0,\, 1]$, el espacio $H^s(\mathbb{R}^n)$ decrece desde $L^2(\mathbb{R}^n)$ hasta $H^1(\mathbb{R}^n)$.

Se puede también definir una norma equivalente en el espacio físico obteniéndose que

$$H^s(\mathbb{R}^n) = \left\{ u \in L^2(\mathbb{R}^n) : \frac{\mid u(x) - u(y) \mid}{\mid x - y \mid^{s+\frac{n}{2}}} \in L^2(\mathbb{R}^n \times \mathbb{R}^n) \right\} \tag{5.4.2}$$

(véase [2]).

Pero en esta sección utilizaremos sólo la definición de $H^s(\mathbb{R}^n)$ en términos de la transformada de Fourier.

Basta que probemos que la traza de cualquier función $H^1\left(\mathbb{R}^n_+\right)$ pertenece a $H^{1/2}(\mathbb{R}^{n-1})$. Esto, mediante el uso de cartas locales, conduce al mismo resultado en el caso de un abierto acotado Ω de clase C^1 arbitrario.

Para ello, dada $u = u(x',\, x_n)$ consideramos su traza $u(x',\, 0)$. Debemos probar que

$$\int_{\mathbb{R}^{n-1}} \mid \xi' \mid \left|\hat{u}(\xi,\, 0)\right|^2 d\xi < \infty \tag{5.4.3}$$

donde $\hat{\cdot}$ denota la transformada de Fourier en las variables x', siendo ξ' la variable dual correspondiente.

Como la propiedad de traza es de naturaleza local podemos suponer que u es de soporte compacto de modo que $u = 0$ cuando $x_n \geqslant 1$.

Entonces

$$\left|\hat{u}(\xi',\, 0)\right|^2 = -\int_0^1 \partial_n \left(\left|\hat{u}(\xi',\, x_n)\right|^2\right) dx_n \;\; = -2\int_0^1 \hat{u}(\xi',\, x_n)\partial_{x_n}\hat{u}(\xi',\, x_n)dx_n.$$

Por tanto

$$\mid \xi' \mid \left|\hat{u}(\xi',\, 0)\right|^2 \leqslant \int_0^1 \left[\mid \xi' \mid^2 \left|\hat{u}(\xi',\, x_n)\right|^2 + \left|\partial_n \hat{u}(\xi',\, x_n)\right|^2\right] dx_n \tag{5.4.4}$$

y entonces

$$\int_{\mathbb{R}^{n-1}} \mid \xi \mid \left|\hat{u}(\xi,\, 0)\right|^2 d\xi \leqslant \int_{\mathbb{R}^{n-1}} \int_0^1 |\xi'|^2 \left|\hat{u}(\xi', x_n)\right|^2 dx_n d\xi'$$

$$+\int_{\mathbb{R}^{n-1}}\int_0^1 |\partial_n\hat{u}(\xi,\, x_n)|^2\, dx_n d\xi = \int_{\mathbb{R}^n}\left[\mid \partial_{x'}u \mid^2 + \mid \partial_n u \mid^2\right] dx' dx_n = \int_{\mathbb{R}^n} \mid \nabla u \mid^2\, dx. \tag{5.4.5}$$

El resultado recíproco es también cierto. Es decir, toda función de $H^{1/2}(\partial\Omega)$ es la traza de una función de $H^1(\Omega)$. Como consecuencia de este resultado cuya prueba vamos a esbozar se deduce que la solución de

$$\begin{cases} -\Delta u = 0 & \text{en} \quad \Omega \\ u = g & \text{en} \quad \partial\Omega \end{cases} \tag{5.4.6}$$

con $g \in H^{1/2}(\partial\Omega)$ pertenece al espacio de Sobolev $H^1(\Omega)$.

Cuando Ω es de clase C^1, para probar este resultado, nuevamente, gracias a las cartas locales, basta con probarlo en el caso de $\mathbb{R}^{n-1}$ y $\mathbb{R}^n_+$.

Definimos entonces la extensión de $g = g(x')$ dada en $\mathbb{R}^{n-1}$ resolviendo el problema

$$\begin{cases} -\Delta u = 0 & \mathbb{R}^n_+ \\ u = g & \mathbb{R}^{n-1}. \end{cases} \tag{5.4.7}$$

Aplicando la transformada de Fourier en x' vemos que el problema es equivalente a la familia de ecuaciones diferenciales lineales de segundo orden parametrizadas por $\xi' \in \mathbb{R}^{n-1}$:

$$\begin{cases} -\partial_n^2\hat{u} + \mid \xi' \mid^2 \hat{u} = 0 \\ \hat{u}(\xi',\, 0) = \hat{g}(\xi') \end{cases} \tag{5.4.8}$$

cuya solución viene dada por

$$\hat{u}(\xi',\, x_n) = \hat{g}(\xi')e^{-|\xi'|x_n}. \tag{5.4.9}$$

Basta entonces comprobar que si

$$\int_{\mathbb{R}^{n-1}} \mid \hat{g}(\xi') \mid^2 (1 + \mid \xi \mid) < \infty, \tag{5.4.10}$$

lo cual equivale a que $g \in H^{1/2}(\mathbb{R}^{n-1})$, entonces la solución $\hat{u} = \hat{u}(\xi',\, x_n)$ obtenida en (5.4.9) satisface

$$\int_{\mathbb{R}^n}\left[\mid \xi' \mid^2 \mid \hat{u}(\xi',\, x_n) \mid^2 + |\partial_n\hat{u}(\xi',\, x_n)|^2\right] d\xi' dx_n < \infty. \tag{5.4.11}$$

Este resultado es fácil de probar.

Verifiquemos por ejemplo la finitud de la primera integral en (5.4.11), puesto que la otra puede estimarse del mismo modo. En virtud de la expresión explícita de la solución en (5.4.9) tenemos

$$\int_{\mathbb{R}^n} \mid \xi' \mid^2 \mid \hat{u}(\xi',\, x_n) \mid^2 d\xi' dx_n = \int_{\mathbb{R}^n} \mid \hat{g}(\xi') \mid^2 \mid \xi' \mid^2 e^{-2|\xi'|x_n} d\xi' dx_n$$

$$= \int_{\mathbb{R}^{n-1}} |\, \hat{g}(\xi')\, |^2 |\, \xi'\, |^2 \int_{\mathbb{R}} e^{-2|\xi'|x_n} dx_n = \int_{\mathbb{R}^n} \frac{|\, \hat{g}(\xi')\, |^2 |\, \xi'\, |}{2} d\xi' < \infty \quad (5.4.12)$$

por la hipótesis $g \in H^{1/2}(\mathbb{R}^{n-1})$.

5.5. Principio del máximo

En esta subsección recordamos una versión básica del

principio del máximo que permite comparar soluciones de ecuaciones elípticas con datos distintos y que juega un papel decisivo en la prueba de la convergencia del MDD, además de aplicarse en muchos otros contextos. La versión que aquí reproducimos ha sido extraida de la sección IX.7 del libro de Brézis [2].

Teorema. *(Principio del máximo para el problema de Dirichlet). Sean* $f \in L^2(\Omega)$ *y* $u \in H^1(\Omega) \cap C(\bar{\Omega})$ *tales que*

$$\int_\Omega \nabla u \cdot \nabla \varphi dx + \int_\Omega u\varphi u dx = \int_\Omega f\varphi,\ \forall \varphi \in H_0^1(\Omega).$$

Entonces

$$\text{mín}\left[\inf_\Gamma u,\ \inf_\Omega f\right] \leqslant u(x) \leqslant \text{máx}\left[\sup_\Gamma u,\ \sup_\Omega f\right],\ \forall x \in \Omega. \quad (5.5.1)$$

Demostración.

Utilizamos el método de truncatura de Stampacchia. Para ello introducimos una función $G = G(s) \in C^1(\mathbb{R})$ tal que:

- $|\, G'(s)\, | \leqslant M,\ \forall s \in \mathbb{R}$.
- G es estríctamente creciente en $(0,\ \infty)$.
- $G(s) = 0$ en $(-\infty,\ 0)$.

Sea

$$k = \text{máx}\left\{\sup_\Gamma u,\ \sup_\Omega f\right\}$$

que suponemos finito.

Pretendemos probar que $u \leqslant k$ p.c.t. $x \in \Omega$. La otra desigualdad se prueba de manera análoga.

La prueba consiste esencialmente en utilizar $v = G(u-k)$ como función test.

Distinguimos dos casos:

a) $\mid \Omega \mid < \infty$.

En este caso $v = G(u-k) \in H_0^1(\Omega)$. En efecto, v es la composición de u con una función globalmente Lipschitz y, por tanto, como $\mid \Omega \mid < \infty$, $v \in H^1(\Omega)$. Por otra parte $v \in H_0^1(\Omega)$ puesto que v es continua y se anula en el borde.

De la formulación variacional deducimos que

$$\int_\Omega \mid \nabla u \mid^2 G'(u-k) + \int_\Omega uG(u-k) = \int_\Omega fG(u-k).$$

Es decir

$$\int_\Omega \mid \nabla u \mid^2 G'(u-k) + \int_\Omega (u-k)G(u-k) = \int_\Omega (f-k)G(u-k).$$

Deducimos inmediatamente que

$$\int_\Omega (u-k)G(u-k)dx \leqslant 0,$$

y, como $sG(s) \geqslant 0$ para todo $x \in \mathbb{R}$, que $(u-k)G(u-k) = 0$ p. c. t. $x \in \Omega$. Por consiguiente $u \leqslant k$ para todo $x \in \Omega$.

En esta argumentación hemos utilizado que

$$\int_\Omega (f-k)G(u-k) \leqslant 0; \int_\Omega \mid \nabla u \mid^2 G'(u-k)dx \geqslant 0$$

lo cual es cierto obviamente por la elección de f y las propiedades de la función G.

b) El mismo argumento puede adaptarse sin mucha dificultad al caso en que $\mid \Omega \mid = \infty$. Para ello es suficiente elegir como función test $v = G(u-k')$ con $k' > k$.

■

El Teorema anterior proporciona el principio del máximo (PM) para el operador $-\Delta + I$. Pero se verifica en un contexto mucho más general de problemas elípticos de segundo orden (véase la Proposición IX.29 de [2]).

El PM se aplica en particular a las soluciones débiles del problema de Dirichlet asociado al operador de Laplace caracterizadas por la condición:

$$\begin{cases} u \in H^1(\Omega); \\ \displaystyle\int_\Omega \nabla u \cdot \nabla \varphi dx = \int_\Omega f\varphi dx, \, \forall \varphi \in H_0^1(\Omega). \end{cases} \tag{5.5.2}$$

La prueba es semejante a la anterior sólo que en este caso obtenemos, si $f \leq 0$,

$$\int_\Omega \mid \nabla u \mid^2 G'(u-k)dx \leqslant 0.$$

Si definimos la función

$$H(s) = \int_0^s [G'(\sigma - k)]^{1/2}\, ds$$

esta condición puede escribirse simplemente como

$$\int_\Omega \mid \nabla(H(u)) \mid^2 dx \leqslant 0.$$

Por tanto, la función $H(u)$ ha de ser constante en Ω. Ahora bien, habida cuenta de la definición de k tenemos que $u \leqslant k$ en Γ y por lo tanto $H(u) = 0$ en Γ. Deducimos por tanto que $H(u) \equiv 0$ p.c.t. $x \in \Omega$ lo cual implica que $u \leqslant k$ p.c.t. $x \in \Omega$. Estas dos últimas condiciones son efectivamente equivalentes. En efecto, por las propiedades de G tenemos que $G'(\sigma - k) = 0$ cuando $\sigma \leqslant k$ y que $G'(\sigma - k) > 0$ cuando $\sigma > k$. De estas propiedades se deduce que $H(s) = 0$ cuando $s \leqslant k$ y que $H(s) > 0$ cuando $s > k$.

De este modo, para el problema (5.5.2) se deduce que, cuando $f \leq 0$, entonces, el máximo de u se alcanza en la frontera del dominio Ω.

Capítulo 6

Diferencias y volúmenes finitos

6.1. Diferencias finitas para coeficientes variables $1-d$

Sin duda el problema más ejemplar del Análisis Numérico de EDP's es el de resolver el problema de Dirichlet $1-d$:

$$-u_{xx} = f,\ 0 < x < 1;\ u(0) = u(1) = 0. \tag{6.1.1}$$

El método más elemental para hacerlo es el basado en diferencias finitas en un mallado regular. Esto consiste esencialmente en introducir una partición regular del intervalo $(0,\ 1)$, $\{x_j\}_{j=0}^{N+1}$, donde $x_j = jh$, con $h > 0$, y buscar, para cada $h > 0$ y $j \in \{0, \cdots,\ N+1\}$ una aproximación u_j de $u(x_j)$. Para ello utilizamos el esquema de tres puntos

$$\frac{2u_j - u_{j+1} - u_{j-1}}{h^2} = f_j,\ j = 1, \cdots,\ N;\ u_0 = u_{N+1} = 0. \tag{6.1.2}$$

Para cada $h > 0$ (6.1.2) es un sistema lineal de N ecuaciones con N incógnitas que puede escribirse de la forma

$$AU = F \tag{6.1.3}$$

con $U = (u_j, \cdots, u_N)^\perp$, $F = (f_1, \cdots,\ f_N)^\perp$.

La matriz A del sistema es simétrica, definida positiva de modo que, para cada $h > 0$, (6.1.3) admite una única solución.

Con el objeto de que el problema anterior esté formulado sin ninguna ambigüedad es imprescindible que $h > 0$ sea de la forma $h = 1/k$ donde $k \in \mathbb{N}$. Asimismo hemos de precisar el valor f_j de la aproximación de $f(x_j)$ elegida en (6.1.2). Cuando f es continua basta tomar $f(x_j)$ y cuando no lo es puede sustituirse por una media, i.e.

$$f_j = \frac{1}{h}\int_{x_j-h/2}^{x_j+h/2} f(s)ds. \tag{6.1.4}$$

Este esquema es de orden 2. Para ello basta observar que si u es una solución regular de (6.1.1) entonces, la diferencia entre el miembro de la izquierda de (6.1.1), $-u_{xx}$, y el de (6.1.2), $\big[2u_j - u_{j+1} - u_{j-1}\big]/h^2$, es del orden de $O(h^2)$.

Esta ecuación (elíptica en la clasificación clásica de las EDP's) surge al analizar las soluciones estacionarias de algunas de las ecuaciones de evolución más importantes como son las ecuaciones de ondas, del calor o de Schrödinger.

El hecho de que los coeficientes de la ecuación (6.1.1) sean constantes refleja que el medio en el que ésta represente la cantidad física estudiada sea *homogéneo*. Sin embargo en la práctica, con frecuencias, hemos de hacer frente a medios *heterogéneos*. La versión correspondiente de (6.1.1) sería entonces:

$$-\big(a(x)u_x\big)_x = f,\ 0 < x < 1,\ u(0) = u(1) = 0. \tag{6.1.5}$$

El coeficiente variable $a = a(x)$ se supone acotado y medible, i.e. $a \in L^\infty(0,\,1)$. Suponemos además que existen constantes $0 < \alpha < \beta < \infty$ tales que

$$\alpha \le a(x) \le \beta,\ \text{p.c.t. } x \in (0,\,1). \tag{6.1.6}$$

En estas circunstancias el problema (6.1.5) admite una única solución. De manera más precisa, mediante la aplicación del Lema de Lax-Milgram, se deduce que para todo $f \in H^{-1}(0,\,1)$ existe una única solución $u \in H_0^1(0,\,1)$ de (6.1.5). En este caso que nos ocupa la solución puede calcularse de manera explícita. Tenemos que

$$u(x) = -\int_0^x \frac{1}{a(s)} f(\tau)d\tau ds + C_1 b(x) + C_2, \tag{6.1.7}$$

donde

$$b(x) = \int_0^x \frac{1}{a(s)} ds, \tag{6.1.8}$$

y las constantes C_1, C_2 están determinadas de manera única para que las condiciones de contorno se satisfagan. Es decir,

$$C_2 = 0 \tag{6.1.9}$$

y

$$C_1 = \left[\int_0^1 \frac{1}{a(s)} \int_0^s f(\tau) d\tau ds\right] \Big/ \int_0^1 \frac{1}{a(s)} ds. \tag{6.1.10}$$

Cuando $a = a(x)$ es variable (6.1.5) es el modelo para la deformación de una cuerda elástica heterogénea, por ejemplo.

Analicemos ahora los esquemas en diferencias finitas de aproximación de (6.1.5). Para ello introducimos los operadores de diferencias finitas discretos ∂_+ y ∂_- donde

$$\big(\partial_+ U\big)_j = \frac{u_{j+1} - u_j}{h};\ \big(\partial_- U\big)_j = \frac{u_j - u_{j-1}}{h}. \tag{6.1.11}$$

Ambos operadores son una aproximación del operador de derivación continuo $d_x = d \cdot /dx$. Se trata obviamente de aproximaciones de primer orden puesto que, cuando $u_j = u(x_j)$, siendo $u = u(x)$ una función de clase C^2, tenemos

$$\frac{du}{dx}(x_j) = \big(\partial_+ U\big)_j + O(h) = \big(\partial_- U\big)_j + O(h). \tag{6.1.12}$$

Con el objeto de que el esquema numérico de aproximación de (6.1.5) sea simétrico, es conveniente introducir la aproximación

$$\begin{cases} -\dfrac{1}{2}\left[\partial_+\big(a\partial_-\big) + \partial_-\big(a\partial_+\big)\right] u = f_j,\ j = 1, \cdots, N \\ u_0 = u_{N+1} = 0. \end{cases} \tag{6.1.13}$$

Cuando a es constante el esquema (6.1.13) es una generalización de (6.1.2). Este esquema, componente a componente, puede escribirse del siguiente modo:

$$\Big(-\big(a_{j-1} + a_j\big)u_{j-1} + \big(a_{j-1} + 2a_j + a_{j+1}\big)u_j - \big(a_j + a_{j+1}\big)u_{j+1}\Big) \Big/ 2h^2 = f_j, \tag{6.1.14}$$

para los nodos interiores $j = 1, \cdots, N$.

Dejamos como ejercicio para el lector comprobar cual es el orden del esquema cuando el coeficiente $a = a(x)$ es regular.

Pero muchas veces las ecuaciones heterogéneas y los coeficientes variables se presentan en situaciones en que estos no son regulares, ni siquiera continuos.

Este es el caso, por ejemplo, cuando a es constante a trozos en el cual se representa la presencia de dos materiales homogéneos diferentes con constantes materiales distintas.

Consideremos por ejemplo el caso de un coeficiente de la forma

$$a(x) = \begin{cases} \varepsilon, & 0 < x \leq x^* \\ 1, & x^* < x < 1, \end{cases} \tag{6.1.15}$$

con $0 < x^* < 1$. Con el objeto de hacer los cálculos más explícitos suponemos que el segundo miembro f se anula pero tomamos una condición de contorno no homogénea en $x = 1$, i.e. consideramos el problema

$$\begin{cases} -\big(a(x)u_x\big)_x = 0, & 0 < x < 1; \\ u(0) = 0, \qquad u(1) = 1, \end{cases} \tag{6.1.16}$$

donde $a = a(x)$ es como en (6.1.15). La solución explícita de (6.1.16) es conocida:

$$u(x) = \begin{cases} \alpha x, & 0 < x \leq x^*, \\ \varepsilon\alpha x + 1 - \varepsilon\alpha, & x^* \leq x < 1, \end{cases} \tag{6.1.17}$$

con

$$\alpha = \frac{1}{x^* - \varepsilon x^* + \varepsilon}. \tag{6.1.18}$$

Supongamos ahora que el punto de discontinuidad x^* del coeficiente es tal que $x_k < x^* \leq x_{k+1}$. La solución obtenida en este caso es

$$u_j = \alpha_j,\ 0 \leq j \leq k;\ u_j = \beta_j - \beta(N+1) + 1,\ k+1 \leq j \leq N+1 \tag{6.1.19}$$

con

$$\beta = \varepsilon\alpha;\ \alpha = \left(\varepsilon\frac{1-\varepsilon}{1+\varepsilon} + \varepsilon(N+1-k) + k\right)^{-1}.$$

Por tanto

$$u_k = \frac{x_k}{\varepsilon h(1-\varepsilon)/(1+\varepsilon) + (1-\varepsilon)x_k + \varepsilon}.$$

Cuando $x^* = x_{k+1}$ la solución exacta en el punto $x = x_k$ es,

$$u(x_k) = \frac{x_k}{(1-\varepsilon)x_{k+1} + \varepsilon}.$$

El error satisface por tanto

$$u_k - u(x_k) = O\left(\varepsilon\frac{1-\varepsilon}{1+\varepsilon}h\right).$$

Cuando $x^* = x_k + h/2$ se comprueba que el error es más bien

$$u_k - u(x_k) = O\left(\frac{(1-\varepsilon)^2}{\varepsilon(1+\varepsilon)}h\right).$$

Vemos entonces que, cuando $\varepsilon \neq 1$, el error es $O(h)$, lo cual contrasta con el hecho de que el método sea de orden 2 cuando a es regular.

6.2. Volúmenes finitos

En esta sección vamos a desarrollar un método de volúmenes finitos que, con respecto a las diferencias finitas, tiene la ventaja de que el método es de orden $O(h^2)$ tanto para coeficientes regulares como discontinuos.

El método de volúmenes finitos está basado en la formulación variacional de problema (6.1.5)

$$\begin{cases} u \in H_0^1(0,\,1) \\ \displaystyle\int_0^1 a(x)u_x\varphi_x dx = \int_0^1 f\varphi dx, \quad \forall \varphi \in H_0^1(0,\,1), \end{cases} \tag{6.2.1}$$

en la cual no se distingue si a es regular o no, siempre y cuando se satisfaga la condición (6.1.6) de elipticidad. En efecto, bajo la condición (6.1.6) la existencia de una única solución de (6.2.1) está garantizada por el lema de Lax-Milgram.

Denotemos mediante Ω el dominio $(0,\,1)$ en el que la ecuación está formulada y mediante Ω_j los intervalos centrados en los puntos del mallado $\Omega_j = \big(x_j - h/2,\, x_j + h/2\big)$, $j = 1, \cdots, N$. Denotamos por $\psi^j = \psi^j(x)$ la función característica del intervalo Ω_j. Mediante a_j denotamos una aproximación de a en el intervalo Ω_j que puede ser tomada como la media

$$a_j = \frac{1}{h}\int_{\Omega_j} a dx. \tag{6.2.2}$$

Una manera natural de aproximar (6.2.1) es utilizar la formulación variacional

$$\begin{cases} u \in H_0^1(0,\,1) \\ \displaystyle\sum_{j=1}^{N} \int_{\Omega_j} a_j u_x v_x dx = \int_0^1 f v dx,\ \forall v \in H_0^1(0,\,1) \end{cases} \tag{6.2.3}$$

en la cual hemos sustituido el valor del coeficiente $a = a(x)$ por su aproximación constante a trozos. Pero, obviamente no hay ninguna razón de que la solución de (6.2.3) sea constante o lineal a trozos.

Con el objeto de introducir una tal aproximación utilizamos que

$$-\int_{\Omega_j} \big(au_x\big)_x dx = -au_x\Big|_{x_j-\frac{h}{2}}^{x_j+\frac{h}{2}}.$$

Necesitamos ahora aproximar $au_x\big(x_j+h/2\big)$. En la medida en que au_x es regular (pues es una primitiva de f), si a tiene una discontinuidad justo en $x_j + \frac{h}{2}$, au_x no puede ser regular en ese punto. Por tanto aproximar u_x no es una buena idea. Es más bien conveniente escribir

$$u\Big|_{x_j}^{x_{j+1}} = \int_{x_j}^{x_{j+1}} u_x dx = \int_{x_j}^{x_{j+1}} \frac{1}{a} au_x dx \sim \big(au_x\big)_{j+1/2} \int_{x_j}^{x_{j+1}} \frac{1}{a} dx.$$

De la aproximación constante a trozos de a deducimos que

$$\int_{x_j}^{x_{j+1}} \frac{1}{a} dx = \frac{h}{w_j}$$

con

$$w_j \equiv 2a_j a_{j+1} / \left(a_j + a_{j+1}\right), \tag{6.2.4}$$

es decir la media armónica de a_j y a_{j+1}.

Obtenemos por tanto

$$\left(au_x\right)_{j+1/2} \sim w_j \frac{(\varphi_{j+1} - \varphi_j)}{h},$$

y, por tanto, la siguiente discretización:

$$w_{j-1}\frac{(u_j - u_{j-1})}{h} - w_j \frac{(u_{j+1} - u_j)}{h} = hf_j,\ j = 1, \cdots, N, \tag{6.2.5}$$

con

$$f_j = \frac{1}{h}\int_{\Omega_j} f dx. \tag{6.2.6}$$

Se trata de un esquema de aproximación que difiere del de diferencias finitas (6.1.14).

Puede sin embargo verse que ambos son muy próximos cuando a es regular. En efecto, cuando a es regular

$$w_j \sim \frac{a_j + a_{j+1}}{2} \tag{6.2.7}$$

y, en caso de realizar esta sustitución, en (6.2.5) recuperaríamos el esquema (6.1.14).

Apliquemos ahora este método, denominado de *volúmenes finitos*, al caso en que a es discontinuo discutido en la sección anterior. Supongamos que $x^* = x_k + h/2$. Entonces

$$w_j = \varepsilon,\ 1 \le j < k;\ w_k = \frac{2\varepsilon}{1+\varepsilon};\ w_j = 1,\ k < j \le N.$$

Obtenemos en este caso (6.1.19) con

$$\beta = \alpha\varepsilon,\ \alpha = h/\left(x^* - \varepsilon x^* + \varepsilon\right).$$

Comparando con la solución exacta, vemos que el error en los puntos del mallado es nulo. En circunstancias más generales sería $O(h^2)$. El método de volúmenes finitos es por tanto más preciso que el de diferencias finitas.

El método puede también adaptarse fácilmente al caso en que el coeficiente $a = a(x)$ es discontinuo en un punto del mallado x_j, i.e. $x^* = x_j$.

En este caso tendríamos

$$-\int_{\Omega_j}\left(au_x\right)_x dx = -au_x\Big|_{x_j-h/2}^{x_j+h/2} + \lim_{x\uparrow x_j} au_x - \lim_{x\downarrow x_j} au_x.$$

Ahora bien, como au_x es continuo, los dos últimos términos se cancelan. Aproximando ahora u_x por diferencias finitas obtendríamos

$$-\int_{\Omega_j}\left(au_x\right)dx \sim -a_{j+1/2}\left(\frac{u_{j+1}-u_j}{h}\right) + a_{j-1/2}\left(\frac{u_j-u_{j-1}}{h}\right).$$

Esto da lugar al esquema

$$\left[-a_{j-1/2}u_{j-1} + \left(a_{j-1/2}+a_{j+1/2}\right)u_j - a_{j+1/2}u_{j+1/2}\right] = hf_j.$$

Se trata de un esquema semejante al clásico de diferencias finitas pero es mucho más eficaz cuando a es discontinuo puesto que el error de este nuevo esquema es del orden de $O(h^2)$, incluso para coeficientes discontinuos.

6.3. Diferencias finitas para coeficientes variables: varias dimensiones espaciales

Siendo e_α el $\alpha-$ésimo vector de la base canónica de $\mathbb{R}^n$ definimos los operadores de diferencias

$$\partial_{\alpha,+} = \left(\varphi_{j+e_\alpha} - \varphi_j\right)/h;\ \partial_{\alpha_j} = \left(\varphi_j - \varphi_{j-e_\alpha}\right)/h. \tag{6.3.1}$$

En esta ocasión $j = \left(j_1, \cdots, j_n\right)$ es un multi-índice que denota los puntos de un mallado regular de $\mathbb{R}^n$ de nodos

$$x_j = \left(x_{j_1}, \cdots, x_{j_n}\right),$$

donde $x_j = jh$.

Consideramos ahora la ecuación elíptica

$$\sum_{j=1}^{n}\sum_{i=1}^{n}\partial_j\left(a_{ij}\partial_j u\right) = f \tag{6.3.2}$$

con coeficientes variables $a_{ij} = a_{ij}(x)$, medibles y acotados que satisfacen además la condición de elipticidad

$$\beta \mid \xi \mid^2 \geq a_{ij}(x)\xi_i\xi_j \geq \alpha \mid \xi \mid^2,\ \forall \xi \in \mathbb{R}^n, \tag{6.3.3}$$

con $0 < \alpha < \beta < \infty$.

Aquí y en lo sucesivo hemos adoptado la convención de sumación de índices repetidos.

Con el objeto de simplificar la presentación suponemos además que la matriz de coeficientes es simétrica de modo que

$$a_{ij} \equiv a_{ji},\ \forall i,\ j = 1, \cdots, n.$$

Inspirándonos en el caso $1-d$ podemos ahora introducir el siguiente esquema de discretización

$$-\frac{1}{2}\left[\partial_{\beta,+}\left(a_{\alpha\beta}\partial_{\alpha,-}\right) + \partial_{\beta,-}\left(a_{\alpha\beta}\partial_{\alpha,+}\right)\right] u_j = f_j. \tag{6.3.4}$$

Se trata de un esquema en diferencias de orden 2 para coeficientes variables.

Obviamente, en la práctica, la ecuación (6.3.2) normalmente se estudia en un dominio Ω de $\mathbb{R}^n$ y por tanto la ecuación (6.3.4) habrá de satisfacerse en el conjunto de multi-índices correspondiente. La ecuación (6.3.2) y su versión discreta (6.3.4) han normalmente de ser también completadas con condiciones de contorno.

Conviene observar que en la discretización (6.3.4) los puntos del mallado en los que se apoya el esquema en cada nodo no es simétrico. Por ejemplo, en dos dimensiones espaciales, la aproximación (6.3.4) en el punto de coordenadas $\left(j_1 h,\, j_2 h\right)$ sólo involucra los siguientes puntos, además de sí mismo: $((j_1+1)h,\, (j_2-1)h)$, $(j_1 h, (j_2-1)h)$, $(j_1 h,\, (j_2+1)h)$ $((j_1-1)h)$, $(j_2+1)h)$, $((j_1-1)h,\, j_2 h)$, $((j_1+1)h,\, j_2 h)$. Pero, por ejemplo, los nodos $((j_1+1)h,\, (j_2+1)h)$ y $((j_1-1)h,\, (j_2-1)h)$ no intervienen en la discretización. Pero esta no es la única elección posible y pueden construirse esquemas en diferencias que involucren otros nodos. En particular puede encontrarse un esquema simétrico que involucre todos los nodos pero en este caso la matriz del sistema es menos hueca y su resolución es por tanto más costosa.

El tratamiento de las condiciones de contorno no siempre es sencillo. Una vez que el dominio se ha aproximado mediante otro cuadriculado cuya frontera esté integramente en el mallado tenemos que establecer una condición de contorno en cada nodo de la frontera. Cuando la condición de contorno es de tipo Dirichlet, i.e.

$$u = g \text{ en } \partial\Omega$$

la cuestión se resuelve imponiendo la condición $u_j = g_j$ en cada nodo $x_j \in \partial\Omega$. Esto determina completamente la incógnita u_j y reduce la dimensión del vector solución. El valor g_j de u_j interviene entonces en la ecuación que se satisface en los nodos interiores vecinos.

La situación es algo más compleja para la ecuación de contorno de Neumann. Recordemos brevemente el tratamiento que se le da en una dimensión, cuando el operador involucrado es de coeficientes constantes. En este caso se procede por extensión par. Es decir, se usa el hecho de que si $u = u(x)$ es solución de

$$-u_{xx} = f,\ x > 0;\ u_x(0) = 0,$$

entonces su reflexión par

$$\widetilde{u}(x) = \begin{cases} u(x), & x > 0 \\ u(-x), & x < 0 \end{cases}$$

es solución de

$$-\widetilde{u}_{xx} = \widetilde{f},\ x \in \mathbb{R}$$

donde $\widetilde{f}$ es la reflexión par de f.

Aplicando el mismo criterio en el caso discreto escribimos la ecuación correspondiente al nodo $j = 0$:

$$\frac{2\widetilde{u}_0 - \widetilde{u}_1 - \widetilde{u}_{-1}}{h^2} = \widetilde{f}_0.$$

Como $\widetilde{u}_1 = \widetilde{u}_{-1}$ se deduce que

$$\frac{2\,(u_0 - u_1)}{h^2} = f_0.$$

De este modo obtenemos una aproximación de orden dos.

Otra manera natural de tratar la condición de Neumann en el extremo $j = 0$ es escribir que

$$u_0 = u_1$$

puesto que

$$u_x(0) \sim \frac{u_1 - u_0}{h}.$$

En este caso la ecuación escrita en el nodo $j = 1$ proporciona la ecuación

$$\frac{u_1 - u_2}{h^2} = f_1.$$

Pero en este caso el esquema es de orden 1.

Consideremos ahora el esquema (6.3.4). Supongamos que el punto frontera x_j tiene como coordenada x_1, $x_1 = 1$. Al escribir la ecuación (6.3.4) en este punto hacemos aparecer el valor en algunos nodos situados fuera del dominio Ω. Se trata de nodos virtuales. Como la condición de contorno es en este caso de la forma

$$a_{1\alpha}\partial_\alpha\varphi(1, x_2) = g(x_2)$$

es esta la condición que hemos de utilizar para eliminar las contribuciones de los nodos virtuales.

La ecuación discreta es de la forma

$$\begin{aligned}&-\frac{1}{2}\left[\partial_{\beta,+}\left(a_{\alpha\beta}\partial_{\alpha,-}\right)+\partial_{\beta,-}\left(a_{\alpha\beta}\partial_{\alpha,+}\right)\right]u\\&=q_j^{-3}u_{j-e_2}+q_j^{-2}u_{j+e_1-e_2}+q_j^{-1}u_{j-e_1}+q_j^0u_j\\&\quad+q_j^1u_{j+e_1}+q_j^2u_{j-e_1+e_2}+q_j^3u_{j+e_2}\end{aligned}$$

con

$$\begin{aligned}q_j^{-3}&=-\frac{\left(a_{22,\,j-e_2}+a_{22,j}\right)}{2h^2}-\frac{\left(a_{12,\,j-e_2}+a_{12,\,j}\right)}{2h^2},\\q_j^{-2}&=\left(a_{12,\,j-e_2}+a_{12,\,j+e_1}\right)/2h^2,\\q_j^{-1}&=-\left(a_{11,\,j-e_1}+a_{11,\,j}\right)/2h^2-\left(a_{12,\,j-e_1}+a_{12,\,j}\right)/2h^2,\\q_j^1&=q_{j+e_1}^{-1},\ q_j^2=q_{j-e_1+e_2}^{-2},\ q_j^3=q_{j+e_2}^{-3}\\q_j^0&=-\sum_{m\neq 0}q_j^m.\end{aligned}$$

Por la fórmula de Taylor tenemos que

$$\begin{aligned}&-q_j^{-1}\left(u_j-u_{j-e_1}\right)+q_j^1\left(u_{j+3_1}-u_j\right)-q_j^2\left(u_j-u_{j-e_1+e_2}\right)\\&+q_j^{-2}\left(u_{j+e_1-e_2}-u_j\right)\sim\frac{2}{h}a_{1\alpha}\partial_\alpha u(x_j)=\frac{2}{h}f(x_2).\end{aligned}$$

Esta identidad permite eliminar la contribución de los nodos virtuales.

Tal y como ocurría en $1-d$, en caso en que los coeficientes de la ecuación sean discontinuos el método de volúmenes finitos permite obtener mejores aproximaciones.

Para implementarlo descomponemos Ω en celdas Ω_j centradas en los puntos x_j del mallado.

Integrando la ecuación en Ω_j obtenemos

$$-\int_{\Omega_j}\partial_\beta\left(a_{\alpha\beta}\partial_\alpha u\right)dx=\int_{\Omega_j}f\,dx. \tag{6.3.5}$$

Ahora bien

$$-\int_{\Omega_1}\partial_\beta\left(a_{\alpha\beta}\partial_\alpha u\right)dx=-\int_{\Gamma_j}a_{\alpha\beta}\partial_\alpha u h_\beta d\sigma \tag{6.3.6}$$

donde $n=(n_1,\,n_2)$ es el vector exterior unitario a Ω_j y Γ_j es su frontera. La clave reside por tanto en la aproximación de las integrales de frontera sobre Γ_j.

Consideramos en primer lugar la integral sobre el segmento AB con $A = x_j + \left(\frac{h}{2}, -\frac{h}{2}\right)$ y $B = x_j + \left(\frac{h}{2}, \frac{h}{2}\right)$. Hacemos entonces la aproximación

$$\int_A^B a_{\alpha 1}\partial_\alpha u d\sigma \sim h\left(a_{\alpha 1}\partial_\alpha u\right)_C \tag{6.3.7}$$

donde C es el centro de AB, i.e. $C = x_j + (h/2, 0)$. Debemos ahora aproximar el segundo miembro de (6.3.7).

Consideramos en primer lugar el caso de los coeficientes continuos. Realizamos las siguientes aproximaciones :

$$\int_A^B a_{11}\partial_1 u\, dx_2 \sim h\, a_{11}(C)\left(\partial_{1,+}u\right)_j$$
$$\int_A^B a_{12}\partial_2 u\, dx_2 \sim \frac{h\, a_{12}(C)}{2}\left(\partial_{2,-}u_{j+e_1} + \partial_{2,+}u_j\right),$$

o, lo que es lo mismo,

$$\int_A^B a_{12}\partial_2 u\, dx_2 \sim \frac{h\, a_{12}(C)}{4}\left(\partial_{2,+} + \partial_{2,-}\right)\left(u_j + u_{j+e_1}\right)$$

que también coincide con

$$\int_A^B a_{12}\partial_2 u\, dx_2 \sim \frac{h\, a_{12}(C)}{2}\left(\partial_{2,+}u_{j+e_1} + \partial_{2,-}u_j\right).$$

Cuando los coeficientes son discontinuos sobre Γ_j la condición de salto ha de ser tenida en cuenta. Por ejemplo, procediendo como en el caso $1-d$, realizamos la siguiente aproximación

$$\int_A^B a_{11}\partial_1 dx_2 \sim h\, w_j \partial_{1,+}u_j$$

con

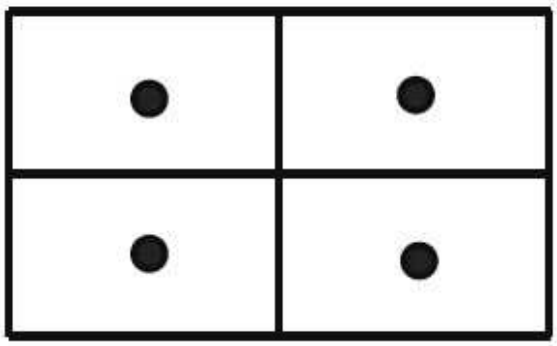
$$w_j = \frac{2a_{11,j}a_{11,j+e_1}}{a_{11,j} + a_{11,j+e_1}}.$$

Las condiciones de contorno pueden tratarse de manera semejante. Cuando se trata de la condición de Dirichlet simplemente sustituimos el valor de la incógnita correspondiente. En el caso de las condiciones de Neumann procedemos promediando el valor de contorno. Por ejemplo:

$$\int_A^B a_{\alpha_1}\partial_\alpha u dx_2 \sim hf(x_{2,j}).$$

Los métodos de discretización presentados hasta ahora están centrados en los nodos del mallado de modo que Ω_j es el cuadrado centrado en x_j y con lados de longitud h en cada una de las direcciones espaciales.

Se pueden hacer desarrollos semejantes en el caso en que la partición Ω_j cubra todo Ω, y sus centros sean los nodos en los que vamos a aproximar la solución tal y como se indica en la siguiente figura.



Los nodos son por tanto en este caso de la forma $x_j = (j-p)h$ con $p = (1/2,\, 1/2)$.

Las discretizaciones por diferencias finitas pueden realizarse del mismo modo que antes. Sólo las condiciones de contorno necesitan de un tratamiento diferenciado.

Capítulo 7

Métodos de descomposición

7.1. El método de las direcciones alternadas

7.1.1. Motivación

La complejidad de los modelos matemáticos que se precisan para analizar los cada vez más sofisticados problemas que se plantean en Ciencia y Tecnología crece sin cesar. Es por eso que, frecuentemente, en dichos problemas podemos encontrar varios sistemas de naturaleza distinta acoplados: Ecuaciones Diferenciales Ordinarias (EDO), Ecuaciones en Derivadas Parciales (EDP) de diferentes tipos, sistemas discretos, estocásticos, etc. Cabe decir que se trata de modelos heterogéneos o de carácter multi-físico.

Es por eso que, a menudo, el análisis de los modelos no puede realizarse mediante el uso de uno de los métodos matemáticos o numéricos específicos de determinado tipo de sistemas sino que es preciso combinar varios de ellos. Lo mismo puede decirse en lo que respecta a los métodos de aproximación y simulación numérica.

Resulta por tanto natural desarrollar y utilizar *métodos de descomposición* que permitan descomponer o dividir el sistema en subsistemas más simples con características bien precisas e identificadas en los que podamos aplicar un método específico.

Pero los métodos de descomposición precisan de un ensamblaje posterior para poder obtener las propiedades globales del sistema. Esto puede realizarse mediante el *método de direcciones alternadas* en el que se resuelven de manera iterada los diversos subsistemas en los que el sistema global se ha descompuesto. Este método iterativo permite aplicar en cada subsistema un método específico

pero al mismo tiempo recuperar las propiedades globales del sistema.

Este método de direcciones alternadas es muy versátil y puede aplicarse en muy diversas situaciones. En particular puede ser utilizado para descomponer sistemas de EDP en el que coexisten subsistemas de naturaleza diversa. Es el caso por ejemplo del sistema de la termoelasticidad que acopla una ecuación de ondas con la ecuación del calor o de las ecuaciones de Navier-Stokes donde coexisten los efectos difusivos de la ecuación del calor (y más precisamente del sistema de Stokes lineal) y los propios de las ecuaciones hiperbólicas no-lineales (ecuaciones de Euler) en las que está presente un término convectivo no-lineal cuadrático.

En la siguiente sección presentamos el método en el contexto de los sistemas de ecuaciones diferenciales lineales con coeficientes constantes.

7.1.2. Sistemas de EDO lineales. El teorema de Lie

Consideremos el sistema de EDO:

$$\begin{cases} \dot{x}(t) = (A+B)x(t) \\ x(0) = x_0, \end{cases} \tag{7.1.1}$$

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

Para cada dato inicial $x_0 \in \mathbb{R}^N$ el sistema (7.1.1) admite una única solución global $x(t) \in C^\omega\left(\mathbb{R};\, \mathbb{R}^N\right)$, donde mediante C^ω denotamos la clase de funciones analíticas.

Además la solución viene dada por la fórmula de representación:

$$x(t) = e^{(A+B)t}x_0. \tag{7.1.2}$$

Recordemos que dada una matriz cuadrada C, su exponencial puede definirse mediante el desarrollo en serie de potencias

$$e^C = \sum_{k=1}^{\infty} \frac{C^k}{k!} \tag{7.1.3}$$

cuya convergencia puede garantizarse mediante el criterio de la mayorante de Weierstrass. En efecto[1]

$$\| e^C \| = \left\| \sum_{k=0}^{\infty} \frac{C^k}{k!} \right\| \leqslant \sum_{k=0}^{\infty} \frac{\| C^k \|}{k!} \leqslant \sum_{k=0}^{\infty} \frac{\| C \|^k}{k!} = e^{\|C\|}. \tag{7.1.4}$$

[1] El mismo argumento permite definir e^L donde $L \in \mathcal{L}(X;\, Y)$ es cualquier operador lineal y acotado entre dos espacios de Banach X e Y.

El siguiente resultado, debido a Lie, es la base de los métodos de descomposición y de direcciones alternadas:

Theorem 7.1.1 *Dadas dos matrices cuadradas $N \times N$ A y B se tiene*

$$e^{A+B} = \lim_{n\to\infty} \left(e^{\frac{A}{n}} e^{\frac{B}{n}}\right)^n. \tag{7.1.5}$$

Antes de proceder a su demostración analicemos su significado.

Es bien sabido que, en general, salvo que las matrices A y B conmuten, no es cierto que e^{A+B} coincida con el producto $e^A e^B$. Analicemos este hecho. De acuerdo con (7.1.3):

$$\begin{aligned} e^{A+B} = \sum_{k=0}^{\infty} \frac{(A+B)^k}{k!} &= I + [A+B] + \frac{[A+B]^2}{2} + \cdots \\ &= I + [A+B] + \frac{A^2 + AB + BA + B^2}{2} + \cdots \end{aligned} \tag{7.1.6}$$

Por otra parte

$$\begin{aligned} e^A e^B &= \left[\sum_{k=0}^{\infty} \frac{A^k}{k!}\right] \left[\sum_{j=0}^{\infty} \frac{B^j}{j!}\right] \\ &= \left[I + A + \frac{A^2}{2} + \cdots\right] \cdot \left[I + B + \frac{B^2}{2} + \cdots\right] \\ &= I + [A+B] + AB + \frac{A^2}{2} + \frac{B^2}{2} + \cdots \end{aligned} \tag{7.1.7}$$

En las expresiones (7.1.6) se constata efectivamente una diferencia al nivel de los términos cuadráticos. En efecto, mientras que en (7.1.6) los términos cuadráticos son $\left[A^2 + B^2 + AB + BA\right]\Big/2$ en (7.1.7) son $\left[A^2 + B^2 + 2AB\right]\Big/2$ de modo que la diferencia entre uno y otro es el término $[BA - AB]\Big/2$. Obviamente esta diferencia se anula cuando A y B conmutan pero no en caso contrario. El término $BA - AB$ que interviene en esta diferencia se denomina *conmutador* y será denotado del modo siguiente:

$$[A,\, B] = BA - AB. \tag{7.1.8}$$

Se trata de una medida del grado de conmutatividad de las matrices A y B.

En virtud de la identidad (7.1.5) tenemos que

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

Para n fijo, se tiene

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

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

Analicemos ahora el significado de la expresión $e^{A/n}e^{B/n}$. Conviene observar que al aplicar $e^{A/n}e^{B/n}$ a un elemento $x_0 \in \mathbb{R}^N$ lo que se obtiene es

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

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

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

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

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

Al aplicar por tanto el Teorema 7.1.1 a la resolución de (7.1.1) obtenemos que

$$x(t) = e^{(A+B)t}x_0 = \lim_{n\to\infty}\left(e^{\frac{At}{n}}e^{\frac{Bt}{n}}\right)^n x_0, \tag{7.1.9}$$

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

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

que se obtiene del modo siguiente:

- Iteramos n veces un procedimiento en el que, arrancando del valor del paso anterior, se avanza un paso temporal del tamaño t/n en la resolución del sistema (7.1.1).
- En cada paso lo que hacemos es resolver de manera consecutiva cada uno de los dos sistemas involucrados en (7.1.1):

$$x' = Ax \tag{7.1.10}$$

y

$$x' = Bx \tag{7.1.11}$$

Conviene observar que, según este procedimiento, en ningún caso resolvemos el sistema completo (7.1.1) sino que siempre resolvemos los subsistemas (7.1.10) y (7.1.11).

En cada intervalo temporal de longitud t/n resolvemos ambos sistemas (7.1.10) y (7.1.11), lo cual indica que recorremos el intervalo temporal en dos ocasiones, una por cada subsistema.

Estas características del método iterativo hacen que se denomine de direcciones alternadas.

Conviene también distinguir el método de direcciones alternadas descrito con los métodos numéricos de aproximación de ecuaciones diferenciales. Aquí hemos descrito un método iterativo en el que suponemos que las soluciones de los subsistemas (7.1.10) y (7.1.11) pueden calcularse de manera exacta. Nos hemos mantenido por tanto en el contexto de las Ecuaciones Diferenciales Ordinarias (EDO). En la práctica, este método de direcciones alternadas ha de ser combinado con un método numérico para la resolución de ecuaciones diferenciales. Es importante señalar, y esta es una de las virtudes del método de direcciones alternadas, que cada uno de los sitemas (7.1.10) y (7.1.11) pueden resolverse mediante métodos distintos, de manera que podemos utilizar métodos mejor adaptados a las características de cada subsistema.

7.1.3. Demostración del Teorema de Lie

En esta sección probamos el Teorema 7.1.1.

En virtud de (7.1.6) y (7.1.7) tenemos que

$$e^{(\frac{A}{n}+\frac{B}{n})t} = e^{\frac{A}{n}t}e^{\frac{B}{n}t} + O\left(\frac{t}{n^2}\right). \tag{7.1.12}$$

En (7.1.12) el término principal del resto es, tal y como veíamos,

$$[A,\, B]\Big/2n^2. \tag{7.1.13}$$

La fórmula (7.1.13) es rigurosa en el sentido que

$$\left|\left|e^{\frac{At}{n}+\frac{Bt}{n}} - e^{\frac{At}{n}}e^{\frac{Bt}{n}}\right|\right| \leqslant \frac{C}{n^2} \mid t \mid \tag{7.1.14}$$

donde $C > 0$ es una constante independiente de n. La prueba de (7.1.14) puede realizarse utilizando el criterio de la mayorante de Werierstrass. Obviamente, la constante C en (7.1.14) depende de A y B pero conviene subrayar que es independiente del parámetro n.

En virtud de (7.1.12) tenemos

$$e^{(A+B)t} = \left[e^{(A+B)\frac{t}{n}}\right]^n = \left[e^{\frac{At}{n}}e^{\frac{Bt}{n}} + O\left(\frac{t^2}{n^2}\right)\right]^2. \tag{7.1.15}$$

Por otra parte,

$$\left[e^{\frac{At}{n}}e^{\frac{Bt}{n}} + O\left(t^2\Big/n^2\right)\right]^n = \left\{e^{\frac{At}{n}}e^{\frac{Bt}{n}}\left[1 + e^{-\frac{Bt}{n}}e^{-\frac{At}{n}}O\left(t^2\Big/n^2\right)\right]\right\}^n. \tag{7.1.16}$$

Basta por tanto concluir que

$$\lim_{n\to\infty}\left(e^{\frac{At}{n}}e^{\frac{Bt}{n}}\right)^n=\lim_{n\to\infty}\left\{e^{\frac{At}{n}}e^{\frac{Bt}{n}}\left[1+e^{-\frac{Bt}{n}}e^{-\frac{At}{n}}O\left(t^2\Big/n^2\right)\right]\right\}^n. \quad (7.1.17)$$

Denotando

$$C_n=e^{\frac{At}{n}}e^{\frac{Bt}{n}} \quad (7.1.18)$$

se trata de probar que

$$\lim_{n\to\infty}C_n^n=\lim_{n\to\infty}\left[C_n\left(1+C_n^{-1}O\left(\frac{t^2}{n^2}\right)\right)\right]^n. \quad (7.1.19)$$

Este hecho puede comprobarse con facilidad utilizando el Teorema del valor medio

$$\begin{aligned}\left[C_n\left(1+C_n^{-1}O\left(\frac{t^2}{n^2}\right)\right)\right]^n-C_n^n&=\ C_n^{n-1}O\left(\frac{t^2}{n^2}\right)n\left(1+\xi_n\right)^{n-1}\\&=\ t^2O\left(\frac{1}{n}\right)C_n^{n-1}\left(1+\xi_n\right)^{n-1}, \quad (7.1.20)\end{aligned}$$

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

Por tanto

$$\begin{aligned}&\left\|\left[C_n\left(1+C_n^{-1}O\left(t^2\Big/n^2\right)\right)\right]^n-C_n^n\right\|\leqslant\frac{Ct^2}{n}\parallel C_n^{n-1}\parallel[1+\parallel\xi_n\parallel]^{n-1}\\&\leqslant\frac{Ct^2}{n}\parallel C_n\parallel^{n-1}(1+\parallel\xi_n\parallel)^{n-1}\leqslant\frac{Ct^2}{n}e^{\|A\|(n-1)t/n}e^{\|B\|(n-1)t/n}(1+\parallel\xi_n\parallel)^{n-1} \quad (7.1.21)\end{aligned}$$

Ahora bien, como

$$\parallel\xi_n\parallel\leqslant\parallel C_n^{-1}\parallel O\left(t^2\Big/n^2\right)\leqslant e^{\|A\|t/n}e^{\|B\|t/n}O\left(t^2\Big/n^2\right)=O\left(t^2\Big/n^2\right) \quad (7.1.22)$$

tenemos que

$$(1+\parallel\xi_n\parallel)^{n-1}=\left[\left(1+1\Big/\parallel\xi_n\parallel^{-1}\right)^{\|\xi_n\|^{-1}}\right]^{\|\xi_n\|(n-1)}\to e^0=1 \quad (7.1.23)$$

y, por tanto, volviendo a (7.1.21) deducimos que

$$\left\|\left[C_n\left(1+C_n^{-1}O\left(t^2\Big/n^2\right)\right)\right]^n-C_n^n\right\|=O(1/n),\ 0\leqslant t\leqslant T. \quad (7.1.24)$$

Se obtiene por tanto (7.1.19) y (7.1.17) que es el resultado que precisábamos probar.■

Conviene hacer las consideraciones siguientes:

• La prueba que hemos realizado es válida no sólo para matrices $N\times N$ A y B. Se aplica también a operadores acotados entre espacios de Banach.

• Este resultado puede también generalizarse al marco de los operadores m-disipativos que generan semigrupos que permiten abordar la resolución de EDP. Se trata de la denominada fórmula del producto de Trotter ([21], vol. 1, sec. VIII.8).

7.1.4. Algunos ámbitos de aplicación

Tal y como hemos mencionado en la introducción de la sección el método de descomposición o de direcciones alternadas descrito es sumamente versátil y puede aplicarse en muy diferentes formas, no solamente en el ámbito de la resolución de un sistema lineal de EDO de la forma (7.1.1).

De manera general, el método puede aplicarse en cualquier sistema en el que estén presentes subsistemas de naturaleza diversa.

Mencionemos algunos ejemplos:

- **Resolución de sistemas algebraicos**

  Buena parte de los métodos de descomposición o "splitting" para la resolución de sistemas lineales de la forma $Ax = b$ están basados en la idea de descomponer la matriz A de un modo u otro y proceder a aproximar la solución x mediante la resolución iterada de un sistema simplificado. Es el caso por ejemplo de los clásicos métodos iterativos de Gauss-Seidel, Jacobi, etc. ([14], [19])

- **Las ecuaciones de Navier-Stokes**

  Otro de los ejemplos más notables lo constituyen las ecuaciones de Navier-Stokes para un fluido incompresible

  $$\begin{cases} u_t - \nu\Delta u + u \cdot \nabla u = \nabla p \\ \operatorname{div} u = 0. \end{cases}$$

  En este caso el sistema es una superposición de las ecuaciones de Stokes lineales

  $$\begin{cases} u_t - \nu\Delta u = \nabla p \\ \operatorname{div} u = 0 \end{cases}$$

  y de las ecuaciones de Euler para un fluido perfecto

  $$\begin{cases} u_t + u \cdot \nabla u = \nabla p \\ \operatorname{div} u = 0. \end{cases}$$

  El método de las direcciones alternadas puede también ser aplicado en este caso, [26].

- **El sistema de la termoelasticidad**

  En este caso el sistema es de la forma

$$\begin{cases} u_{tt} - \mu\Delta u - (\lambda+\mu)\nabla \operatorname{div} u + \alpha\nabla\theta = 0 \\ \theta_t - \Delta\theta + \beta \operatorname{div} u_t = 0 \end{cases}$$

  en el que se observa el acoplamiento de una ecuación de tipo ondas (el sistema de Lamé en elasticidad) con la ecuación del calor.

- **El método de la descomposición de dominios**

  Una de las variantes más útil del método de las direcciones alternadas es el de la descomposición de dominios. Se trata de un método muy versátil aplicable de manera sistemática en "todo" sistema de EDP definido en un dominio de $\mathbb{R}^N$. La idea de base consiste simplemente en descomponer el dominio en subdominios más pequeños y de propiedades geométricas más homogéneas, de modo que la resolución del subsistema correspondiente a cada subdominio sea más sencilla. Nuevamente es preciso establecer una iteración alternada y eso exige introducir condiciones de contorno adecuadas en las interfases entre uno y otro dominio que garanticen la convergencia. Este método es debido en sus orígenes a H.A. Schwarz. En la próxima sección haremos una breve presentación del mismo inspirándonos en el capítulo IV, 4 del volumen II del libro de R. Courant y D. Hilbert [6].

7.2. Descomposición de dominios en $1-d$

Consideremos un dominio acotado y regular Ω de $\mathbb{R}^N$ y una descomposición en dos subdominios Ω_- y Ω_+ con intersección no vacía, con fronteras regulares compuestas por un número finito de componentes y que se intersequen con ángulo no nulo.

La idea es probar que a partir de la resolubilidad de una ecuación, por ejemplo el problema de Dirichlet para el Laplaciano, en los subdominios Ω^- y Ω^+, mediante un método iterativo de direcciones alternadas, el mismo problema puede resolverse en el dominio total Ω.

Este método fue ideado en un principio para calcular las soluciones en otros dominios que no fuesen esferas, cuadrados, etc., donde las funciones de Green (y por tanto las soluciones) podían calcularse explícitamente. En la actualidad el método se utiliza de manera profusa como herramienta básica de aproximación y simulación numérica.

Con el objeto de presentar el método consideremos el caso en que el dominio $\Omega = (-1, 1)$ y los subdominios son, por ejemplo, $\Omega^- = (-1, 1/2)$ y $\Omega^+ = (-1/2, 1)$.

Consideramos la ecuación

$$\begin{cases} -u_{xx} = f & -1 < x < 1 \\ u(-1) = u(1) = 0 \end{cases} \tag{7.2.1}$$

y las ecuaciones en los subdominios

$$\begin{cases} -u_{-,xx} = f, & -1 < x < 1/2 \\ u_-(-1) = 0, & u_-(1/2) = \alpha \end{cases} \tag{7.2.2}$$

$$\begin{cases} -u_{+,xx} = f, & -1/2 < x < 1 \\ u_+(-1/2) = \beta, & u_+(1) = 0. \end{cases} \tag{7.2.3}$$

El método consiste en resolver de manera alternada los sistemas reducidos (7.2.2) y (7.2.3) para obtener una sucesión cuyo límite sea la solución de (7.2.1). Obviamente, para definir la iteración tenemos que establecer el valor de α y β que imponemos en los extremos $x = 1/2$ y $x = -1/2$ respectivamente.

Una posible iteración es la siguiente:

• **Etapa 1.** Inicialización.

Resolvemos en primer lugar (7.2.2) con $\alpha = 0$ y posteriormente (7.2.3) con dato $\beta = u_-^1(-1/2)$, siendo u_-^1 la solución obtenida de (7.2.2).

• **Etapa 2.**

Con el valor u_+^1 obtenido en la solución de (7.2.3) de la primera etapa resolvemos (7.2.2) con $\alpha = u_+^1(1/2)$ y posteriormente (7.2.3) con $\beta = u_-^2(-1/2)$, siendo u_-^2 la nueva solución de (7.2.2). De este modo

calculamos u_-^2 y u_+^2.

• **Siguientes etapas**

Iterando este proceso obtenemos una sucesión u_-^k de soluciones de (7.2.2) y otra u_+^k de soluciones de (7.2.3).

La clave de la convergencia del método está en probar que u_-^k y u_+^k convergen cuando $k \to \infty$ a la solución de (7.2.1) en $(-1, 1/2)$ y $(-1/2, 1)$ respectivamente cuando $k \to \infty$.

El método que acabamos de describir no es más que una de las numerosas variantes del método de descomposición de dominios. En este caso los subdominios

se superponen o solapan teniendo como intersección el subintervalo $(-1/2,\, 1/2)$. En otras de las variantes los dominios no se superponen sino que dividen al dominio Ω compartiendo únicamente una interfase. Sería el caso por ejemplo si $\Omega_- = (-1,\, 0)$ y $\Omega_+ = (0,\, 1)$. En este caso sin embargo es preciso elegir las condiciones de transmisión en la interfase ($x = 0$ en el presente caso) de manera adecuada para garantizar la convergencia del método.

El principio del máximo es muy útil para probar la convergencia del método en este caso particular.

Suponiendo que $f \in L^\infty(-1,\, 1)$ y denotando por M su norma tenemos que la solución de (7.2.3) satisface

$$\mid u(x) \mid \leqslant \frac{M}{2}(1 - x^2), \tag{7.2.4}$$

lo mismo ocurre para cada una de las soluciones u_-^k y u_+^k en los subdominios

$$\left|u_-^k\right|,\, \left|u_+^k\right| \leqslant \frac{M}{2}(1 - x^2). \tag{7.2.5}$$

Estas acotaciones permiten garantizar que, extrayendo subsucesiones, las sucesiones u_-^k y u_+^k convergen cuando $k \to \infty$ en la topología débil de $L^2(-1,\, 1)$ por ejemplo.

Pero la convergencia puede analizarse de manera mucho más precisa.

En efecto, de (7.2.5) es fácil deducir que $\left\{u_-^k\right\}$ (resp. $\left\{u_+^k\right\}$) está acotada en $H^1(-1,\, 1/2)$ (resp. $H^1(-1/2,\, 1)$). Sus límites débiles son obviamente soluciones de la ecuación en los subintervalos correspondientes. El punto clave de la prueba de la convergencia consiste en probar que los dos límites coinciden en el subintervalo común $(-1/2,\, 1/2)$, puesto que es ésto lo que garantiza que el solapamiento de los dos límites sea la solución de (7.2.1).

Con el objeto de resolver este último problema sustraemos a la solución u de (7.2.1) una solución cualquiera del problema en toda la recta real[2]

$$-v_{xx} = \tilde{f} \text{ en } \mathbb{R}, \tag{7.2.6}$$

donde $\tilde{f}$ es la extensión por cero de f fuera del intervalo $(-1,\, 1)$.

Entonces

$$w = u - v \tag{7.2.7}$$

ha de satisfacer

$$\begin{cases} -w_{xx} = 0 & -1 < x < 1 \\ w(-1) = \alpha_-, & w(1) = \alpha_+, \end{cases} \tag{7.2.8}$$

[2]Este argumento puede también aplicarse en varias dimensiones espaciales.

con valores a y b unívocamente determinados por la solución v en la recta real.

Es en este marco en el que aplicamos el método de descomposición de dominios.

Conviene señalar que el cambio de variables (7.2.6)-(7.2.7) que permite pasar del problema (7.2.1) a (7.2.8) puede también aplicarse en varias dimensiones espaciales puesto que la solución del problema de Laplace en $\mathbb{R}^N$ puede calcularse fácilmente por convolución con la solución fundamental.

Con el objeto de aplicar la iteración en direcciones alternadas a (7.2.8) consideramos los dos subsistemas

$$\begin{cases} -w_{+,xx} = 0, & -1 < x < 1/2 \\ w_+(-1) = \alpha_-, & w_+(1/2) = \beta_+ \end{cases} \tag{7.2.9}$$

y

$$\begin{cases} -w_{-,xx} = 0, & -1/2 < x < 1 \\ w_-(-1/2) = \beta_-, & w_-(1) = \alpha_+. \end{cases} \tag{7.2.10}$$

Iterando y alternando la resolución de (7.2.9) y (7.2.10) tal y como indicamos más arriba, obtenemos dos sucesiones $\left\{w_-^k\right\}_{k\geqslant 1}$ y $\left\{w_+^k\right\}_{k\geqslant 1}$.

El punto clave de la demostración de convergencia consiste en observar que

$$\left|w_+^{k+1}(1/2) - w_+^k(1/2)\right| \leqslant \ell \left|w_-^{k+1}(-1/2) - w_-^k(-1/2)\right| \tag{7.2.11}$$

con

$$0 < \ell < 1, \tag{7.2.12}$$

y que, simultáneamente,

$$\left|w_-^{k+1}(-1/2) - w_-^k(-1/2)\right| \leqslant \ell \left|w_+^k(1/2) - w_+^{k-1}(1/2)\right|. \tag{7.2.13}$$

De (7.2.11) y (7.2.13) deducimos que

$$\left|w_+^{k+1}(1/2) - w_+^k(1/2)\right| \leqslant \ell^2 \left|w_+^k(1/2) - w_+^{k-1}(1/2)\right|.$$

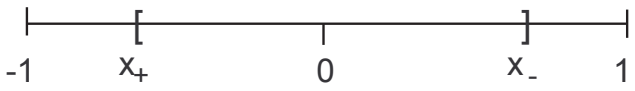


Figura 7.1: Descomposición de dominios $1-d$: $\Omega_- = (-1,\, x_-)$, $\Omega_+ = (x_+,\, 1)$.

Procedemos entonces con la iteración

$$\begin{cases} -\left(u_-^k\right)'' = 0 \\ u_-^k(-1) = \alpha_-;\; u_-^k(x_-) = u_+^{k-1}(x_-) \end{cases} \tag{7.2.14}$$

$$\begin{cases} -\left(u_+^k\right)'' = 0 \\ u_+^k(x_+) = u_-^k(x_+),\ u_+^k(1) = \alpha_+. \end{cases} \tag{7.2.15}$$

Analizamos ahora las diferencias

$$v_-^k = u - u_-^k;\ v_+^k = u - u_+^k \tag{7.2.16}$$

que satisfacen

$$\begin{cases} -\left(v_-^k\right)'' = 0 \\ v_-^k(-1) = 0,\ v_-^k(x_-) \quad = \quad u(x_-) - u_+^{k-1}(x_-) = v_+^{k-1}(x_-) \end{cases} \tag{7.2.17}$$

$$\begin{cases} -\left(v_+^k\right)'' = 0 \\ v_+^k(x_+) = v_-^k(x_+),\ v_+^k(1) = 0. \end{cases} \tag{7.2.18}$$

Las soluciones de estos problemas pueden calcularse de manera explícita aunque ésto no sea necesario para la prueba de la convergencia de las soluciones, puesto que, tal y como veremos más adelante, puede desarrollarse también en el caso de un dominio general en varias dimensiones espaciales.

Tenemos

$$v_-^k(x) = \frac{v_+^{k-1}(x_-)(x+1)}{x_- + 1} \tag{7.2.19}$$

$$v_+^k(x) = \frac{v_-^k(x_+)(x-1)}{x_+ - 1}. \tag{7.2.20}$$

Por tanto

$$\left|v_-^k(x_+)\right| = \frac{\left|v_+^{k-1}(x_-)\right|}{1 + x_-} \mid 1 + x_+ \mid = \left|v_+^{k-1}(x_-)\right| \gamma_1 \tag{7.2.21}$$

con

$$\gamma_1 = \frac{\mid 1 + x_+ \mid}{\mid 1 + x_- \mid} \tag{7.2.22}$$

y

$$\left|v_+^k(x_-)\right| = \left|v_-^k(x_+)\right| \frac{\mid x_- - 1 \mid}{1 - x_+} = \left|v_-^k(x_+)\right| \gamma_2 \tag{7.2.23}$$

con

$$\gamma_2 = \frac{1 - x_-}{1 - x_+}. \tag{7.2.24}$$

De este modo obtenemos

$$\left|v_-^k(x_+)\right| = \left|v_-^{k-1}(x_+)\right| \gamma_1\gamma_2 \tag{7.2.25}$$

y por tanto

$$\left|v_-^k(x_+)\right| = (\gamma_1\gamma_2)^{k-1} \left|v_-^1(x_+)\right| = \gamma_1 (\gamma_1\gamma_2)^{k-1} \left|v_+^0(x_-)\right|$$

siendo $v_+^0(x_-)$ un valor cualquiera de inicialización del método.

De modo análogo

$$\left|v_+^k(x_-)\right| = (\gamma_1\gamma_2)^k \left|v_+^0(x_-)\right|. \tag{7.2.26}$$

Basta entonces comprobar que

$$\gamma_1\gamma_2 < 1 \tag{7.2.27}$$

para asegurarse de que $v_+^k(x_-)$ y $v_-^k(x_+)$ convergen exponencialmente a cero cuando $k \to \infty$.

A partir de este hecho es fácil deducir la convergencia de la solución puesto que

$$\left|\left|v_-^k\right|\right|_{L^\infty(-1,\,x_-)} \leqslant \left|v_+^{k-1}(x_-)\right| \tag{7.2.28}$$

y

$$\left|\left|v_+^k\right|\right|_{L^\infty(x_+,\,1)} \leqslant \left|v_-^k(x_+)\right|. \tag{7.2.29}$$

Comprobamos finalmente (7.2.27). Basta para ello comprobar que $\gamma_1,\ \gamma_2 < 1$. Pero esto es obvio puesto que γ_1 y γ_2 no representan más que la longitud relativa de los subintervalos que en cada subdominio Ω_- y Ω_+ quedan fuera del intervalo de solapamiento $(x_+,\ x_-)$. Así, a medida que el intervalo en el que se tiene solapamiento aumenta, la velocidad de convergencia del método tse incrementa.

Conviene señalar que esta prueba, mediante la utilización del Principio del Máximo, puede adaptarse para abordar el caso de varias dimensiones espaciales y el caso de aproximaciones numéricas.

Consideramos en primer lugar las aproximaciones numéricas finitas $1-d$.

7.3. Descomposición de dominios para las diferencias finitas $1-d$

Consideramos ahora el esquema en diferencias finitas de tres puntos para la aproximación del problema de Dirichlet (7.2.1):

$$\begin{cases} \dfrac{2u_j - u_{j+1} - u_{j-1}}{h^2} = f_j, \quad j = -N, \cdots, N \\ u_{-N-1} = u_{N+1} = 0 \end{cases} \tag{7.3.1}$$

donde $h > 0$ es un parámetro destinado a tender a cero, $N \in \mathbb{N}$ es tal que $Nh = 1$, u_j representa una aproximación numérica de la solución u de (7.2.1) en x_j y f_j de $f(x_j)$. Cuando f es continua podemos tomar $f_j = f(x_j)$ mientras

que si f es simplemente integrable podemos elegir la media de f en el intervalo $(-x_j - h/2, x_j + h/2)$ como aproximación de f en f_j, es decir, como valor de f_j.

El mismo esquema puede utilizarse para la aproximación u de (7.2.2). Tenemos en este caso

$$\begin{cases} \dfrac{2u_j - u_{j+1} - u_{j-1}}{h^2} = 0, & j = -N, \cdots, N \\ u_{-N-1} = \alpha_-, & u_{N+1} = \alpha_+. \end{cases} \tag{7.3.2}$$

Es bien conocido que el esquema de tres puntos es consistente de orden 2 de modo que

$$\| u - \vec{u}_h \|_\infty \leqslant Ch^2. \tag{7.3.3}$$

Aquí y en lo sucesivo $u = u(x)$ denota la solución del problema continuo, $\vec{u}_h$ es el vector que representa la solución del problema numérico

$$\vec{u}_h = (u_{-N}, \cdots, u_0, \cdots, u_N) \tag{7.3.4}$$

y $\| \cdot \|_\infty$ denota la norma ℓ^∞ discreta sobre los puntos $\{x_j\}_{-N \leqslant j \leqslant N}$ del mallado.

En virtud de (7.3.3), $\vec{u}_h$, la solución discreta, proporciona una buena aproximación de la solución continua.

El método de descomposición de dominios se adapta sin dificultades al caso de la ecuación discreta (7.3.2). Para ello descomponemos el conjunto de índices discretos en dos subdominios. Lo hacemos del siguiente modo.

Sean $\ell_{h,-}$ y $\ell_{h,+}$ los índices tales que

$$x_- = \ell_{h,-}h,\ x_+ = \ell_{h,+}h, \tag{7.3.5}$$

siendo x_- y x_+ los extremos interiores de los subdominios Ω_- y Ω_+ en los que descomponemos el dominio global $(-1, 1)$ en la aplicación del método de descomposición de dominios.

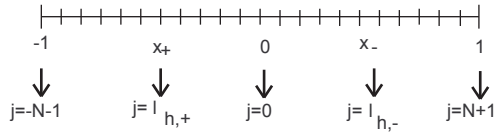


Figura 7.2: Descomposición de dominios en el mallado discreto.

Consideramos ahora los dos subproblemas

$$\begin{cases} \dfrac{2u^k_{-,j} - u^k_{-,j+1} - u^k_{-,j-1}}{h^2} = 0, & -N \leqslant j \leqslant \ell_{h,-} - 1 \\ u^k_{-,-N-1} = \alpha_-,\ u_{-,\ell_{h,-}} = u^{k-1}_{+,\ell_{h,-}} \end{cases} \tag{7.3.6}$$

y

$$\begin{cases} \dfrac{2u^k_{+,j} - u^k_{+,j+1} - u^k_{+,j-1}}{h^2} = 0, & \ell_{h,+} \leqslant j \leq N \\ u^k_{+,\ell_{h,+}} = u^k_{-,\ell_{h,+}}, u^k_{+,N+1} = \alpha_+. \end{cases} \tag{7.3.7}$$

Se trata de los análogos discretos de (7.2.14) y (7.2.15).

Comprobamos ahora la convergencia exponencial de u^k_- y u^k_+ a los valores correspondientes de la solución numérica $\vec{u}_h$ en ℓ^∞. Procedemos como en el caso continuo. Para ello omitimos en lo sucesivo el subíndice h en la notación y definimos $\vec{v}^k_-$ y $\vec{v}^k_+$ como la diferencia entre la solución discreta $\vec{u}$ de (7.3.2) y las soluciones $\vec{u}^k_-$ y $\vec{u}^k_+$ de los subproblemas (7.3.6) y (7.3.7).

El mismo argumento del caso continuo se aplica y obtenemos exactamente la misma tasa de convergencia cuando $k \to \infty$, independiente de h. Esto es así porque tanto $\vec{v}^k_-$ como $\vec{v}^k_+$ son soluciones de los problemas

$$\frac{2v_j - v_{j+1} - v_{j-1}}{h^2} = 0 \tag{7.3.8}$$

y estas a su vez son funciones afines

$$v_j = ajh + b, \tag{7.3.9}$$

donde las constantes a y b son las que permiten ajustar las condiciones de Dirichlet en los extremos del intervalo en cuestión.

De este modo deducimos la existencia de una constante $C > 0$ y otra $\gamma < 1$ de modo que

$$\| \vec{u}_h - \vec{u}^k_{-,h} \|_{\infty,-} \leqslant Ch^2 \tag{7.3.10}$$

$$\| \vec{u}_h - \vec{u}^k_{+,h} \|_{\infty,+} \leqslant Ch^2, \tag{7.3.11}$$

donde $\| \cdot \|_{\infty,\pm}$ denotan las normas ℓ^∞ discretas en los subdominios $\pm$ correspondientes.

Combinando la estimación (7.3.3) y (7.3.10) y (7.3.11) obtenemos

$$\| u - \vec{u}^k_{h,-} \|_{\infty,-} + \| u - \vec{u}^k_{h,+} \|_{\infty,+} \leqslant Ch^2 + C\gamma^k. \tag{7.3.12}$$

Esto permite cuantificar la convergencia del genuino método numérico que consiste en aplicar el método de descomposición de dominios en el esquema discreto de aproximación numérica.

Así si el error admisible es $\varepsilon > 0$, en virtud de (7.3.12) elegimos $h_0 > 0$ suficientemente pequeño tal que

$$Ch_0^2 \leqslant \varepsilon/2. \tag{7.3.13}$$

Una vez realizada esta elección de h_0 elegimos k_0 tal que

$$C\gamma^{k_0} \leqslant h_0.$$

De este modo nos aseguramos que el par $\left\{\vec{u}_{h_{0,-}}^{k_0}, \vec{u}_{h_{0,+}}^{k_0}\right\}$ proporciona una aproximación de la solución $u = u(x)$ del problema continuo $u = u(x)$ con error inferior a ε.

7.4. "Splitting"

En esta sección analizaremos brevemente los métodos de splitting para la resolución de la ecuación de Burgers viscosa

$$u_t - \nu u_{xx} + \left(u^2\right)_x = 0. \tag{7.4.1}$$

A lo largo de esta sección suponemos que $\nu > 0$ está fijado.

Buena parte de las ideas que aquí introduzcamos serán útiles también en el contexto de las ecuaciones más realistas y complejas de las ecuaciones de Navier-Stokes con viscosidad.

En vista de la propia estructura de la ecuación (7.4.1) es obvio que en ella subyacen dos modelos. Por una parte la ecuación del calor

$$u_t - \nu u_{xx} = 0 \tag{7.4.2}$$

y por otra, la ecuación de Burgers sin viscosidad

$$u_t + \left(u^2\right)_x = 0. \tag{7.4.3}$$

De hecho, a través de la transformación de Hopf-Cole, esto ha quedado claramente de manifiesto habiéndose comprobado que la solución de (7.4.1) incorpora tanto efectos viscosos como convectivos.

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

En las notas [34] describíamos algunas ideas básicas de los métodos de descomposición. En particular presentamos el Teorema de Lie que demuestra que para dos matrices $n \times n$ A_1 y A_2 cualesquiera se tiene

$$e^{A_1+A_2} = \lim_{j\to\infty} \left(e^{A_1/j} e^{A_2/j}\right)^j, \tag{7.4.4}$$

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

$$x' = \Big(A_1 + A_2\Big)x,\ t \in \mathbb{R},\ x(0) = x_0, \tag{7.4.5}$$

puesto que su solución es de la forma

$$x(t) = e^{(A_1+A_2)t} x_0. \tag{7.4.6}$$

La ecuación de Burgers viscosa (7.4.1) puede también entenderse como un problema de la forma (7.4.5), si bien en este caso la incógnita $u = u(x,\, t)$ es, para cada $t > 0$, una función que depende de x y que por tanto pertenece a un espacio de dimensión infinita, A_1 es el operador diferencial $A_1 u = \partial_x^2 u$ y A_2 es el operador no-lineal $A_2 u = -(u^2)_x$.

En esta sección vamos a introducir esquemas de descomposición para las versiones discretas de estas ecuaciones.

Consideramos por tanto el modelo (7.4.5) que puede discretizarse en tiempo con paso Δt sustituyendo, como es habitual, la derivada $x'(t)$ por un cociente incremental en el intervalo $[n\Delta t,\, (n+1)\Delta t]$. El tipo de esquemas de descomposición que vamos a utilizar se basan en la utilización de un nodo adicional (o varios), por ejemplo en el punto medio $(n+1/2)\Delta t$, de modo que en cada subintervalo $[n\Delta t,\, (n+1/2)\Delta t]$ y $[(n+1/2)\Delta t,\, (n+1)\Delta t]$, veamos (7.4.5) como un sistema en el que domina A_1 o A_2.

En la sección siguiente analizaremos la validez de las discretizaciones temporales a la hora de aproximar las soluciones de EDP de evolución. Tal y como veremos, este tipo de métodos, clásicos y bien comprendidos en el marco de las EDOs, es también de aplicación en EDPs, incluso en el marco no-lineal en el que la teoría clásica de semigrupos lineales no se aplica, haciéndose necesaria la utilización de ideas propias de los semigrupos no-lineales.

7.4.1. Peaceman-Rachford

Introducimos en primer lugar el esquema de Peaceman-Rachford:

$$\begin{cases} x^0 = x^0 \\ \dfrac{x^{n+1/2} - x^n}{\Delta t/2} = A_1 x^{n+1/2} + A_2 x^n \\ \dfrac{x^{n+1} - x^{n+1/2}}{\Delta t/2} = A_1 x^{n+1/2} + A_2 x^{n+1}. \end{cases} \tag{7.4.7}$$

Vemos que el esquema así obtenido es doblemente implícito, una primera vez al pasar de x^n a $x^{n+1/2}$ y después al pasar de $x^{n+1/2}$ a x^{n+1}. En cada caso sin embargo, nos apoyamos principalmente en uno de los operadores A, B considerando al otro como una perturbación que incorpora la información del paso anterior.

Con el objeto de entender el efecto que este esquema de descomposición tiene sobre un sistema consideremos el caso del sistema más sencillo posible

$$x'(t) = Ax(t),\ t > 0;\ x(0) = x_0, \tag{7.4.8}$$

siendo A una matriz simétrica definida negativa para la que la solución es de la forma

$$x(t) = e^{At} x_0. \tag{7.4.9}$$

Si $\lambda_1, \cdots, \lambda_N$ son los autovalores de la matriz A y $e_1, \cdots, e_N$ los correspondientes autovectores, las soluciones son combinaciones lineales de soluciones elementales de la forma

$$x(t) = e^{\lambda_j t} e_j,\ j = 1, \cdots, N. \tag{7.4.10}$$

Como $\lambda_j \leq 0$, $j = 1, \cdots, N$ vemos, obviamente que

$$\mid x(t) \mid \leq \mid x_0 \mid,\ \forall t \geq 0, \tag{7.4.11}$$

para toda solución.

Descomponemos ahora la matriz A como

$$A = \alpha A + \beta A \tag{7.4.12}$$

de modo que

$$A_1 = \alpha A,\ A_2 = \beta A, \tag{7.4.13}$$

con

$$\alpha > 0,\ \beta > 0,\ \alpha + \beta = 1. \tag{7.4.14}$$

Aplicando el esquema de descomposición (7.4.7) obtenemos

$$x^{k+1} = \left(I - \frac{\Delta t}{2}\beta A\right)^{-1}\left(I + \frac{\Delta t}{2}\alpha A\right)\left(I - \frac{\Delta t}{2}\alpha A\right)^{-1}\left(I + \frac{\Delta t}{2}\beta A\right)x^k.$$

Iterando esta identidad obtenemos

$$x^k = \left(I - \frac{\Delta t}{2}\beta A\right)^{-k}\left(I + \frac{\Delta t}{2}\alpha A\right)^{k}\left(I - \frac{\Delta t}{2}\alpha A\right)^{-k}\left(I + \frac{\Delta t}{2}\beta A\right)^{k}x^0.$$

Cuando aplicamos esta identidad en la dirección de cada uno de los vectores propios e_j, i.e. con $x^0 = e_j$, deducimos que

$$x_j^k = \left(\frac{1 + \frac{\Delta t}{2}\alpha\lambda_j}{1 - \frac{\Delta t}{2}\alpha\lambda_j}\right)^k\left(\frac{1 + \frac{\Delta t}{2}\beta\lambda_j}{1 - \frac{\Delta t}{2}\beta\lambda_j}\right)^k e_j.$$

Teniendo en cuenta que

$$\left|\frac{1+\xi}{1-\xi}\right| \leq 0,\ \forall \xi < 0$$

deducimos la estabilidad del esquema puesto que

$$\left|x_j^k\right| \leq |\ e_j\ |,\ \forall k \geq 0,\ \forall j = 1, \cdots, N.$$

Además en el caso en que $\lambda_j < 0$, $j = 1, \cdots, N$, en el que $x(t) \to 0$ cuando $t \to \infty$ exponencialmente, la solución discreta también reproduce este comportamiento puesto que

$$\left|\frac{1 + \frac{\Delta t}{2}\alpha\lambda_j}{1 - \frac{\Delta t}{2}\lambda_j}\right| < 1,\ \left|\frac{1 + \frac{\Delta t}{2}\beta\lambda_j}{1 - \frac{\Delta t}{2}\beta\lambda_j}\right| < 1,\ j = 1, \cdots, N.$$

Con el objeto de estudiar más en detalle este esquema introducimos la función racional

$$R_1(\xi) = \left(\frac{1 + \frac{\alpha\xi}{2}}{1 - \frac{\alpha\xi}{2}}\right)\left(\frac{1 + \frac{\beta\xi}{2}}{1 - \frac{\beta\xi}{2}}\right)$$

que admite, en un entorno de $\xi = 0$, el desarrollo de Taylor

$$R_1(\xi) = 1 + \xi + \frac{\xi^2}{2} + \left(\alpha^2 + \beta^2 + \alpha\beta\right)\frac{\xi^3}{4} + O(1)\xi^4$$

mientras que la función exponencial que interviene en la resolución de la ecuación diferencial tiene el desarrollo

$$e^{\xi} = 1 + \xi + \frac{\xi^2}{2} + \frac{\xi^3}{6} + O(\xi^4).$$

De estos dos desarrollos de Taylor se deduce que el esquema es consistente de orden dos para cualquier elección de los parámetros $0 < \alpha, \beta < 1$ tal que

$\alpha + \beta = 1$. La menor diferencia en el término de orden 3 se produce para la elección $\alpha = \beta = 1/2$.

El esquema es por tanto convergente de orden 2.

El único inconveniente de este método es su falta de estabilidad absoluta o de estabilidad stiff que se pone de manifiesto en el hecho que $R_1(\xi) \to 1$ cuando $\xi \to -\infty$. Como ξ juega el papel de $\Delta t\lambda$, esto indica que la velocidad exponencial de convergencia se pierde a medida que $|\Delta t|$ aumenta. Pero esto es inevitable en el marco de las ecuaciones parabólicas en las que típicamente λ recorre una sucesión infinita que tiene a $-\infty$. Sea cual sea el valor de $\Delta t > 0$ elegido, por pequeño que este sea, el esquema numérica es incapaz de reproducir la dinámica de la EDP en la que a medida que $|\lambda|$ aumenta, la estabilidad exponencial de las soluciones de acentúa.

El método de Peacema-Rachford que acabamos de describir es uno de la amplia familia de métodos de descomposición o de direcciones alternadas cuyo abanico de aplicaciones es muy amplio.

Consideremos por ejemplo el problema de Dirichlet para la ecuación del calor:

$$\begin{cases} u_t - \Delta u = 0 & \text{en} \quad \Omega,\, t > 0 \\ u = 0 & \text{sobre} \quad \partial\Omega,\, t > 0 \\ u(x,\, 0) = u_0(x) & \text{en} \quad \Omega. \end{cases} \tag{7.4.15}$$

Analicemos el caso en que $\Omega \subset \mathbb{R}^2$.

El esquema semi-discreto más elemental para la aproximación numérica de este sistema es el que se obtiene al utilizar el esquema de aproximación de cinco puntos para el Laplaciano. Obtenemos así:

$$\begin{cases} u_{i,j} + \dfrac{2u_{i,j} - u_{i+1,j} - u_{i-1,j}}{h_1^2} + \dfrac{2u_{i,j} - u_{i,j+1} - u_{i,j-1}}{h_2^2} = 0, & (i,j) \in \Omega, t > 0 \\ u_{i,j} = 0, & (i,j) \in \partial\Omega,\, t > 0 \\ u_{i,j}(0) = u_{0,i,j}, & (i,j) \in \Omega. \end{cases} \tag{7.4.16}$$

En estas expresiones utilizamos las notaciones habituales, de modo que h_1 denota el paso del mallado en la variable x_1, h_2 el del mallado en la dirección x_2, $u_{i,j}$ la aproximación de la solución u en el punto $(h_1 i,\, h_2 j)$, $(i,j) \in \Omega$ los índices tales que el punto correspondiente del mallado $(h_1 i,\, h_2 j)$ pertenece al dominio Ω y $(i,\, j) \in \partial\Omega$ para los puntos del mallado que están sobre la frontera $\partial\Omega$. Esto es rigurosamente válido cuando toda la frontera $\partial\Omega$ está constituida por puntos del mallado. En el caso general, es preciso aproximar la frontera $\partial\Omega$ mediante una frontera embebida en el mallado.

Denotando mediante U_h el vector incógnita $(u_{i,\,j})$ que contiene todos los

valores numéricos, el sistema (7.4.16) puede escribirse en la forma

$$\begin{cases} \dfrac{dU_h}{dt} + A_{1h}U_h + A_{2h}U_h = 0,\ t > 0 \\ U_h(0) = U_{h,\,0}, \end{cases} \tag{7.4.17}$$

donde A_{1h} y A_{2h} son los análogos discretos de $\partial^2/\partial x_1^2$ y $\partial^2/\partial x_2^2$, respectivamente.

Este sistema puede aproximarse mediante los esquemas de discretización temporal clásicos como son los esquemas de Euler explícitos e implícitos o el de Crank-Nicolson. Pero podemos también aplicar el esquema de Peaceman-Rachford que se puede escribir de manera sintética como sigue

$$\begin{cases} \dfrac{U_h^{n+1/2} - U_h^n}{\Delta t/2} = A_{1h}U^{n+1/2} + A_{2h}U^n \\ \dfrac{U_h^{n+1} - U_h^{n+1/2}}{\Delta t/2} = A_{1h}U_h^{n+1/2} + A_{2h}U_h^{n+1}. \end{cases} \tag{7.4.18}$$

El sistema (7.4.18) consiste esencialmente en dos sistemas desacoplados, semejantes a los que se obtienen al discretizar la ecuación del calor $1-d$. En efecto, tanto la matriz A_{1h} como A_{2h} que intervienen en cada uno de ellos son las matrices habituales en la discretización numérica de la ecuación del calor $1-d$. Se trata por tanto de sistemas tridiagonales fáciles de resolver.

Este método se denomina de *direcciones alternadas* puesto que la ecuación del calor $2-d$ se aproxima mediante la resolución iterada de ecuaciones del calor alternadamente en las variables x_1 y x_2.

Pero volvamos al sistema abstracto general (7.4.8) y a su discretización (7.4.7). En (7.4.5) los operadores A_1 y A_2 juegan papeles totalmente simétricos, pero puede también utilizarse de manera asimétrica. Por ejemplo si aplicamos (7.4.7) para la aproximación de (7.4.8) con $A_1 = A$ y $A_2 = 0$ obtenemos

$$\frac{x^{n+1/2} - x^n}{\Delta t/2} = Ax^{n+1/2}$$

$$\frac{x^{n+1} - x^{n+1/2}}{\Delta t/2} = Ax^{n+1/2}.$$

Se observa por tanto que $x^{n+1/2} = \left(x^{n+1} + x^n\right)\Big/2$, lo cual a su vez implica que

$$\frac{x^{n+1} - x^n}{\Delta t} = A\left(\frac{x^{n+1} + x^n}{2}\right).$$

Tomando $A_1 = 0$ y $A_2 = A$ en la descomposición llegamos exactamente a la misma expresión. Pero esto sólo es cierto cuando el sistema considerado es

autónomo. En efecto, cuando el operador A depende también de t, mientras que la primera elección conduce al sistema

$$\frac{x^{n+1}-x^n}{\Delta t} = A\left(\left(n+\frac{1}{2}\right)\Delta t\right)\left(\frac{x^{n+1}+x^n}{2}\right) \tag{7.4.19}$$

la segunda proporciona

$$\frac{x^{n+1}-x^n}{\Delta t} = \frac{1}{2}\left[A(nt)x^n + A((n+1)\Delta t)x^{n+1}\right]. \tag{7.4.20}$$

Un análisis más cuidadoso indica que, para que ambos esquemas coincidan, no sólo es necesario que A sea independiente de t, i.e. que el sistema sea autónomo, sino también que A sea lineal. En el caso en que A sea no-lineal la diferencia entre (7.4.19) y (7.4.20) es obvia.

Ambos esquemas son esquemas de tipo Crank-Nicolson y son de orden dos cuando el operador A es suficientemente regular tanto en t como en x.

Estos esquemas pueden también escribirse como

$$\left\{\begin{array}{rcl} x^{n+1/2} &=& x^n + \dfrac{\Delta t}{2}A\left(x^{n+1/2},\,(n+1/2)\Delta t\right) \\ x^{n+1} &=& x^n + \Delta t A\left(x^{n+1/2},\,\left(n+\dfrac{1}{2}\right)\Delta t\right) \\ &=& x^n + 2\left(x^{n+1/2}-x^n\right) \end{array}\right. \tag{7.4.21}$$

y

$$x^{n+1} = x^n + \frac{\Delta t}{2}\left[A\left(x^n,\,n\Delta t\right) + A\left(x^{n+1},\,(n+1)\Delta t\right)\right], \tag{7.4.22}$$

respectivamente. Se trata pues de esquemas semi-implícitos de Runge-Kutta de orden 2.

7.4.2. Douglas-Rachford

Existen otras variantes del esquema de descomposición considerado. Tenemos por ejemplo el esquema de Douglas-Rachford que, conocido x^n, calcula el valor de $\hat{x}^{n+1}$ y x^{n+1}, siendo x^{n+1} la aproximación del estado x en el tiempo posterior y $\hat{x}^{n+1}$ un valor auxiliar. El esquema es de la forma

$$\left\{\begin{array}{rcl} \dfrac{\hat{x}^{n+1}-x^n}{\Delta t} &=& A_1\left(\hat{x}^{n+1},\,(n+1)\Delta t\right) + A_2\left(x^n,\,n\Delta t\right) \\ \dfrac{x^{n+1}-x^n}{\Delta t} &=& A_1\left(\hat{x}^{n+1},\,(n+1)\Delta t\right) + A_2\left(x^{n+1},\,(n+1)\Delta t\right). \end{array}\right. \tag{7.4.23}$$

La convergencia de este esquema es conocida en un contexto muy general de operadores monótonos A_1 y A_2. Analicemos, como en el caso anterior, la convergencia en el caso más sencillo en que A es una matriz simétrica $N \times N$.

En este caso obtenemos, con la descomposición (7.4.12),

$$x^{n+1} = (I - \beta\Delta t A)^{-1}(I - \alpha\Delta t A)^{-1}(I - \alpha\beta \mid \Delta t \mid^2 A^2)x^n,$$

o, lo que es lo mismo,

$$x^n = (I - \beta\Delta t A)^{-n}(I - \alpha\Delta t A)^{-n}(I - \alpha\beta \mid \Delta t \mid^2 A^2)x^0.$$

Aplicando el esquema en cada una de las direcciones propias de la matriz A deducimos que

$$x_j^k = \frac{\left(1 + \alpha\beta \mid \Delta t \mid^2 \lambda_j^2\right)^k}{(1 + \alpha\Delta t\lambda_j)^k(1 + \beta\Delta t\lambda_j)^k} x_{0j},\ j = 1, \cdots, N,\ k \geq 1.$$

La funcional racional asociada al esquema es por tanto en este caso

$$R_2(\xi) = \frac{1 + \alpha\beta\xi^2}{(1 - \alpha\xi)(1 - \beta\xi)}.$$

Como $0 < R_2(\xi) < 1$ para todo $\xi < 0$, se deduce que

$$\left|x_j^k\right| \leq \left|x_j^0\right|,\ \forall j = 1, \cdots, N,\ k \geq 1,$$

lo cual implica la estabilidad incondicional del esquema. Sin embargo, como

$$R_2(\xi) = 1 + \xi + \xi^2 + O(\xi^3)$$

se observa que este esquema de Douglas-Rachford es sólo consistente de orden 1.

Nuevamente tenemos que

$$R_2(\xi) \to 1,\ \xi \to -\infty$$

lo cual indica que el esquema tendrá un mal comportamiento para los sistemas "stiff" en lo que respecta la estabilidad absoluta.

Conviene observar que en este esquema los papeles jugados por A_1 y A_2 no son simétricos.

Este esquema es más fácil de generalizar al caso de descomposiciones en un mayor número de operadores que el esquema de Peaceman-Rachford.

Si en (7.4.23) tomamos $A_1 = 0$ y $A_2 = A$ o $A_1 = A$ y $A_2 = 0$, obtenemos en ambos casos el esquema de Euler implícito.

7.4.3. θ-método

Consideramos por último el θ-método introducido por Glowinski. Supuesto conocido x^k, calculamos $x^{k+\theta}$, $x^{k+1-\theta}$ y x^{k+1} del siguiente modo:

$$\left\{\begin{array}{lcl} \dfrac{x^{k+\theta}-x^k}{\theta\Delta t} & = & A_1\left(x^{k+\theta},\,(k+\theta)\Delta t\right)+A_2(x^k,\,k\Delta t),\\ \dfrac{x^{k+1-\theta}-x^{k+\theta}}{(1-2\theta)\Delta t} & = & A_1\left(x^{k+\theta},\,(k+\theta)\Delta t\right)+A_2\left(x^{k+1-\theta},\,(k+1-\theta)\Delta t\right),\\ \dfrac{x^{k+1}-x^{k+1-\theta}}{\theta\Delta t} & = & A_1\left(x^{k+1},\,(n+1)\Delta t\right)+A_2\left(x^{k+1-\theta},\,(k+1-\theta)\Delta t\right). \end{array}\right. \tag{7.4.24}$$

Conviene distinguir este esquema del habitual θ-esquema, intermedio entre los esquemas de Euler explícito e implícito.

Analicemos el esquema (7.4.24) en el caso en que A es una matriz simétrica. Tenemos en este caso

$$x^{k+1}=(I-\alpha\theta\Delta tA)^{-2}(I+\beta\theta\Delta tA)^2(I-\beta\theta'\Delta tA)^{-1}(I+\alpha\theta'\Delta tA)x^k, \tag{7.4.25}$$

donde $\theta'=1-2\theta$, de modo que

$$x_j^k=\frac{(1+\beta\theta\Delta t\lambda_j)^{2k}(1+\alpha\theta'\Delta t\lambda_j)^k}{(1-\alpha\theta\Delta t\lambda_j)^{2k}(1-\beta\theta'\Delta t\lambda_j)^k}e_j. \tag{7.4.26}$$

La función racional característica del esquema es por tanto en este caso

$$R_3(\xi)=\frac{(1+\beta\theta\xi)^2(1+\alpha\theta'\xi)}{(1-\alpha\theta\xi)^2(1-\beta\theta'\xi)}. \tag{7.4.27}$$

Como

$$\lim_{\xi\to-\infty}R_\xi(3)=\beta/\alpha, \tag{7.4.28}$$

para alcanzar la estabilidad del esquema es necesario imponer la condición $\alpha\geq\beta$ y la condición $\alpha>\beta$ para alcanzar la estabilidad absoluta.

La estabilidad incondicional del esquema exige que

$$|R_3(\xi)|\leq 1,\ \forall\xi\in\mathbb{R}^-. \tag{7.4.29}$$

Un análisis más cuidadoso permite probar que

$$|R_3(\xi)|<1,\ \forall\xi\in\mathbb{R}^- \tag{7.4.30}$$

se verifica al menos siempre y cuando θ,α y β están en el rango

$$\theta\in[1/4,\,1/2),\ 0<\beta<\alpha<1,\ \alpha+\beta=1. \tag{7.4.31}$$

Por otra parte

$$R_3(\xi) = 1 - \xi + \frac{\xi^2}{2}\left[1 + (\beta - \alpha)\left(2\theta^2 - 4\theta + 1\right)\right] + O(\xi^3), \tag{7.4.32}$$

de modo que el esquema es consistente de orden dos sí y sólo sí

$$\alpha = \beta = 1/2 \tag{7.4.33}$$

o bien

$$\theta = 1 - 1/\sqrt{2}. \tag{7.4.34}$$

Si se elige $\alpha = \beta = 1/2$ se pierde la estabilidad absoluta. Conviene pues elegir θ según (7.4.34).

A la hora de aplicar el θ-método en el contexto de las EDP, es conveniente que en cada una de las tres ecuaciones de (7.4.24) la ecuación sea la misma. Esto conduce a la condición

$$\alpha\beta = \beta(1 - 2\theta), \tag{7.4.35}$$

lo cual implica

$$\alpha = (1 - 2\theta)/(1 - \theta);\ \beta = \theta\Big/(1 - \theta). \tag{7.4.36}$$

Combinando (7.4.36) con la condición $\alpha > \beta$ deducimos que $0 < \theta < 1/3$.

Un análisis más cuidadoso muestra la existencia de $\theta^*(\theta^* = 0{,}087385580, \cdots,)$ de modo que para $\theta^* < \theta < 1/3$ el esquema es incondicionalmente y absolutamente estable. Con la elección $\theta = 1 - 1/\sqrt{2} = 0{,}292893219\cdots$ se garantiza asimismo que el esquema sea de orden dos.

7.5. Descripción del MDD en varias dimensiones espaciales

Consideremos el problema

$$\begin{cases} -\Delta u = 0 & \text{en} \quad \Omega \\ u = g & \text{en} \quad \partial\Omega, \end{cases} \tag{7.5.1}$$

y describamos el algoritmo de aplicación del MDD.

Descomponemos el dominio Ω en dos subdominios Ω_- y Ω_+ de Ω que cubren Ω con una zona de solapamiento $\Omega_- \cap \Omega_+$ como se indica en la siguiente figura.

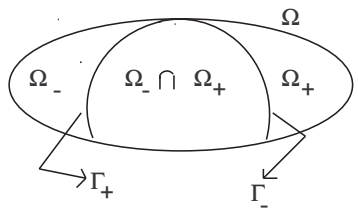


Figura 7.3: Descomposición de dominios en dos dimensiones espaciales.

Definimos entonces la secuencia

$$\begin{cases} -\Delta u_-^k = 0 \text{ en } \Omega_- \\ u_-^k\Big|_{\partial\Omega\cap\partial\Omega_-} = g \\ u_-^k\Big|_{\Gamma_-} = u_+^{k-1}\Big|_{\Gamma_-} \end{cases} \tag{7.5.2}$$

$$\begin{cases} -\Delta u_+^k = 0 \text{ en } \Omega_+ \\ u_+^k\Big|_{\partial\Omega\cap\partial\Omega_+} = g \\ u_+^k\Big|_{\Gamma^+} = u_-^k\Big|_{\Gamma_+} \end{cases} \tag{7.5.3}$$

donde Γ_-y Γ_+ son respectivamente las partes de las fronteras de Ω_- y Ω_+ contenidas en Ω.

A partir de una inicialización

$$u_+^0 \in H^{1/2}(\Gamma_-) \tag{7.5.4}$$

este procedimiento da lugar a una sucesión de soluciones

$$\left\{u_-^k\right\}_{k\geqslant 1} \subset H^1(\Omega_-);\ \left\{u_+^k\right\}_{k\geqslant 1} \subset H^1(\Omega_+). \tag{7.5.5}$$

La inicialización $u_+^0 \in H^{1/2}(\Gamma_-)$ puede por ejemplo obtenerse como restricción o traza sobre Γ_- de la función u^* de $H^1(\Omega)$ que prolonga los valores de frontera g al interior de Ω.

Pretendemos ahora probar la convergencia de u_-^k (resp. u_+^k) a u en Ω_- (resp. Ω_+) cuando $k \to \infty$. Lo hacemos mediante la aplicación del principio del máximo.

Definimos

$$v_-^k = u - u_-^k;\ v_+^k = u - u_+^k \tag{7.5.6}$$

que satisfacen

$$\begin{cases} -\Delta v_-^k = 0 \text{ en } \Omega_- \\ v_-^k\Big|_{\partial\Omega\cap\partial\Omega_-} = 0 \\ v_-^k\Big|_{\Gamma_-} = v_+^{k-1}\Big|_{\Gamma_-} \end{cases} \tag{7.5.7}$$

$$\begin{cases} -\Delta v_+^k = 0 \qquad \text{en} \quad \Omega_+ \\ v_+^k\Big|_{\partial\Omega\cap\partial\Omega_+} = 0 \\ v_+^k\Big|_{\Gamma_+} = v_-^k\Big|_{\Gamma_+}. \end{cases} \tag{7.5.8}$$

Por el principio del máximo probado en la sección anterior tenemos que

$$\| v_-^k \|_{L^\infty(\Omega_-)} \leqslant \| v_+^{k-1} \|_{L^\infty(\Gamma_-)} \tag{7.5.9}$$

$$\| v_+^k \|_{L^\infty(\Omega_+)} \leqslant \| v_-^k \|_{L^\infty(\Gamma_+)} . \tag{7.5.10}$$

Con el objeto de concluir la convergencia es por tanto suficiente probar que

$$\| v_+^k \|_{L^\infty(\Gamma_-)} + \| v_-^k \|_{L^\infty(\Gamma_+)} \to 0,\ k \to \infty. \tag{7.5.11}$$

En este hecho va a jugar un papel determinante el que Γ_+ y Γ_- estén alejadas la una de la otra.

Consideremos el caso particular en que Γ_+ y Γ_- son segmentos paralelos, tal y como se indica en la siguiente figura:

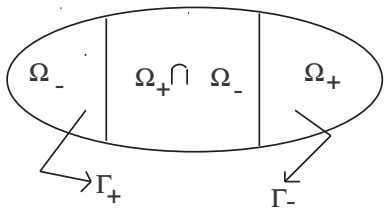


Figura 7.4: Descomposición de dominios con interfases planas paralelas.

Consideramos ahora un problema de la forma (7.5.7) que se verifica en Ω_-:

$$\begin{cases} -\Delta v = 0 & \text{en} \quad \Omega_- \\ v = 0 & \text{en} \quad \partial\Omega \cap \partial\Omega_- \\ v = h & \text{en} \quad \Gamma_-. \end{cases} \tag{7.5.12}$$

La clave de la convergencia del método

$$\| v \|_{L^\infty(\Gamma_+)} \leqslant \gamma \| h \|_{L^\infty(\Gamma_-)},$$

con $0 < \gamma < 1$ independiente de h que calcularemos de manera explícita.

En efecto, por el principio del máximo,

$$v \leqslant V \tag{7.5.13}$$

donde V es la solución de

$$\begin{cases} -\Delta V = 0 & \text{en} \quad \Omega_- \\ V = 0 & \text{en} \quad \partial\Omega \cap \partial\Omega_- \\ V = \| h \|_{L^\infty(\Gamma_-)} & \text{en} \quad \Gamma_-. \end{cases} \tag{7.5.14}$$

A su vez, también por el principio del máximo,

$$V \leqslant W \tag{7.5.15}$$

donde $W = W(x)$ es la función afín que sólo depende de la variable x' perpendicular a Γ_- y tal que $W = 0$ en el punto de Ω_- más alejado de Γ_-.

A partir de (7.5.15) es fácil deducir que se cumple

$$\| v \|_{L^\infty(\Gamma_+)} \leqslant \gamma \| h \|_{L^\infty(\Gamma_-)}$$

con $0 < \gamma < 1$.

En efecto, basta de hecho tomar $\gamma = d/D$ donde d es la distancia entre las dos interfases Γ_- y Γ_+ y D y la distancia de Γ_- al punto de $\bar{\Omega}_-$ más alejado de Γ_-.

7.6. MDD para las diferencias finitas multi-d

En esta sección vamos a extender el estudio realizado de la convergencia del MDD en el marco de las aproximaciones por diferencias finitas en el caso de una dimensión espacial, al caso de varias dimensiones.

Con el objeto de simplificar la presentación vamos a considerar el caso de un dominio cuadrado en el plano, aunque las ideas que aquí vamos a desarrollar se adaptan con facilidad al caso de dominios generales en cualquier dimensión.

Consideramos por tanto el dominio

$$\Omega = (-1, 1) \times (-1, 1). \tag{7.6.1}$$

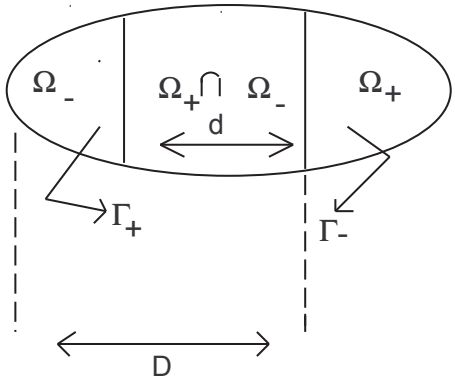


Figura 7.5: Representaciones de las dimensiones d y D que intervienen en determinación de la velocidad de convergencia del MDD.

Introducimos un mallado de paso $h > 0$ y los nodos $x_{j,k}$ de coordenadas

$$x_{j,k} = (jh,\, kh),\ -(N+1) \leqslant j,\, k \leqslant N+1 \tag{7.6.2}$$

con $N + 1 = 1/h$ que supondremos un número entero, lo cual en la práctica supone elegir h de la forma $h = 1/(N+1)$ donde N es un número natural.

Con esta partición, los nodos de la frontera $\partial\Omega$ corresponden a los índices $j = -(N+1),\, N+1$ y $k = -(N+1),\, N+1$.

Introducimos ahora la aproximación por diferencias finitas del problema de Dirichlet en el dominio Ω:

$$\begin{cases} \dfrac{4u_{j,k} - u_{j+1,\,k} - u_{j-1,\,k} - u_{j,\,k+1} - u_{j,\,k-1}}{h^2} = f_{j,k}, \quad -N \leqslant j,\, k \leqslant N \\ u_{j,k} = 0 \text{ si } j = -(N+1),\, N+1 \text{ o } k = -(N+1),\, N+1. \end{cases} \tag{7.6.3}$$

Se trata efectivamente de una aproximación del problema de Dirichlet:

$$\begin{cases} -\Delta u = f & \text{en} \quad \Omega \\ u = 0 & \text{en} \quad \partial\Omega. \end{cases} \tag{7.6.4}$$

Es bien sabido que (7.6.3) proporciona una aproximación convergente de orden dos de las soluciones de (7.6.4). Son diversas las pruebas de este hecho, una de ellas basada en el principio del máximo discreto que presentaremos en breve (véase [13] para más detalles).

El problema de Dirichlet con condiciones de contorno no homogéneas puede aproximarse del mismo modo. En efecto, si la ecuación considerada es

$$\left\{ \begin{array}{rcl} -\Delta u = 0 & \text{en} & \Omega \\ u = g & \text{en} & \partial\Omega \end{array} \right. \tag{7.6.5}$$

entonces el esquema discretizado correspondiente es de la forma

$$\left\{ \begin{array}{l} \dfrac{4u_{j,\,k} - u_{j+1,\,k} - u_{j-1,\,k} - u_{j,\,k+1} - u_{j,\,k-1}}{h^2} = 0, \quad -N \leqslant j,\, k \leqslant N \\ \\ u_{j,\,k} = g_{j,\,k},\ j = -(N+1),\, N+1;\ k = -(N+1),\, N+1. \end{array} \right. \tag{7.6.6}$$

Aquí y en lo sucesivo $f_{j,\,k}$ y $g_{j,\,k}$ proporcionan aproximaciones de f y g en los puntos $x_{j,\,k}$ del mallado. Cuando las funciones en cuestión son continuas basta con tomar su valor en dicho nodo. Cuando no lo son podemos por ejemplo tomar una media de las mismas en un cuadrado de lado h centrado en dicho nodo.

En ambos casos, la función $u_{j,\,k}$ solución del problema discreto proporciona una aproximación de la solución u del problema continuo en los nodos $x_{j,\,k}$.

Los dos sistemas discretos son de hecho sistemas lineales de $N \times N$ ecuaciones con $N \times N$ incógnitas que, escritos en forma matricial, están asociados a matrices simétricas definidas positivas (para un estudio más detallado de estas cuestiones tanto en el marco de la ecuación de Laplace como las de evolución puede consultarse las notas [36]).

En esta sección vamos a probar la convergencia del método de descomposición de dominios en el caso discreto, i.e. con $h > 0$ fijo.

El algoritmo es idéntico al caso de la ecuación de Laplace continua y la demostración de convergencia también. Basta de hecho aplicar el principio del máximo para la ecuación discreta. Su enunciado es idéntico al del caso continuo. Con el objeto de presentarlo de manera más sintética dado un vector $\{u_{j,\,k}\}_{-(N+1)\leqslant j,\,k\leqslant N+1}$ arbitrario introducimos

$$M_\Omega = \max_{-N\leqslant j,\,k\leqslant N} \left[\frac{4u_{j,\,k} - u_{j+1,\,k} - u_{j-1,\,k} - u_{j,\,k+1} - u_{j,\,k-1}}{h^2} \right]$$

$$m_\Omega = \min_{-N\leqslant j,\,k\leqslant N} \left[\frac{4u_{j,\,k} - u_{j+1,\,k} - u_{j-1,\,k} - u_{j,\,k+1} - u_{j,\,k-1}}{h^2} \right]$$

$$M_\Gamma = \max_{\substack{j=-(N+1),\,N+1,\,-(N+1)\leqslant k\leqslant N+1 \\ -(N+1)\leqslant j\leqslant N+1;\,k=-(N+1),\,N+1}} [u_{j,\,k}]$$

$$m_\Gamma = \min_{\substack{j=-(N+1),\,N+1,\,-(N+1)\leqslant k\leqslant N+1 \\ -(N+1)\leqslant j\leqslant N+1;\,k=-(N+1),\,N+1}} [u_{j,\,k}]\,.$$

Entonces, para cualquier vector $u_{j,\,k}$ tenemos

$$\text{mín}\,[m_\Omega,\, m_\Gamma] \leqslant u_{j,\,k} \leqslant \text{máx}\,[M_\Omega,\, M_\Gamma]\,, \quad -(N+1) \leqslant j,\, k \leqslant N+1.$$

Este hecho es independiente del valor de h.

A partir de este principio del máximo discreto, la convergencia del MDD en el marco discreto puede probarse de manera idéntica a como lo hicimos en el caso continuo.

Supongamos por ejemplo que introducimos las interfases $\Gamma_- = \{(1/2,\, y),\, -1 \leqslant y \leqslant 1\}$, y $\Gamma_+ = \{(-1/2,\, y),\, -1 \leqslant y \leqslant 1\}$:

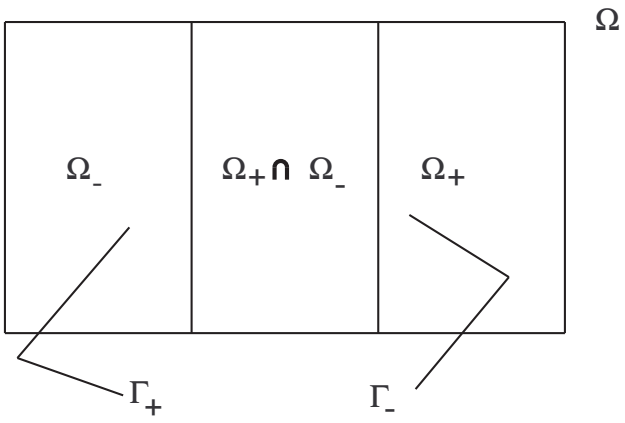


Figura 7.6: Descomposición de un dominio rectangular en subdominios rectangulares.

Es fácil entonces construir funciones discretas $W_{j,\,k}$ tales que

$$\begin{cases} \dfrac{4W_{j,\,k} - W_{j+1,\,k} - W_{j-1,\,k} - W_{j,\,k+1} - W_{j,\,k-1}}{h^2} = 0, \qquad -N \leqslant j,\, k \leqslant N \\ W_{j,\,k} = C \text{ en } \Gamma_- \end{cases}$$

donde $C > 0$ es una constante arbitraria y de modo que

$$W \geqslant 0 \text{ en } \partial\Omega_- \backslash \Gamma_-.$$

Basta por ejemplo considerar una función afín que depende exclusivamente de x_1. Obviamente el valor de esta función W a lo largo de la interfase Γ_+ es entonces Cd/D. Utilizando este tipo de función para obtener mayoraciones de las sucesiones de soluciones obtenidas en la aplicación del MDD deducimos con facilidad la convergencia exponencial del método.

Probamos por último el principio del máximo discreto. Sin pérdida de generalidad, basta probar la cota superior, puesto que la inferior se prueba de manera análoga. Asimismo, podemos suponer que $M_\Omega \leqslant 0$. Basta para ello constatar que la discretización de la función parabólica $v(x,\, y) = x^2 + y^2$ es tal que

$$\frac{4V_{j,\,k} - V_{j+1,\,k} - V_{j-1,\,k} - V_{j,\,k+1} - V_{j,\,k-1}}{h^2} = -4,$$

para todo j y k.

Sin pérdida de generalidades podemos entonces conseguir que $M_\Omega \leq 0$ sumando o sustrayendo a la función u en consideración una constante veces la función parabólica v.

Consideremos por tanto el caso $M_\Omega \leqslant 0$. Basta entonces probar que el máximo se alcanza en la frontera. Esto es fácil de comprobar. Supongamos que se alcanza en un nodo interior $(j,\, k)$. Entonces

$$u_{j,\,k} \geqslant \text{máx}\left\{u_{j+1,\,k}, u_{j-1,\,k}, u_{j,\,k+1},\, u_{j,\,k-1}\right\},$$

de modo que

$$\frac{4u_{j,\,k} - u_{j+1,\,k} - u_{j-1,\,k} - u_{j,\,k+1} - u_{j,\,k-1}}{h^2} \geqslant 0.$$

Como $M_\Omega \leqslant 0$ deducimos que, necesariamente,

$$4u_{j,\,k} = u_{j+1,\,k} + u_{j-1,\,k} + u_{j,\,k+1} + u_{j,\,k-1}.$$

De este modo se obtiene que si el máximo de u se alcanza en un nodo interno se alcanza también en los cuatro vecinos. Iterando el argumento vemos que necesariamente se ha de alcanzar también en el borde, tal y como queríamos probar.

Capítulo 8

Métodos de descenso

8.1. El método directo del Cálculo de Variaciones

En numerosas ocasiones las soluciones de problemas elípticos pueden obtenerse mediante el Método Directo del Cálculo de Variaciones (MDCV). [1]

Recordemos brevemente este principio. Consideramos un funcional convexo y coercivo $J : H \to \mathbb{R}$ en un espacio de Hilbert H. Entonces, el funcional alcanza su mínimo en al menos un punto del espacio:

$$\exists h \in H : \, J(h) = \min_{g \in H} J(g). \tag{8.1.1}$$

Para comprobarlo basta proceder del modo siguiente:

Paso 1. Se define el ínfimo

$$I = \inf_{g \in H} J(g) \tag{8.1.2}$$

que, por la coercividad de J, necesariamente satisface $I > -\infty$.

Paso 2. Se construye una sucesión minimizante

$$(g_n)_{n \in \mathbb{N}} \subset H : \, J(g_n) \searrow I. \tag{8.1.3}$$

Por la coercividad del funcional J se deduce que $(g_n)_{n \in \mathbb{N}}$ está acotada en H.

Paso 3. Como H es un espacio de Hilbert, existe una subsucesión, que seguiremos denotanto $(g_n)_{n \in \mathbb{N}}$, que converge débilmente:

$$g_n \rightharpoonup g \text{ en } H. \tag{8.1.4}$$

[1] Para una introducción más detallada del método podrán consultarse las notas [36].

Paso 4. Como J es continuo en H y convexo, es semicontinuo inferiormente para la topología débil. Por tanto

$$J(g) \leqslant \liminf_{n\to\infty} J(g_n). \tag{8.1.5}$$

Deducimos que $J(g) \leqslant I$ lo cual, a su vez, por la definición de ínfimo, garantiza que $J(g) = I$, lo cual demuestra que el ínfimo se alcanza y que, por tanto, se trata de un mínimo.

Este principio es suficiente para resolver muchos problemas elípticos.

Por ejemplo, el problema de Dirichlet

$$\begin{cases} -\Delta u = f & \text{en} \quad \Omega \\ u = 0 & \text{en} \quad \partial\Omega, \end{cases} \tag{8.1.6}$$

puede resolverse fácilmente de este modo cuando Ω es un dominio acotado. Basta aplicar el MDCV al funcional

$$J(v) = \frac{1}{2}\int_\Omega \mid \nabla v \mid^2 dx - \int_\Omega fv dx \tag{8.1.7}$$

en el espacio de Hilbert $H_0^1(\Omega)$.

Cuando $f \in L^2(\Omega)$, por la desigualdad de Poincaré es fácil comprobar que J es continuo, convexo y coercivo. El mínimo se alcanza en un punto $u \in H_0^1(\Omega)$ que resulta ser una solución débil caracterizada por las condiciones

$$\begin{cases} u \in H_0^1(\Omega) \\ \displaystyle\int_\Omega \nabla u \cdot \nabla v dx = \int_\Omega fv dx, \, \forall v \in H_0^1(\Omega). \end{cases} \tag{8.1.8}$$

El mismo principio se aplica en todo espacio de Banach reflexivo. Esto permite por ejemplo resolver problemas elípticos no lineales de la forma

$$\begin{cases} -\Delta u + \mid u \mid^{p-1} = f & \text{en} \quad \Omega \\ u = 0 & \text{en} \quad \partial\Omega, \end{cases} \tag{8.1.9}$$

minimizando el funcional

$$J(u) = \frac{1}{2}\int_\Omega \mid \nabla u \mid^2 dx + \frac{1}{p+1}\int_\Omega \mid u \mid^{p+1} dx - \int_\Omega fu dx \tag{8.1.10}$$

en el espacio de Banach reflexivo $H_0^1(\Omega) \cap L^{p+1}(\Omega)$.

Este MDCV es más que un método demostración de resultados de existencia y puede dar lugar a métodos iterativos de aproximación de soluciones. Se trata de los *métodos de descenso* que en cada paso hacen decrecer el valor del funcional a minimizar hasta alcanzar el mínimo. En las próximas subsecciones presentamos los métodos de máximo descenso y de gradiente conjugado en el contexto de los sistemas lineales en dimensión finita.

8.2. El método del máximo descenso

Con el objeto de ilustrar estos métodos consideramos el sistema lineal algebraico

$$Ax = b \tag{8.2.1}$$

donde b es un vector dado de $\mathbb{R}^N$, A es una matriz simétrica definida positiva $N \times N$ y x es el vector incógnita, también de $\mathbb{R}^N$.

La solución de (8.2.1) es el mínimo del funcional

$$J(x) = \frac{1}{2}\langle Ax,\, x\rangle - \langle b,\, x\rangle \tag{8.2.2}$$

en $\mathbb{R}^N$.

Su gradiente viene dado por

$$\nabla J(x) = Ax - b. \tag{8.2.3}$$

Al aplicar un método de descenso obtendremos una sucesión de puntos en $\mathbb{R}^N$ que tendrán como objetivo aproximar el punto de mínimo $x \in \mathbb{R}^N$. Dado uno de esos puntos $y \in \mathbb{R}^N$, definimos el residuo

$$r = b - Ay. \tag{8.2.4}$$

Es obvio que x es solución de (8.2.1) cuando el residuo se anula.

La sucesión $\{x_k\}$ que el método de descenso proporciona puede definirse del modo siguiente:

$$x_{k+1} = x_k + \alpha_k r_k \tag{8.2.5}$$

donde

$$r_k = b - Ax_k, \tag{8.2.6}$$

siendo α_k una constante que elegirememos de modo que el funcional J decrezca lo más posible.

Por tanto, en este método de descenso, a partir del punto x_k de la iteración anterior avanzamos en la dirección del residuo, pues se trata de la dirección de máxima pendiente del funcional.

Con el objeto de elegir el valor más adecuado de α_k calculamos $J(x_{k+1})$. Tenemos

$$\begin{aligned} J(x_{k+1}) &= J(x_k) + \frac{\alpha_k^2}{2}\langle Ar_k,\, r_k\rangle - \alpha_k\langle r_k,\, b\rangle + \alpha_k\langle Ax_k,\, r_k\rangle \\ &= J(x_k) - \alpha_k\langle r_k,\, r_k\rangle \; + \frac{\alpha_k^2}{2}\langle Ar_k,\, r_k\rangle. \end{aligned}$$

Calculamos el valor óptimo de α_k imponiendo la condición $\partial J/\partial \alpha_k = 0$. Obtenemos así

$$\alpha_k = \frac{\langle r_k,\, r_k\rangle}{\langle Ar_k,\, r_k\rangle} = \frac{\mid r_k \mid^2}{\langle Ar_k,\, r_k\rangle}. \tag{8.2.7}$$

De esta expresión deducimos a su vez que

$$r_{k+1} = r_k - \alpha_k\, Ar_k, \tag{8.2.8}$$

lo cual permite actualizar el valor del residuo sin tener que computarlo según su definición.

Conviene también observar que de (8.2.7)-(8.2.8) se deduce que

$$\langle r_{k+1},\, r_k\rangle = \langle r_k,\, r_k\rangle - \alpha_k \langle Ar_k,\, r_k\rangle = 0. \tag{8.2.9}$$

Esto significa que el residuo es siempre ortogonal al del paso anterior. Es por eso que en este método del descenso la dirección de avance es siempre ortogonal a la del paso previo, lo cual es coherente con el hecho de descender lo más posible en cada paso.

Obtenemos asímismo

$$J(x_{k+1}) = J(x_k) - \frac{1 \mid r_k \mid^2}{2\langle Ar_k,\, r_k\rangle}, \tag{8.2.10}$$

lo cual confirma que el funcional J decrece.

De manera sintética, el método del descenso puede escribirse como

$$\begin{cases} x_{k+1} = x_k + \alpha_k x_k \\ r_{k+1} = r^k - \alpha_k Ar_k \\ \alpha_k = \mid r_k \mid^2 \Big/ \langle Ar_k,\, r_k\rangle. \end{cases} \tag{8.2.11}$$

La sucesión $\{x_k\}$ que el método del descenso proporciona tiene necesariamente como punto de acumulación la solución de la ecuación (8.2.1).

En efecto, como J decrece a lo largo de la sucesión $\{x_k\}$ y J es coerciva, deducimos que $\{x_k\}$ está acotada. En vista de la expresión (8.2.10) cualquier punto de acumulación de la sucesión $\{x_k\}$ ha de ser de residuo nulo y por tanto solución de (8.2.1). Como la solución de (8.2.1) es única deducimos que toda la sucesión ha de converger a la misma.

Este argumento no es más que una adaptación del clásico principio de invarianza de LaSalle para sistemas dinámicos dotados de una funcional Lyapunov. Recordemos brevemente la argumentación. Como $J(x_k)$ decrece y está acotada inferiormente su límite ℓ existe:

$$\ell = \lim_{k\to\infty} J(x_k).$$

Para cualquier punto de acumulación y de la sucesión, como $x_{k_j} \to y$ tenemos, por la continuidad de J,

$$\ell = J(y) = \lim_{j\to\infty} J\left(x_{k_j}\right).$$

Podemos aplicar una vez más el algoritmo de descenso (8.2.11) a partir de y, para obtener y_1. Ahora bien, y_1 será entonces el límite de la sucesión x_{k_j+1} obtenida a partir de la anterior x_{k_j} trasladando el índice de una unidad o, lo que es lo mismo, aplicando el algoritmo (8.2.11) a la misma.

Entonces $x_{k_j+1} \to y$ y, por consiguiente,

$$J(y_1) = \lim J\left(x_{k_j+1}\right) = \ell.$$

Por lo tanto $J(y) = J(y_1) = \ell$ y, en virtud de (8.2.11), esto sólo es posible si el residuo correspondiente al punto de acumulación y se anula. Es decir, si y es solución de (8.2.1).

Más adelante vamos a dar una prueba más cuantitativa de la convergencia del método. Antes de hacerlo conviene observar que en la implementación del método (8.2.11) sólo hemos de aplicar el operador A una sola vez en el cálculo de Ar_k.

A pesar de que el método ha sido desarrollado en el caso en que A es simétrica definida positiva, el método converge en un marco mucho más general.

Theorem 8.2.1 *Si A es definida positiva y si A^tA^{-1} también lo es, el método del descenso* (8.2.11) *converge a la única solución de* (8.2.1) *para cualquier valor x_0 de la inicialización.*

Observación 8.2.1 Obviamente, si A es simétrica y definida positiva, A^tA^{-1} también lo es puesto que, en este caso, $A^tA^{-1} = I$. ■

Demostración 8.2.1 En primer lugar observamos que si A es definida positiva, A^{-1} también lo es. Por otra parte, como A^tA^{-1} es definida positiva, existe $c_0 > 0$ tal que

$$c_0\langle x,\, A^{-1}x\rangle \leqslant \langle x,\, A^tA^{-1}x\rangle \tag{8.2.12}$$

y $c_1 > 0$ tal que

$$c_1\langle x,\, Ax\rangle \leqslant \langle x,\, x\rangle. \tag{8.2.13}$$

Consideramos ahora la cantidad $\langle r_k,\, A^{-1}r_k\rangle$. Tenemos, por la definición de α_k,

$$\begin{aligned}\langle r_{k+1},\, A^{-1}r_{k+1}\rangle &= \langle r^k,\, A^{-1}r_k\rangle - \alpha_k\langle r_k,\, r_k\rangle - \alpha_k\langle Ar_k,\, A^{-1}r_k\rangle \\ &\quad + \alpha_k\langle r_k,\, Ar_k\rangle = \langle r_k,\, A^{-1}r_k\rangle - \alpha_k\langle r_k, A^tA^{-1}r_k\rangle.\end{aligned} \tag{8.2.14}$$

Por otra parte, de (8.2.13) deducimos que

$$\alpha_k \geqslant c_1, \tag{8.2.15}$$

y, por tanto, por (8.2.12),

$$\langle r_{k+1},\, A^{-1}r_{k+1}\rangle \leqslant (1 - c_0c_1)\,\langle r_k,\, A^{-1}r_k\rangle. \tag{8.2.16}$$

La constante $1 - c_0c_1$ es positiva. Basta para ello observar que, como A y A^{-1} son definidas positivas, si (8.2.12) y (8.2.13) se cumplen, también se cumplen para valores menores de c_0 y c_1.

Iterando (8.2.16) obtenemos

$$\langle r_k,\, A^{-1}r_k\rangle \leqslant (1 - c_0c_1)^k\,\langle r_0,\, A^{-1}r_0\rangle, \tag{8.2.17}$$

de donde se deduce que $\langle r_k,\, A^{-1}r_k\rangle$ tiende a cero. Como A^{-1} es definida positiva, esto implica que r_k tiende a cero. Es decir $b - Ax_k$ tiende a cero, lo cual implica que x_k tiende a $A^{-1}b$, la única solución de (8.2.1). ■

Observación 8.2.2 La demostración anterior muestra que la convergencia del método se acelera a medida que c_0c_1 se aproxima a 1. Esto ocurre cuando A es próxima a la matriz identidad. Pero en general el método puede converger muy lentamente por la tendencia de los residuos a oscilar. En efecto, a pesar de que r_{k+1} es ortogonal a r_k nada impide que r_{k+2} deje de serlo, lo cual puede conducir a las oscilaciones aludidas. ■

Observación 8.2.3 La prueba de la convergencia que hemos desarrollado está basada en el análisis de la cantidad $\left(r_k,\, A^{-1}r_k\right)$. Esta cantidad es relevante puesto que

$$J(x_k) = \frac{1}{2}\langle A^{-1}r_k,\, r_k\rangle + r(0) \tag{8.2.18}$$

donde

$$F(y) = \frac{1}{2}\langle y - x,\, A(y - x)\rangle, \tag{8.2.19}$$

siendo $x = A^{-1}b$ la solución de (8.2.1).■

8.3. El método del gradiente conjugado

El método del gradiente conjugado es una de las maneras más habituales de acelerar la convergencia del método del descenso.

En este caso elegimos la siguiente iteración:

$$x_{k+1} = x_k + \alpha_k\,[r_k + \gamma_k\,(x_k - x_{k-1})]$$

para parámetros α_k y γ_k a determinar.

Según esta fórmula, el nuevo cambio de posición $x_{k+1} - x_k$ es combinación del residuo r_k que es la dirección de pendiente máxima y el cambio de dirección en el paso anterior.

Escribimos la fórmula anterior como

$$x_{k+1} = x_k + \alpha_k p_k \tag{8.3.1}$$

donde

$$\begin{array}{rcl} p_k & = & r_k + \gamma_k \left(x_k - x_{k-1}\right) = r_k + \gamma_k \alpha_{k-1} p_{k-1} \\ & & = r_k + \beta_{k-1} p_{k-1}. \end{array} \tag{8.3.2}$$

Obtenemos así

$$\left\{ \begin{array}{rcl} x_{k+1} & = & x_k + \alpha_k p_k \\ r_{k+1} & = & r^k - \alpha_k \, A p_k \\ p_{k+1} & = & r_{k+1} + \beta_k p_k. \end{array} \right. \tag{8.3.3}$$

El vector p_k se denomina *dirección de búsqueda*.

Necesitamos determinar α_k, β_k y p_0. Nuevamente, hemos de hacerlo de modo que $J(x_{k+1})$ sea mínimo.

Optimizando el valor de α_k obtenemos

$$\alpha_k = \frac{\langle p_k, \, r_k \rangle}{\langle p_k, \, A p_k \rangle}. \tag{8.3.4}$$

De este modo

$$J(x_{k+1}) = J(x_k) - \frac{\langle p_k, \, r_k \rangle^2}{2 \langle p_k, \, A p_k \rangle}. \tag{8.3.5}$$

De esta forma se oberva que r_0 es una buena elección de p_0 pues garantiza el decrecimiento del valor de J.

Combinando (8.3.4) con la segunda identidad de (8.3.3) obtenemos

$$\langle p_k, \, r_{k+1} \rangle = \langle p_k, \, r_k \rangle - \alpha_k \langle p_k, \, A p_k \rangle = 0. \tag{8.3.6}$$

Ahora, de (8.3.6) y de la tercera identidad de (8.3.3) obtenemos

$$\langle p_{k+1}, \, r_{k+1} \rangle = \langle r_{k+1}, \, r_{k+1} \rangle + \beta_k \langle p_k, \, r_{k+1} \rangle = \mid r_{k+1} \mid^2 . \tag{8.3.7}$$

Como hemos elegido $p_0 = r_0$ deducimos que

$$\langle p_k, \, r_k \rangle = \mid r_k \mid^2, \, \forall k \geqslant 0$$

y por lo tanto

$$\alpha_k = \frac{\mid r_k \mid^2}{\langle p_k, \, A p_k \rangle}, \tag{8.3.8}$$

$$J(x_{k+1}) = J(x_k) - \frac{|\ r_k\ |^4}{2\langle p_k,\ Ap_k\rangle}. \tag{8.3.9}$$

En vista de (8.3.9), deberíamos elegir p_k de modo que $\langle p_k,\ Ap_k\rangle$ fuese mínima.

Tenemos

$$\langle p_k,\ Ap_k\rangle = \langle r_k,\ Ar_k\rangle + 2\beta_{k-1}\langle r_k,\ Ap_{k-1}\rangle + \beta_{k-1}^2\langle p_{k-1},\ Ap_{k-1}\rangle.$$

Por tanto, la elección óptima de β_{k-1} es

$$\beta_k = -\frac{\langle r_{k+1},\ Ap_k\rangle}{\langle p_k,\ Ap_k\rangle}.$$

De este modo observamos que

$$\langle p_{k+1},\ Ap_k\rangle = \langle r_{k+1},\ Ap_k\rangle + \beta_k\langle p_k,\ Ap_k\rangle = 0$$

y por consiguiente

$$\langle p_{k+1},\ Ap_k\rangle = 0.$$

Esto significa que las sucesivas direcciones de búsqueda verifican una condición de conjugación.

Por otra parte,

$$\langle p_k,\ Ap_k\rangle = \langle r_k,\ Ap_k\rangle + \beta_{k+1}\langle p_{k-1},\ Ap_k\rangle = \langle r_k,\ Ap_k\rangle$$

y por tanto

$$\langle r_{k+1},\ r_k\rangle = \langle r_k,\ r_k\rangle - \alpha_k\langle Ap_k,\ r_k\rangle = \langle r_k,\ r_k\rangle - \alpha_k\langle p_k,\ Ap_k\rangle = 0. \tag{8.3.10}$$

Esto permite reescribir β_k. En efecto,

$$\langle r_{k+1},\ r_{k+1}\rangle = \langle r_{k+1},\ r_k\rangle - \alpha_k\langle r_{r+1},\ Ap_k\rangle = -\alpha_k\langle r_{k+1},\ Ap_k\rangle$$

y por tanto

$$\beta_k = \frac{1}{\alpha_k}\frac{|\ r_{k+1}\ |^2}{\langle p_k,\ Ap_k\rangle} = \frac{|\ r_{k+1}\ |^2}{|\ r_k\ |^2}.$$

El método del gradiente conjugado puede entonces reescribirse del modo siguiente:

$$\begin{aligned}
p_0 &= r_0 = b - Ax_0\\
x_{k+1} &= x_k + \alpha_k\, p_k\\
r_{k+1} &= r_k - \alpha_k\, Ap_k\\
p_{k+1} &= r_{k+1} + \beta_k\, p_k\\
\alpha_k &= |\ r_k\ |^2\ /\langle p_k,\ Ap_k\rangle\\
\beta_k &= |\ r_{k+1}\ |^2\ /\ |\ r_k\ |^2\ .
\end{aligned}$$

Conviene señalar que cuando β_k es pequeño el método es próximo al de descenso máximo.

Para el método de gradiente conjugado se tiene la siguiente propiedad fundamental:

$$\langle r_k,\, r_j\rangle = \langle p_k,\, Ap_j\rangle = 0,\, \forall k \neq j.$$

Com consecuencia de ésto tenemos que:

Theorem 8.3.1 *Si A es una matriz $N \times N$ simétrica y definida positiva, el método de gradiente conjugado converge en, a lo sumo, N pasos.*

Sin embargo, en las aplicaciones prácticas, el valor de N puede resultar muy grande por lo que puede resultar necesario una estimación de convergencia más precisa.

Theorem 8.3.2 *Si la matriz A es simétrica y definida positiva, el error e_k del método de gradiente conjugado satisface:*

$$\langle e_k,\, Ae_k\rangle^{1/2} \leqslant 2\left(\frac{\sqrt{\lambda_{\text{máx}}}-\sqrt{\lambda_{\text{mín}}}}{\sqrt{\lambda_{\text{máx}}}+\sqrt{\lambda_{\text{mín}}}}\right)^k \langle e_0,\, Ae_0\rangle^{1/2},$$

donde $\lambda_{\text{mín}}$ y $\lambda_{\text{máx}}$ son respectivamente los autovalores mínimo y máximo de la matriz A.

De acuerdo con este resultado, la convergencia del método de gradiente conjugado se acelera cuando los autovalores de A están muy cerca unos de otros.

Para un desarrollo más completo de este método el lector podrá consultar el capítulo 14 del libro de Strikwerda [22] y el capítulo 2 del libro de Quarteroni y Valli [19].

8.4. Sistema gradiente en dimensión finita: Convergencia al equilibrio

Hemos visto que al discretizar las ecuaciones de evolución tipo Navier-Stokes y Burgers en tiempo, en cada paso de la iteración temporal, nos encontramos con una ecuación elíptica no-lineal a resolver. Lo mismo ocurre cuando abordamos la resolución de la ecuación de evolución mediante la teoría de semi-grupos no-lineales, como veremos más adelante.

Pero los problemas elípticos subyacentes tiene importancia más allá de estas consideraciones al nivel de la modelización. En efecto, son muchos los casos en los que, con el objeto de simplificar el modelo en consideración, se sustituye la

ecuación de evolución por una ecuación estacionaria, de modo que, en el ámbito de las ecuaciones analizadas en este curso, esto supone pasar de una ecuación parabólica a una elíptica.

Esto supone una simplificación muy importante también desde el punto de vista numérico pues desaparece la variable temporal.

Hay una razón para que esta reducción se pueda hacer en algunas ecuaciones: las soluciones, a medida que $t \to +\infty$, sufren una simplificación asintótica que hace que se parezcan cada vez más a la (o una) solución estacionaria. En esta sección vamos a analizar algunos casos sencillos en los que este hecho puede probarse de manera rigurosa. Obviamente, en la práctica, esta posibilidad de simplificar el modelo para considerarlo estacionario, se utiliza incluso en casos en cuya validez es dudosa. Esto, evidentemente, puede ser una causa para invalidar los resultados obtenidos. Es precisamente por esta razón que es importante conocer las técnicas básicas que permiten justificar esta simplificación al nivel de la modelización.

Consideramos en primer lugar un sistema gradiente en dimensión finita

$$\begin{cases} x'(t) + \nabla H(x(t)) = 0, \quad t > 0 \\ x(0) = 0, \end{cases} \tag{8.4.1}$$

donde

$$H : \mathbb{R}^N \to \mathbb{R} \tag{8.4.2}$$

es una función de clase C^2, convexa y que alcanza su valor mínimo en un único punto $x^* \in \mathbb{R}^N$:

$$H(x^*) = \min_{x \in \mathbb{R}^N} H(x). \tag{8.4.3}$$

Obviamente, se tiene,

$$\nabla H(x^*) = 0, \tag{8.4.4}$$

por lo que x^* es una solución estacionaria de (8.4.1). En realidad, x^* es la única solución estacionaria de (8.4.1) puesto que serlo es equivalente a ser un punto crítico de H y sólo existe uno: el mínimo global x^*.

Multiplicando en (8.4.1) por $x'(t)$ obtenemos que

$$| x'(t) |^2 + \langle \nabla H(x(t)), x'(t) \rangle = 0. \tag{8.4.5}$$

Aquí y en lo sucesivo, $| \cdot |$ denota la norma euclidea en $\mathbb{R}^N$ y $\langle \cdot, \rangle$ el producto escalar asociado.

Por otra parte,

$$\langle \nabla H(x(t)), x'(t) \rangle = \frac{d}{dt} H(x(t)).$$

Deducimos por tanto la identidad,

$$\frac{d}{dt}H(x(t)) = - \mid x'(t) \mid^2, \tag{8.4.6}$$

de donde se obtiene que

$$H(x(t)) + \int_0^t \mid x'(s) \mid^2 ds = H(x_0). \tag{8.4.7}$$

En particular,

$$H(x(t)) \leq H(x_0). \tag{8.4.8}$$

Suponiendo que H es coerciva, i.e.

$$\lim_{|x|\to\infty} H(x) = \infty, \tag{8.4.9}$$

lo cual es una hipótesis natural a la hora de minimizar el funcional H, obtenemos que la trayectoria $t \to x(t)$ está acotada.

Definimos ahora el conjunto ω-límite:

$$\omega(x_0) = \{y_0 \in \mathbb{R}^N : \exists t_j \to \infty,\, x(t_j) \to y_0\} \tag{8.4.10}$$

que es no vacío.

Del principio de invarianza de La Salle[2], en vista de la ley de disipación (8.4.6) es fácil comprobar que $y(t)$ solución de (8.4.1) con dato inicial y_0, es tal que $\nabla H(y_0) = 0$. Por la unicidad del punto crítico de H deducimos que $y_0 = x^*$. Esto demuestra que

$$\omega(x_0) = \{x^*\} \tag{8.4.11}$$

y, por tanto, que

$$x(t) \to x^*,\, t \to \infty. \tag{8.4.12}$$

El resultado que acabamos de demostrar muestra que, bajo condiciones bastante generales sobre el potencial H, todas las soluciones de (8.4.1) convergen a la única solución de equilibrio x^*. Esto permite justificar la sustitución del modelo de evolución (8.4.1) por el estacionario (8.4.4). Pero ésto ha de hacerse con prudencia puesto que la demostración de convergencia que hemos realizado no proporciona ninguna estimación sobre la velocidad con la que esto se produce.

[2] $H(x(t))$ está acotada inferiormente y es decreciente. Tiene por tanto un límite $\lim_{t\to\infty} H(x(t)) = L$. Por otra parte, como $x(t_j) \to y_0$, por la propiedad de semigrupo, $x(t_j+t) \to y(t)$. Como $H(x(t_j+t)) \to L$, $j \to \infty$ para todo $t > 0$, deducimos que $H(y(t)) = L$, para todo $t > 0$. Aplicando la identidad de energía (8.4.6) deducimos que $y'(t) = 0$ lo cual equivale a $y(t) \equiv y_0$. Como la única solución estacionaria es x^*, deducimos que $y_0 = x^*$.

Bajo hipótesis adicionales sobre el potencial H, se puede además estimar la velocidad de convergencia. Dada la solución $x = x(t)$ de (8.4.1) y la solución estacionaria x^* de (8.4.4), consideramos la diferencia

$$y(t) = x(t) - x^*.$$

Tenemos entonces

$$\begin{aligned} y'(t) &= -\Big[\nabla H(x(t)) - \nabla H(x^*)\Big] \\ &= -\Big[\nabla H(y(t)) + x^*) - \nabla H(x^*)\Big]. \end{aligned}$$

Multiplicando escalarmente por $y(t)$ en esta ecuación deducimos que

$$\frac{1}{2}\frac{d}{dt} \mid y(t) \mid^2 = -\Big[\langle \nabla H(y(t) + x^*) - \nabla H(x^*),\, y(t)\rangle\Big].$$

Si H es de clase C^2, utilizando el desarrollo de Taylor, deducimos que

$$\langle \nabla H(y(t) + x^*) - \nabla H(x^*),\, y(t)\rangle = \langle D^2 H(\xi(t))y(t),\, y(t)\rangle,$$

donde D^2H denota la matriz Hessiana de H.

Suponiendo que H es estrictamente uniformemente convexa, deducimos la existencia de $\alpha > 0$ tal que

$$D^2 H(\xi) \geq \alpha I,\ \forall \xi \in \mathbb{R}^N, \tag{8.4.13}$$

de modo que

$$\langle \nabla H(y(t) + x^*) - \nabla H(x^*),\, y(t)\rangle \geq \alpha \mid y(t) \mid^2,$$

es decir,

$$\frac{1}{2}\frac{d}{dt} \mid y(t) \mid^2 \leq -\alpha \mid y(t) \mid^2, \tag{8.4.14}$$

de donde se deduce la convergencia exponencial

$$\mid y(t) \mid \leq e^{-\alpha t} \mid y_0 \mid = e^{-\alpha t} \mid x_0 - x^* \mid . \tag{8.4.15}$$

8.5. Sistemas gradiente y métodos de descenso

En la sección 8.4 hemos probado que por sistemas gradiente de la forma

$$\begin{cases} x' + \nabla H(x) = 0, \quad t > 0 \\ x(0) = x_0, \end{cases} \tag{8.5.1}$$

bajo condiciones adecuadas de convexidad y de coercividad del potencial H, las soluciones de (8.5.1) convergen, cuando $t \to \infty$, a la solución estacionaria x^*:

$$\nabla H(x^*) = 0. \tag{8.5.2}$$

Podemos interpretar este resultado, como lo hemos hecho hasta ahora, como el de la simplificación asintótica que nos permite pasar de (8.5.1) a (8.5.2).

Pero puede ser interpretado también de manera distinta. En efecto, bajo las hipótesis de convexidad y coercividad impuestas sobre el funcional H, este posee un único punto crítico x^*, que es el mínimo global, solución de (8.5.2). El hecho de que las soluciones de (8.5.1) converjan, cuando $t \to \infty$, a este punto crítico nos permiten interpretar la ecuación de evolución (8.5.1) como una manera de aproximar el mínimo del funcional. La ecuación (8.5.1) puede por tanto entenderse como un algoritno continuo en t para la aproximación del mínimo del funcional H.

De hecho, se trata de un algoritmo de descenso en la medida que, tal y como probábamos en (8.4.6), se verifica la identidad de energía

$$\frac{d}{dt}H(x(t)) = - \mid x'(t) \mid^2, \tag{8.5.3}$$

que demuestra que $H(x(t))$ decrece a medida que t aumenta. La ecuación diferencial (8.5.1) constituye por tanto un mecanismo continuo de minimización del funcional H que conduce la trayectoria desde el punto inicial x_0 de energía y nivel $H(x_0)$ hasta el punto de mínimo x^* de energía $H(x^*)$, de manera monótona.

El mismo tipo de algoritmo puede reproducirse de manera discreta en tiempo. Para hacerlo, introducimos una discretización en tiempo de la ecuación (8.5.1) de paso Δt,

$$\frac{x^{k+1} - x^k}{\Delta t} = -\nabla H(x^{k+1}) \tag{8.5.4}$$

que puede también escribirse como

$$x^{k+1} + \Delta t \nabla H(x^{k+1}) = x^k, \tag{8.5.5}$$

o, de otro modo, como

$$x^{k+1} = (I + \Delta t \nabla H)^{-1}(x^k). \tag{8.5.6}$$

La aplicación $h = (I + \Delta t \nabla H)^{-1}$ está bien definida. En efecto, para resolver

$$h(x) = y \Leftrightarrow x + \Delta t \nabla H(x) = y \tag{8.5.7}$$

basta con minimizar el funcional

$$J(x) = \frac{1}{2} \mid x \mid^2 + \Delta t H(x) - y \cdot x \tag{8.5.8}$$

en $\mathbb{R}^N$. Este funcional posee en efecto un único mínimo $x \in \mathbb{R}$ para cada $y \in \mathbb{R}^N$ puesto que es continuo, convexo y coercivo.

Este punto de mínimo es precisamente la única solución de (8.5.7) por ser J también estrictamente convexo.

La iteración discreta (8.5.4) es por tanto la misma, escrita en la forma (8.5.6) que se utilizaría en la búsqueda de un punto fijo de la aplicación $J = (I + \Delta t \nabla H)^{-1}$ si esta fuese contractiva. Veamos que en realidad lo es.

Consideramos dos puntos $y_1, y_2 \in \mathbb{R}^N$ y las correspondientes soluciones $x_1, x_2 \in \mathbb{R}^N$:

$$x_j + \Delta t \nabla H(x_j) = y_j,\ j = 1, 2.$$

Resultando las ecuaciones para $j = 1, 2$ obtenemos

$$x_1 - x_2 = \Delta t\big(\nabla H(x_1) - \nabla H(x_2)\big) = y_1 - y_2.$$

Multiplicando esta identidad por $x_1 - x_2$ (haciendo el producto escalar en $\mathbb{R}^N$) obtenemos que

$$\mid x_1 - x_2 \mid^2 + \Delta t \langle \nabla H(x_1) - \nabla H(x_2),\, x_1 - x_2 \rangle = \langle y_1 - y_2,\, x_1 - x_2 \rangle \leq \mid y_1 - y_2 \mid \mid x_1 - x_2 \mid .$$

Ahora bien, si H es uniforme y estrictamente convexa, existe $\alpha > 0$ tal que

$$\langle \nabla H(x_1) - \nabla H(x_2),\, x_1 - x_2 \rangle \geq \alpha \mid x_1 - x_2 \mid^2,$$

de donde deducimos que

$$(1 + \alpha \Delta t) \mid x_1 - x_2 \mid^2 \leq \mid y_1 - y_2 \mid \mid x_1 - x_2 \mid .$$

Es decir,

$$\mid x_1 - x_2 \mid \leq k \mid y_1 - y_2 \mid$$

con

$$k = \frac{1}{1 + \alpha \Delta t} < 1.$$

Vemos pues que $J = (I + \Delta t \nabla H)^{-1}$ es estrictamente contractiva. La iteración (8.5.4) construida discretizando en tiempo el sistema dinámico gradiente, por lo tanto, converge y converge al punto fijo de J que satisface

$$x + \nabla t \nabla H(x) = x.$$

Se trata evidentemente de una ecuación equivalente a

$$\nabla H(x) = 0$$

cuya única solución es el mínimo absoluto x^* de H.

Hemos comprobado por tanto que el sistema dinámico discreto (8.5.4) proporciona también un algoritmo iterativo de aproximación del mínimo del funcional H.

8.6. Mínimo cuadrados

Hemos visto que muchos de los problemas que se plantean en el contexto de la Mecánica de Fluidos admiten una formulación variacional. Esto permite, desde un punto de vista teórico, resolverlos mediante el Método Directo del Cálculo de Variaciones y, desde un punto de vista computacional, hacerlo mediante un método iterativo de descenso como el Método del Gradiente Conjugado.

Pero hay otros muchos problemas que no pueden ser abordados de este modo, ya sea porque no admiten una formulación variacional o bien porque el funcional involucrado no es convexo.

En el caso por ejemplo de los sistemas lineales

$$Ax = b, \tag{8.6.1}$$

en los que, si A no es simétrica, $Ax - b$ no es el gradiente de un funcional $J : \mathbb{R}^N \to \mathbb{R}$.

Por otra parte, el contexto de las ecuaciones de reacción-difusión de la teoría de la combustión surgen ecuaciones elípticas con no-linealidad exponencial de la forma

$$\begin{cases} -\Delta u = e^u + f & \text{en} \quad \Omega \\ u = 0 & \text{en} \quad \partial\Omega. \end{cases}$$

En este caso, las soluciones son puntos críticos del funcional

$$J(u) = \frac{1}{2}\int_\Omega \mid \nabla u \mid^2 dx - \int_\Omega e^u dx - \int_\Omega f u dx,$$

que está bien definido si $\Omega \subset \mathbb{R}^2$. Pero al no ser convexo, los métodos de descenso habituales no garantizan la convergencia a un punto crítico.

En estos casos los métodos inspirados en los mínimos cuadrados pueden ser de gran utilidad.

Para ilustrar este tipo de métodos consideramos en primer lugar un sistema de N ecuaciones no-lineales en $\mathbb{R}^N$:

$$f_i\big(x_1, \ldots, x_N\big) = 0,\ i = 1, \ldots, N, \tag{8.6.2}$$

que escribimos en forma vectorial como

$$f(x) = 0. \tag{8.6.3}$$

Sea Σ una matriz $N \times N$ simétrica y definida positiva y consideremos la función

$$j(y) = \frac{1}{2}\langle \Sigma f(y),\, f(y)\rangle \tag{8.6.4}$$

donde $< \cdot, \cdot >$ denota el producto escalar en $\mathbb{R}^N$.

La solución mínimos cuadrados asociada a la matriz M es aquella que resuelve el siguiente problema de minimización:

$$\begin{cases} x \in \mathbb{R}^N \\ j(x) \leq j(y), \quad \forall y \in \mathbb{R}^N. \end{cases} \tag{8.6.5}$$

Para ilustrar el significado de esta reducción consideramos la ecuación lineal en la que

$$f(y) = Ay - b. \tag{8.6.6}$$

La ecuación original a resolver es entonces

$$Ax = b, \tag{8.6.7}$$

que tiene una solución sí y sólo sí

$$b \in R(A) = \{q :\ q = Ay,\, y \in \mathbb{R}^N\}. \tag{8.6.8}$$

Tomamos, para simplificar los cálculos $\Sigma = I$. En este caso las soluciones obtenidas para mínimos cuadrados son las correspondientes a la *forma normal*

$$A^t A x = A^t b. \tag{8.6.9}$$

Lo sorprendente de este hecho es que la matriz $A^t A$ involucrada en la ecuación normal es simétrica y semi-definida positiva. Además, este nuevo sistema tiene siempre al menos una solución puesto que $b \in R(A^t) = R(A^t A)$.

Cuando $\ker(A^t A) = 0$ (i.e. $\operatorname{rank}(A) = N$) el sistema (8.6.9) admite una única solución. Sin embargo, cuando $\operatorname{rank}(A) < N$, el sistema admite una infinidad de soluciones de la forma

$$x = \hat{x} + z,\ \hat{x} \in R(A^t),\ z \in \ker A. \tag{8.6.10}$$

De las soluciones

$$\mathbb{R}^N = R(A^t) \oplus \ker(A);\ (R(A^t))\pm = \ker(A)), \tag{8.6.11}$$

deducimos la relación

$$\| x \|^2 = \| \hat{x} \|^2 + \| z \|^2 \geq \| \hat{x} \|^2 . \tag{8.6.12}$$

Vemos pues que $\hat{x}$ es la única solución de norma mínima de (8.6.9) son los puntos críticos del funcional

$$J(x) = \frac{1}{2} \| A^t x \|^2 - A^t b \cdot x, \tag{8.6.13}$$

que es convexo. El método de mínimos cuadrados juega pues el papel de convexificador del sistema (8.6.7).

Consideramos ahora el caso en que f no es afín. Suponemos que $f \in C^2$. Sea x una solución (8.6.3), entonces

$$\begin{cases} j'(x) = f'(x)^t \Sigma f(x) = 0 \\ j''(x) = f'(x)^t \Sigma f'(x). \end{cases} \tag{8.6.14}$$

La matriz $j''(x)$ es semi-definida positiva y, cuando $f(x)$ es regular (i.e. $\det(f'(x)) \neq 0$), es definida positiva. De este modo observamos que $j''(\cdot)$ es definida positiva en un entorno de x y por tanto j es estrictamente convexa. Vemos por tanto que el método de mínimos cuadrados tiene propiedades de convexificación locales.

El papel de la matriz Σ elegida para aplicar el método de mínimos cuadrados es múltiple. Por ejemplo, permite privilegiar algunas de las ecuaciones $f_i(x) = 0$ frente a otras. La matriz Σ puede también contribuir a reducir el número de condición de $j'(y)$, lo cual hace el problema de los mínimos cuadrados (8.6.5) más robusto.

En el marco lineal, la ecuación correspondiente a cada matriz Σ es de la forma

$$A^t \Sigma A x = A^t \Sigma b. \tag{8.6.15}$$

Se trata de la ecuación normal generalizada. Este sistema tiene siempre una solución que puede ser única o no en función de que $\ker(A)$ se reduzca o no al $\{0\}$.

Cuando las dimensiones del sistema no son excesivas el sistema (8.6.15) puede resolver mediante un método directo (Cholesky, por ejemplo). Cuando $R(A^t) < N$ conviene introducir una regularización del sistema de la forma

$$(\varepsilon S + A^t \Sigma A) x_\varepsilon = A^t \Sigma b, \tag{8.6.16}$$

siendo $\varepsilon > 0$ y S una matriz simétrica y definida positiva ($S = I$, por ejemplo). Se tiene entonces que

$$x_\varepsilon - \hat{x}_s = O(\varepsilon), \tag{8.6.17}$$

donde $\hat{x}_s$ es la única solución de (8.6.15) en $R(S^{-1}A^t)$. Por otra parte, la ecuación (8.6.16) puede resolverse mediante el método de gradiente conjugado.

Capítulo 9

Métodos de Galerkin

9.1. El lema de Lax-Milgram y sus variantes

Muchos de los problemas de la Mecánica de Medios Continuos y sus aproximaciones numéricas admiten una formulación variacional semejante a la siguiente:

$$\begin{cases} v \in V \\ A(u,\, v) = F(v),\, \forall v \in V \end{cases} \tag{9.1.1}$$

donde V es un espacio de Hilbert, A una forma bilineal y F una forma lineal.

El siguiente resultado de Lax-Milgram es una herramienta básica y fundamental para su resolución:

Lema de Lax-Milgram.

Sea V un espacio de Hilbert real de norma $\| \cdot \|$, $A(u,\, v) : V \times V \to \mathbb{R}$ una forma bilineal y $F : V \to \mathbb{R}$ un funcional lineal y continuo. Supongamos que $A(\cdot,\, \cdot)$ es continua, i.e.,

$$| A(u,\, v) | \leqslant \gamma \| u \| \| v \|,\, \forall u,\, v \in V \tag{9.1.2}$$

y coerciva,

$$A(u,\, u) \geqslant \alpha \| u \|^2,\, \forall u \in V. \tag{9.1.3}$$

Entonces, existe una única $u \in V$ solución de (9.1.1) *y satisface*

$$\| u \| \leqslant \frac{1}{\alpha} \| F \|_{V'} \,. \tag{9.1.4}$$

Por otra parte, si A es simétrica, la solución u de (9.1.1) *es el único punto de mínimo del funcional*

$$J(v) = \frac{1}{2} A(v,\, v) - F(v). \tag{9.1.5}$$

La siguiente generalización es debida a Nečas:

Theorem 9.1.1 *Sean W y V dos espacios de Hilbert reales, con normas $|||\cdot|||$ y $\|\cdot\|$ respectivamente. Supongamos que existen dos constantes positivas γ y α tales que la forma bilineal $A : W \times V \to \mathbb{R}$ satisface*

$$| A(w, v) |\leqslant \gamma|||w||| \parallel v \parallel, \forall w \in W, \forall v \in V \tag{9.1.6}$$

$$\sup_{\substack{v\in V\\ v\neq 0}} \frac{A(w, v)}{\parallel v \parallel} \geqslant \alpha|||w|||, \forall w \in W \tag{9.1.7}$$

$$\sup_{w\in W} A(w, v) > 0, \forall v \in V, v \neq 0. \tag{9.1.8}$$

Entonces, para todo $F \in V'$, existe una única solución de

$$\begin{cases} w \in W \\ A(w, v) = F(v), \forall v \in V. \end{cases} \tag{9.1.9}$$

Conviene observar que la diferencia fundamental entre (9.1.1) y (9.1.9) es que, en la segunda, la solución w pertenece a un espacio distinto (W) al que pertenecen las funciones test (V).

Demostración 9.1.1 Gracias al teorema de representación de Riesz existe un operador continuo $\mathcal{A} : W \to V$ tal que

$$A(w, v) = (\mathcal{A}w, v)_V, \forall v \in V \tag{9.1.10}$$

donde $(\cdot, \cdot)_V$ denota el producto escalar en V. Además, en virtud de (9.1.6) tenemos

$$\parallel \mathcal{A}w \parallel\leqslant \gamma|||w|||, \forall w \in W. \tag{9.1.11}$$

El problema se reduce a probar que, para cada $F \in V'$, existe una única $w \in W$ tal que

$$\mathcal{A}w = RF \tag{9.1.12}$$

donde $R : V' \to V$ es la isometría asociada al teorema de representación de Riesz.

El operador $\mathcal{A}$ es inyectivo, i.e., $\mathcal{A}w = 0$ implica que $w = 0$.

Por otra parte el rango de $\mathcal{A}$ es cerrado. En efecto, si $\mathcal{A}w_n \to v$ en V tenemos

$$|||w_n - w_m||| \leqslant \frac{1}{\alpha} \sup_{\substack{v\in V\\ v\neq 0}} \frac{(\mathcal{A}(w_n - w_m), v)_V}{\parallel v \parallel} \leqslant \frac{1}{\alpha} \parallel \mathcal{A}(w_n - w_m) \parallel .$$

Por tanto $w_n \to w$ en W y entonces $\mathcal{A}w_n \to \mathcal{A}w$ en V. Entonces $v = \mathcal{A}w$.

Por último, si $z \in R(A)^{\perp}$, *i.e.*, si

$$(\mathcal{A}w,\, z)_V = A(w,\, z) = 0, \forall w \in W,$$

necesariamente $z = 0$. Por tanto $\mathcal{A}$ es sobreyectivo.

Por consiguiente, deducimos que si $F \in V'$ existe una única solución de (9.1.12), tal y como se pretendía probar.

Además,

$$\alpha|||u||| \leqslant \sup_{\substack{v\in V\\ v\neq 0}} \frac{(\mathcal{A}u,\, v)_V}{\parallel v \parallel} = \sup_{\substack{v\in V\\ v\neq 0}} \frac{(RF,\, v)_V}{\parallel v \parallel} = \parallel F \parallel_{V'} .$$

■

Acabamos de probar una generalización del Lema de Lax-Milgram. En el caso en que la forma bilineal A es simétrica, es fácil comprobar que la solución de (9.1.1) se puede obtener por minimización del funcional J de (9.1.5).

Para probar que J alcanza su mínimo basta aplicar el Método Directo del Cálculo de Variaciones pues J es un funcional continuo, coercivo y convexo en el espacio de Hilbert V. Para comprobar que en el mínimo de J es la solución de (9.1.1) basta constatar que el mínimo $u \in V$ se tiene

$$DJ(u) = 0 \text{ en } V', \tag{9.1.13}$$

es decir,

$$\langle DJ(u),\, v\rangle = 0,\, \forall v \in V. \tag{9.1.14}$$

Es fácil comprobar que

$$\langle DJ(u),\, v\rangle = A(u,\, v) - F(v). \tag{9.1.15}$$

9.2. El método de Galerkin

El método Galerkin proporciona una forma sistemática de obtener aproximaciones finito-dimensionales convergentes de problemas variacionales de la forma (9.1.1).

Para ello consideramos una familia $\{V_h\}_{h>0}$ de subespacios de dimensión finita de V.

Supongamos que para todo $v \in V$ existe una sucesión $v_h \in V_h$ tal que

$$v_h \to v \text{ en } V, \text{ cuando } h \to 0. \tag{9.2.1}$$

Dado $F \in V'$, la aproximación de Galerkin de (9.1.1) es:

$$u_h \in V_h,\, A(u_h,\, v_h) = F(v_h),\, \forall v_h \in V_h. \tag{9.2.2}$$

Este sistema puede reescribirse en forma matricial. Si $\{\varphi_j : j = 1, \cdots, N_h\}$ es una base de V_h y

$$u_h = \sum_{j=1}^{N_h} \xi_j \varphi_j, \tag{9.2.3}$$

el problema (9.2.2) puede reescribirse de la forma siguiente

$$\tilde{A}\xi = F \tag{9.2.4}$$

donde $\tilde{A}$ es la matriz de componentes $\tilde{A}_{ij} = A(\varphi_i, \varphi_j)$, ξ es el vector incógnita de componentes ξ_j, y F es el dato con componentes $F_j = F(\varphi_j)$.

La matriz A se denomina matriz de rigidez.

Tenemos el siguiente resultado:

Theorem 9.2.1 *Bajo las hipótesis del Lema de Lax-Milgram, para todo $h > 0$ existe una única solución $u_h \in V_h$ de* (9.2.2) *tal que*

$$\| u_h \|_V \leqslant \frac{\| F \|_{V'}}{\alpha}. \tag{9.2.5}$$

Además,

$$\| u - u_h \|_V \leqslant \frac{\gamma}{\alpha} \inf_{v_h \in V_h} \| u - v_h \|_V, \tag{9.2.6}$$

donde u y u_h son respectivamente las soluciones de (9.1.1) *y de* (9.2.2).

En particular, bajo la hipótesis adicional (9.2.1) el método de Galerkin converge.

Demostración 9.2.1 La aplicación del Lema de Lax-Milgram en (9.2.2) proporciona la existencia de una única solución $u_h \in V_h$. Tomando la función test $v_h = u_h$ obtenemos

$$\alpha \| u_h \|^2 \leqslant A(u_h, u_h) = F(u_h) \leqslant \| F \|_{V'} \| u_h \|, \tag{9.2.7}$$

lo cual garantiza que la sucesión u_h está acotada en V.

Sustrayendo las formulaciones variacionales (9.1.1) y (9.2.2) obtenemos

$$A(u - u_h, v_h) = 0, \forall v_h \in V_h. \tag{9.2.8}$$

Tomando como función test $v_h = w_h - u_h$ con $w_h \in V_h$ obtenemos

$$\begin{aligned} \alpha \| u - u_h \|^2 &\leqslant A(u - u_h, u - u_h) = A(u - u_h, u - w_h) \\ &\leqslant \gamma \| u - u_h \| \| u - w_h \|, \forall w_h \in V_h \end{aligned}$$

de donde se obtiene (9.2.6).

Observación 9.2.1 Cuando A es simétrica, la solución u_h de (9.2.2) se obtiene minimizando la restricción de J a $V_h \times V_h$. En este caso la matriz $\tilde{A}$ del sistema (9.2.4) es también simétrica definida positiva.

Observación 9.2.2 En la estimación (9.2.6) que es válida sea cual sea la familia de espacios $\{V_h\}_{h>0}$ de dimensión finita considerada se observa con claridad la importancia de suponer que (9.2.1) se satisface. En efecto, es sólo bajo esta hipótesis que (9.2.1) garantiza que $u_h \to u$ en V cuando $h \to 0$.

Por otra parte, es natural que, cuando (9.2.1) no se cumple, el método de Galerkin no converja. En efecto, cuando (9.2.1) no se cumple, los subespacios V_h elegidos "no llenan.$^{\text{el}}$ espacio de Hilbert V donde el problema variacional ortiginal está formulado y por tanto no cabe esperar la convergencia de la aproximación. De todas maneras, de (9.2.6) se deduce que el método converge siempre que la solución u sea límite de funciones u_h de V_h. En otras palabras, siempre que u sea límite de funciones de V_h, u es en realidad límite de las aproximaciones de Galerkin.

9.2.1. Interpretación geométrica del método de Galerkin

Cuando la forma bilineal A es simétrica, la solución $u_h \in V_h$ del problema aproximado de Galerkin es la proyección ortogonal de la solución (9.1.1) sobre V_h con respecto al producto escalar correspondiente a la forma bilineal $A(\cdot, \cdot)$.

En efecto, como $A(\cdot, \cdot)$ es una forma bilineal simétrica, continua y coerciva en V, induce unn producto escalar que es completamente equivalente al producto escalar canónico de V.

Por tanto, dado $u \in V$ cualquiera, su proyección ortogonal $u_h \in V_h$ con respecto a este producto escalar está caracterizado por las dos propiedades equivalentes siguientes:

- Distancia mínima entre u y u_h:

$$A(u - u_h,\, u - u_h) = \min_{v_h \in V_h} A(u - v_h,\, u - v_h); \tag{9.2.9}$$

- Ortogonalidad:

$$A(u - u_h,\, v_h) = 0,\ \forall v_h \in V_h. \tag{9.2.10}$$

La segunda propiedad de ortogonalidad puede reescribirse de la siguiente forma

$$A(u,\, v_h) = A(u_h,\, v_h),\ \forall v_h \in V_h \tag{9.2.11}$$

Vemos por tanto que, cuando u es solución de (9.1.1), u_h es solución de (9.2.1).

En virtud de esta propiedad es más fácil entender la estimación de error (9.2.6). En la medida en que u_h es el elemento de V_h más próximo a u en la norma inducida por la forma bilineal A, para estimar el error, es decir, la distancia de u a u_h podemos tomar como cota superior la distancia de u a cualquier elemento $\tilde{u}_h$ de V_h. Este hecho se utiliza de manera profusa en la práctica para obtener estimaciones explícitas de error o órdenes de convergencia. Esto se hace mediante la obtención de una aproximación $\tilde{u}_h$ de u más o menos explícita para la que seamos capaces de estimar la distancia a u de una manera sencilla. Cada vez que somos capaces de hacer ésto hemos obtenido una estimación del error entre u y u_h, la solución del problema original (9.1.1) y de la aproximación de Galerkin (9.2.1).

9.2.2. Orden de convergencia

En la terminología clásica del Análisis Numérico se dice que el método de aproximación es de orden p si $\| u - u_h \| = O(h^p)$ cuando $h \to 0$.

En vista de (9.2.6) para obtener un método de Galerkin de orden p bastaría con encontrar unos subespacios V_h de V de dimensión finita de modo que para cada $u \in V$ existiese una sucesión aproximante $v_h \in V_h$ tal que

$$\| u - v_h \| = O(h^p).$$

Desafortunadamente esto no puede ocurrir en dimensión infinita.[1]

Sin embargo esto es perfectamente factible para determinados subespacios $\mathcal{V}$ densos en V.

Este último hecho nos va a permitir en la práctica obtener estimaciones de la velocidad de convergencia para los problemas elípticos más clásicos asociados al Laplaciano, al operador de Stokes, al sistema de elasticidad, etc. Esto es así puesto que los resultados clásicos de la teoría de regularidad elíptica garantizan que, a pesar de que en un principio las soluciones pertenecen a V, en la práctica pertenecen a un subespacio de funciones más regulares que podemos identificar y que permite establecer una tasa explícita de convergencia del método. Por ejemplo, en el caso del problema de Dirichlet para el Laplaciano, esto es así puesto que, cuando el segundo miembro es de cuadrado integrable y el dominio de clase C^2 las soluciones débiles de $H_0^1(\Omega)$ pertenecen en realidad a $H^2(\Omega)$.

[1] El lector interesado podrá reflexionar sobre el ejemplo canónico en que $V = \ell^2$ y V_h es el subespacio de dimensión finita constituido por las sucesiones con términos nulos a partir del $[1/h]$-ésimo. Si bien no se puede obtener un orden de aproximación uniforme en todo el espacio ℓ^2 si que puede hacerse, por ejemplo, para el subespacio h^1 del espacio ℓ^2 (véase el ejercicio 12).

A partir de esta formulación abstracta y general del método de Galerkin pueden darse un sinfin de aplicaciones importantes. En el contexto del problema de Dirichlet para la ecuación de Laplace las variantes más importantes provienen de hacer las diversas posibles elecciones de los subespacios aproximantes V_h. Hay dos casos particularmente importantes y que analizaremos brevemente en las proximas secciones: los métodos espectrales y el método de elementos finitos.

9.2.3. Métodos espectrales

Se trata de un caso particular del método de Galerkin en el que elegimos V_h a partir de las autofunciones del operador de Laplace.

Recordemos brevemente los resultados básicos de descomposición espectral.

Consideramos por tanto un dominino Ω acotado de $\mathbb{R}^n$, $f \in H^{-1}(\Omega)$ y estudiamos el problema de Dirichlet

$$\begin{cases} -\Delta u = f \quad \text{en} \quad \Omega \\ u\Big|_{\partial\Omega} = 0 \end{cases} \tag{9.2.12}$$

que admite la formulación variacional

$$\begin{cases} u \in H_0^1(\Omega), \\ \displaystyle\int_\Omega \nabla u \cdot \nabla \varphi dx = \langle f,\, \varphi \rangle_{H^{-1}(\Omega)\times H_0^1(\Omega)},\ \forall \varphi \in H_0^1(\Omega). \end{cases} \tag{9.2.13}$$

Consideramos ahora el problema espectral

$$\begin{cases} -\Delta \varphi = \lambda \varphi \quad \text{en } \Omega \\ \varphi\Big|_{\partial\Omega} = 0. \end{cases} \tag{9.2.14}$$

Los resultados clásicos de teoría espectral garantizan que existe una sucesión de autovalores positivos y de multiplicidad finita

$$0 < \lambda_1 < \lambda_2 \leqslant \lambda_3 \leqslant \cdots \leqslant \lambda_N \leqslant \cdots \to \infty \tag{9.2.15}$$

que tiende a infinito y cuyas autofunciones asociadas $\{\varphi_j\}_{j\in\mathbb{N}}$ pueden ser elegidas de modo que constituyan una base ortonormal de $L^2(\Omega)$.

En estas condiciones, si $f \in L^2(\Omega)$, puede descomponerse como

$$f(x) = \sum_{j=1}^{\infty} f_j \varphi_j(x), \tag{9.2.16}$$

donde f_j son los coeficientes de Fourier que vienen dados por

$$f_j = \int_\Omega f \varphi_j dx, \tag{9.2.17}$$

y la solución correspondiente del problema (9.2.11) (ó (9.2.12)) es entonces

$$u(x) = \sum_{j=1}^{\infty} \frac{f_j}{\lambda_j} \varphi_j(x). \tag{9.2.18}$$

Utilicemos ahora la base de autofunciones para construir un método de Galerkin.

Para cualquier $N \geqslant 1$ introducimos el subespacio V_N de V generado por las N primeras funciones propias $\varphi_1, \cdots, \varphi_N$, i.e.

$$V_N = \text{span}\left[\varphi_1, \cdots, \varphi_N\right]. \tag{9.2.19}$$

Aunque las autofunciones constituyen una base ortonormal de $L^2(\Omega)$ son también funciones de $H_0^1(\Omega)$ y, de hecho,

$$\int_\Omega \mid \nabla \varphi_j \mid^2 dx = \lambda_j \int_\Omega \varphi_j^2 dx = \lambda_j, \, \forall j \geqslant 1. \tag{9.2.20}$$

La aproximación de Galerkin se define entonces del modo siguiente:

$$\begin{cases} u_N \in V_N \\ \displaystyle\int_\Omega \nabla u_N \cdot \nabla v_N dx = \langle f, \, v_N \rangle, \, \forall v_N \in V_N. \end{cases} \tag{9.2.21}$$

Como V_N es un espacio de dimensión N, (9.2.20) equivale en realidad a un sistema lineal de N ecuaciones con N incógnitas.

Buscamos la solución u_N de (9.2.21) en la forma

$$u_N = \sum_{j=1}^{N} u_{j,\,N} \varphi_j(x)$$

y escribimos las N ecuaciones que codifican (9.2.21) consistentes en tomar las funciones test $v_N = \varphi_{1,\cdots,}\varphi_N$. Usando la ortogonalidad de $\{\varphi_j\}$ en $H_0^1(\Omega)$ y en $L^2(\Omega)$ deducimos que (9.2.21) equivale a

$$\lambda_j u_{j,\,N} = f_j, \, j = 1, \cdots, N \tag{9.2.22}$$

lo cual muestra que el sistema (9.2.21) es diagonal y admite por solución

$$u_{j,\,N} = f_j/\lambda_j, \, j = 1, \cdots, N. \tag{9.2.23}$$

Es decir

$$u_N(x) = \sum_{j=1}^{N} \frac{f_j}{\lambda_j} \varphi(x) \tag{9.2.24}$$

que es en realidad la truncatura de la solución del problema de Dirichlet $u \in H_0^1(\Omega)$ calculada explícitamente en (9.2.18).

Esto era efectivamente previsible. El producto escalar inducido por la forma bilineal de este problema es precisamente el canónico de $H_0^1(\Omega)$. Por tanto u_N no es más que la proyección ortogonal sobre V_N, subespacio cerrado de $H_0^1(\Omega)$, de la solución u de (9.2.18). Como las autofunciones $\{\varphi_j\}_{j\geqslant 1}$ son ortogonales en $H_0^1(\Omega)$ se deduce entonces que u_N es necesariamente de la forma (9.2.24).

Observamos por tanto que el método de Galerkin en la base de las funciones propias del Laplaciano es en realidad el método de Fourier consistente en aproximar la solución mediante la truncatura del desarrollo de la serie de Fourier de la misma.

En la argumentación anterior hemos utilizado en un par de ocasiones la ortogonalidad de las funciones en $H_0^1(\Omega)$. Esto necesita justificación pues, en principio, las autofunciones son ortonormales en $L^2(\Omega)$. La ortogonalidad en $H_0^1(\Omega)$ se deduce del siguiente sencillo cálculo. Multiplicando por φ_k en la ecuación satisfecha por φ_j e integrando por partes se deduce que

$$\int_\Omega \nabla\varphi_j \cdot \nabla\varphi_k dx = \lambda_j \int_\Omega \varphi_j\varphi_k dx.$$

Como

$$\int_\Omega \varphi_j\varphi_k dx = 0$$

cuando $j \neq k$, deducimos que

$$\int_\Omega \nabla\varphi_j \cdot \nabla\varphi_k dx = 0, \tag{9.2.25}$$

si $j \neq k$. Además, cuando $j = k$ tenemos

$$\int_\Omega \mid \nabla\varphi_j \mid^2 dx = \lambda_j \int_\Omega \varphi_j^2 dx = \lambda_j. \tag{9.2.26}$$

Por tanto $\parallel \varphi_j \parallel_{H_0^1(\Omega)} = \sqrt{\lambda_j}$. Se deduce por tanto que $\left\{\varphi_j \Big/ \sqrt{\lambda_j}\right\}_{j\geqslant 1}$ constituye una base ortonormal de $H_0^1(\Omega)$.

Hemos comprobado que el método de Galerkin en la base de autofunciones del Laplaciano conduce al método de Fourier. Por otra parte, el método es convergente. Nos proponemos ahora estudiar el orden de convergencia del método.

Tal y como habíamos comentado anteriormente, bajo la nueva hipótesis de que $f \in H^{-1}(\Omega)$ o, equivalentemente, $u \in H_0^1(\Omega)$, no cabe esperar ninguna estimación precisa sobre el orden convergencia del método. Veamos sin embargo

que una hipótesis adicional sobre los datos del problema permite obtener tasas explícitas.

Supongamos que $f \in L^2(\Omega)$. En este caso la solución u no sólo pertenece a la clase $H_0^1(\Omega)$ sino que verifica además que $\Delta u \in L^2(\Omega)$. De este modo, habida cuenta de (9.2.18) tenemos que

$$\sum_{j=1}^{\infty} \mid f_j \mid^2 = \sum_{j=1}^{\infty} \mid u_j \mid^2 \lambda_j^2 < \infty. \tag{9.2.27}$$

Por otra parte,

$$\begin{aligned} \| u - u_N \|^2_{H_0^1(\Omega)} &= \sum_{j=N+1}^{\infty} \lambda_j \mid u_j \mid^2 \\ &\leqslant \frac{1}{\lambda_{N+1}} \sum_{j=N+1}^{\infty} \lambda_j \mid u_j \mid^2 \leqslant \frac{\| f \|^2_{L^2(\Omega)}}{\lambda_{N+1}}. \end{aligned} \tag{9.2.28}$$

Deducimos pues que

$$\| u - u_N \|_{H_0^1(\Omega)} = O\left(\frac{1}{\sqrt{\lambda_{N+1}}}\right). \tag{9.2.29}$$

En este caso, es decir, bajo la hipótesis adicional (9.2.27), sí que deducimos por tanto una estimación sobre la velocidad de convergencia, (9.2.29).

Veamos ahora lo que (9.2.29) significa. En dimensión $n = 1$, cuando $\Omega = (0,\, L)$, tenemos $\lambda_j = \pi^2 j^2 / L^2$, de donde se deduce que

$$\| u - u_N \|_{H_0^1(0,\, L)} = O(1/N). \tag{9.2.30}$$

Pero esta estimación se deteriora a medida que aumentamos en dimensión. En efecto, por el Teorema de Weyl (véase [6]), sabemos que cuando Ω es un abierto acotado de $\mathbb{R}^n$,

$$\lambda_N(\Omega) \sim c(\Omega) N^{2/n},\ N \to \infty. \tag{9.2.31}$$

A partir de (9.2.29) y (9.2.31) se deduce que

$$\| u - u_N \|_{H_0^1(\Omega)} = O\left(N^{-1/n}\right). \tag{9.2.32}$$

Es decir, bajo la hipótesis de que $f \in L^2(\Omega)$ obtenemos una convergencia del método de Fourier de orden $1/n$, siendo n la dimensión espacial.

Hemos comprobado la convergencia del método de Galerkin-Fourier y la posibilidad de obtener estimaciones explícitas sobre el orden de convergencia, mediante hipótesis adicionales sobre el segundo miembro f de la ecuación.

El método de Fourier-Galerkin sin embargo no es fácil de manejar puesto que, salvo en caso en que la geometría de Ω sea particularmente sencilla (un intervalo de $\mathbb{R}$, un cuadrado de $\mathbb{R}^2$, etc.), el propio problema del cálculo de las autofunciones φ_j que constituyen la base del método es más complejo que el cálculo de la propia solución del problema (9.2.12).

Es por eso que conviene diseñar otros métodos, donde la base pueda ser calculada de manera simple. Uno de los más relevantes es sin duda el Método de los Elementos Finitos (MEF) que describimos brevemente en la siguiente sección en el caso de una dimensión espacial.

9.2.4. El método de Elementos Finitos $1D$

Con el objeto de ilustrar el MEF comenzamos por el caso de una sola dimensión espacial, donde los cálculos son particularmente sencillos y explícitos. El lector interesado en un desarrollo más detallado del método y su aplicación a las ecuaciones del calor y de ondas $1D$ podrá consultar las notas [36].

Consideramos por tanto el problema de Dirichlet en el intervalo $\Omega = (0,\, L)$:

$$\begin{cases} -u'' = f, & 0 < x < L \\ u(0) = u(L) = 0. \end{cases} \tag{9.2.33}$$

La solución de este problema puede calcularse explícitamente pero es un ejemplo excelente para elaborar y desarrollar el MEF.

La formulación variacional de (9.2.33) es

$$\begin{cases} u \in H_0^1(0,\, L) \\ \displaystyle\int_0^L u'v'dx = \langle f,\, v\rangle,\ \forall v \in H_0^1(0,\, L). \end{cases} \tag{9.2.34}$$

Dado $h > 0$ de la forma $h = 1/(N+1)$ con N natural, introducimos el espacio V_h de las funciones continuas, lineales a trozos en los segmentos de la forma $[x_j,\, x_{j+1}]$, $x_j = jh$, y que se anulan en los extremos $x = 0,\, L$. Este espacio está generado por las funciones de base $\phi_j = \phi_j(x)$, $j = 1, \cdots, N$ que satisfacen, para cada $j = 1, \cdots, N$:

- $\phi_j(x_j) = 1$,
- $\phi_j(x_k) = 0$ si $k \neq j$,
- ϕ_j es lineal en cada intervalo $[x_k,\, x_{k+1}]$.

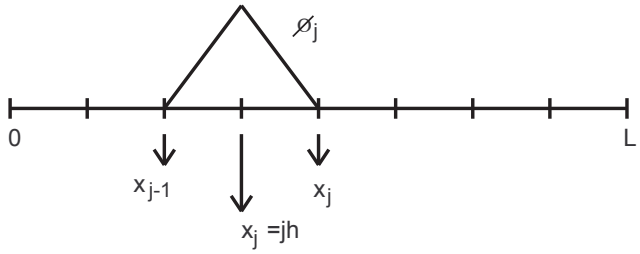


Figura 9.1: Gráfica de las funciones de base Φ_j del MEF P1 en una dimensión espacial.

Estas funciones están representadas en la siguiente figura

La aproximación de Galerkin se expresa, como es habitual, de la siguiente forma:

$$\begin{cases} u_h \in V_h \\ \displaystyle\int_0^L u_h' v_h' dx = \langle f,\, v_h \rangle, \quad \forall v_h \in V_h \end{cases} \tag{9.2.35}$$

Se trata del MEF P_1, puesto que los elementos utilizados son polinomios de grado uno en cada subintervalo.

Veamos cual es la representación de (9.2.35) en tanto que sistema finito-dimensional.

Tenemos

$$u_h(x) = \sum_{j=1}^{N} u_j \phi_j(x). \tag{9.2.36}$$

Por otra parte, (9.2.35) se verifica si y sólo si

$$\int_0^L u_h' \phi_j' dx = \langle f,\, \phi_j \rangle,\ \forall j = 1, \cdots, N. \tag{9.2.37}$$

Con el objeto de reescribir (9.2.37) suponemos que $f \in L^2(0,\, L)$. Entonces

$$\langle f,\, \phi_j \rangle = \int_0^L f(x)\phi_j(x) dx = f_j. \tag{9.2.38}$$

Además

$$\Gamma_{jk} = \int_0^L \phi_j' \phi_k' dx = \begin{cases} \dfrac{2L}{h} & \text{si} \quad j = k \\ -\dfrac{L}{h} & \text{si} \quad \mid j - k \mid = 1 \\ 0 & \text{si no .} \end{cases} \tag{9.2.39}$$

Introduciendo la matriz

$$R_h = \frac{L}{h}\begin{pmatrix} 2 & -1 & 0 & \cdots \\ -1 & 2 & -1 & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ 0 & -1 & 2 & -1 \\ 0 & \cdots & -1 & 2 \end{pmatrix} \tag{9.2.40}$$

y los vectores dato e incógnita respectivamente

$$F_h = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_N \end{pmatrix}, U_h = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{pmatrix}, \tag{9.2.41}$$

el sistema se puede escribir en la forma

$$R_h\, U_h = F_h. \tag{9.2.42}$$

En el contexto de la teoría de MEF la matriz R_h se denomina *matriz de rigidez*.

Es fácil comprobar que R_h es una matriz simétrica definida positiva. Por tanto (9.2.42) admite una única solución $U_h \in \mathbb{R}^N$.

En realidad, el hecho que (9.2.35), y por tanto (9.2.42), admite una única solución es consecuencia de la teoría general desarrollada para los métodos de Galerkin.

En la práctica frecuentemente se toma, en lugar de un segundo miembro general $f \in L^2(0,\, L)$ su interpolación lineal a trozos

$$f(x) = \sum_{j=1}^{N} \tilde{f}_j \phi_j. \tag{9.2.43}$$

En ese caso

$$f_j = \int_0^L f(x)\phi_j(x)dx \quad = \quad \int_0^L \left(\sum_{k=1}^{N} \tilde{f}_j \phi_j(x)\right) \phi_j(x)dx = M_h \tilde{F}_h \tag{9.2.44}$$

donde

$$\tilde{F}_h = \begin{pmatrix} \tilde{f}_1 \\ \vdots \\ \tilde{f}_N \end{pmatrix} \tag{9.2.45}$$

y M_h es la denominada *matriz de masa* de elementos

$$m_{jk} = \int_0^L \phi_j \phi_k dx = \begin{cases} \dfrac{2Lh}{3} & \text{si} \quad j = k \\ \dfrac{Lh}{6} & \text{si} \quad \mid j-k \mid = 1 \\ 0 & \text{si no .} \end{cases} \tag{9.2.46}$$

En este caso el sistema (9.2.42) se escribe de la forma

$$R_h U_h = M_h \tilde{F}_h. \tag{9.2.47}$$

Las soluciones de la aproximación de Galerkin (9.2.35), en la medida en que la forma bilineal asociada es simétrica, se puede caracterizar mediante un problema de minimización. Con la notación de (9.2.47) la solución $U_h \in \mathbb{R}^N$ es el mínimo de la función cuadrática $J_h : \mathbb{R}^N \to \mathbb{R}$ definida como

$$J_h(U_h) = \frac{1}{2}\langle R_h U_h,\, U_h\rangle - \langle M_h \tilde{F}_h,\, U_h\rangle. \tag{9.2.48}$$

Ocupémonos ahora de la estimación del error. La teoría general de los métodos de Galerkin para garantizar la convergencia exige que para cada $v \in H_0^1(0,\, L)$ exista una sucesión $v_h \in V_h$ tal que $v_h \to v$ en $H_0^1(0,\, L)$ cuando $h \to 0$. Comprobemos esta propiedad.

Como las funciones $v \in H_0^1(0,\, L)$ son continuas y por tanto el valor $v(x_j)$ en los puntos $x_j = jh$ del mallado está bien definido es natural tomar como $v_h \in V_h$ la función de interpolación que toma los valores $v(x_j)$ en los puntos del mallado, es decir,

$$v_h(x) = \sum_{j=1}^{N} v(x_j)\phi_j(x). \tag{9.2.49}$$

Veamos pues que $v_h \to v$ en $H_0^1(0,\, L)$ con esta elección de V_h.

Tenemos

$$\| v - v_h \|^2_{H_0^1(0,\, L)} = \int_0^L \mid v'(x) - v_h'(x) \mid^2 dx = \sum_{j=0}^{N} \int_{x_j}^{x_j+1} \mid v'(x) - v_h'(x) \mid^2 dx. \tag{9.2.50}$$

Calculemos las integrales en cada uno de los subintervalos.

Tenemos

$$\int_{x_j}^{x_j+1} \mid v' - v_h' \mid^2 dx = \int_{x_j}^{x_j+1} \left| \frac{v(x_{j+1}) - v(x_j)}{h} \right|^2 dx.$$

Ahora bien

$$v'(x)-\frac{(v(x_{j+1})-v(x_j))}{h}=\frac{1}{h}\int_{x_j}^{x_{j+1}}(v'(x)-v'(s))ds=\frac{1}{h}\int_{x_j}^{x_{j+1}}\int_{s}^{x}v''(\sigma)d\sigma ds.$$

Por tanto, suponiendo que $v''\in L^2(0,\,L)$, obtenemos

$$\begin{aligned}\left|v'(x)-\frac{(v(x_{j+1})-v(x_j))}{h}\right| &\leqslant \frac{1}{h}\int_{x_j}^{x_{j+1}}\int_{s}^{x}\mid v''(\sigma)\mid d\sigma ds\\ &\leqslant \frac{1}{h}\int_{x_j}^{x_{j+1}}\int_{x_j}^{x_{j+1}}\mid v''(\sigma)\mid d\sigma ds\\ &\leqslant \int_{x_j}^{x_{j+1}}\mid v''(s)\mid ds\leqslant\sqrt{h}\left(\int_{x_j}^{x_{j+1}}\mid v''(s)\mid^2 ds\right)^{1/2}.\end{aligned}$$

Por consiguiente:

$$\left|v'(x)-\frac{(v(x_{j+1})-v(x_j))}{h}\right|^2\leqslant h\int_{x_j}^{x_{j+1}}\mid v''(s)\mid^2 ds.$$

Volviendo a (9.2.50) obtenemos que

$$\parallel v-v_h\parallel^2_{H_0^1(0,\,L)}\leqslant h^2\sum_{j=0}^{N}\int_{x_j}^{x_{j+1}}\mid v''(s)\mid^2 ds=h^2\parallel v''\parallel^2_{L^2(0,\,L)}. \qquad (9.2.51)$$

Hemos por tanto probado el siguiente resultado.

Lemma 9.2.1 *Si $v\in H^2\cap H_0^1(0,\,L)$, la función de interpolación $v_h\in V_h$ es tal que*

$$\parallel v-v_h\parallel_{H_0^1(0,\,L)}\leqslant C\,h\parallel v''\parallel_{L^2(0,\,L)}. \qquad (9.2.52)$$

Pero este lema no responde por completo a la cuestión puesto que suponemos que $v\in H^2\cap H_0^1(0,\,L)$. En el caso general en que v meramente pertenezca al espacio $v\in H_0^1(0,\,L)$, por densidad de $H^2\cap H_0^1(0,\,L)$ en $H_0^1(0,\,L)$, se deduce la existencia de $v_h\in V_h$ tal que $v_h\to v$ en $H_0^1(0,\,L)$.

Obtenemos así el siguiente resultado.

Theorem 9.2.2 *El método de elementos finitos P_1 diseñado converge. Es decir, para todo $f\in H^{-1}(0,\,L)$ la solución de la aproximación de Galerkin* (9.2.35) *es tal que*

$$u_h\to u\ en\ H_0^1(0,\,L), \qquad (9.2.53)$$

siendo u la solución del problema de Dirichlet (9.2.33).

Además, bajo la hipótesis adicional $f\in L^2(0,\,L)$ el método es convergente de orden 1*, es decir,*

$$\parallel u-u_h\parallel_{H_0^1(0,\,L)}=O(h). \qquad (9.2.54)$$

Demostración. La convergencia del método se deduce de los resultados generales sobre aproximaciones de Galerkin y del hecho que, tal y como hemos comprobado, todo elemento v de V se puede aproximar mediante elementos de V_h.

La cota (9.2.54) sobre el orden de convergencia se deduce de la combinación de las cotas (9.2.6) y del Lema 9.2.1 de aproximación puesto que, cuando $f \in L^2(0,\, L)$, la solución del problema de Dirichlet (9.2.33) no sólo pertenece a la clase $H_0^1(0,\, L)$ sino que satisface la propiedad de regularidad adicional $u \in H^2(0,\, L)$.

■

En la siguiente sección vamos a extender estas ideas al caso de dos dimensiones espaciales, aunque, como veremos, el análisis de la convergencia del método va a ser más complejo en aquel caso. Una vez de haber desarrollado la teoría del MEF P_1 en el caso de dos dimensiones espaciales, su adaptación y extensión al caso de más dimensiones es más o menos inmediata si bien computacionalmente es bastante más complejo debido a las dificultades derivadas de "triangular" o mallar dominios en dimensiones de espacio $n \geqslant 3$.

9.2.5. El método de Elementos Finitos $2D$

A lo largo de esta sección suponemos que Ω es un abierto acotado de $\mathbb{R}^2$ de clase C^2 y consideramos el problema de Dirichlet

$$\left\{ \begin{array}{rcl} -\Delta u = f & \text{en} & \Omega \\ u = 0 & \text{en} & \partial\Omega \end{array} \right. \tag{9.2.55}$$

que admite la formulación variacional

$$\left\{ \begin{array}{l} u \in H_0^1(\Omega) \\ \displaystyle\int_\Omega \nabla u \cdot \nabla v dx = \langle f,\, v\rangle,\ \forall v \in H_0^1(\Omega). \end{array} \right. \tag{9.2.56}$$

Con el objeto de introducir una aproximación de la solución u en primer lugar introducimos una aproximación polinomial Ω_h del dominio Ω tal y como se refleja en la figura:

Sustituimos entonces la solución u de (9.2.55) en Ω por la solución correspondiente en Ω_h. Esto corresponde de hecho a sustituir el dominio Ω por su aproximación Ω_h lo cual en la práctica es lo mismo que suponer que el dominio Ω considerado es poligonal.

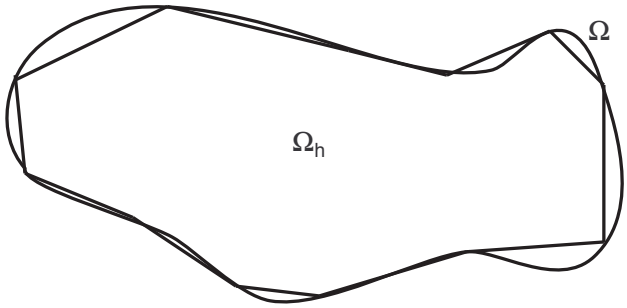


Figura 9.2: Aproximación poligonal de un dominio curvo regular.

Obviamente, al sustituir Ω por Ω_h en la resolución del problema de Dirichlet, introducimos ya una primera fuente de error que estimamos en el Apéndice A al final de estas notas. Pero continuemos por el momento, puesto que este error efectivamente tiende a cero cuando $h \to 0$ si la aproximación polinomial que Ω_h proporciona de Ω es más y más fina.

En lo que sigue supondremos que Ω es pues un dominio poligonal.

Introducimos ahora una triangulación $\mathcal{T}_h$ del dominio Ω constituida por triángulos de tamaño del orden de h. Como es habitual en el MEF supondremos que si dos triángulos de la triangulación se tocan lo hacen en un punto o bien a lo largo de todo un lado común. Obtenemos así una triangulación de la forma:

Más adelante indicaremos lo que entendemos por un mallado en el que los triángulos son de tamaño del orden de h. Por el momento podemos suponer que existen constantes positivas α y $\beta > 0$ tales que

$$\alpha \leqslant \frac{\operatorname{diam} T_h}{h} \leqslant \beta, \, \forall T_h \in \mathcal{T}_h \tag{9.2.57}$$

y para todo $h > 0$.

Por T_h denotamos los triángulos del mallado $\mathcal{T}_h$.

Subrayamos que hemos excluido expresamente las situaciones descritas en las dos siguientes figuras en las que, respectivamente, dos triángulos comparten parte de un lado sin compartirlo íntegramente, o se tocan en un sólo punto sin que este sea un vértice.

Denotamos mediante $\{x_j\}_{j\in\{1,\cdots,N\}}$ los nodos internos del mallado. Evidentemente, a medida que el mallado $\mathcal{T}_h$ es más y más fino de forma que $h \to 0$, el número N de nodos internos tiende a infinito.

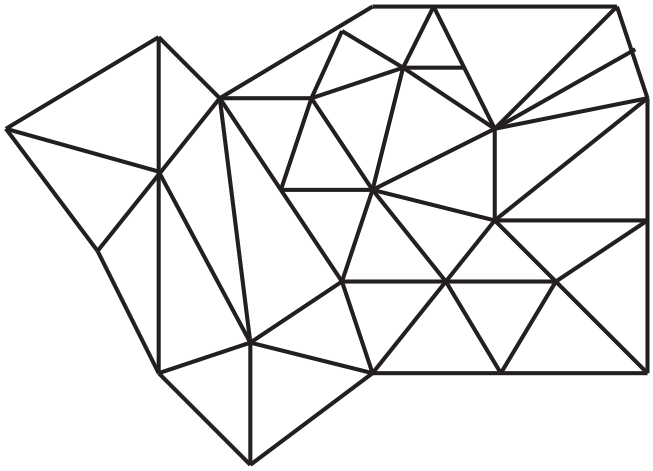


Figura 9.3: Triangulación de un dominio poligonal.

Figura 9.4: Configuraciones no admisibles en la triangulación.

Introducimos el espacio V_h constituido por las funciones lineales a trozos y continuas, que se anulan en $\partial\Omega$, generada por las funciones de base $\{\phi_j\}_{j\in\{1,\cdots,N\}}$ definidas de modo que

$$\phi_j(x_j) = 1,\ \phi_j(x_k) \neq 0,\ \text{ si } k \neq 0. \tag{9.2.58}$$

Se trata obviamente de funciones piramidales. Nuevamente consideramos el MEF P_1.

La aproximación de Galerkin adopta la forma habitual:

$$\begin{cases} u_h \in V_h \\ \displaystyle\int_\Omega \nabla u_h \cdot \nabla v_h \, dx = \langle f,\, v_h \rangle,\ \forall v_h \in V_h. \end{cases} \tag{9.2.59}$$

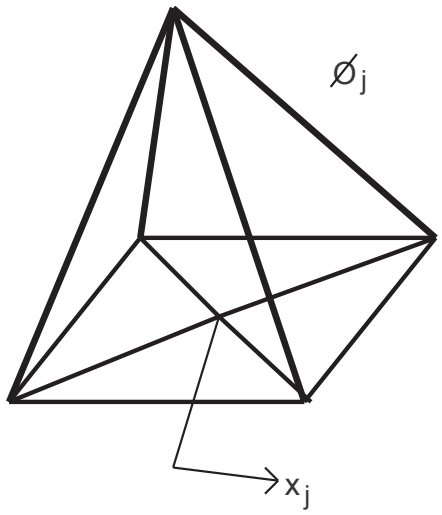


Figura 9.5: Funciones de base para el MEF P1 en dos dimensiones espaciales.

La solución de este problema existe y es única. Se trata de un sistema de N ecuaciones lineales con N incógnitas.

Tal y como ocurría en el caso $1D$, el sistema puede escribirse en la forma

$$R_h U_h = M_h F_h \tag{9.2.60}$$

donde R_h y M_h son las matrices de rigidez y de masa respectivamente.

Nuestro objetivo ahora es probar que, a medida que la talla característica h del mallado $\mathcal{T}_h$ tiende a cero, es decir, a medida que el mallado se hace más y más fino, la solución u_h de (9.2.59) converge en $H_0^1(\Omega)$ a la solución del problema continuo (9.2.55).

Para ello, según la teoría general de los métodos de Galerkin (Teorema 9.2.1) basta probar que para cada $u \in H_0^1(\Omega)$ existe una sucesión de elementos $w_h \in V_h$ tales que

$$w_h \to u \text{ en } H_0^1(\Omega),\ h \to 0. \tag{9.2.61}$$

La prueba de esta propiedad se inspira en las ideas del caso $1D$, salvo que en este caso resulta técnicamente algo más complicada.

Por densidad podemos suponer que $u \in C_c^\infty(\Omega)$. Entonces, la elección más

natural de $w_h = w_h(x)$ es la familia interpolante[2]

$$w_h(x) = \sum_{j=1}^{N} u(x_j)\phi_j(x). \tag{9.2.62}$$

Evidentemente

$$\int_\Omega \mid \nabla(u - w_h) \mid^2 dx = \sum_{j=1}^{N} \int_{T_j} \mid \nabla(w - u_h) \mid^2 dx. \tag{9.2.63}$$

Tenemos por tanto que estudiar la norma en cada uno de los triángulos:

$$I_j = \int_{T_j} \mid \nabla(w_h - u) \mid^2 dx. \tag{9.2.64}$$

Con el objeto de estudiar la norma (9.2.64) procedemos en dos etapas:

- Primero lo hacemos en el caso del triángulo $T_j = hK$ donde K es el triángulo de vértices (0, 0), (0, 1), (1, 0).
- Después, mediante un cambio de variables abordamos el caso general.

Paso 1: El caso $T_j = hK$.

En este caso, mediante un simple cambio de variables vemos que

$$I_j = \frac{1}{h}\tilde{I} \tag{9.2.65}$$

donde $\tilde{I}$ es la integral correspondiente al triángulo de referencia K.

Por otra parte

$$\tilde{I} = \int_K \mid \nabla(u - w) \mid^2 dx, \tag{9.2.66}$$

siendo w la función de interpolación de una función regular definida en K puede estimarse del modo siguiente.[3]

En lo sucesivo denotamos mediante r_k el operador de interpolación de modo que $w = r_k u$. El problema consiste pues en estimar la norma

$$\| (I - r_k)u \|_{\dot{H}^1(K)} . \tag{9.2.67}$$

Teniendo que $I - r_k$ se anula sobre las funciones afines en K, i.e. sobre P_1, tenemos que

$$\| (I - r_k)u \|_{\dot{H}^1(K)} \leqslant \| I - r_k \|_{\mathcal{L}(H^s(K),\, H^1(K))} \inf_{p \in P_1} \| u + p \|_{\dot{H}^1(K)} . \tag{9.2.68}$$

[2]Nótese que en dimensión espacial $n = 2$ basta con suponer que $u \in H^2(\Omega)$ para poder definir la función interpolante (9.2.62) puesto que en las funciones de H^2 son continuas.

[3]En este punto observamos que la restricción a K de la interpolación de u en Ω coincide con la interpolación en K de la restricción de u sobre K.

Por otra parte tenemos:

Lema de Deny-Lions: *Para cada $\ell \geqslant 0$ existe una constante $C = C(\ell)$, tal que*

$$\inf_{p \in P_\ell} \| u + p \|_{H^{\ell+1}(K)} \leqslant C \| u \|_{\dot{H}^{\ell+1}(K)} . \tag{9.2.69}$$

Observación 9.2.3 *Nosotros sólo utilizaremos este lema en el caso $\ell = 1$ pero se cumple para $\ell \geqslant 0$ arbitrario. Recordemos que $\| \cdot \|_{\dot{H}^{\ell+1}(K)}$ denota la semi-norma en $H^{\ell+1}(K)$ que sólo involucra las normas L^2 de las derivadas de orden máximo $\ell + 1$ y P_ℓ el espacio de polinomios de grado ℓ.*

Demostración del lema de Deny-Lions ($\ell = 1$)

Como sólo lo utilizaremos cuando $\ell = 1$ nos limitamos a probarlo en este caso.

En primer lugar observamos que la norma en $H^2(K)$ es equivalente a la semi-norma

$$\| v \|_* = \left[\| v \|^2_{\dot{H}^2(K)} + \sum_{|\alpha| \leqslant 1} \left(\int_K D^\alpha v \right)^2 \right]^{1/2} . \tag{9.2.70}$$

En otras palabras, existe $C > 0$ tal que

$$\| v \|^2_{H^2(K)} \leqslant C \left[\| v \|^2_{\dot{H}^2(K)} + \sum_{|\alpha| \leqslant 1} \left(\int_K D^\alpha v \right)^2 \right]^{1/2} . \tag{9.2.71}$$

La prueba se realiza mediante el argumento clásico de reducción al absurdo. Como la seminorma $\| \cdot \|_{\dot{H}^2(K)}$ involucra las normas de todas las derivadas puras de orden 2 basta con probar que

$$\| v \|^2_{H^1(K)} \leqslant C \left[\| v \|^2_{\dot{H}^2(K)} + \sum_{|\alpha| \leqslant 1} \left(\int_K D^\alpha v \right)^2 \right]^{1/2} . \tag{9.2.72}$$

Argumentamos ahora por reducción al absurdo. Si (9.2.72) no se cumple existe una sucesión de funciones $\{v_j\}_{j \geqslant 1}$ en $H^2(K)$ tales que

$$\| v_j \|_{\dot{H}^2(K)} \to 0, \; j \to \infty \tag{9.2.73}$$

$$\int_K D^\alpha v_j dx \to 0, \; \forall \mid \alpha \mid \leqslant 1, \; j \to \infty \tag{9.2.74}$$

$$\| v_j \|_{H^1(K)} = 1. \tag{9.2.75}$$

De (9.2.73) y (9.2.75) deducimos que $\{v_j\}_{j \geqslant 1}$ está acotada en $H^2(K)$. Por la compacidad de la inclusión de $H^2(K)$ en $H^1(K)$ deducimos la existencia de una

subsucesión, que seguimos denotando como $\{v_j\}$, tal que

$$v_j \rightharpoonup v \quad \text{débilmente en} \quad H^2(K) \tag{9.2.76}$$

$$v_j \to v \quad \text{fuertemente en} \quad H^1(K). \tag{9.2.77}$$

De (9.2.75) y (9.2.77) deducimos que

$$\| v \|_{H^1(K)} = 1. \tag{9.2.78}$$

Por otra parte, por la semicontinuidad inferior de la seminorma $\| \cdot \|_{\dot{H}^2(K)} = 0$ con respecto a la topología débil de $H^2(K)$ y (9.2.73) tenemos que $\| v \|_{\dot{H}^2(K)} = 0$, lo cual implica que v es una función lineal. Este hecho, combinado con (9.2.74) implica que, como $D^\alpha v \equiv 0$ para todo α tal que $| \alpha | \leqslant 1$, entonces $v \equiv 0$. Esto último está en contradicción con (9.2.78).

Esto concluye la demostración de (9.2.72).

Por otra parte, dada cualquier función $v \in H^2(K)$ existe un único polinomio $\hat{p} \in P_1$ tal que

$$\int_K D^\alpha v \, dx = -\int_K D^\alpha \hat{p} \, dx, \; \forall \, | \alpha | \leqslant 1. \tag{9.2.79}$$

En efecto, (9.2.79) constituye un sistema de tres ecuaciones algebraicas lineales con tres incógnitas que admite una única solución.

Aplicando (9.2.71) a $v + \hat{p}$ obtenemos que

$$\begin{aligned} \inf_{p \in P_1} \| v + p \|_{H^2(K)} \leqslant \| v + \hat{p} \|_{H^2(K)} &\leqslant C \| v + \hat{p} \|_* \\ &= C \| v + \hat{p} \|_{\dot{H}^2(K)} \\ &= C \| v \|_{\dot{H}^2(K)}, \end{aligned}$$

lo cual concluye la demostración del Lema de Deny-Lions en el caso $\ell = 1$.

■

Combinando (9.2.68) y (9.2.69) deducimos que

$$\| (I - r_K)u \|_{\dot{H}^1(K)} \leqslant C \| u \|_{\dot{H}^2(K)},$$

o, lo que es, lo mismo, para la integral $\tilde{I}$ de (9.2.66) tenemos la estimación

$$| \tilde{I} | \leqslant C \| u \|^2_{\dot{H}^2(K)} \, . \tag{9.2.80}$$

Volviendo ahora al intervalo original $T_j = hk$ observamos que si u está definida en T_j, entonces $v(x) = u(hx)$ está definida en K. Además

$$\tilde{I} = \int_K | \nabla(v - r_K v) |^2 = \int_{T_j} | \nabla(u - w_h) |^2 \tag{9.2.81}$$

y por otra parte

$$\| v \|^2_{\dot{H}(K)} = h^2 \| u \|^2_{\dot{H}^2(T_j)} . \tag{9.2.82}$$

De (9.2.80)-(9.2.82) se deduce que

$$| I_j | \leqslant C h^2 \| u \|^2_{\dot{H}^2(T_j)} . \tag{9.2.83}$$

Sumando las estimaciones (9.2.82) sobre todos los triángulos T_j de la triangulación τ_h deducimos que

$$\int_\Omega | \nabla(u - w_h) |^2 \, dx \leqslant C h \| u \|^2_{H^2(\Omega)}, \tag{9.2.84}$$

tal y como queríamos probar.

Paso 2: El caso general.

El caso considerado en el paso anterior en el que cada triángulo de la triangulación se considera una contracción de tamaño del triángulo de referencia K no es realista pues en general las triangulaciones tienen una estructura mucho más homogénea.

Para abordar el caso general necesitamos analizar el cambio de variables que nos permite pasar de un triángulo arbitrario T al triángulo de referencia K.

Consideramos la aplicación afín

$$\begin{cases} L_T : K \to T \\ L_T(x) = B_T x + b_T \end{cases} \tag{9.2.85}$$

que transforma el triángulo o elemento de referencia K en T, siendo B_T una matriz 2×2 y b_T un vector de traslación.

Tenemos el siguiente resultado:

Lemma 9.2.2 *Sea $v \in H^m(T)$, $m \geqslant 0$ y $w = v_0 L_T$ la correspondiente función definida en el elemento de referencia K.*

Entonces, existe una constante $C = C(m) > 0$ tal que

$$\| w \|_{\dot{H}^m(K)} \leqslant C \| B_T \|^m (\det B_T)^{-1/2} \| v \|_{\dot{H}^m(T)}, \forall v \in H^m(T) \tag{9.2.86}$$

y

$$\| v \|_{\dot{H}^m(T)} \leqslant C \| B_T^{-1} \|^m (\det B_T)^{1/2} \| w \|_{\dot{H}^m(K)}, \forall w \in H^m(K). \tag{9.2.87}$$

Demostración. La demostración es una sencilla combinación de la regla de la cadena para la derivación de la función compuesta y del teorema de cambio de variables en la integral.■

El último ingrediente que necesitamos es una estimación de las normas de B_T y de su inversa en función de la geometría del triángulo T.

Introducimos la siguiente notación

$$h_T := \text{diam}\,(T);\ \rho_T = \sup\{\text{diam}\,(B) : B \subset T,\ B = \text{ bola }\}. \tag{9.2.88}$$

Tenemos el siguiente resultado:

Lemma 9.2.3 *Se verifican las siguientes estimaciones*

$$\| B_T \| \leqslant \frac{h_T}{\rho_k};\ \| B_T^{-1} \| \leqslant \frac{h_k}{\rho_T}. \tag{9.2.89}$$

Demostración.

Tenemos

$$\| B_T \| = \frac{1}{\rho_k} \sup\{\| B_T\xi \| : \| \xi \| = \rho_k\}. \tag{9.2.90}$$

Por otra parte, para cualquier vector ξ tal que $\| \xi \| = \rho_k$, existen puntos $x,\ y \in K$ tales que $x - y = \xi$. Además, $B_T\xi = L_T x - L_T y$, puesto que L_T es una aplicación afín. Deducimos por tanto que

$$\| B_T\xi \| = \| L_T x - L_T y \| \leqslant h_T \tag{9.2.91}$$

puesto que tanto $L_T x$ como $L_T y$ pertenecen a T.

De (9.2.90) y (9.2.91) deducimos la primera desigualdad de (9.2.89). La segunda se prueba de manera semejante.

■

Estamos ya en condiciones de probar la convergencia del MEF en el caso general. En vista de las estimaciones del Lema 9.2.3 introducimos el concepto de *triangulación regular*. Diremos que las triangulaciones $\{\mathcal{T}_\langle\}_{h>0}$ del dominio Ω son regulares si existen constantes $\alpha,\ \beta > 0$ tales que

$$\alpha h^2 \leqslant | T | \leqslant \beta h^2,\ \forall T \in \tau_h,\ \forall h > 0 \tag{9.2.92}$$

y además existe $\gamma > 0$ de modo que

$$\inf_{\substack{T \in \mathcal{T}_h \\ h>0}} \{ \text{ ángulos del triángulo } T\} \geqslant \gamma. \tag{9.2.93}$$

La condición (9.2.92) asegura que todos los triángulos involucrados en la triangulación son esencialmente de un diámetro del orden de h o de área del orden de h^2. La condición (9.2.93) garantiza que los triángulos involucrados en las triangulaciones no son excesivamente excéntricos. Esta condición asegura

que ρ_T es también uniformemente del orden h de modo que existen constantes μ, $\nu > 0$ tales que

$$\nu h \leqslant \rho_T \leqslant \mu h, \, \forall T \in \tau_h, \, \forall h > 0. \tag{9.2.94}$$

Bajo estas hipótesis tenemos el siguiente resultado:

Theorem 9.2.3 *Supongamos que Ω es un dominio poligonal plano y consideremos una familia regular de triangulaciones $\{\tau_h\}_{h>0}$.*

Entonces se verifican las siguientes propiedades:

(a) *Para todo $v \in H_0^1(\Omega)$ existe una sucesión $v_h \in V_h$ tal que*

$$v_h \to v \text{ en } H_0^1(\Omega), \, h \to 0. \tag{9.2.95}$$

(b) *Existe una constante $C > 0$ tal que*

$$\| \, v - r_h v \, \|_{H_0^1(\Omega)} \leqslant C \, h \, \| \, v \, \|_{H^2(\Omega)}, \tag{9.2.96}$$

para todo $v \in H^2(\Omega) \cap H_0^1(\Omega)$ y $h > 0$.

(c) *Para todo $f \in H^{-1}(\Omega)$ las soluciones $u_h \in V_h$ obtenidas mediante el MEF-P_1 asociado a la triangulación $\{\mathcal{T}_h\}_{h>0}$ convergen en $H_0^1(\Omega)$ a la solución u del problema de Dirichlet* (9.2.55) (ó (9.2.56)).

(d) *Cuando la solución de* (9.2.55) (ó (9.2.56)) *u es tal que $u \in H^2 \cap H_0^1(\Omega)$,*

$$\| \, u - u_h \, \|_{H_0^1(\Omega)} \leqslant C \, h \, \| \, u \, \|_{H^2(\Omega)}, \tag{9.2.97}$$

para todo $h > 0$.

Observación.

• El Teorema 9.2.3(d) establece una convergencia de orden h del método para las soluciones u de clase $H^2(\Omega)$. Conviene entonces analizar bajo qué condiciones se puede garantizar que la solución del problema continuo (9.2.55) pertenece a la clase $H^2(\Omega)$. En este punto hemos de ser cuidadosos pues el dominio Ω en cuestión es poligonal y por tanto no se pueden aplicar los teoremas clásicos de regularidad de soluciones de problemas elípticos.

Sin embargo, es bien sabido que cuando Ω es un *polígono convexo* de $\mathbb{R}^2$, las soluciones u de (9.2.55) (ó (9.2.56)) pertenecen a $H^2(\Omega)$ para todo $f \in L^2(\Omega)$.

Por tanto, bajo la hipótesis de convexidad de Ω, el MEF-P_1 converge con orden 1 para todo $f \in L^2(\Omega)$.

• Conviene observar que, en la práctica, el dominio Ω poligonal en el que aplicamos el MEF-P_1 es una aproximación poligonal del dominio original que es

un dominio regular. Es obvio que si el dominio original regular es convexo las aproximaciones Ω_h de Ω se pueden tomar de modo que sean también convexas. Por tanto, el orden 1 del método queda garantizado cuando el dominio regular de entrada es convexo.

• A pesar que la estimación (9.2.97) sobre el orden de convergencia exige que el dominio sea convexo esto no es así en lo que respecta a la propiedad (c) de convergencia del método que se cumple para cualquier dominino poligonal y por tanto sin hipótesis adicional alguna sobre la convexidad del dominio regularidad sobre el que se formula el producto de Dirichlet (9.2.57) (ó (9.2.56)).■

Demostración.

Basta en realidad que completemos la prueba de (9.2.96), i.e. de (b) del Teorema 9.2.3.

En efecto, una vez que (9.2.96) se cumple, por densidad, se verifica también (a). A partir de (c) se deduce inmediatamente la convergencia del MEF-P_1, propiedad (c). Asimismo (9.2.96) implica inmediatamente (9.2.97).

Concluyamos por tanto la prueba de (9.2.96). Volviendo a (9.2.63) tenemos

$$\int_\Omega \mid \nabla(v - r_h v) \mid^2 dx = \sum_{T \in \tau_h} \int_T \mid \nabla(v - r_h v) \mid^2 dx. \tag{9.2.98}$$

Bajo las hipótesis de regularidad del mallado, del Lema 9.2.2 y del Lema 9.2.3, deducimos que

$$\int_T \mid \nabla(v - r_h v) \mid^2 dx \leqslant C\, h^2 \parallel v \parallel_{\dot{H}^2(T)} \tag{9.2.99}$$

para cada $T \in \tau_h$ y cada $h > 0$.

Combinando (9.2.98) y (9.2.99) deducimos (9.2.97). Esto concluye la demostración del Teorema 9.2.3.

■

Capítulo 10

Breve introducción al control óptimo

A lo largo de estas notas hemos desarrollado un cierto número de técnicas numéricas que tienen como objetivo principal el proporcionar métodos eficaces de aproximación de las soluciones de EDP. Pero, muchas veces, en la práctica, la obtención de la solución de una EDP no es más que uno de los pasos en la resolución del problema completo que se plantea en algún ámbito del I+D+i. Un tipo de problema, sumamente importante por su ubicuidad, es el de Control Óptimo o la Optimización. En estos problemas se trata de diseñar una estrategia óptima que nos asegure el mejor rendimiento posible del procedimiento o mecanismo en consideración. En este tipo de situaciones la EDP es el modelo que sustituye o representa la realidad en consideración, y su solución se denomina la variable de estado.

Al analizar este tipo de problemas no se trata sólo de resolver la EDP, cosa que con frecuencia se puede hacer mediante las técnicas analíticas y numéricas que hemos presentado, sino que en este caso hemos de realizar una elección óptima de algunos de los parámetros que intervienen en el modelo: los controles. Dependiendo de la aplicación en consideración el tipo de problemas de control puede ser muy diverso. Por ejemplo, son frecuentes los problemas de Diseño Óptimo en los que el parámetro a optimizar puede ser la forma del dominio en el que la EDP se verifica. Pero puede también tratarse de optimizar los coeficientes de la ecuación, lo cual correspondería al caso en que lo que hay que elegir son las propiedades materiales del medio que ocupa la geometría considerada. Puede también tratarse de optimizar una fuerza o fuente externa que puede intervenir

en la EDP tanto como un segundo miembro como una condición de contorno no homogénea.

En esta sección vamos a ilustrar las ideas básicas que subyacen en esta rica teoría considerando el caso más sencillo de una ecuación elíptica formulada en un dominio fijo y en la que el parámetro a optimizar es una fuente o fuerza externa que interviene en el segundo miembro de la EDP. En primer lugar discutiremos la existencia y la caracterización del control óptimo en el marco del modelo continuo. Después analizaremos una versión discreta basada en una aproximación por elementos finitos. Mostraremos también la existencia del óptimo en el marco discreto. Por último probaremos la convergencia del control óptimo discreto al continuo. Este último es un resultado importante pues la praxis en ingeniería consiste precisamente en, a la hora de calcular los controles óptimos, sustituir el modelo continuo por el discreto.

Con el objeto de presentar el problema concreto que vamos a abordar vamos pues a considerar un dominio acotado Ω de $\mathbb{R}^d$, con $d \geq 1$. Sea ω un subdominio de Ω que es lugar en el que el control va a poder ser aplicado. Sea 1_ω la función característica del subconjunto ω y consideremos la ecuación de estado:

$$\begin{cases} -\Delta u = f1_\omega & \text{en} \quad \Omega \\ u = 0 & \text{en} \quad \partial\Omega. \end{cases} \tag{10.0.1}$$

Este modelo nos permite describir por ejemplo la vibración de una membrana sometida a una fuerza $f = f(x)$ localizada en ω o la difusión de un contaminante concentrado en dicho conjunto.

A pesar de que el modelo considerado es particularmente sencillo la mayoría de las ideas que vamos a desarrollar en esta sección son de aplicación en un contexto mucho más amplio.

En (10.0.1) el segundo miembro $f = f(x)$ que usa una fuente o fuerza externa que se ejerce en el subconjunto ω es el control.

Para cada $f \in L^2(\omega)$, el problema (10.0.1) admite una única solución que admite la caracterización variacional:

$$\begin{cases} u \in H_0^1(\Omega) \\ \displaystyle\int_\Omega \nabla u \cdot \nabla \varphi dx = \int_\omega f\varphi dx, \quad \forall \varphi \in H_0^1(\Omega). \end{cases} \tag{10.0.2}$$

El problema de control óptimo que nos planteamos resolver es el siguiente. Dada una configuración $u_d = u_d(x)$ deseada dada, buscamos el valor del control $f = f(x)$ que hace que la solución o estado u coincida o se parezca a u_d. Hay muchas maneras de expresar esta proximidad. Una es la de minimizar un

funcional de la forma

$$J(f) = \frac{1}{2}\int_{\Omega} \mid u - u_d \mid^2 dx, \tag{10.0.3}$$

para toda función $f \in L^2(\omega)$.

Es fácil coprobar que

$$J : L^2(\omega) \to \mathbb{R} \tag{10.0.4}$$

es una función continua y convexa definida en un espacio de Hilbert. Con el objeto de probar la existencia del control óptimo, i.e. para poder aseguar que el ínfimo

$$I = \inf_{f \in L^2(\omega)} J(f) \tag{10.0.5}$$

es en realidad un mínimo, bastaría pues probar su coercividad.

En otras palabras, deseamos probar que el funcional J es coercivo

La coercividad del funcional equivale a probar que $J(f) \to \infty$ cuando $\| u \|_{L^2(\Omega)} \to \infty$. En vista de la estructura del funcional J esto es lo mismo que probar que $\| u \|_{L^2(\omega)} \to \infty$ cuando $\| f \|_{L^2(\omega)} \to \infty$. Pero esto no necesariamente ocurre pues la norma de la solución u que corresponde a la del segundo miembro f en $L^2(\omega)$ es la norma en $H^2(\Omega)$. Para convencerse de ello basta considerar el problema uni-dimensional

$$\begin{cases} -u'' = f, \quad 0 < x < \pi \\ u(0) = u(\pi) = 0 \end{cases} \tag{10.0.6}$$

en el que la solución se puede calcular de manera explícita usando series de Fourier:

$$u(x) = \sum_{k \geq 1} \frac{f_k}{k^2} \sin(kx), \tag{10.0.7}$$

siendo $\{f_k\}_{k \geq 1}$ los coeficientes de Fourier de f:

$$f(x) = \sum_{k \geq 1} f_k \sin(kx). \tag{10.0.8}$$

Mientras que la norma de f en $L^2(0, \pi)$ es del orden de $\left(\sum \mid f_k \mid^2 \right)^{1/2}$ la de u es del orden de $\left(\sum_{k \geq 1} \mid f_k \mid^2 \Big/ k^4 \right)^{1/2}$. Es obvio que esta última puede mantenerse acotada mientras que la primera diverge.

El funcional considerado, por tanto, no es coercivo y no podemos aplicar el Método Directo del Cálculo de Variaciones puesto que la sucesiones minimizantes de (10.0.5) no están necesariamente acotadas.

Desde un punto de vista analítico hay maneras sencillas de evitar esta dificultad. De hecho, en la práctica, es natural considerar que los controles f admisibles no son arbitrarios sino imponer alguna restricción sobre la talla de los mismos.

Una posibilidad por tanto consiste en, dado $M > 0$, restringir la búsqueda del óptimo a los controles f de norma no superior a M. En este caso nos encontramos entonces ante el problema de optimización:

$$I_M = \inf_{\substack{f \in L^2(\omega) \\ \|f\|_{L^2(\omega)} \leq M}} J(f). \tag{10.0.9}$$

En este caso la existencia del mínimo es obvia puesto que estamos minimizando un funcional continuo y convexo en un conjunto acotado, cerrado y convexo y por tanto cerrado de la topología débil. De este modo se puede asegurar la existencia del óptimo:

$$\exists f_M \in L^2(\omega) : \| f_M \|_{L^2(\omega)} \leq M;\, J(f_M) = \min_{\substack{f \in L^2(\omega) \\ \|f\|_{L^2(\omega)} \leq M}} J(f). \tag{10.0.10}$$

Otra manera de asegurar la existencia del óptimo es reforzar la coercividad del funcional J. Dado $N > 0$ esto puede hacerse considerando el funcional

$$J_N(f) = \frac{1}{2} \int_\Omega |\, u - u_d \,|^2 \, dx + \frac{N}{2} \int_\omega f^2 dx. \tag{10.0.11}$$

Este funcional $J_N : L^2(\omega) \to \mathbb{R}$ es continuo, convexo y coercivo en el espacio de Hilbert $L^2(\omega)$. Admite por tanto un mínimo. De hecho el funcional es estrictamente convexo, lo cual asegura la unicidad del mínimo:

$$\exists! f_N \in L^2(\omega) :\, J_N(f) = \min_{f \in L^2(\omega)} J_N(f). \tag{10.0.12}$$

El parámetro $N > 0$ del funcional J_N permite regular el coste del control. Así, cuando N es grande, el control f tenderá a tener una norma menor que cuando N es muy pequeño, caso en el que la utilización del control f no se penaliza.

De hecho, el control óptimo f_N se puede caracterizar fácilmente en este caso mediante las ecuaciones de Euler-Lagrange del problema de minimización (10.0.12) que en este caso se reducen a:

$$\langle DJ_N(f_N),\, g \rangle = 0, \quad \forall g \in L^2(\omega). \tag{10.0.13}$$

En (10.0.13) $DJ_N(f_N)$ denota la diferencial de J_N en el punto f_N.

En vista de la estructura del funcional J_N de (10.0.11), el sistema (10.0.13) se reduce a:

$$\int_\Omega (u_N - u_d)\, v dx + N \int_\omega f_N g dx = 0, \quad \forall g \in L^2(\omega), \tag{10.0.14}$$

siendo v la solución de (10.0.1) correspondiente al segundo miembro g que representa la dirección en la que se está calculando la derivada direccional, i.e.

$$\begin{cases} -\Delta v = g1_\omega & \text{en} \quad \Omega \\ v = 0 & \text{en} \quad \partial\Omega. \end{cases} \tag{10.0.15}$$

En (10.0.14) u_N representa la solución óptima asociada al control óptimo.

Si bien el MDCV garantiza la existencia del mínimo, en la práctica éste ha de ser calculado mediante la utilización de un método de descenso que necesita del cálculo de gradiente de J. De manera general tenemos que

$$\langle DJ_N(f),\, g\rangle = \int_\Omega (u-u_d)v dx + N\int_\omega fg dx. \tag{10.0.16}$$

En el práctica este modo de calcular es muy costoso pues exige, para cada función g, el cálculo de la solución (10.0.15). Una técnica habitual para la simplificación del cálculo de gradientes es la basada en el estado adjunto. El estado adjunto ϕ se define en este caso como la solución del problema

$$\begin{cases} -\Delta\phi = u - u_d & \text{en} \quad \Omega \\ \phi\Big|_{\partial\Omega} = 0, \end{cases} \tag{10.0.17}$$

siendo u la solución de la ecuación de estado (10.0.1). Obviamente, como $u-u_d \in L^2(\Omega)$, la ecuación de estado adjunto admite una única solución $\phi \in H_0^1(\Omega)$.

Multiplicando en (10.0.17) por v y en (10.0.15) por ϕ tenemos:

$$\int_\Omega (u-u_d)v dx = \int_\Omega \nabla\phi\cdot\nabla v dx = \int_\omega g\phi dx. \tag{10.0.18}$$

De este modo la expresión (10.0.16) del gradiente se reduce a

$$\langle DJ_N,\, g\rangle = \int_\omega g(\phi + Nf)dx. \tag{10.0.19}$$

De esta identidad se deduce la principal propiedad del estado adjunto ϕ: A condición de resolver el problema adjunto (10.0.17), que tiene una complejidad semejante a la ecuación de estado (10.0.1), todas las derivadas direccionales de J_N se reducen al cálculo de integrales en ω, como las que aparecen en el segundo miembro de (10.0.19). Este hecho contrasta de manera considerable con la primera expresión (10.0.16) que exige la resolución de la ecuación de estado (10.0.15) para cada valor de g.

Este mismo procedimiento nos permite caracterizar el control óptimo f_N que verifica (10.0.14). En efecto, combinando (10.0.14) y (10.0.19) deducimos que

$$\phi_N + Nf_N = 0 \text{ en } \omega$$

i.e.

$$f_N = -\frac{\phi_N}{N} \tag{10.0.20}$$

donde ϕ_N es el estado adjunto correspondiente. Escribiendo conjuntamente la ecuación de estado y la de estado adjunto correspondientes al control óptimo f_N obtenemos lo que se denomina el *sistema de optimalidad* (SO):

$$\begin{cases} \begin{cases} -\Delta u_N = -\dfrac{\phi_N}{N} 1_\omega & \text{en} \quad \Omega \\ u_N\Big|_{\partial\Omega} = 0 \end{cases} \\ \begin{cases} -\Delta \phi_N = u_N - u_d & \text{en} \quad \Omega \\ \phi_N\Big|_{\partial\Omega} = 0. \end{cases} \end{cases} \tag{10.0.21}$$

De los desarrollos anteriores sabemos que el SO (10.0.21) admite una única solución (u_N, ϕ_N) y que el control óptimo viene entonces dado por (10.0.20).

Los métodos de descenso habituales lo que hacen es, precisamente, de manera iterativa, acercamos al control óptimo (10.0.20). Para hacerlo, en vista de la expresión (10.0.19) del gradiente, se constata que en cada punto $f \in L^2(\omega)$, la dirección g de máximo descenso es aquella en que

$$g = -\frac{(\phi + Nf)}{\| \phi + Nf \|_{L^2(\omega)}} \tag{10.0.22}$$

siendo ϕ la solución correspondiente de (10.0.17).

Una vez que hemos desarrollado el cálculo de gradientes en el marco del funcional continuo (10.0.11) asociado a la ecuación de estado (10.0.1), lo cual permite a su vez la implementación de métodos de descenso eficaces para el cálculo del mínimo, hemos de desarrollar el mismo programa en el marco de una aproximación numérica por elementos finitos. Vamos primeramente a describir cómo esto puede ser desarrollado, probando la existencia del óptimo en este marco discreto y después la convergencia del óptimo discreto al óptimo continuo.

Los cambios a realizar al pasar del funcional y marco continuo al discreto son los siguientes:

- Sustituir Ω por una aproximación poligonal Ω_h y triangularla $\{T\}_{T\in\mathcal{T}_h}$, siguiendo los criterios habituales de regularidad de los mallados que garantizan la convergencia de orden 1 del método de elementos finitos en $H_0^1(\Omega)$.

- Sustituir ω por una aproximación adaptada al mallado $\mathcal{T}_h$. Esto puede hacerse de diversas maneras y, en particular, definiendo ω_h como la unión de los triángulos de $\mathcal{T}_h$ contenidos en ω.

- Sustituir el conjunto $L^2(\omega)$ de controles admisibles por $\mathcal{U}_{ad}^h$, el conjunto de funciones del espacio de elementos finitos que sólo involucra a los nodos de ω_h.

Una vez realizadas estas adaptaciones geométricas sustituimos la ecuación de estado continua (10.0.1) por su versión discreta:

$$\begin{cases} u_h \in V_h \\ \displaystyle\int_\Omega \nabla u_h \cdot \nabla \varphi dx = \int_\omega f\varphi dx, \; \forall \varphi \in V_h, \end{cases} \tag{10.0.23}$$

donde V_h es el subespacio de elementos finitos $P1$ asociado a la triangulación $\mathcal{T}_h$.

El funcional J_N de (10.0.11) ha de entonces ser reemplazado por su versión discreta:

$$J_N^h(f) = \frac{1}{2}\int_{\Omega_h} \mid u_h - u_d \mid^2 dx + \frac{N}{2}\int_{\omega_h} f^2 dx. \tag{10.0.24}$$

Se trata en esta ocasión de un funcional continuo y convexo definido en un espacio de dimensión finita. El funcional es también coercivo en este caso.

Como consecuencia de estas propiedades deducimos que el funcional J_h alcanza su valor mínimo en un único punto. Con el objeto de analizar el comportamiento en el límite cuando $h \to 0$, denotamos este punto de mínimo por $\hat{f}_h$ y la solución correspondiente de (10.0.23) por $\hat{u}_h$. Omitimos aquí la dependencia con respecto al parámetro N del funcional J_N^h con el objeto de simplificar la notación.

Procediendo como en el marco del funcional continuo J_N es fácil comprobar que el mínimo $\hat{f}_h$ se caracteriza por la versión discreta de (10.0.14), i.e.

$$\int_\Omega (\hat{u}_h - u_d)\, v_h dx + N\int_\omega f_h g dx = 0, \quad \forall g \in \mathcal{U}_{ad}^h. \tag{10.0.25}$$

Esto, a su vez, permite caracterizar el control óptimo $\hat{f}_h$ a través del sistema adjunto:

$$\begin{cases} \phi_h \in V_h \\ \displaystyle\int_\Omega \nabla \phi_h \cdot \nabla \varphi = \int_\Omega (u_h - u_d)\varphi dx, \; \forall \varphi \in V_h. \end{cases} \tag{10.0.26}$$

El sistema adjunto (10.0.26) admite una única solución $\phi_h \in V_h$ y el control óptimo $\hat{f}_h$ está entonces caracterizado por la ecuación

$$\hat{f}_h = -\phi_h \Big/ N \text{ en } \omega_h. \tag{10.0.27}$$

En este punto conviene señalar que, en el ámbito de las aplicaciones, lo habitual es precisamente utilizar este tipo de aproximaciones mediante elementos finitos (u otros métodos numéricos) y minimizar el funcional J_h con h suficientemente pequeño mediante un método descenso como, por ejemplo, el método de gradiente conjugado.

En el caso particular que nos ocupa, por su sencillez, no es difícil probar que el mínimo de funcional J_N^h converge, cuando $h \to 0$, al del funcional J_N. Este procedimiento, el de minimizar una versión discreta del funcional continuo en consideración, sin embargo, se utiliza en un contexto mucho más amplio, y con frecuencia en casos en que la prueba de la convergencia no es posible con los métodos de los que se dispone en la actualidad.

Probemos ahora la convergencia de los mínimos discretos $\hat{f}_h$.

En primer lugar, denotando mediante I_h, el mínimo del funcional J_N^h en V_h, es fácil de comprobar que

$$I_h \leq \frac{1}{2} \int_\Omega \mid u_d \mid^2 dx = J_h(0). \tag{10.0.28}$$

De la desigualdad (10.0.28) se deduce que

$$\parallel \hat{f}_h \parallel_{L^2(\omega_h)} \leq C, \tag{10.0.29}$$

lo cual a su vez implica que

$$\parallel \hat{u}_h \parallel_{H_0^1(\Omega)} \leq C. \tag{10.0.30}$$

Extrayendo subsucesiones tenemos que

$$\hat{f}_h \rightharpoonup f^* \text{ en } L^2(\omega);\ \hat{u}_h \rightharpoonup u^* \text{ en } H_0^1(\Omega), \tag{10.0.31}$$

siendo $f \in L^2(\omega)$ y u^* la solución correspondiente de (10.0.1). En la primera convergencia de (10.0.31), en la medida en que el conjunto ω_h de soporte de $\hat{f}_h$ no necesariamente coincide con (10.0.31), podemos entender que se trata de una convergencia débil en $L^2(\Omega)$ a la función $f^* 1_\omega$, de las funciones $\hat{f}_h 1_{\omega_h}$.

Debemos ahora probar que f^* es el mínimo de J. Una vez que esto haya sido probado habremos obtenido la convergencia de toda la familia f_h, cuando $h \to 0$, sin necesidad de extraer subsucesiones. Más adelante veremos que, de hecho, la convergencia tiene lugar en la topología fuerte de L^2.

Para identificar f^* con el mínimo de J procedemos mediante un argumento habitual de Γ-convergencia. Para ello basta probar que f^* es un minimizador de J, es decir que

$$J_N(f^*) \leq J_N(g), \quad \forall g \in L^2(\omega). \tag{10.0.32}$$

Para obtener (10.0.32) observamos en primer lugar que

$$J_N(f^*) \leq \frac{\lim}{h \to 0} J_N^h(\hat{f}^h), \tag{10.0.33}$$

gracias a las convergencias (10.0.31) y a la semicontinuidad inferior débil de los dos términos que intervienen en la definición de J_N^h.

De hecho, de (10.0.31) tenemos

$$\int_\omega \mid f^* \mid^2 dx \leq \frac{\lim}{h \to 0} \int_{\omega_h} \mid f_h \mid^2 dx, \tag{10.0.34}$$

y, por la compacidad de la inclusión de $H_0^1(\Omega)$ en $L^2(\Omega)$,

$$\int_\Omega \mid u^* - u_d \mid^2 dx = \frac{\lim}{h \to 0} \int_{\Omega_h} \mid \hat{u}_h - u_d \mid^2 dx. \tag{10.0.35}$$

Por otra parte, dado $g \in L^2(\omega)$ arbitrario, es fácil probar la existencia de una familia $g_h \in L^2(\omega_h)$ tal que

$$g_h \to g \text{ en } L^2(\omega) \tag{10.0.36}$$

y, por consiguiente, de modo que las soluciones correspondientes del problema discreto y continuo, v_h y v, respectivamente, satisfagan

$$v_h \to v \text{ en } H_0^1(\Omega). \tag{10.0.37}$$

De estas convergencias se deduce que

$$J_N^h(g_h) \to J_N(g). \tag{10.0.38}$$

Como, por otra parte

$$J_N^h\left(\hat{f}_h\right) \leq J_N^h(g_h), \quad \forall h > 0, \tag{10.0.39}$$

combinando (10.0.33), (10.0.38) y (10.0.39), deducimos que (10.0.32) se verifica.

Esto nos permite identificar f^* con el mínimo $\hat{f}$, y probar que es toda la sucesión $\hat{f}_h$ la que converge débilmente a $\hat{f}$ en $L^2(\omega)$. Además obtenemos

$$J_N^h\left(\hat{f}_h\right) \to J_N(\hat{f})$$

de lo cual, en virtud de (10.0.34) y (10.0.35), obtenemos asimismo que

$$\int_{\omega_h} \left|\hat{f}_h\right|^2 dx \to \int_\omega \hat{f}^2 dx.$$

Deducimos así, gracias a la convergencia débil y a la convergencia de las normas, que

$$\hat{f}_h \to \hat{f} \text{ en } L^2(\omega) \text{ fuertemente,}$$

y, por consiguiente,

$$\hat{u}_h \to \hat{u} \text{ en } H_0^1(\Omega) \text{ fuertemente.}$$

Esto concluye la prueba de la convergencia fuerte de los minimizadores.

De hecho, si usamos los resultados sobre el orden de convergencia del método de elementos finitos podemos también deducir resultados sobre el orden de convergencia de los controles.

Capítulo 11

Ecuaciones de evolución

11.1. Resolución de la ecuación del calor mediante técnicas de semigrupos

Si bien el método de semigrupos no es un método numérico de aproximación en el sentido estricto del concepto, antes de proceder al análisis numérico de la ecuación del calor, conviene aplicarlo pues nos proporciona un marco funcional adecuado para aproximar soluciones.

Consideramos por tanto un abierto Ω de $\mathbb{R}^n$ que, en principio, no es necesario suponer que sea ni acotado ni regular.

Consideramos la ecuación del calor con condiciones de contorno de Dirichlet:

$$\begin{cases} u_t - \Delta u = 0 & \text{en} \quad \Omega \times (0, \infty) \\ u = 0 & \text{en} \quad \partial\Omega \times (0, \infty) \\ u(x, 0) = u_0(x) & \text{en} \quad \Omega. \end{cases} \tag{11.1.1}$$

Con el objeto de situar el problema en el marco abstracto de la Teoría de Hille-Yosida definimos el espacio de Hilbert $X = L^2(\Omega)$ y el operador $A = D(A)$ con dominio

$$D(A) = \left\{ \varphi \in H_0^1(\Omega) : \ \Delta\varphi \in L^2(\Omega) \right\}. \tag{11.1.2}$$

El problema de valores iniciales (11.1.1) entra de este modo en el marco de los problemas abstractos de evolución que el Teorema de Hille-Yosida permite resolver. Pero para poder aplicarlo necesitamos comprobar que A es un operador maximal-disipativo. Comprobamos en primer lugar que es maximal. Para ello, dado $f \in L^2(\Omega)$ hemos de probar la existencia de una solución de

$$u \in D(A), \ (I - A)u = f. \tag{11.1.3}$$

La formulación en términos de EDP de (11.1.3) es

$$\begin{cases} u - \Delta u = f & \text{en } \Omega \\ u = 0 & \text{en } \partial\Omega \end{cases} \tag{11.1.4}$$

y su formulación variacional

$$\begin{cases} u \in H_0^1(\Omega) \\ \displaystyle\int_\Omega [\nabla u \cdot \nabla \varphi + u\varphi]\, dx = \int_\Omega f\varphi dx,\ \forall \varphi \in H_0^1(\Omega). \end{cases} \tag{11.1.5}$$

Como consecuencia del Teorema de Lax-Milgram es fácil comprobar que, efectivamente, (11.1.5) admite una única solución. Por otra parte, la solución $u \in H_0^1(\Omega)$ así obtenida es tal que

$$\Delta u = u - f \text{ en } \mathcal{D}'(\Omega) \tag{11.1.6}$$

de donde deducimos que $\Delta u \in L^2(\Omega)$ y por tanto $u \in D(A)$.

Comprobamos ahora que el operador A es disipativo. Formalmente esto es fácil de hacer utilizando integración por partes:

$$(Au,\, u)_{L^2(\Omega)} = \int_\Omega \Delta u\, u dx = -\int_\Omega |\, \nabla u\, |^2\, dx \leqslant 0. \tag{11.1.7}$$

En la práctica esto exige justificar la fórmula de integración por partes en el dominio Ω para funciones $u \in D(A)$. Cuando el dominio Ω es de clase C^2 esto no supone ninguna dificultad pues la fórmula de Green es válida para funciones de $H^2(\Omega)$ y por otra parte, por los resultados clásicos de regularidad elíptica, $D(A) = H^2 \cap H_0^1(\Omega)$. Cuando Ω no es regular la justificación de (11.1.7) exige un argumento adicional de densidad.

Como A es un operador maximal disipativo, como consecuencia directa de la aplicación del Teorema de Hille-Yosida obtenemos el siguiente resultado de existencia y unicidad de soluciones de la ecuación del calor.

Theorem 11.1.1 (a) **Soluciones débiles.** *Para todo $u_0 \in L^2(\Omega)$ la ecuación* (11.1.1) *admite una única solución $u \in C\left([0,\, \infty);\, L^2(\Omega)\right)$ que además es tal que*

$$\int_0^\infty \int_\Omega |\, \nabla u\, |^2\ dxdt \leqslant \frac{1}{2}\int_\Omega u_0^2(x)dx. \tag{11.1.8}$$

(b) **Soluciones fuertes.** *Para todo $u_0 \in H_0^1(\Omega)$ con $\Delta u_0 \in L^2(\Omega)$, la ecuación* (11.1.1) *admite una única solución*

$$u \in C\left([0,\, \infty);\, H_0^1(\Omega)\right) \cap C^1([0,\, \infty);\, L^2(\Omega) \tag{11.1.9}$$

tal que

$$\Delta u \in C([0,\, \infty);\, L^2(\Omega)). \tag{11.1.10}$$

Además, cuando el dominio Ω *es de clase* C^2 *la solución pertenece a la clase*

$$u \in C\left([0,\, \infty);\, H^2 \cap H_0^1(\Omega)\right). \tag{11.1.11}$$

Demostración.

Los resultados de existencia y unicidad de soluciones débiles y fuertes son consecuencia directa de la aplicación de los resultados abstractos de semigrupos, gracias a que hemos comprobado que A es un operador maximal-disipativo.

La estimación (11.1.8) se deduce fácilmente de la fórmula de disipación de la energía. En efecto, multiplicando la ecuación del calor de (11.1.1) por u e integrando en Ω deducimos fácilmente que

$$\frac{1}{2}\frac{d}{dt}\int_\Omega u^2(x,\, t)dx + \int_\Omega \mid \nabla u(x,\, t) \mid^2 dx = 0. \tag{11.1.12}$$

Integrando esta identidad en tiempo deducimos que

$$\frac{1}{2}\int_\Omega u^2(x,\, t)dx + \int_0^t\int_\Omega \mid \nabla u(x,\, s) \mid^2 dxds = \frac{1}{2}\int_\Omega u_0^2(x)dx,\ \forall t > 0. \tag{11.1.13}$$

De (11.1.13) deducimos que

$$\int_0^t\int_\Omega \mid \nabla u(x,\, s) \mid^2 dxds \leqslant \frac{1}{2}\int_\Omega u_0^2(x)dx. \tag{11.1.14}$$

Pasando al límite cuando $t \to \infty$ deducimos (11.1.8).

El hecho de que cuando Ω es de clase C^2 las soluciones fuertes pertenecen a la clase (11.1.11) es consecuencia inmediata de que, en ese caso, por los resultados clásicos de regularidad elíptica $D(A) = H^2 \cap H_0^1(\Omega)$.

■

Observación. Como consecuencia de (11.1.8) cuando el dominio Ω está acotado en una dirección de modo que la desigualdad de Poincaré se cumple, deducimos que $u \in L^2\left(0,\, \infty;\, H_0^1(\Omega)\right)$. Cuando la desigualdad de Poincaré no se cumple sólo podemos garantizar que $u \in L^2_{loc}\left(0,\, \infty;\, H_0^1(\Omega)\right)$. En efecto, en este caso, para acotar la norma L^2 hemos de utilizar que, como consecuencia de la identidad de energía (11.1.13),

$$\parallel u(t) \parallel_{L^2(\Omega)} \leqslant \parallel u_0 \parallel_{L^2(\Omega)},\ \forall t > 0. \tag{11.1.15}$$

Por tanto

$$\int_0^T \int_\Omega u^2 \, dxdt \leqslant T \parallel u \parallel^2_{L^\infty(0,\, T;\, L^2(\Omega))} \leqslant T \parallel u_0 \parallel^2_{L^2(\Omega)} . \tag{11.1.16}$$

Combinando (11.1.14) y (11.1.16) deducimos que

$$\int_0^T \int_\Omega \left[u^2 + \mid \nabla u \mid^2\right] dxdt \leqslant \left(T + \frac{1}{2}\right) \parallel u_0 \parallel^2_{L^2(\Omega)}, \tag{11.1.17}$$

de donde se deduce la estimación en $L^2_{loc}\big(0,\, \infty;\, H^1_0(\Omega)\big)$.

Estas propiedades establecen un efecto regularizante de las soluciones de la ecuación del calor puesto que para datos iniciales $u_0 \in L^2(\Omega)$ la solución pertenece a $L^2_{loc}\big(0,\, \infty;\, H^1_0(\Omega)\big)$. Este efecto regularizante puede cuantificarse de manera más precisa aún.

En efecto, multiplicando la ecuación del calor por $-\Delta u$ e integrando en Ω deducimos que

$$-\int_\Omega (u_t - \Delta u)\Delta u = 0$$

o, lo que es lo mismo,

$$\frac{1}{2}\frac{d}{dt}\int_\Omega \mid \nabla u \mid^2 dx + \int_\Omega \mid \Delta u \mid^2 dx = 0.$$

Por tanto,

$$\frac{d}{dt}\int_\Omega \mid \nabla u \mid^2 dx \leqslant 0$$

y, por consiguiente, la función $t \to \parallel \nabla u(t) \parallel^2_{L^2(\Omega)}$ es decreciente.

Deducimos por tanto que

$$T \parallel \nabla u(T) \parallel^2_{L^2(\Omega)} \leqslant \int_0^T \int_\Omega \mid \nabla u \mid^2 \; dxdt \leqslant \frac{1}{2} \parallel u_0 \parallel^2_{L^2(\Omega)} .$$

De este modo podemos cuantificar el efecto regularizante que muestra que las soluciones de la ecuación del calor que parten de un dato inicial en $L^2(\Omega)$ entran instantáneamente en $H^1_0(\Omega)$:

$$\parallel \nabla u(t) \parallel_{L^2(\Omega)} \leqslant \frac{1}{\sqrt{2t}} \parallel u_0 \parallel_{L^2(\Omega)} . \tag{11.1.18}$$

La estimación (11.1.18) es singular en $t = 0$. Esto no podría ser de otro modo si suponemos que $u_0 \in L^2(\Omega)$ pues, de lo contrario, el dato inicial debería de pertenecer a $H^1_0(\Omega)$.

En realidad el efecto regularizante de la ecuación del calor es mucho más fuerte aún. Cuando el dominio Ω es de clase C^∞ las soluciones de la ecuación del calor pertenecen a $C^\infty(\Omega \times (0,\, \infty))$.

11.2. Aproximación de Galerkin de la ecuación del calor

Las soluciones débiles de la ecuación del calor pueden caracterizarse mediante la siguiente formulación variacional

$$\left\{\begin{array}{ll} & u \in C\left([0,\,\infty);\, L^2(\Omega)\right) \cap L^2_{loc}\left(0,\,\infty;\, H^1_0(\Omega)\right) \\ \dfrac{d}{dt} & \displaystyle\int_\Omega u(x,\,t)\varphi(x)dx + \int_\Omega \nabla u(x,\,t)\cdot\nabla\varphi(x)dx = 0,\ \forall t>0 \\ & \displaystyle\int_\Omega u(x,\,0)\varphi(x)dx = \int_\Omega u_0(x)\varphi(x)dx,\ \forall\varphi\in H^1_0(\Omega) \end{array}\right. \tag{11.2.1}$$

En (11.2.1) hemos tomado funciones test $\varphi = \varphi(x)$. La ecuación de (11.2.1) ha por tanto de entenderse en el sentido de las distribuciones en tiempo, i.e.

$$\frac{d}{dt}\int_\Omega u(x,\,t)\varphi(x)dx + \int_\Omega \nabla u(x,\,t)\cdot\nabla\varphi(x)dx = 0,\ \text{en } \mathcal{D}'(0,\,\infty), \tag{11.2.2}$$

lo cual significa que, para cualquier función test $\psi = \psi(t) \in \mathcal{D}(0,\,\infty)$ se tiene

$$-\int_0^\infty\int_\Omega u(x,\,t)\psi(t)\varphi(x)dx + \int_0^\infty\int_\Omega \nabla u(x,\,t)\cdot\psi(t)\nabla\varphi(x)dx = 0. \tag{11.2.3}$$

Como las combinaciones lineales finitas de funciones test en variables separadas de la forma $\psi(t)\varphi(x)$ son densas en $\mathcal{D}(\Omega\times(0,\,\infty))$ deducimos entonces que la ecuación del calor se satisface en el sentido de las distribuciones:

$$-\int_0^\infty\int_\Omega u\phi_t\,dxdt + \int_0^\infty\int_\Omega \nabla u\cdot\nabla\phi\,dxdt = 0,\ \forall\phi\in\mathcal{D}(\Omega\times(0,\,\infty)). \tag{11.2.4}$$

Mediante un argumento de densidad se deduce de esta expresión que, como $u \in C\left([0,\,\infty);\, L^2(\Omega)\right) \cap L^2_{loc}\left(0,\,\infty;\, H^1_0(\Omega)\right)$, entonces

$$-\int_0^T\int_\Omega u\phi_t\,dxdt + \int_0^T\int_\Omega \nabla u\cdot\nabla\phi\,dxdt = 0 \tag{11.2.5}$$

para toda función test $\phi \in H^1(\Omega\times(0,\,T))$ tal que $\phi(x,\,T)\equiv\phi(x,\,0)\equiv 0$.

De esta expresión se puede deducir mediante un nuevo argumento de aproximación una formulación variacional de la ecuación del calor que incorpora el dato inicial:

$$-\int_0^T\int_\Omega u\phi_t\,dxdt + \int_\Omega u_0(x)\phi(x,\,0)dx + \int_0^T\int_\Omega \nabla u\cdot\nabla\phi dxdt = 0, \tag{11.2.6}$$

para todo $\phi \in H^1(\Omega\times(0,\,T))$ tal que $\phi(x,\,T)\equiv 0$.

Acabamos de ver que (11.2.1) implica (11.2.6). En realidad (11.2.6) también implica (11.2.1). Podríamos entonces adoptar, en lugar de (11.2.1), la siguiente formulación variacional de la solución de la ecuación calor, totalmente equivalente a (11.2.1):

$$\begin{cases} u \in C\left([0,\,\infty);\, L^2(\Omega)\right) \cap L^2_{loc}\left(0,\,\infty;\, H^1_0(\Omega)\right), \\ -\displaystyle\int_0^T \int_\Omega u\phi_t \, dxdt + \int_\Omega u_0(x)\phi(x,\,0)dx + \int_0^T \int_\Omega \nabla u \cdot \nabla \phi \, dxdt = 0, \\ \forall \phi \in H^1(\Omega \times (0,\,T));\, \phi(x,\,T) \equiv 0,\, \forall T > 0. \end{cases} \tag{11.2.7}$$

Los resultados que presentaremos en esta scción podrían también obtenerse trabajando con la formulación (11.2.7), pero nosotros lo haremos basándonos en (11.2.1).

Antes de continuar conviene comentar la relación existente entre la solución obtenida mediante el método de semigrupos como en la sección anterior y la solución en el sentido variacional (11.2.1). Obviamente las dos soluciones son la misma. Veámos por ejemplo que la solución obtenida mediante el método de semigrupos es también solución en el sentido débil (11.2.1).

En el caso en que el dato inicial pertenece al dominio del operador esto es obvio pues se trata de una solución fuerte que satisface la ecuación en casi todo punto. Cuando el dato inicial sólo pertenece a $L^2(\Omega)$ la solución obtenida mediante el método de semigrupos es límite de soluciones con datos iniciales en el dominio del operador. Por tanto, por densidad,[1] las soluciones en el sentido de los semigrupos satisfacen (11.2.1).

Una cuestión fundamental que cabe plantearse antes de continuar trabajando con la formulación débil es si las soluciones de (11.2.1) son únicas. En efecto lo son. Formalmente para demostrarlo basta utilizar $\varphi = u(t)$ como función test en (11.2.1) de donde se deduce la ley de energía

$$\frac{1}{2}\frac{d}{dt}\int_\Omega u^2(x,\,t)dx + \int_\Omega \mid \nabla u(x,\,t) \mid^2 dx = 0 \tag{11.2.8}$$

[1] Cuando A es un operador m-disipativo el dominio del operador es denso en X, si X es un espacio de Banach reflexivo. Para comprobarlo basta tomar un elemento $f \in X$ y resolver las ecuaciones $x - \lambda Ax = f$ para $\lambda > 0$ arbitrario. Para cada $\lambda > 0$ la solución $x_\lambda \in X$ de este problema existe y es única. Además $x_\lambda \in D(A)$ y $x_\lambda \to f$ débilmente cuando $\lambda \to 0$. Esto permite garantizar que las soluciones de problemas de evolución obtenidas mediante el método de semigrupos para operadores m-disipativos, son límites de soluciones con datos iniciales en el dominio del operador.

que implica en particular que

$$\int_\Omega u^2(x,\, t)dx \leqslant \int_\Omega u_0^2(x)dx. \tag{11.2.9}$$

De (11.2.9) se obtiene la unicidad con facilidad puesto que si u y $\tilde{u}$ son soluciones con el mismo dato inicial, la diferencia $u - \tilde{u}$ es solución de (11.2.1) con el dato inicial $u_0 \equiv 0$. Al aplicar (11.2.9) a $u - \tilde{u}$ deduciríamos que $u \equiv \tilde{u}$.

Pero, en la práctica, este argumento necesita de desarrollos técnicos adicionales puesto que la utilización de $u(t)$ como función test no esté del todo justificada puesto que: a) $u(t)$ depende de t y en (11.2.1) sólo se admiten funciones test independientes de t; b) la regularidad de u que (11.2.1) proporciona no permite garantizar que $u(t) \in H_0^1(\Omega)$ para todo $t \geqslant 0$.

Pero hemos visto ya que las soluciones débiles en el sentido de (11.2.1) lo son también en el sentido de (11.2.7). A partir de (11.2.7) la unicidad se deduce fácilmente por el método de dualidad o transposición. Para ello consideramos el problema adjunto:

$$\begin{cases} -\phi_t - \Delta\phi = f & \text{en} \quad \Omega \times (0,\, T) \\ \phi = 0 & \text{en} \quad \partial\Omega \times (0,\, T) \\ \phi(x,\, T) = 0 & \text{en} \quad \Omega. \end{cases} \tag{11.2.10}$$

La teoría de semigrupos, mediante la fórmula de variación de constantes garantiza que cuando $f \in \mathcal{D}(\Omega \times (0,\, T))$, (11.2.10) admite una única solución ϕ que satisface todos los requerimientos de las funciones test de (11.2.7). De (11.2.7) y usando la ecuación $\phi_t + \Delta\phi = -f$ que ϕ satisface se deduce que

$$\int_\Omega u_0(x)\phi(x,\, 0)dx + \int_0^T \int_\Omega f\, udxdt = 0.$$

Si aplicamos esta identidad a dos posibles soluciones débiles u y $\tilde{u}$ con el mismo dato inicial $u_0 \in L^2(\Omega)$ deducimos que

$$\int_0^T \int_\Omega f(u - \tilde{u})dxdt = 0,\, \forall f \in \mathcal{D}(\Omega \times (0,\, T)) \tag{11.2.11}$$

lo cual garantiza que $u \equiv \tilde{u}$.

A partir de este momento adoptamos (11.2.1) como formulación débil de las soluciones de la ecuación del calor sabiendo que la solución de (11.2.1) existe, es única y coincide con la obtenida en la sección anterior por la técnica de semigrupos.

Procedemos ahora a introducir la aproximación de Galerkin.

Es muy fácil introducir ahora una aproximación de Galerkin. Dada una familia de subespacios V_h de $H_0^1(\Omega)$ que cubren todo el espacio $H_0^1(\Omega)$ cuando

$h \to 0$ de modo que cada uno de los subespacios V_h son de dimensión finita $\dim(V_h) = N_h$ con $N_h \to \infty$, $h \to 0$ consideramos el problema

$$\begin{cases} & u_h \in C([0,\,\infty);\, V_h) \\ \dfrac{d}{dt} & \displaystyle\int_\Omega u_h(x,\,t)\varphi(x)dx + \int_\Omega \nabla u_h(x,\,t)\cdot\nabla\varphi(x)dx = 0, \quad \forall t > 0 \\ & \displaystyle\int_\Omega u_h(x,\,0)\varphi(x)dx = \int_\Omega u_0(x)\varphi(x)dx, \qquad\qquad \forall \varphi \in V_h. \end{cases} \tag{11.2.12}$$

La diferencia entre la formulación variacional de la ecuación del calor continuo (11.2.1) y su aproximación de Galerkin es que si bien en la primera consideramos todas las funciones test de $H_0^1(\Omega)$, en la aproximación de Galerkin sólo consideramos las funciones test en el espacio V_h de dimensión N_h.

La aproximación de Galerkin (11.2.12) constituye por tanto un sistema de N_h ecuaciones diferenciales acopladas de orden uno. Escribamos el sistema de manera más explícita.

Sea $\{\phi_1, \cdots, \phi_{N_h}\}$ una base de V_h de modo que

$$V_n = \operatorname{span}\{\Phi_1, \cdots, \phi_{N_h}\}. \tag{11.2.13}$$

La función u_h solución de (11.2.12), por pertenecer a la clase $C([0,\,\infty);\, V_h)$, es por tanto necesariamente de la forma

$$u_h(x,\,t) = \sum_{j=1}^{N_h} u_j(t)\phi_j(x) \tag{11.2.14}$$

Podemos por tanto identificar la solución con el vector incógnita

$$U_h(t) = \begin{pmatrix} u_1(t) \\ \vdots \\ u_{N_h}(t) \end{pmatrix}. \tag{11.2.15}$$

El sistema (11.2.12) puede entonces escribirse en la forma

$$\begin{cases} M_h \dfrac{d\vec{U}_h}{dt}(t) + R_h\vec{U}_h(t) + 0, \quad t > 0 \\ \vec{U}_h(0) = \vec{U}_{0,\,h}, \end{cases} \tag{11.2.16}$$

donde M_h y R_h son las matrices de masa y de rigidez asociadas a la aproximación de Galerkin. Los elementos de M_h y R_h son respectivamente:

$$M_h = (m_{jk})_{1\leqslant j,\,k\leqslant N_h}\,; \qquad m_{jk} = \int_\Omega \phi_j\phi_k dx \tag{11.2.17}$$

$$R_h = (r_{jk})_{1\leqslant j,\,k\leqslant N_h}\,; \quad r_{jk} = \int_\Omega \nabla\phi_j\cdot\nabla\phi_k dx. \tag{11.2.18}$$

Ambas matrices son simétricas y definidas positivas. La simetría es obvia. Para comprobar el carácter definido positivo lo hacemos por ejemplo en el caso M_h, siendo la prueba semejante para R_h. Dado un vector $\vec{W}_h$ tenemos

$$\langle M_h \vec{W}_h, \vec{W}_h \rangle = \int_\Omega w^2(x)dx$$

donde

$$w(x) = \sum_{j=1}^{N_h} w_j \phi_j(x),$$

siendo $\{w_j\}_{1 \leqslant j \leqslant N_h}$ las componentes del vector $\vec{W}_h$; donde se deduce que M_h es definida positiva.

El dato inicial $\vec{U}_{0,\,h}$ del sistema (11.2.16) es de la forma

$$\vec{U}_{0,\,h} = \begin{pmatrix} u_{0,\,1} \\ \vdots \\ u_{0,\,N_h} \end{pmatrix} \tag{11.2.19}$$

donde

$$u_{0,\,j} = \int_\Omega u_0(x)\phi_j(x)dx,\ j = 1, \cdots, N_h. \tag{11.2.20}$$

La solución $\vec{U}_h = \vec{U}_h(t)$ de (11.2.6) existe y es única y puede representarse mediante la siguiente fórmula exponencial

$$\vec{U}_h(t) = e^{-M_h^{-1} R_h t} \vec{U}_{0,\,h}. \tag{11.2.21}$$

El problema que se plantea es analizar de qué modo las soluciones de (11.2.12) o, equivalentemente, de (11.2.16), convergen cuando $h \to 0$ a la solución de la ecuación del calor. Para hacerlo es más sencillo trabajar con las funciones $u_h = u_h(x,\, t)$ obtenidas como soluciones de las aproximaciones de Galerkin que con los vectores $\vec{U}_h$ de las componentes de la solución en la base V_h.

Antes de pasar al límite es preciso obtener estimaciones uniformes, independientes de h, de las soluciones. El modo más natural de hacerlo es utilizar el método de la energía. Para ello hemos de emplear en la formulación de Galerkin (11.2.12) la propia solución u_h como función test. Pero esto no es posible, al menos de manera inmediata, puesto que la solución u_h depende en particular de t y en (11.2.12) sólo se pueden admitir funciones test independientes de t. Con el objeto de justificar la utilización de u_h como función test es conveniente considerar la formulación (11.2.16) de la aproximación de Galerkin como sistema de EDO de orden uno. Las soluciones $\vec{U}_h = \vec{U}_h(t)$ son funciones regulares, incluso

analíticas de t a valores en $\mathbb{R}^{N_h}$. Podemos entonces multiplicar escalarmente en (11.2.16) por $\vec{U}_h$ y deducir

$$\frac{1}{2}\frac{d}{dt}\langle M_h\vec{U}_h(t),\, \vec{U}_h(t)\rangle + \langle R_h\vec{U}_h(t),\, \vec{U}_h(t)\rangle = 0. \tag{11.2.22}$$

Utilizando las definiciones de las matrices M_hy R_h y la propia estructura de las soluciones $u_h = u_h(x,\, t)$ de la aproximación de Galerkin deducimos que

$$\frac{1}{2}\frac{d}{dt}\int_\Omega u_h^2(x,\, t)dx + \int_\Omega \mid \nabla u_h(x,\, t) \mid^2 dx = 0.$$

Vemos por tanto que las aproximaciones de Galerkin de la ecuación del calor satisfacen la misma identidad de energía que las soluciones del propio problema continuo.

Integrando en tiempo deducimos que

$$\frac{1}{2}\int_\Omega u_h^2(x,\, t)dx + \int_0^t\int_\Omega \mid \nabla u_h(x,\, t) \mid^2 dxdt = \frac{1}{2}\int_\Omega u_{0,\,h}^2(x)dx \tag{11.2.23}$$

de donde deducimos inmediatamente que

$$\{u_h\}_{h>0} \text{ está acotada en } L^\infty(0,\, \infty;\, L^2(\Omega)) \cap L^2_{loc}\left(0,\, \infty;\, H_0^1(\Omega)\right). \tag{11.2.24}$$

Una vez que tenemos la cota uniforme (11.2.24) podemos pasar al límite cuando $h \to 0$. La cuestión central es si el límite de la sucesión $\{u_h\}_{h>0}$ es la solución de la ecuación del calor continua.

A partir de (11.2.24), extrayendo subsucesiones que seguimos denotando mediante el índice h para simplificar la notación, tenemos

$$u_h \rightharpoonup u \text{ débil } * \text{ en } L^\infty(0,\, \infty;\, L^2(\Omega) \text{ y débilmente en } L^2_{loc}\left(0,\, \infty;\, H_0^1(\Omega)\right). \tag{11.2.25}$$

Debemos probar que u es la solución de la ecuación del calor. En este punto es esencial la hipótesis de que los subespacios V_h llenan todo el espacio $H_0^1(\Omega)$ o, en otras palabras, que para todo $\varphi \in H_0^1(\Omega)$ existe $\varphi_h \in V_h$ tal que

$$\varphi_h \to \varphi \quad \text{en} \quad H_0^1(\Omega). \tag{11.2.26}$$

Combinando las convergencias (11.2.25) y (11.2.26) deducimos que

$$\int_\Omega^T u_h(x,\, t)\varphi_h(x)dx \to \int_\Omega u(x,\, t)\varphi(x)dx \text{ en } L^\infty(0,\, \infty) \text{ débil}* \tag{11.2.27}$$

y

$$\int_\Omega \nabla u_h(x,\, t)\cdot\nabla\varphi_h dx \to \int_\Omega \nabla u(x,\, t)\cdot\nabla\varphi(x)dx \text{ débilmente en } L^2_{loc}(0,\, \infty). \tag{11.2.28}$$

Deducimos así que el límite $u = u(x,\, t)$ satisface

$$\begin{cases} u \in L^\infty\left(0,\, \infty;\, L^2(\Omega)\right) \cap L^2_{loc}\left(0,\, \infty;\, H^1_0(\Omega)\right) \\ \dfrac{d}{dt}\displaystyle\int_\Omega u(x,\, t)\varphi(x)dx + \int_\Omega \nabla u(x,\, t)\cdot\nabla\varphi(x)dx = 0,\ \forall \varphi \in H^1_0(\Omega). \end{cases} \tag{11.2.29}$$

Con el objeto de poder garantizar que el límite u obtenido de este modo es la solución de (11.2.1) hemos de comprobar que:

(a) $u \in C([0,\, \infty);\, L^2(\Omega))$,

(b) $u(x,\, 0) = u_0(x)$ en Ω.

Vamos ahora a comprobar estas dos propiedades del límite u. Pero antes de hacerlo señalamos que, como el límite que hemos obtenido está identificado de manera única, es toda la sucesión $\{u_h\}_{h>0}$ de soluciones aproximadas de Galerkin la que converge en el sentido de (11.2.25) sin necesidad de extraer subsucesiones.

Tenemos por tanto el siguiente resultado:

Theorem 11.2.1 *Supongamos que los subespacios V_h de $H^1_0(\Omega)$ de dimensión finita considerados son tales que para todo $\varphi \in H^1_0(\Omega)$ existe $\varphi_h \in V_h$ tal que*

$$\varphi_h \to \varphi \quad \text{en} \quad H^1_0(\Omega) \quad \text{cuando} \quad h \to 0. \tag{11.2.30}$$

Entonces, para todo $u_0 \in L^2(\Omega)$ las soluciones de Galerkin $\{u_h\}_{h>0}$ obtenidas resolviendo (11.2.12) *son tales que* (11.2.25) *se satisface.*

Concluyamos por tanto probando las dos propiedades del límite u que hemos dejado pendientes.

• **Continuidad en tiempo de u**

Como $u = u(x,\, t)$ verifica (11.2.29) satisface la ecuación del calor en el sentido de las distribuciones:

$$u_t - \Delta u = 0 \text{ en } \mathcal{D}'(\Omega \times (0,\, \infty)).$$

Como $u \in L^2\left(0,\, T;\, H^1_0(\Omega)\right)$ para todo $0 < T < \infty$, $\Delta u \in L^2\left(0,\, T;\, H^{-1}(\Omega)\right)$. Por tanto $u_t \in L^2\left(0,\, T;\, H^{-1}(\Omega)\right)$.

Es un resultado clásico que si $u \in L^2\left(0,\, T;\, H^1_0(\Omega)\right)$ y $u_t \in L^2\left(0,\, T;\, H^{-1}(\Omega)\right)$, entonces $u \in C\left([0,\, T];\, L^2(\Omega)\right)$.

La idea de la prueba es la siguiente. En primer lugar constatamos que $u \in C_w\left([0,\, \infty);\, L^2(\Omega)\right)$ es débilmente continua de manera que

$$t \to \int_\Omega u(x,\, t)\varphi dx$$

es continua para cada $\varphi \in L^2(\Omega)$.

Basta entonces probar la continuidad de las normas. Consideramos entonces

$$\begin{aligned}\left|\int_\Omega u^2(x,\,t_2)dx - \int_\Omega u^2(x,\,t_1)dx\right| &= \left|\int_\Omega \left[u^2(x,\,t_2) - u^2(x,\,t_1)\right]dx\right| \\ &= \left|\int_\Omega\int_{t_1}^{t_2} \partial_t[u^2(x,\,t)]dtdx\right| = \left|2\int_{t_1}^{t_2}\int_\Omega u\,u_t\,dxdt\right| \\ &\leqslant 2\int_{t_1}^{t_2} \parallel u(t) \parallel_{H_0^1(\Omega)} \parallel u_t(t) \parallel_{H^{-1}(\Omega)} dt \to 0,\ t_1 \to t_2\end{aligned} \tag{11.2.31}$$

puesto que $u \in L^2\left(0,\,T;\,H_0^1(\Omega)\right)$ y $u_t \in L^2\left(0,\,T;\,H^{-1}(\Omega)\right)$.

• Identificación del dato inicial

Como sabemos ya que $u \in C\left([0,\,\infty);\,L^2(\Omega)\right)$ su dato inicial $u(x,\,0)$ está bien definido y es un elemento de $L^2(\Omega)$. Basta ver que se trata del dato $u_0 \in L^2(\Omega)$ de (11.2.1).

Para ello en primer lugar observamos que las soluciones de (11.2.12) satisfacen también

$$\begin{aligned}-\int_0^T\int_\Omega u_h\psi_t\varphi_h(x)dxdt + \int_0^T\int_\Omega \nabla u_h\cdot\psi\nabla\varphi_h dxdt \\ -\int_\Omega u_{h,\,0}\psi(0)\varphi_h(x)dx = 0,\end{aligned} \tag{11.2.32}$$

para todo $0 < T < \infty$, para todo $\varphi_h \in V_h$ y para todo $\psi \in C^1([0,\,T])$ tal que $\psi(T) = 0$.

La convergencia (11.2.25) permite pasar al límite en (11.2.32). Obtenemos así

$$-\int_0^T\int_\Omega u\psi_t\varphi dxdt + \int_0^T\int_\Omega \nabla u\cdot\psi\nabla\varphi dxdt - \int_\Omega u_0(x)\psi(0)\varphi(x)dx = 0. \tag{11.2.33}$$

En el paso al límite en los datos iniciales de (11.2.32) hemos utilizado la hipótesis (11.2.30) que garantiza que φ_h converge a φ fuertemente en $H_0^1(\Omega)$. Pero necesitamos entonces saber que los datos iniciales $u_{h,\,0}$ elegidos son tales que

$$u_{h,\,0} \to u_0 \text{ en } H^{-1}(\Omega). \tag{11.2.34}$$

Recordemos que dado $u_0 \in L^2(\Omega)$ hemos elegido los datos iniciales aproximantes $u_{0,\,h}$ de modo que

$$u_{0,\,h}(x) = \sum_{j=1}^{N_h}\int_\Omega u_0(x)\phi_j(x)dx\phi_j(x). \tag{11.2.35}$$

Obviamente, cuando $\{\phi_j\}$ es una base ortogonal de V_h, $u_{0,\,h}$ es la proyección ortogonal de $u_0 \in L^2(\Omega)$ sobre V_h. En general, $u_{0,h}$ se obtiene de la proyercción a través del producto con la matriz de masa correspondiente.En cualquier caso tenemos

$$\| u_{0,\,h} \|_{L^2(\Omega)} \leqslant C \| u_0 \|_{L^2(\Omega)}, \tag{11.2.36}$$

con $C = 1$ en el caso de la base ortogonal.

Como $H_0^1(\Omega)$ es denso en $L^2(\Omega)$ basta con que probemos que (11.2.34) se cumple para todo $u_0 \in H_0^1(\Omega)$. Ahora bien, como por definición $u_{0,\,h}$ es el elemento de V_h más próximo a u_0 en la norma $L^2(\Omega)$ y (11.2.30) se cumple (lo cual implica, obviamente, que $\varphi_h \to \varphi$ en $L^2(\Omega)$), deducimos que (11.2.34) se cumple para todo $u_0 \in H_0^1(\Omega)$.

Esto concluye la prueba de la convergencia del método de Galerkin.

11.3. Breve introducción a la Teoría de Semigrupos

En las secciones anteriores hemos abordado el problema de la aproximación numérica de ecuaciones elípticas. Pero la mayoría de problemas relevantes del ámbito de las Ciencias y de la Tecnología son problemas de evolución en los que interviene la variable temporal.

Los métodos numéricos para la aproximación de EDP son esencialmente una superposición o combinación de los métodos numéricos propios de las ecuaciones elípticas, que conducen a ecuaciones semi-discretas, y los métodos de discretización temporal de EDO. El lector interesado en una introducción elemental a estas técnicas en el marco de los problemas en una dimensión espacial podrá consultar las notas [36]. En [35] abordamos con más profundidad los métodos en diferencias finitas para ecuaciones de tipo ondas.

A la hora de realizar un estudio sistemático de estos problemas es preciso disponer de un marco funcional adecuado. En este sentido, una de las mejores alternativas es el uso de la teoría de semigrupos que, de manera muy versátil, permite abordar el problema de la existencia, unicidad y regularidad de soluciones de numerosos problemas de EDP de evolución.

La teoría de semigrupos garantiza la existencia y unicidad de soluciones de ecuaciones abstractas de la forma

$$\begin{cases} U_t = AU, & t > 0 \\ U(0) = U_0, \end{cases} \tag{11.3.1}$$

siendo A un operador lineal no acotado en un espacio de Banach.

En estas notas vamos a recordar las nociones y resultados más básicos de la Teoría de semigrupos y en particular el Teorema de Hille-Yosida en su versión más elemental (véase, por ejemplo, el capítulo VII del libro [2]).

Con el objeto de enunciar este importante Teorema conviene recordar la noción del operador maximal disipativo.[2]

Definition 11.3.1 *Un operador $A : D(A) \subset H \to H$ lineal, no acotado, en el espacio de Hilbert H se dice disipativo si*

$$(Av,\, v)_H \leqslant 0,\, \forall v \in D(A). \tag{11.3.2}$$

Se dice que es maximal-disipativo si, además, satisface la siguiente condición de maximalidad:

$$R(I - A) = H \Leftrightarrow \forall f \in H,\, \exists \in D(A) \quad \text{t.q.} \quad u - Au = f. \tag{11.3.3}$$

Theorem 11.3.1 *(de Hille-Yosida).*

Sea A un operador maximal-disipativo en un espacio de Hilbert H. Entonces, para todo $u_0 \in D(A)$ existe una función

$$u \in C\Big([0,\, \infty);\, D(A)\Big) \cap C^1\Big([0,\, \infty);\, H\Big) \tag{11.3.4}$$

única tal que

$$\begin{cases} \dfrac{du}{dt} = Au \qquad en \quad [0,\, \infty) \\ u(0) = u_0. \end{cases} \tag{11.3.5}$$

Además se tiene

$$\| u(t) \| \leqslant \| u_0 \|_H,\, \left\| \frac{du}{dt}(t) \right\|_H = \| Au(t) \|_H \leqslant \| Au_0 \|_H,\, \forall t > 0. \tag{11.3.6}$$

En virtud del Teroema de Hille-Yosida, cuando A es un operador maximal disipativo, es el *generador de un semigrupo* $S(t) : H \to H$ que a cada $u_0 \in H$ asocia el valor $u(t) = S(t)u_0$ de la solución del problema abstracto (11.3.5) en cada instante de tiempo $t > 0$.

[2] En el contexto de los sistemas de la Mecánica la palabra "disipativo" tiene un sentido preciso: Se dice que un sistema de evolución es disipativo si la energía de las soluciones decrece en tiempo. Es este precisamente el sentido del término en el marco abstracto en la Teoría de Operadores que se desprende de (11.3.2). En efecto, multiplicando escalarmente en H la primera ecuación (11.3.5) por u, en virtud de (11.3.2) deducimos que $(\frac{d}{dt}u(t), u(t))_H = \frac{1}{2}\frac{d}{dt} \| u(t) \|_H^2 = \langle Au(t),\, u(t) \rangle \leqslant 0$.

El semigrupo $S(t)$ también se denota habitualmente como e^{At}, en vista de la analogía del sistema abstracto (11.3.5) con el clásico sistema lineal de ecuaciones diferenciales lineales con coeficientes constantes en el que A es una matriz.

El semigrupo $\{S(t)\}_{t\geqslant 0} = \{e^{At}\}_{t\geqslant 0}$ es una familia uniparamétrica de operadores lineales acotados en H. En realidad, en virtud de (11.3.6), $S(t)$ es una contracción para cada $t \geqslant 0$. Por otra parte, el semigrupo verifica las siguientes propiedades:

- $S(0) = I$,
- $t \to S(t)u_0$ es continua de $[0, \infty)$ en H para cada $u_0 \in H$,
- $S(t) \circ S(s) = S(t + s)$.

La última propiedad, denominada propiedad de semigrupo, es debida al carácter autónomo (o invariante por traslaciones temporales) de la ecuación (11.3.1).

En el teorema anterior hemos enunciado el resultado de existencia y unicidad de soluciones fuertes en el dominio $D(A)$ del operador. Pero el semigrupo $S(t)$ se extiende a todo el espacio H de modo que para todo U_0 en H existe una única solución $U = U(t)$ en la clase $C([0, \infty); H)$ que viene dada por $U(t)) = S(t)U_0$ y que constituye una solución débil del problema.

La posibilidad de obtener soluciones débiles a partir de funciones fuertes puede explicarse fácilmente en el marco del problema abstracto (11.3.5). En efecto, suponiendo que A es un operador maximal disipativo, consideremos el problema abstracto y definamos la función

$$v(t) = \int_0^t u(s)ds + v_0. \tag{11.3.7}$$

Integrando a su vez la ecuación de (11.3.5) con respecto al tiempo obtenemos

$$u(t) - u_0 = A \int_0^t u ds$$

que podemos reescribir de la siguiente manera:

$$u(t) = A \int_0^t u ds + u_0 \Leftrightarrow v_t = Av - Av_0 + u_0.$$

Por lo tanto, para poder garantizar que también v es una solución del problema abstracto (11.3.5) basta con elegir v_0 de modo que

$$Av_0 = u_0. \tag{11.3.8}$$

Supongamos que, dado $u_0 \in H$, (11.3.8) admite una única solución $v_0 \in D(A)$. Entonces la función v definida en (11.3.7) satisface

$$\begin{cases} v_t = Av, & t > 0 \\ v(0) = v_0. \end{cases} \tag{11.3.9}$$

En virtud del Teorema de Hille-Yosida, como $v_0 \in D(A)$, la ecuación (11.3.9) admite una única solución fuerte

$$v \in D\Big([0,\, \infty);\, D(A)\Big) \cap C^1\Big([0,\, \infty);\, H\Big). \tag{11.3.10}$$

De (11.3.10) deducimos que

$$u = v_t \in C\Big([0,\, \infty);\, H\Big). \tag{11.3.11}$$

Vemos de este modo que, cuando $u_0 \in H$, la ecuación abstracta admite una única solución débil en la clase (11.3.11).

En la definición de operador maximal disipativo se garantiza que $I - A$ es un operador con rango pleno. Pero nada se dice del operador A. Conviene sin embargo señalar que ésto es irrelevante a la hora de resolver el problema abstracto (11.3.5). En efecto, introduzcamos el clásico cambio de variables

$$w(t) = e^{\lambda t} u(t), \tag{11.3.12}$$

donde $\lambda \in \mathbb{R}$.

Entonces $w_t = e^{\lambda t}[u_t + \lambda u]$. Por tanto, u es solución de (11.3.5) si y sólo si w es solución de

$$\begin{cases} w_t = Aw + \lambda w, & t > 0 \\ w(0) = u_0. \end{cases} \tag{11.3.13}$$

Esto indica que el cambio de variable permite transformar soluciones fuertes (resp. débiles) de (11.3.5) en soluciones fuertes (resp. débiles) de (11.3.13) y viceversa.

Por otra parte, cuando A es maximal-disipativo, para $\lambda = -1$, el operador $A - I$ de (11.3.13) es de rango pleno. Esto permite utilizar el argumento anterior de integración en tiempo para obtener soluciones débiles a partir de las soluciones fuertes directamente en (11.3.13) cuando $\lambda = -1$ (porque el problema correspondiente a (11.3.8) podría efectivamente garantizarse que tiene una única solución $v_0 \in D(A)$ para cada $u_0 \in H$).

El Teorema de Hille-Yosida se extiende con una prueba muy semejante y con pequeños desarrollos técnicos adicionales al caso de espacios de Banach X. En

ese caso la mayor diferencia radica en la definición de operador disipativo. La condición a imponer es que

$$\| (I - \lambda A)x \|_X \geqslant \| x \|_X, \, \forall x \in D(A), \, \lambda > 0. \tag{11.3.14}$$

Es fácil comprobar que si A es un operador disipativo en un espacio de Hilbert tal y como lo hemos definido anteriormente, también lo es según esta definición.

Veamos ahora algunas de las ideas fundamentales de la demostración del Teorema fundamental de esta teoría: El Teorema de Hille-Yosida.

La idea central de la demostración de Hille-Yosida, que permite utillizar la propiedad del operador A de ser maximal-disipativo, es introducir y usar la *regularización de Yosida* del operador A:

$$A_\lambda = -\frac{1}{\lambda}(I - (I - \lambda A)^{-1}). \tag{11.3.15}$$

Es fácil comprobar que, formalmente, A_λ converge a A cuando $\lambda \to 0$. Para ello basta analizar la expresión algebraica de la derecha de (11.3.15) en el caso de números reales:

$$-\frac{1}{\lambda}\left[1 - \frac{1}{1 - \lambda_x}\right] = -\frac{1}{\lambda}\left[\frac{1 - \lambda x - 1}{1 - \lambda_x}\right] = \frac{x}{1 - \lambda_x} \to x, \, \lambda \to 0.$$

Pero para que la definición (11.3.15) tenga rigurosamente sentido es primeramente preciso probar que el operador $I - \lambda A$ es inversible. La hipótesis de maximalidad sobre A garantiza que esto es así cuando $\lambda = 1$. Veamos que ésto permite probar que $I - \lambda A$ es inversible para todo $\lambda > 1/2$. En efecto, reescribimos la ecuación

$$x - \lambda A x = y \tag{11.3.16}$$

como

$$x - Ax = \frac{1}{\lambda} y + \left(1 - \frac{1}{\lambda}\right) x,$$

o, lo que es lo mismo,

$$x = (I - A)^{-1}\left[\frac{1}{\lambda} y + \left(1 - \frac{1}{\lambda}\right) x\right]. \tag{11.3.17}$$

Es fácil comprobar que cuando $| \, 1 - 1/\lambda \, | < 1$ el segundo miembro de (11.3.17) admite una única solución para el Teorema de punto fijo de Banach.

Iterando este argumento se puede comprobar que $(I - \lambda A)^{-1}$ está bien definido para todo $\lambda > 0$. Además

$$\| (I - \lambda A) \|^{-1}_{\mathcal{L}(H, H)} \leqslant 1. \tag{11.3.18}$$

En efecto, como A es disipativo, $\langle Ax,\, x\rangle \leqslant 0$ y por tanto, si $x = (I-\lambda A)^{-1}y$ tenemos

$$\langle (I-\lambda A)^{-1}y,\, y\rangle = \langle x,\, (I-\lambda A)x\rangle = \langle x, x\rangle - \lambda\langle x,\, Ax\rangle \geqslant \langle x,\, x\rangle = \| x \|_H^2 = \| (I-\lambda A)^{-1}y \|_H^2$$

de donde se deduce que, efectivamente,

$$\| (I-\lambda A)^{-1}y \|_H \leqslant \| y \|_H, \quad \forall y \in H, \tag{11.3.19}$$

lo cual equivale a (11.3.18).

Deducimos por tanto que A_λ, para cada $\lambda > 0$, es un operador lineal y acotado de H en H.

Esto nos permite resolver la ecuación abstracta

$$\begin{cases} u' = A_\lambda u, \quad t > 0 \\ u(0) = u_0, \end{cases} \tag{11.3.20}$$

como si se tratase de una EDO.

En efecto, como A_λ es un operador lineal y acotado, $e^{A_\lambda t}$ se puede definir, como en el caso matricial, mediante el desarrollo en serie de potencias de la exponencial

$$e^{A_\lambda t} = \sum_{k=0}^{\infty} \frac{(A_\lambda t)^k}{k!}. \tag{11.3.21}$$

Es fácil comprobar que $e^{A_\lambda t}$ define un operador lineal y acotado de H en H para cada $\lambda > 0$ y $t > 0$. Además

$$u_\lambda(t) = e^{A_\lambda t}u_0 \in C^\infty([0,\, \infty);\, H) \tag{11.3.22}$$

y es la única solución de (11.3.20).

La regularización de Yosida genera entonces un semigrupo $S_\lambda(t) = e^{A_\lambda t}$.

Además, para todo $\lambda > 0$, A_λ hereda la propiedad de A de ser disipativo, de modo que

$$\langle A_\lambda x,\, x\rangle \leqslant 0,\ \forall x \in H,\ \forall \lambda > 0. \tag{11.3.23}$$

Entonces

$$\| u_\lambda(t) \|_H \leqslant \| u_0 \|,\ \left\| \frac{du_\lambda(t)}{dt} \right\|_H = \| A_\lambda u_\lambda(t) \|_H \leqslant \| A_\lambda u_0 \|_H,\ \forall t > 0,\ \forall \lambda > 0. \tag{11.3.24}$$

Para comprobar (11.3.23) basta proceder del modo siguiente

$$\begin{aligned} \langle A_\lambda x,\, x\rangle &= \langle A_\lambda x,\, x - (I - A\lambda)^{-1}x\rangle + \langle A_\lambda x,\, (I-\lambda A)^{-1}x\rangle \\ &= -\lambda \| A_\lambda x \|^2 + \langle A_\lambda x,\, (I-\lambda A)^{-1}x\rangle \\ &= -\lambda \| A_\lambda x \|^2 + \langle A(I-\lambda A)^{-1}x,\, (I-\lambda A)^{-1}x\rangle \leqslant -\lambda \| A_\lambda x \|^2 \leqslant 0. \end{aligned}$$

Como, al menos formalmente, $A_\lambda \to A$ cuando $\lambda \to 0$, en virtud de las cotas uniformes (11.3.24) de las soluciones de las ecuaciones aproximadas (11.3.20) en las que el operador A ha sido sustituido por su regularización de Yosida A_λ cabe esperar que la solución u de (11.3.5) en el Teorema de Hille-Yosida se obtenga como límite cuando $\lambda \to 0$ de las soluciones aproximadas u_λ.

La clave de la demostración del Teorema de Hille-Yosida consiste en ver que, cuando $u_0 \in D(A)$, $\{u_\lambda(t)\}_{\lambda>0}$ constituye una sucesión de Cauchy en $C([0,\infty);H)$ cuando $\lambda \to 0$.

En efecto, tenemos

$$\frac{du_\lambda}{dt} - \frac{du_\mu}{dt} = A_\lambda u_\lambda - A_\mu u_\mu$$

y por tanto

$$\frac{1}{2}\frac{d}{dt} \parallel u_\lambda - u_\mu \parallel_H^2 = \langle A_\lambda u_\lambda - A_\mu u_\mu,\, u_\lambda - u_\mu\rangle.$$

Pero

$$\begin{aligned}
&\langle A_\lambda u_\lambda - A_\mu u_\mu,\, u_\lambda - u_\lambda\rangle = \\
&= \langle A_\lambda u_\lambda - A_\mu u_\mu,\, u_\lambda - (I-\lambda A)^{-1}u_\lambda + (I-\lambda A)^{-1}u_\lambda - (I-\mu A)^{-1}u_\mu + (I-\mu A)^{-1}u_\mu - u_\mu\rangle \\
&= \langle A_\lambda u_\lambda - A_\mu u_\mu,\, -\lambda A_\lambda u_\lambda + \mu A_\mu u_\mu\rangle \\
&\quad + \langle A((I-\lambda A)^{-1}u_\lambda - (I-\mu A)^{-1}u_\mu),\, (I-\lambda A)^{-1}u_\lambda - (I-\mu A)^{-1}u_\mu\rangle \\
&\leqslant \langle A_\lambda u_\lambda - A_\mu u_\mu,\, -\lambda A_\lambda u_\lambda + \mu A_\mu x_\mu\rangle.
\end{aligned}$$

Por tanto

$$\frac{1}{2}\frac{d}{dt} \parallel u_\lambda - u_\mu \parallel_H^2 \leqslant 2(\lambda+\mu) \parallel Au_0 \parallel_H^2 . \tag{11.3.25}$$

En este punto hemos utilizado la segunda cota de (11.3.24) junto con $\parallel A_\lambda x \parallel_H \leqslant \parallel Ax \parallel_H$, para todo $x \in H$ y $\lambda > 0$, propiedad esta que se deduce fácilmente de la identidad $A_\lambda = (I-\lambda A)^{-1}A$.

La estimación (11.3.25) garantiza la propiedad de Cauchy de la sucesión u_λ en $C([0,\infty);H)$ que habíamos enunciado.

El mismo argumento permite probar que si $u_0 \in D(A^2)$, entonces du_λ/dt es también de Cauchy en $C([0,\infty);H)$. Esto permite pasar al límite en (11.3.20) cuando $\lambda \to 0$ y obtener la solución del problema abstracto (11.3.5) que el Teorema de Hille-Yosida enuncia cuando $u_0 \in D(A)$. Como $D(A^2)$ es denso en $D(A)$, un argumento de densidad permite concluir la existencia de solución para datos $u_0 \in D(A)$, tal y como se enuncia en el Teorema 11.3.1.

El lector interesado en una demostración completa del Teorema de Hille-Yosida puede consultar el capítulo VII del libro de H. Brezis [2]. En el libro de T. Cazenave y A. Haraux [3] se da también una extensión de este resultado a espacios de Banach y diversas aplicaciones a ecuaciones de evolución semilineales entre las que se incluyen la ecuación del calor, de ondas y de Schrödinger.

11.4. La ecuación de ondas continua

En esta sección vamos a indicar el modo en que la ecuación de ondas puede enmarcarse en el contexto de la teoría de semigrupos que se muestra mucho más flexible a la hora de abordar sus variantes que, por ejemplo, el método de Fourier ([35]).

Consideremos pues la ecuación de ondas

$$\begin{cases} u_{tt} - \Delta u = 0 & \text{en} \quad \Omega \times (0,\, \infty) \\ u = 0 & \text{en} \quad \partial\Omega \times (0,\, \infty) \\ u(x,\, 0) = u_0,\, u_t(x,\, 0) = u_1(x) & \text{en} \quad \Omega, \end{cases} \tag{11.4.1}$$

donde Ω es un abierto de $\mathbb{R}^n$, $n \geqslant 1$, que supondremos acotado para simplificar la presentación, si bien esta hipótesis no es en absoluto esencial.

Conviene escribir la ecuación de ondas como un sistema de orden uno:

$$\begin{cases} u_t &= v \\ v_t &= \Delta u. \end{cases} \tag{11.4.2}$$

De este modo la incógnita genuina del sistema es el par $U = (u,\, v) = (u,\, u_t)$, lo cual coincide con nuestra intuición según la cual la verdadera incógnita no es sólamente la *posición* u sino también la *velocidad* u_t. Por otra parte, esto explica que en (5.0.1) tomemos los datos iniciales u_0 y u_1 para u y u_t respectivamente.

En la variable vectorial U el sistema (11.4.2) puede escribirse formalmente como[3]

$$U_t = AU \tag{11.4.3}$$

donde A es el operador lineal

$$A = \begin{pmatrix} 0 & I \\ \Delta & 0 \end{pmatrix}, \tag{11.4.4}$$

siendo I el operador identidad y Δ el operador de Laplace.

Pero la escritura (11.4.3)-(11.4.4) es puramente formal. En efecto, como es bien sabido, en el marco de los espacios de Hilbert (o de Banach) de dimensión infinita, una definición rigurosa del opeador exige no solamente que indiquemos el modo en que actúa sino también su dominio.

El espacio natural para resolver la ecuación de ondas es el espacio de Hilbert

$$H = H_0^1(\Omega) \times L^2(\Omega). \tag{11.4.5}$$

[3]En este punto abusamos de la notación, pues U se trataría del vector columna $\binom{u}{u_t}$ si bien, para simplificar la escritura a veces lo escribiremos como vector fila.

La elección de este espacio es efectivamente natural en vista de los siguientes hechos:

- *La energía*

$$E(t) = \frac{1}{2}\int_\Omega \left[\mid \nabla u(x,\, t) \mid^2 + \mid u_t(x,\, t) \mid^2 \right] dx \tag{11.4.6}$$

  se conserva en tiempo, lo cual puede ser comprobado formalmente multiplicando la ecuación de ondas por u_t e integrando en Ω.

  La conservación de la energía sugiere que, efectivamente, es natural buscar soluciones tales que $u \in H^1(\Omega)$ y $u_t \in L^2(\Omega)$.

- La condición de contorno de Dirichlet, $u = 0$ en $\partial\Omega$, sugiere la necesidad de buscar soluciones que se anulen en la frontera. Es bien conocido que, en el marco del espacio de Sobolev H^1, la manera más natural de interpretar esta condición es exigir que $u \in H_0^1(\Omega)$.

  El espacio de la energía H *es un espacio de Hilbert* dotado de la norma:

$$\parallel (f,\, g) \parallel_H = \left[\Big|\Big| f \Big|\Big|^2_{H_0^1(\Omega)} + \Big|\Big| g \Big|\Big|^2_{L^2(\Omega)} \right]^{1/2}. \tag{11.4.7}$$

Por otra parte, las normas $\parallel \cdot \parallel_{L^2(\Omega)}$, $\parallel \cdot \parallel_{H_0^1(\Omega)}$ están definidas de la manera usual[4]:

$$\Big|\Big| f \Big|\Big|_{H_0^1(\Omega)} = \left[\int_\Omega \mid \nabla f \mid^2 dx \right]^{1/2};\ \Big|\Big| g \Big|\Big|_{L^2(\Omega)} = \left[\int_\Omega g^2 dx \right]^{1/2}. \tag{11.4.8}$$

Definimos el operador A como un operador lineal no-acotado en H. Para ello establecemos que el dominio del operador A es precisamente el subespacio de los elementos $V \in H$ para los que $AV \in H$. En vista de la estructura de A ésto da como resultado el dominio:

$$\begin{aligned} D(A) &= \{(u,\, v) \in H_0^1(\Omega) \times L^2(\Omega) :\, v \in H_0^1(\Omega),\, \Delta u \in L^2(\Omega)\} \\ &= \{(u,\, v) \in H_0^1(\Omega) \times H_0^1(\Omega) :\, \Delta u \in L^2(\Omega)\}. \end{aligned} \tag{11.4.9}$$

Cuando el dominio Ω es de clase C^2 el resultado clásico de regularidad elíptica que garantiza que las funciones $u \in H_0^1(\Omega)$ tales que $\Delta u \in L^2(\Omega)$ pertenecen en realidad a $H^2(\Omega)$, permite reescribir el dominio de la manera siguiente

$$D(A) = \Big[H^2(\Omega) \cap H_0^1(\Omega) \Big] \times H_0^1(\Omega). \tag{11.4.10}$$

[4]En este punto utilizamos implícitamente el hecho que Ω sea *acotado*. En efecto, si no lo fuese (o si, de manera más general, si Ω no fuese acotado en una dirección) no se podría garantizar que la desigualdad de Poincaré se verifica, lo cual a su vez no permitiría garantizar que la norma definida en (11.4.8) fuese equivalente a la inducida por $H^1(\Omega)$ sobre el subespacio $H_0^1(\Omega)$.

En este punto conviene subrayar que la hipótesis de que Ω sea regular de clase C^2 no es en absoluto esencial. Todo lo que vamos a decir en lo sucesivo identificando el dominio con (11.4.10) es también cierto, sin la hipótesis de regularidad del abierto Ω, tomando (11.4.9) como definición del dominio del operador.

En lo sucesivo supondremos por tanto que Ω, además de ser acotado, es de clase C^2.

Es fácil comprobar que A es un operador anti-adjunto, i.e.

$$A^* = -A. \tag{11.4.11}$$

Basta para ello utilizar el hecho de que el operador de Laplace A con dominio $H^2(\Omega) \cap H_0^1(\Omega)$ en el espacio de Hilbert $L^2(\Omega)$ es un operador autoadjunto.

Pero, de hecho, para comprobar la antisimetría que (11.4.11) indica basta con realizar el siguiente cálculo elemental:

$$\begin{aligned} \Big(AV,\, \tilde{U}\Big)_H &= \Big(v,\, \tilde{u}\Big)_{H_0^1(\Omega)} + \Big(\Delta u,\, \tilde{v}\Big)_{L^2(\Omega)} \\ &= \int_\Omega [\nabla v \cdot \nabla \tilde{u} + \Delta u \tilde{v}]\, dx = -\int_\Omega [v \Delta \tilde{u} + \nabla u \cdot \nabla \tilde{v}]\, dx = -\Big(U,\, A\tilde{U}\Big)_H \end{aligned} \tag{11.4.12}$$

para todo U, $\tilde{U} \in D(A)$.

En (11.4.12) y en lo sucesivo mediante $(\cdot,\, \cdot)_H$ denotamos el producto escalar en H. En vista de la estructura de H como espacio producto, el producto escalar en H es la suma de los productos escalares en $H_0^1(\Omega)$ y $L^2(\Omega)$ de las primeras y segundas componentes del vector V respectivamente.

Con esta definición del operador A podemos ahora escribir la ecuación de ondas (11.4.1) en la forma del problema de Cauchy abstracto (11.3.1).

Como veíamos, tenemos dos tipos de soluciones (11.3.1). Aquellas que denominaremos *soluciones fuertes* tales que[5]

$$U \in C\Big([0,\, \infty);\, D(A)\Big) \cap C^1([0,\, \infty);\, H). \tag{11.4.13}$$

En este caso tanto el término de la izquierda como de la derecha (11.3.1) son funciones bien definidas que pertenecen al espacio $C([0,\, \infty);\, H)$ y, por tanto, la ecuación de (11.3.1) tiene sentido en el espacio H para todo valor de $t > 0$. La segunda ecuación (11.3.1) relativa al dato inicial tiene también sentido pues, por la continuidad de U en tiempo a valores en $D(A)$, $U(0)$ está bien definida

[5]El dominio $D(A)$ de un operador se puede dotar de estructura Hilbertiana a través de la norma

$$\| u \|_{D(A)} = [\| u \|_H^2 + \| Au \|_H^2]^{1/2}$$

.

en $D(A)$. Es por este hecho precisamente que sólo cabe esperar la existencia de soluciones fuertes cuando el dato inicial U_0 de (11.3.1) pertenece a $D(A)$.

En términos de la posición u y velocidad u_t de la solución de la ecuación de ondas (11.4.1), la regularidad (11.4.13) equivale a

$$u \in C\Big([0,\,\infty);\, H^2 \cap H_0^1(\Omega)\Big) \cap C^1\Big([0,\,\infty);\, H_0^1(\Omega)\Big) \cap C^2\Big([0,\,\infty);\, L^2(\Omega)\Big). \tag{11.4.14}$$

Es también claro que (11.4.14) permite dar un sentido a todas las ecuaciones de (11.4.1). En particular, la ecuación de ondas se verifica, para cada $t > 0$, en $L^2(\Omega)$ y, por tanto, en particular, para casi todo $x \in \Omega$.

Las *soluciones débiles* de (11.3.1) son menos regulares. Son en realidad aquéllas que pertenecen al espacio de la energía, i.e.

$$U \in C([0,\,\infty);\, H) \tag{11.4.15}$$

o bien

$$u \in C\Big([0,\,\infty);\, H_0^1(\Omega)\Big) \cap C^1\Big([0,\,\infty);\, L^2(\Omega)\Big). \tag{11.4.16}$$

Cabe preguntarse por el sentido de (11.3.1) bajos las condiciones de regularidad (11.4.15). En efecto, este sentido no está a priori claro pues (11.4.15) no permite definir, en principio, AU, al no pertenecer U a $D(A)$ ni permite calcular la derivada temporal de U.

A pesar de ello, tiene efectivamente sentido hablar de soluciones débiles de (11.4.1) o (11.3.1) y esto se puede ver con más claridad en el contexto de (11.4.1) y bajo la condición de regularidad (11.4.16). En efecto, es bien sabido que el operador $-\Delta$ define un isomorfismo de $H_0^1(\Omega)$ en su dual $H^{-1}(\Omega)$. Por tanto, como $u \in C\Big([0,\,\infty);\, H_0^1(\Omega)\Big)$, tenemos también que $\Delta u \in C\Big([0,\,\infty);\, H^{-1}(\Omega)\Big)$. Por otra parte, como $u \in C\Big([0,\,\infty);\, H_0^1(\Omega)\Big)$, se trata en particular de una distribución por lo que su derivada segunda temporal u_{tt} está bien definida en el espacio de las distribuciones $\mathcal{D}'(\Omega \times (0,\,\infty))$. La ecuación de ondas (11.4.1) tiene por tanto sentido en el marco de las distribuciones. Ahora bien, como $\Delta u \in C\Big([0,\,\infty);\, H^{-1}(\Omega)\Big)$, de la propia ecuación de ondas deducimos que $u_{tt} \in C\Big([0,\,\infty);\, H^{-1}(\Omega)\Big)$ y entonces la ecuación de ondas tiene sentido, para todo $t > 0$ en $H^{-1}(\Omega)$. Vemos por tanto que las soluciones débiles de la ecuación de ondas, en la clase (11.4.16), por ser soluciones de la ecuación de ondas, tienen la propiedad de regularidad adicional

$$u \in C^2\Big([0,\,\infty);\, H^{-1}(\Omega)\Big). \tag{11.4.17}$$

Es fácil comprobar que el operador A asociado a la ecuación de ondas (11.4.1) definido anteriormente es maximal disipativo. El hecho de que A sea anti-adjunto

($A^* = -A$) garantiza que tanto A como $-A$ son disipativos en el sentido de la Definición 11.3.1.

En efecto, de (11.4.12) deducimos que

$$(AU,\, U)_H = 0 \tag{11.4.18}$$

lo cual garantiza la disipatividad[6] de A y $-A$.

Por otra parte, el operador A de la ecuación de ondas verifica también la condición de maximalidad (11.3.3). En efecto, para comprobarlo basta ver que para todo par $(f,\, g) \in H$, i.e. $f \in H_0^1(\Omega)$, $g \in L^2(\Omega)$, existe al menos una solución de la ecuación

$$(I - A)\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} f \\ g \end{pmatrix} \tag{11.4.19}$$

con $(u,\, v) \in D(A) = \left[H^2 \cap H_0^1(\Omega)\right] \times H_0^1(\Omega)$. Esto es efectivamente cierto. Dada la forma explícita del operador A el sistema (11.4.19) se escribe del siguiente modo

$$u - v = f,\; v - \Delta u = g. \tag{11.4.20}$$

La primera ecuación de (11.4.20) puede reescribirse como

$$v = u - f \tag{11.4.21}$$

y entonces la segunda adquiere la forma

$$u - \Delta u = g + f. \tag{11.4.22}$$

Como $g + f \in L^2(\Omega)$, la segunda ecuación (11.4.22), que puede escribirse de manera más precisa como

$$\begin{cases} u - \Delta u = f + g & \text{en} \quad \Omega \\ u = 0 & \text{en} \quad \partial\Omega, \end{cases} \tag{11.4.23}$$

admite una única solución $u \in H^2 \cap H_0^1(\Omega)$ por los resultados clásicos de existencia, unicidad y regularidad para el problema de Dirichlet. Como $f \in H_0^1(\Omega)$

[6] Conviene observar que cuando A satisface (11.4.18) las soluciones de la ecuación abstracta

$$\frac{du}{dt}(t) = Au(t)$$

conservan la energía puesto que

$$\frac{1}{2}\frac{d}{dt}\Big|\Big|u(t)\Big|\Big|_H^2 = (Au(t),\, u(t)) = 0.$$

y $u \in H^2 \cap H_0^1(\Omega)$ la solución v de (11.4.21) satisface entonces $v \in H_0^1(\Omega)$. Deducimos entonces que (11.4.19) admite una única solución en $D(A)$, lo cual garantiza la maximalidad de A.

El Teorema 11.3.1, aplicado a la versión abstracta (11.3.1) de la ecuación de ondas (11.4.1) proporciona de manera inmediata la existencia y unicidad de soluciones fuertes. En efecto, se tiene:

Theorem 11.4.1 *Si Ω es un dominio acotado de clase C^2, para cada par de datos iniciales $(u_0,\, u_1) \in \Big[H^2 \cap H_0^1(\Omega)\Big] \times H_0^1(\Omega)$, la ecuación de ondas posee una única solución fuerte en la clase*

$$u \in C\Big([0,\, \infty);\, H^2 \cap H_0^1(\Omega)\Big) \cap C^1\Big([0,\, \infty);\, H_0^1(\Omega)\Big) \cap C^2\Big([0,\, \infty);\, L^2(\Omega)\Big). \tag{11.4.24}$$

Sólo nos resta deducir la existencia y unicidad de soluciones en el espaico de la energía. Tenemos para ello varias opciones. Una de ellas consiste en analizar el operador de ondas como operador no acotado en el espacio de Hilbert $\tilde{H} = L^2(\Omega) \times H^{-1}(\Omega)$ con dominio $H \subset \tilde{H}$. Es fácil comprobar que el operador A antes definido es también un operador maximal disipativo en este marco funcional. De este modo, como consecuencia del Teorema de Hille-Yosida deducimos que:

Theorem 11.4.2 *En las hipótesis del Teorema* 11.4.1*, si los datos iniciales $(u_0,\, u_1) \in H_0^1(\Omega) \times L^2(\Omega)$ la ecuación de ondas* (11.4.1) *admite una única solución en*

$$u \in C\Big([0,\, \infty);\, H_0^1(\Omega)\Big) \cap C^1\Big([0,\, \infty);\, L^2(\Omega)\Big) \cap C^2\Big([0,\, \infty);\, H^{-1}(\Omega)\Big). \tag{11.4.25}$$

Conviene observar que ambos teoremas de existencia y unicidad (Teoremas 11.4.1 y 11.4.2) proporcionan resultados semejantes pero en espacios que difieren en una derivada en su regularidad.

En el caso de la ecuación de ondas, (11.3.8) puede resolverse directamente sin apelar al cambio de variables. En efecto, en este caso, el problema (11.3.8) puede reescribirse de la siguiente manera: Dadp $(u_0,\, u_1) \in H_0^1(\Omega) \times L^2(\Omega)$ hallar $(v_0,\, v_1) \in \Big[H^2 \cap H_0^1(\Omega)\Big] \times H_0^1(\Omega)$ tal que

$$v_1 = u_0;\, \Delta v_0 = u_1. \tag{11.4.26}$$

La primera ecuación de (11.4.26) proporciona inmediatamente la solución $v_1 \in H_0^1(\Omega)$. Por otra parte, como $u_1 \in L^2(\Omega)$, sabemos que el problema elíptico

$$-\Delta v_0 = -u_1 \text{ en } \Omega;\, v_0 = 0 \text{ en } \partial\Omega, \tag{11.4.27}$$

admite una única solución $v_0 \in H^2 \cap H_0^1(\Omega)$.

Por lo tanto, en el marco de la ecuación de ondas (11.3.8) admite una única solución, la cual permite obtener soluciones débiles de la ecuación de ondas a partir de las soluciones fuertes, a través del cambio de variable (11.3.7).

Como ya hemos indicado anteriormente, el operador de ondas A es antiadjunto, y ésto equivale a la ley de conservación de energía (11.4.6). Vemos por tanto cómo la Teoría semigrupos permite recuperar todos los resultados obtenidos mediante series de Fourier o separación de variables, pero con la ventaja de ofrecer un marco mucho más flexible para abordar otras ecuaciones.

Consideramos por último la ecuación de ondas no-homogénea:

$$\begin{cases} u_{tt} - \Delta u = f & \text{en} \quad \Omega \times (0, \infty) \\ u = 0 & \text{en} \quad \partial\Omega \times (0, \infty) \\ u(0) = u_0,\, u_t(0) = u_1 & \text{en} \quad \Omega. \end{cases} \tag{11.4.28}$$

En este caso (11.4.28) describe las vibraciones del cuerpo Ω sometido a una fuerza exterior $f = f(x, t)$.

El problema (11.4.28) también puede ser escrito en el marco de los problemas abstractos que se pueden abordar en el contexto de la Teoría de Semigrupos. En efecto, la primera ecuación (11.4.28) puede escribirse como el sistema

$$\begin{cases} u_t &= v \\ v_t &= \Delta u + f, \end{cases} \tag{11.4.29}$$

que puede también reformularse como el problema abstracto

$$\begin{cases} U_t + AU = F, \quad t > 0 \\ U(0) = U_0 \end{cases} \tag{11.4.30}$$

donde $U = (u,\, u_t)$, A es el generador del semigrupo de la ecuación de ondas que acabamos de estudiar y

$$F(t) = \begin{pmatrix} 0 \\ f(t) \end{pmatrix}. \tag{11.4.31}$$

Vemos por tanto que la fuerza externa F aplicada en la versión abstracta (11.4.30) del sistema (11.4.28) tiene una primera componente nula mientras que la función f de (11.4.28) interviene sólo en su segunda componente.

Inspirándonos en la fórmula de variación de las constantes para la resolución de ecuaciones diferenciales no homogéneas, el problema abstracto (11.4.30) puede escribirse en la forma integral siguiente

$$U(t) = S(t)U_0 + \int_0^t S(t-s)F(s)ds = e^{At}U_0 + \int_0^t e^{A(t-s)}F(s)ds, \tag{11.4.32}$$

siendo $S(t) = e^{At}$ el semigrupo generado por el operador maximal disipativo A.

En virtud de los resultados anteriores sobre las soluciones fuertes y débiles del sistema abstracto (11.3.5) asociado al operador A, es fácil deducir que

- Si $F \in L^2(0,\, T;\, D(A))$, entonces $e^{A(t-s)}F(s) \in L^1(0,\, t;\, D(A))$ para todo $0 < t < T$.

  Basta para ello utilizar las estimaciones (11.3.6) que, con las notaciones presentes, garantizan que

  $$\left|\left|e^{A(t-s)}F(s)\right|\right|_H \leqslant \left|\left|F(s)\right|\right|_H,\ \left|\left|Ae^{A(t-s)}F(s)\right|\right|_H \leqslant \left|\left|AF(s)\right|\right|_H,$$

  para todo $t \geqslant s$ y casi todo $s \in [0,\, T]$.

  Deducimos entonces que

  $$\int_0^t e^{A(t-s)}F(s)ds \in C([0,\, T];\, D(A)).$$

  Sin embargo, para que podamos garantizar que se tiene una solución fuerte es necesario también que

  $$\int_0^t e^{A(t-s)}F(s)ds \in C^1([0,\, T];\, H)$$

  para lo que es también necesario que $F \in C([0,\, T];\, H)$

- Si $F \in L^1(0,\, T;\, H)$, entonces $e^{A(t-s)}F(s) \in L^1(0,\, t;\, H)$ para todo $0 \leqslant t \leqslant T$ y por tanto

  $$\int_0^t e^{A(t-s)}F(s)ds \in C([0,\, T];\, H).$$

De estos hechos deducimos los siguientes resultados de existencia y unicidad para el sistema abstracto no homogéneo (11.4.30):

- Si $U_0 \in D(A)$ y $F \in C([0,\, T];\, H) \cap L^1(0,\, T;\, D(A))$ entonces (11.4.30) admite una única solución fuerte en la clase

  $$U \in C([0,\, T];\, D(A)) \cap C^1([0,\, T];\, H).$$

  El mismo resultado es válido bajo la hipótesis de que $F \in W^{1,\,1}(0,\, T;\, H)$.

- Si $U_0 \in H$ y $F \in L^1(0,\, T;\, H)$, entonces (11.4.30) admite una única solución débil $U \in C([0,\, T];\, H)$.

Aplicando estos resultados a la ecuación de ondas no-homogénea (11.4.28) obtenemos los siguientes resultados de existencia y unicidad:

- Si $(u_0,\, u_1) \in [H^2 \cap H_0^1(\Omega)] \times H_0^1(\Omega)$ y $f \in C([0,\, T];\, L^2(\Omega)) \cap L^1(0,\, T;\, H_0^1(\Omega))$, entonces (11.4.28) admite una única solución fuerte u en la clase (11.4.24).

- Si $(u_0,\, u_1) \in H_0^1(\Omega) \times L^2(\Omega)$ y $f \in L^1(0,\, T;\, L^2(\Omega))$, entonces (11.4.28) admite una solución débil

$$u \in C([0,\, T];\, H_0^1(\Omega)) \cap C^1([0,\, T];\, L^2(\Omega)). \tag{11.4.33}$$

En este punto conviene subrayar que, salvo que impongamos condiciones adicionales al segundo miembro f, no podemos garantizar que

$$u \in C^2([0,\, T];\, H^{-1}(\Omega)). \tag{11.4.34}$$

En efecto, como $u \in C([0,\, T];\, H_0^1(\Omega))$ y $-\Delta$ es un isomorfismo de $H_0^1(\Omega)$ en $H^{-1}(\Omega)$, tenemos $-\Delta u \in C([0,\, T];\, H^{-1}(\Omega))$. Por tanto, para que (11.4.34) pueda cumplirse, en vista de la ecuación $f = u_{tt} - \Delta u$, es imprescindible que $f \in C([0,\, T];\, H^{-1}(\Omega))$.

El cambio de variable (11.3.7) también puede ser aplicado en el marco de la ecuación abstracta (11.4.30) y permite nuevamente establecer una correspondencia biunívoca entre soluciones fuertes y débiles.

Las mismas técnicas que las desarrolladas en el caso homogéneo pueden ser también utilizadas en el no homogéneo. Esto nos permite, por ejemplo, construir soluciones ultradébiles de (11.4.28). De este modo obtenemos que si $(u_0,\, u_1) \in L^2(\Omega) \times H^{-1}(\Omega)$ y $F \in L^1(0,\, T;\, H^{-1}(\Omega))$, la ecuación (11.4.28) admite una única solución ultra-débil en la clase

$$u \in C([0,\, T];\, L^2(\Omega)) \cap C^1([0,\, T];\, H^{-1}(\Omega)).$$

Además, si $f \in C\Big([0,\, T];\, \Big[H^2 \cap H_0^1(\Omega)\Big]'\Big)$, esta solución pertenece a

$$u \in C^2\Big([0,\, T];\, \Big(H^2 \cap H_0^1(\Omega)\Big)'\Big).$$

Pero, hasta ahora, todos los resultados que hemos obtenido sobre la ecuación de ondas mediante técnicas de teoría de semigrupos, pueden también ser obtenidos mediante series de Fourier. Sin embargo, como habíamos mencionado anteriormente, la teoría de semigrupos es indispensable si deseamos abordar ecuaciones más generales con coeficientes variables dependientes de $(x,\, t)$, no

lineales, etc. Ilustramos este hecho analizando el ejemplo de una ecuación de ondas con un potencial $p = p(x,\, t) \in L^\infty(\Omega \times (0,\, T))$, i.e.

$$\begin{cases} u_{tt} - \Delta u + p(x,\, t)u = 0 & \text{en} \quad \Omega \times (0,\, T) \\ u = 0 & \text{en} \quad \partial\Omega \times (0,\, T) \\ u(x,\, 0) = u_0(x),\; u_t(x,\, 0) = u_1(x) & \text{en} \quad \Omega. \end{cases} \tag{11.4.35}$$

Nuevamente la ecuación (11.4.35) puede ser escrita en la forma de un sistema

$$\begin{cases} u_t = v \\ v_t = \Delta u - p(x,\, t)u, \end{cases} \tag{11.4.36}$$

o, en su versión abstracta,

$$U_t + AU + B(t)U = 0 \tag{11.4.37}$$

donde A es el operador maximal-disipativo asociado a la ecuación de ondas $B(t) : H \to H$ es un operador lineal acotado dependiente del tiempo:

$$B(t) = B(t) \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} 0 \\ -p(x,\, t)u \end{pmatrix} \tag{11.4.38}$$

la ecuación abstracta (11.4.37) puede escribirse como una ecuación integral

$$U(t) = e^{At}U_0 + \int_0^t e^{A(t-s)}B(s)U(s)ds. \tag{11.4.39}$$

Introduciendo la aplicación

$$[\phi(U)](t) = e^{At}U_0 + \int_0^t e^{A(t-s)}B(s)U(s)ds,\; 0 \leqslant t \leqslant T \tag{11.4.40}$$

la ecuación integral (11.4.39) puede también ser reescrita como un problema de punto fijo

$$U(t) = [\phi(U)](t),\; 0 \leqslant t \leqslant T \tag{11.4.41}$$

que puede ser resuelto mediante la aplicación del Teorema de punto fijo de Banach para aplicaciones contractivas.

En efecto, si utilizamos que $B(t)$ es un operador lineal y acotado de H en H, con una cota independiente de $0 \leqslant t \leqslant T$, es fácil comprobar que la aplicación (11.4.40) constituye una contradicción estricta en $C([0,\, \tau);\, H)$ para un τ suficientemente pequeño ($0 \leqslant \tau \leqslant T$). De este modo obtenemos una única solución $U \in C([0,\, \tau];\, H)$ que, mediante un argumento de continuación puede ser extendido a una solución global única $U \in C([0,\, T];\, H)$[7].

[7]Esto es así puesto que la amplitud $\tau > 0$ del intervalo temporal en el que podemos aplicar el Teorema de punto fijo a Φ para deducir la existencia local de soluciones, depende exclusivamente de la cota de la que dispongamos sobre la norma del operador B.

Aplicando este resultado abstracto en el caso de la ecuación de ondas (11.4.35) con potencial obtenemos inmediatamente que: Si $(u_0,\, u_1) \in H_0^1(\Omega) \times L^2(\Omega)$, y $p \in L^\infty(\Omega \times (0,\, T))$, la ecuación de ondas con potencial (11.4.35) admite una única solución

$$u \in C\Big([0,\, T];\, H_0^1(\Omega)\Big) \cap C^1\Big([0,\, T];\, L^2(\Omega)\Big).$$

En realidad la estructura (11.4.38) del operador permite debilitar la hipótesis sobre el potencial p para que el resultado anterior sea válido. En efecto, en la práctica es suficiente que el operador de multiplicación $u \to p(t)u$ envie de manera acotada $H_0^1(\Omega)$ en $L^2(\Omega)$. Si utilizamos las inclusiones de Sobolev es fácil comprobar que esto es así cuando:

- Si $n = 1$, $p \in L^\infty(0,\, T;\, L^2(\Omega))$;
- Si $n = 2$, $p \in L^\infty(0,\, T;\, L^r(\Omega))$, para algún $r > 2$;
- Si $n \geqslant 3$, $p \in L^\infty(0,\, T;\, L^n(\Omega))$.

Más aún, basta analizar con un poco más de cuidado la prueba del carácter contractivo de la aplicación Φ para observar que las hipótesis L^∞ en la variable t pueden ser debilitadas y sustituidas por hipótesis L^1. Así, el resultado anterior de existencia y unicidad de soluciones débiles para la ecuación de ondas con potencial (11.4.35) es cierto en cuanto el potencial p satisface las condiciones:

- $p \in L^1(0,\, T;\, L^2(\Omega))$, si $n = 1$.
- $p \in L^1(0,\, T;\, L^r(\Omega))$, con $r > 2$, si $n = 2$.
- $p \in L^1(0,\, T;\, L^n(\Omega))$, si $n \geqslant 3$.

Los mismos argumentos permiten obtener soluciones fuertes. Pero en este caso habremos de comprobar si el operador abstracto $B(t)$ envia $D(A)$ en $D(A)$. En el marco de la ecuación de ondas con potencial esto supone imponer hipótesis sobre el potencial $p = p(x,\, t)$ de modo que, para cada t, el operador de multiplicación mediante $p(t)$ envie $H^2 \cap H_0^1(\Omega)$ en $H_0^1(\Omega)$ y que haga ésto de modo que la cota resultante pertenezca a $L^1(0,\, T)$. Esto, evidentemente, exige hipótesis adicionales sobre la regularidad del potencial p.

Estos argumentos permiten en realidad obtener resultados de existencia y unicidad tanto de soluciones fuertes como débiles para ecuaciones más generales con potenciales de la forma

$$u_{tt} - \Delta u + a(x,\, t) \cdot \nabla u + b(x,\, t)u_t + p(x,\, t)u = 0. \tag{11.4.42}$$

11.5. La ecuación de ondas semilineal

En estas notas no hemos reproducido más que los elementos más elementales de la Teoría de Semigrupos. Esta, inspirándose en la teoría clásica de EDO y usando en particular la fórmula de variación de las constantes puede adaptarse fácilmente para abordar ecuaciones de evolución semilineales. Existe también una versión no-lineal de la Teoría de Semigrupos que permite abordar problemas genuinamente no lineales como la ecuación de medios porosos, por ejemplo, pero este tema queda fuera del ámbito de estas notas.

Consideremos ahora brevemente una ecuación de ondas semilineal

$$\begin{cases} u_{tt} - \Delta u = f(u) & \text{en} \quad \Omega \times (0, \infty) \\ u = 0 & \text{en} \quad \partial\Omega \times (0, \infty) \\ u(x, 0) = u_0(x),\ u_t(x, 0) = u_1(x) & \text{en} \quad \Omega. \end{cases} \tag{11.5.1}$$

En esta ocasión $f : \mathbb{R} \to \mathbb{R}$ es una función no lineal. Nuevamente, la ecuación (11.5.1) puede ser reescrita en la forma de un sistema

$$\begin{cases} u_t = v \\ v_t = \Delta u + f(u) \end{cases} \tag{11.5.2}$$

que, a su vez, puede ser enmarcado en un sistema semilineal abstracto

$$U_t = AU + F(U) \tag{11.5.3}$$

donde

$$F(U) = F \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} 0 \\ f(u) \end{pmatrix}. \tag{11.5.4}$$

El problema puede entonces ser reducido a la ecuación integral

$$U(t) = e^{At}U_0 + \int_0^t e^{A(t-s)}F(U(s))ds \tag{11.5.5}$$

que, a su vez, es equivalente al problema de punto fijo

$$U(t) = [\phi(U)](t), \tag{11.5.6}$$

para la función

$$[\phi(U)](t) = e^{At}U_0 + \int_0^t e^{A(t-s)}F(U(s))ds. \tag{11.5.7}$$

Sea $R = \| U_0 \|_H$ y B_{2R} la bola de radio $2R$ en H. Supongamos que la no-linealidad F envía H en H de modo que se trate de una función Lipschitziana

sobre conjuntos acotados de H, es decir

$$\begin{aligned} \forall k > 0,\, \exists L_k > 0 :\| F(U_1) - F(U_2) \|_H \quad &\leqslant L_k \| U_1 - U_2 \|_H \\ &\forall U_1,\, U_2 \in H \,:\| U_1 \|_H,\, \| U_2 \|_H \leqslant k. \end{aligned} \tag{11.5.8}$$

Bajo estas hipótesis es fácil comprobar que si $\tau > 0$ es suficientemente pequeño, Φ es una contracción estricta en $C([0,\, \tau];\, B_{2R})$. Esto permite deducir la existencia y unicidad de una solución local (en tiempo) de (11.5.5) en $C([0,\, \tau];\, B_{2R})$.

Veamos lo que la hipótesis (11.5.8) supone sobre la no-linealidad de la ecuación de ondas (11.5.1). En vista de la forma particular (11.5.4) de la no-linealidad del modelo abstracto correspondiente basta en realidad con comprobar que f envía $H_0^1(\Omega)$ en $L^2(\Omega)$ de manera Lipschitz sobre conjuntos acotados. Supongamos que la función f se comporta esencialmente como una potencia $p \geqslant 1$. Es decir supongamos que

$$\Big| f(x) - f(y) \Big| \leqslant C(1+ \mid x \mid^{p-1} + \mid y \mid^{p-1}) \mid x - y \mid, \forall x,\, y \in \mathbb{R} \tag{11.5.9}$$

para algún $p \geqslant 1$ y $C > 0$[8].

Necesitamos comprobar si para todo $k > 0$ existe $L_k > 0$ tal que

$$\begin{aligned} \| f(u_1) - f(u_2) \|_{L^2(\Omega)} \leqslant L_k \| u_1 - u_2 \|_{H_0^1(\Omega)}, \quad &\forall u_1,\, u_2 \in H_0^1(\Omega) \,: \\ &\| u_1 \|_{H_0^1(\Omega)},\, \| u_2 \|_{H_0^1(\Omega)} \leqslant k. \end{aligned} \tag{11.5.10}$$

En vista de la hipótesis (11.5.9) y usando las inclusiones de Sobolev es fácil comprobar que (11.5.10) se cumple bajo las siguientes restricciones sobre p:

$$\begin{cases} \bullet \quad \text{Para todo} \quad 1 \leqslant p < \infty, \qquad \text{si} \quad n = 1,\, 2. \\ \bullet \quad \text{Para todo} \quad 1 \leqslant p \leqslant \frac{n}{n-2}, \quad \text{si} \quad n \geqslant 3. \end{cases} \tag{11.5.11}$$

Deducimos por tanto que: "Bajo estas condiciones sobre el exponente p, si la no-linealidad f satisface la condición de Lipschitz (11.5.9), para cada par de datos iniciales $(u_0,\, u_1) \in H_0^1(\Omega) \times L^2(\Omega)$ existe un $\tau > 0$ y una única solución $u \in C([0,\, \tau];\, H_0^1(\Omega)) \cap C^1([0,\, \tau];\, L^2(\Omega))$".

Una vez que la solución local en tiempo ha sido obtenida, mediante los mismos argumentos de prolongación que se utilizan en el marco de las Ecuaciones Diferenciales Ordinarias (EDO), esta solución local puede ser prolongada al máximo intervalo de existencia $T_{\max}$ de modo que la solución única de (11.5.1) se obtiene finalmente en la clase

$$C([0,\, T_{\max});\, H_0^1(\Omega)) \cap C^1([0,\, T_{\max});\, L^2(\Omega)).$$

[8] Esta hipótesis se cumple, por ejemplo, si $f \in C^1(\mathbb{R};\, \mathbb{R}$ y $\limsup_{|x| \to \infty} \frac{\mid f'(x) \mid}{\mid x \mid^{p-1}} < \infty$.

Además, para el tiempo máximo de existencia se verifica la siguiente alternativa: O bien $T_{\max} = \infty$ (*existencia global*) y por lo tanto la solución está definida para todo tiempo, o bien $T_{\max} < \infty$ (*explosión en tiempo finito*) y en este caso

$$\lim_{t \nearrow T_{\max}} \| u(t) \|_{H_0^1(\Omega)} + \| u_t(t) \|_{L^2(\Omega)} = \infty. \tag{11.5.12}$$

El fenómeno que subyace a esta alternativa es fácil de entender. Mientras que la solución se mantiene acotada puede ser prolongada en el tiempo, con un paso temporal que depende continuamente de la cota de la solución. Por lo tanto, la única manera en que la solución pueda no ser prolongada a todos los tiempos es si explota en tiempo finito.

Mediante una mera hipótesis de crecimiento del tipo (11.5.9) sobre la no-linealidad es imposible determinar si se produce explosión en tiempo finito o no y para ésto son necesarias hipótesis adicionales sobre el "signo" de la no-linealidad.

Consideremos en primer lugar el caso en que la no-linealidad f tiene el "buen signo":

$$\begin{cases} u_{tt} - \Delta u + \mid u \mid^{p-1} u = 0 & \text{en} \quad \Omega \times (0, \infty) \\ u = 0 & \text{en} \quad \partial\Omega \times (0, \infty) \\ u(x, 0) = u_0(x),\ u_t(x, 0) = u_1(x) & \text{en} \quad \Omega. \end{cases} \tag{11.5.13}$$

En este caso, evidentemente, la condición (11.5.9) se cumple y bajo la hipótesis (11.5.11) se deduce la existencia y unicidad local (en tiempo) de soluciones de energía finita de (11.5.13). Además, mientras la solución existe, su energía se conserva. En este caso la energía viene dada por

$$E(t) = \frac{1}{2} \int_\Omega \Big[\mid \nabla u(x, t) \mid^2 + \mid u_t(x, t) \mid^2 \Big] dx + \frac{1}{p+1} \int_\Omega \mid u(x, t) \mid^{p+1} dx. \tag{11.5.14}$$

Como la energía $E(t)$ se conserva y claramente mayora al cuadrado de la norma de (u, u_t) en $H_0^1(\Omega) \times L^2(\Omega)$ deducimos inmediatamente que (11.5.12) es imposible. De este modo concluimos que, bajo la condición (11.5.11), para cada par de datos iniciales $(u_0, u_1) \in H_0^1(\Omega) \times L^2(\Omega)$ existe una única solución global

$$u \in C([0, \infty); H_0^1(\Omega)) \cap C^1([0, \infty); L^2(\Omega)) \tag{11.5.15}$$

y que la energía $E(t)$ de la solución definida en (11.5.14) se conserva para todo $t \geqslant 0$.

La situación cambia completamente para no-linealidades con "mal-signo":

$$\begin{cases} u_{tt} - \Delta u = \mid u \mid^{p-1} u & \text{en} \quad \Omega \times (0, \infty) \\ u = 0 & \text{en} \quad \partial\Omega \times (0, \infty) \\ u(x, 00) = u_0(x),\ u_t(x, 0) = u_1(x) & \text{en} \quad \Omega. \end{cases} \tag{11.5.16}$$

La existencia y unicidad de soluciones locales (en tiempo) es igualmente cierta en este caso. Pero no se puede decir lo mismo acerca de la existencia global. Para el sistema (11.5.16), la energía, que se observa mientras las soluciones existen, es

$$E(t) = \frac{1}{2}\int_\Omega [\mid \nabla u(x,\, t) \mid^2 + \mid u_t(x,\, t) \mid^2] dx - \frac{1}{p+1}\int_\Omega \mid u(x,\, t) \mid^{p+1} dx \tag{11.5.17}$$

pero, el que esta energía permanezca constante o acotada es perfectamente compatible con la explosión (11.5.12) de las soluciones en tiempo finito. De hecho, en este caso, las soluciones pueden efectivamente explotar en tiempo finito. Para convencerse de este hecho basta ver que existen soluciones de la EDO

$$x'' = \mid x \mid^{p-1} x \tag{11.5.18}$$

que, cuando $p > 1$, explotan en tiempo finito, en un tiempo que tiende a cero cuando el tamaño de los datos iniciales tiende a infinito. El hecho de que las soluciones de la ecuación de ondas dependan exclusivamente de los datos iniciales en la base del cono característico permite entonces construir datos iniciales, independientes de x en una bola de Ω, y de modo que en el interior del cono correspondiente coinciden con la solución de la ODE (11.5.18) y por tanto explotan en tiempo finito.

Esta construcción permite efectivamente probar que, para todo $p > 1$ y todo abierto no vacío Ω de $\mathbb{R}^n$, existen datos iniciales $(u_0,\, u_1) \in H_0^1(\Omega) \times L^2(\Omega)$ para los que la solución local de (11.5.16) explota en tiempo finito.

11.6. El problema elíptico

Hemos aplicado métodos de "splitting" o de descomposición para ecuaciones discretas en tiempo, inspirándonos en la teoría clásica de aproximación numérica de Ecuaciones Diferenciales Ordinarias.

Son dos las razones principales para considerar esquemas discretos en tiempo. Por una parte, la resolución numérica efectiva siempre pasa por una discretización temporal. Es pues natural considerar esquemas discretos en tiempo. Pero además, en muchas ocasiones la utilización de discretizaciones temporales es también un método para obtener soluciones para el modelo continuo.

En el marco de las EDP lineales son muchas las técnicas que se pueden utilizar para resolverlas: soluciones fundamentales, análisis de Fourier, separación de variables y métodos espectrales, semigrupos,.... Sin embargo, las ecuaciones no-lineales son mucho más complejas y se dispone de menos métodos sistemáticos

para resolverlas. En esta sección ilustraremos la posible utilización de métodos de discretización temporal en el caso modelo de una ecuación parabólica no-lineal que involucra al operador p-Laplaciano:

$$\begin{cases} u_t - \operatorname{div}\left(|\ \nabla u\ |^{p-2}\ \nabla u\right) = 0 & \text{en} \quad \Omega \times (0,\ \infty) \\ u = 0 & \text{en} \quad \partial\Omega \times (0,\ \infty), \\ u(x,\ 0) = u_0(x) & \text{en} \quad \Omega. \end{cases} \tag{11.6.1}$$

Supondremos que Ω es un abierto acotado y regular de $\mathbb{R}^n$ y que $p \geq 2$, si bien el modelo tiene también sentido para $p \geq 1$, siendo $p = 1$ un caso límite. Pero, para evitar algunas dificultades técnicas adicionales, supondremos que $p \geq 2$.

Conviene observar que cuando $p = 2$, (11.6.1) se reduce a la ecuación del calor lineal para la que, como decíamos, disponemos de diversos métodos. El caso no-lineal, como veremos, es bastante más complejo.

Desde un punto de vista numérico, el modo más natural para abordar el problema es o bien mediante un método de Galerkin o a través de una discretización temporal. En esta sección analizaremos ambos métodos pero antes estudiaremos brevemente el problema elíptico subyacente.

Dado $f = f(x)$ consideramos el problema elíptico:

$$\begin{cases} -\operatorname{div}\left(|\ \nabla u\ |^{p-2}\ \nabla u\right) = f & \text{en} \quad \Omega \\ u = 0 & \text{en} \quad \partial\Omega. \end{cases} \tag{11.6.2}$$

Su formulación variacional es

$$\begin{cases} u \in W_0^{1,\,p}(\Omega), \\ \displaystyle\int_\Omega |\ \nabla u\ |^{p-2}\ \nabla u \cdot \nabla\varphi dx = \int_\Omega f\varphi dx,\ \forall \varphi \in W_0^{1,\,p}(\Omega). \end{cases} \tag{11.6.3}$$

Para que esta formulación variacional tenga sentido basta que f pertenezca al dual de $W_0^{1,\,p}(\Omega)$. Para ello es suficiente que $f \in L^q(\Omega)$ con $q > np\Big/(np+p-1)$.

Para obtener una solución débil consideramos el funcional

$$J :\ W_0^{1,\,p}(\Omega) \longrightarrow \mathbb{R}, \tag{11.6.4}$$

tal que

$$J(v) = \frac{1}{p}\int_\Omega |\ \nabla v\ |^p\ dx - \int_\Omega fv dx. \tag{11.6.5}$$

El espacio $W_0^{1,\,p}(\Omega)$ es reflexivo. Por otra parte, J es continuo, convexo y coercivo. Por tanto, el funcional J alcanza su mínimo en un punto $u \in W_0^{1,\,p}(\Omega)$. Es fácil comprobar que u es una solución de (11.6.2) en el sentido de (11.6.3).

Por otra parte la solución débil es única, puesto que J es estrictamente convexo.

Otra manera de comprobar la unicidad es, como es habitual, suponer que hay dos soluciones u_1 y u_2 y definir $v = u_1 - u_2$. Entonces, v satisface

$$\begin{cases} -\operatorname{div}\left(|\nabla u_1|^{p-2}\nabla u_1 - |\nabla u_2|^{p-2}\nabla u_2\right) = 0 & \text{en} \quad \Omega \\ u_1 - u_2 = 0 & \text{en} \quad \partial\Omega, \end{cases} \tag{11.6.6}$$

o, lo que es lo mismo, su versión variacional

$$\int_\Omega \left[|\nabla u_1|^{p-2}\nabla u_1 - |\nabla u_2|^{p-2}\nabla u_2\right]\cdot\nabla\varphi = 0 \quad \forall\varphi \in W_0^{1,p}(\Omega). \tag{11.6.7}$$

Tomando como función test $\varphi = u_1 - u_2 = v$ obtenemos

$$\int_\Omega \left[|\nabla u_1|^{p-2}\nabla u_1 - |\nabla u_2|^{p-2}\nabla u_2\right]\cdot\nabla\left(u_1 - u_2\right)dx = 0. \tag{11.6.8}$$

Por otra parte, se tiene la desigualdad en $\mathbb{R}^n$:

$$\left(|a|^{p-2}a - |b|^{p-2}b\right)\cdot(a-b) \geq 0,\ \forall a,\ b \in \mathbb{R}^n, \tag{11.6.9}$$

de donde deducimos que, necesariamente,

$$\left(|\nabla u_1|^{p-2}\nabla u_1 - |\nabla u_2|^{p-2}\nabla u_2\right)\cdot(\nabla u_1 - \nabla u_2) = 0 \,\text{p.c.t. } x \in \Omega. \tag{11.6.10}$$

Pero en realidad podemos decir más aún y garantizar que

$$\left(|a|^{p-2}a - |b|^{p-2}b\right)\cdot(a-b) > 0 \text{ si } a \neq b \tag{11.6.11}$$

lo que asegura, en virtud de (11.6.8), que

$$\nabla u_1 = \nabla u_2 \,\text{p.c.t. } x \in \Omega. \tag{11.6.12}$$

Teniendo en cuenta que $u_1,\, u_2 \in W_0^{1,p}(\Omega)$, deducimos entonces que $u_1 = u_2$ p.c.t. $x \in \Omega$.

11.7. El método Galerkin

El método de Galerkin para aproximar la ecuación (11.6.1) se introduce del mismo modo que en el caso lineal. Introducimos una aproximación de $W_0^{1,p}(\Omega)$ mediante subespacios de dimensión finita V_h de elementos finitos P_1 a trozos y continuos. El lector interesado en los aspectos básicos del método de elementos finitos podrá consultar [33] o [34].

Suponemos que

$$\dim(V_h) = N_h,\ V_h = \operatorname{span}[e_1, \cdots, e_{N_h}]. \tag{11.7.1}$$

Buscamos entonces aproximaciones de la solución de (11.6.1) tales que

$$u_h \in C([0,\infty);\, V_h) \tag{11.7.2}$$

de modo que

$$u_h(x,\, t) = \sum_{j=1}^{N_h} u_j(t) e_j(x). \tag{11.7.3}$$

Con el objeto de introducir esta aproximación conviene previamente introducir la formulación variacional de (11.6.1):

$$\begin{cases} u \in C\left([0,\infty);\, L^2(\Omega)\right) \cap L^p\left(0,\infty;\, W_0^{1,\,p}(\Omega)\right) \\ \dfrac{d}{dt}\displaystyle\int_\Omega u(x,\, t)\varphi(x)dx + \int_\Omega |\ \nabla u\ |^{p-2}\ \nabla u \cdot \nabla\varphi dx = 0,\ \forall \varphi \in W_0^{1,\,p}(\Omega), \\ \displaystyle\int_\Omega u(x,\, t)\varphi(x)dx \to \int_\Omega u_0(x)\varphi(x)dx,\ t \to 0,\ \forall \varphi \in W_0^{1,\,p}(\Omega). \end{cases} \tag{11.7.4}$$

El espacio en el que buscamos la solución está inspirado en la estimación de energía para las soluciones de (11.6.1) que, formalmente, consiste en multiplicar la ecuación (11.6.1) por u, e integrar por partes en Ω. Se obtiene así:

$$\frac{1}{2}\frac{d}{dt}\int_\Omega u^2(x,\, t)dx = -\int_\Omega |\ \nabla u\ |^p\ dx. \tag{11.7.5}$$

Integrando en tiempo obtenemos

$$\int_\Omega u^2(x,\, t)dx + \int_0^t \int_\Omega |\ \nabla u\ |^p\ dxdt = \int_\Omega u_0^2(x)dx. \tag{11.7.6}$$

De esta estimación formal deducimos que, si el dato inicial $u_0 \in L^2(\Omega)$, cabe esperar que la solución verifique que $u \in L^\infty\left(0,\infty;\, L^2(\Omega)\right)$ y además $u \in L^p\left(0,\infty;\, W_0^{1,\,p}(\Omega)\right)$.

A partir de (11.7.4) es fácil introducir una aproximación de Galerkin:

$$\begin{cases} u_h \in C([0,\infty);\, V_h) \\ \dfrac{d}{dt}\displaystyle\int_\Omega u_h\varphi dx + \int_\Omega |\ \nabla u_h\ |^{p-2}\ \nabla u_h \cdot \nabla\varphi dx = 0,\ \forall \varphi \in V_h \\ u_h(0) = u_{0,\,h}, \end{cases} \tag{11.7.7}$$

donde $u_{0,\,h}$ es una aproximación de u_0 en V_h que verifica

$$u_{0,\,h} \to u_0 \text{ en } L^2(\Omega). \tag{11.7.8}$$

Como es habitual en el marco de las aproximaciones de Galerkin, tenemos que resolver dos cuestiones. En primer lugar tenemos que probar que (11.7.7) admite una única solución u_h y en segundo que converge a una solución de (11.6.1).

Con el objeto de verificar si (11.7.7) admite una solución, escribimos esta formulación variacional como un sistema de N_h ecuaciones diferenciales ordinarias no lineales con N_h incógnitas. En virtud de la estructura (11.7.3) y teniendo en cuenta que (11.7.7) se satsiface para toda función test $\varphi \in V_h$ sí y sólo sí se cumple para toda función e_j, $j = 1, \cdots, N_h$ de la base de V_h, deducimos que (11.7.7) equivale a

$$\begin{cases} MU' + F(U) = 0, \quad t > 0 \\ U(0) = U_{0,\,h}, \end{cases} \tag{11.7.9}$$

donde el vector columna $U = (u_1, \cdots, u_{N_h})^t$, codifica las N_h incógnitas del sistema, $U_{0,\,h}$ representa del mismo modo el dato inicial $u_{0,\,h} \in V_h$, M es la matriz de masa del método de elementos finitos $M = (m_{ij})_{1 \le i,\, j \le N_h}$ con

$$m_{ij} = \int_\Omega e_i(x) e_j(x) dx, \tag{11.7.10}$$

y F es una función no-lineal que está definida del siguiente modo:

$$F(U) = (F_j(U))_{1 \le j \le N_h}, \tag{11.7.11}$$

donde

$$F_j(U) = \int_\Omega \left| \nabla \left(\sum_{k+1}^{N_h} u_{j_k} e_k(x) \right) \right|^{p-2} \nabla \left(\sum_{k+1}^{N_h} u_k e_k(x) \right) \cdot \nabla e_j(x) dx. \tag{11.7.12}$$

En el caso en que $p = 2$, $F(U) = RU$, donde R es la matriz de rigidez del método de elementos finitos.

Cuando $p \ge 2$, la función F es no-lineal y de clase C^1 de modo que (11.7.9) admite una única solución local. Con el objeto de probar que la solución es global precisamos de una estimación a priori. Aplicando de nuevo el método de energía que consiste en tomar en (11.7.7) la propia solución u_h como función test o multiplicar en (11.7.9) escalarmente con la propia incógnita U, obtenemos que (11.7.6) se satisface también para cada aproximación u_h. Deducimos pues por tanto que la solución de (11.7.7) está globalmente definida.

Esto proporciona también una cota uniforme sobre las soluciones aproximadas:

$$\frac{1}{2} \parallel u_h \parallel^2_{L^\infty(0,\,\infty;\,L^2(\Omega))} + \parallel \nabla u_h \parallel^p_{L^p(\Omega\times(0,\,\infty))} \leq \frac{1}{2} \parallel u_{0,\,h} \parallel^2_{L^2(\Omega)},\ \forall h > 0. \tag{11.7.13}$$

De esta estimación podemos deducir que, extrayendo subsucesiones,

$$\begin{cases} u_h \rightharpoonup u & \text{débil -}* \text{ en} \quad L^\infty\left(0,\,\infty;\,L^2(\Omega)\right) \\ u_h \rightharpoonup u & \text{débil en} \quad\ L^p\left(0,\,\infty;\,W_0^{1,\,p}(\Omega)\right). \end{cases} \tag{11.7.14}$$

Pero estas convergencias débiles no son suficientes para pasar al límite en (11.7.7) puesto que se trata de un problema no-lineal.

El paso al límite necesita de estimaciones adicionales. Con el objeto de obtenerlas conviene volver por un momento a la ecuación continua y comprobar qué otras estimaciones se cumplen. En primer lugar observamos que si u_1 y u_2 son soluciones de (11.6.1) multiplicando por $u_1 - u_2$ la ecuación que $u_1 - u_2$ satisface se obtiene

$$\frac{1}{2}\frac{d}{dt}\int_\Omega |u_1 - u_2|^2\,dx + \int_\Omega \left(|\nabla u_1|^{p-2}\,\nabla u_1 - |\nabla u_2|^{p-2}\,\nabla u_2\right)\cdot(\nabla u_1 - \nabla u_2)\,dx = 0. \tag{11.7.15}$$

Utilizamos ahora la existencia de $c > 0$ tal que[9]

$$\left(\mid a \mid^{p-2} a - \mid b \mid^{p-2} b\right)\cdot(a-b) \geq c \mid a - b \mid^p,\ \forall a,\, b \in \mathbb{R}^n. \tag{11.7.16}$$

Obtenemos así que

$$\frac{1}{2}\frac{d}{dt}\int_\Omega |u_1 - u_2|^2\,dx + c\int_\Omega |\nabla\,(u_1 - u_2)|^p\,dx \leq 0. \tag{11.7.17}$$

Deducimos entonces que

$$\frac{1}{2}\int_\Omega |u_1(x,\,t) - u_2(x,\,t)|^2\,dx + c\int_{\Omega\times(0,\,t)} |\nabla u_1 - \nabla u_2|^p\,dxds \leq \frac{1}{2}\int_\Omega |u_{1,\,0} - u_{2,\,0}|^2\,dx. \tag{11.7.18}$$

De esta desigualdad deducimos que si (u_j) es una sucesión de soluciones de (11.6.1) tales que sus datos iniciales $(u_{j,\,0})$ constituyen una sucesión de Cauchy en $L^2(\Omega)$ entonces u_j es también una sucesión de Cauchy en $L^\infty\left(0,\,\infty;\,L^2(\Omega)\right) \cap L^p\left(0,\,\infty;\,W_0^{1,\,p}(\Omega)\right)$.

Este argumento, aplicado en el contexto de las aproximaciones de Galerkin, permite probar que ∇u_h es una sucesión de Cauchy en $L^p(\Omega\times(0,\,\infty))$ lo cual, en virtud de (11.7.14), proporciona la convergencia fuerte

$$\nabla u_h \longrightarrow \nabla u \text{ en } L^p(\Omega\times(0,\,\infty)), \tag{11.7.19}$$

[9] La demostración de esta desigualdad se deja como ejercicio al lector.

que se precise para pasar al límite en la aproximación de Galerkin (11.7.7) y obtener en el límite la formulación variacional (11.7.4) de la ecuación (11.6.1).

Esto es rigurosamente cierto cuando los espacios de Galerkin V_h crecen a medida que h decrece. Esto no es así para una triangulación arbitraria en el contexto de los elementos finitos, pero ocurre efectivamente cuando, a medida que h decrece, el nuevo espacio V_h se obtiene como el correspondiente a un mallado proveniente del anterior por refinamiento.

Hemos comprobado por tanto que el método de Galerkin proporciona una manera de construir soluciones de (11.6.1).

11.8. Discretización temporal

Dado un paso temporal $\Delta t > 0$, destinado a tender a cero, nos proponemos obtener las soluciones de (11.6.1) como límite cuando $\Delta t \longrightarrow 0$ de soluciones de sistemas discretos en tiempo.

En lo sucesivo, con el objeto de simplificar la notación, denotamos mediante A_p el operador p-Laplaciano, de modo que

$$A_p(u) = -\operatorname{div}\left(|\nabla u|^{p-2}\nabla u\right). \tag{11.8.1}$$

A la hora de discretizar (11.6.1) tenemos dos opciones básicas:

• El esquema de Euler explícito:

$$\begin{cases} u^{k+1} + \Delta t A_p(u^k) = u^k, & \text{en} & \Omega, \\ u^{k+1} = 0 & \text{sobre} & \partial\Omega,\ k \geq 0, \\ u^0 = u_0 & \text{en} & \Omega. \end{cases} \tag{11.8.2}$$

• El esquema de Euler implícito:

$$\begin{cases} u^{k+1} + \Delta t A_p\left(u^{k+1}\right) = u^k, & \text{en} & \Omega, \\ u^{k+1} = 0 & \text{sobre} & \partial\Omega,\ k \geq 0 \\ u^0 = u_0, & \text{en} & \Omega. \end{cases} \tag{11.8.3}$$

En ambos casos u^k representa una aproximación de la solución de (11.6.1) en el instante $t = k\Delta t$. También en ambos esquemas el dato inicial u_0 del sistema (11.6.1) se toma como valor inicial de la iteración para $k = 0$.

La ventaja principal del esquema explícito (11.8.2) frente al implícito (11.8.3) es que los sucesivos valores de u^k se obtienen sin necesidad de resolver ninguna ecuación, "leyendo" su valor a partir de los del paso anterior. Así, de (11.8.2) tenemos

$$u^{k+1} = u^k - \Delta t A_p\left(u^k\right) = B_{\Delta_t}\left(u^k\right). \tag{11.8.4}$$

Iterando en esta aplicación no-lineal B_{Δ_t}, podemos escribir el valor de la solución discreta a partir del dato inicial de manera explícita:

$$u^k = \left(B_{\Delta_t}\right)^k (u_0)\,. \tag{11.8.5}$$

Conviene sin embargo señalar que la expresión (11.8.5), además de ser fuertemente no-lineal, involucra derivadas del dato inicial u_0 del orden de $2k$. Teniendo en cuenta que el número de pasos k necesarios para aproximar el valor de la solución en un instante fijo T tiende a infinito cuando $\Delta t \longrightarrow 0$, este método sólo puede ser útil cuando el dato inicial u_0 es de clase C^∞. Pero incluso en este caso la convergencia del método no está garantizada.

En este contexto es mucho más natural utilizar el método implícito que, como sabemos, en el marco de las EDOs es incondicionalmente estable y convergente. Ahora bien, la aplicación del esquema implícito (11.8.3) exige, en cada paso, resolver el problema elíptico no-lineal:

$$\begin{cases} u^{k+1} + \Delta t A_p\left(u^{k+1}\right) = u^k, & \text{en} \quad \Omega, \\ u^{k+1} = 0 & \text{en} \quad \partial\Omega. \end{cases} \tag{11.8.6}$$

La solución de este problema puede obtenerse por técnicas análogas a las empleadas en la sección 11.6. Más concretamente, para probar la existencia de la solución de (11.8.6) basta minimizar el funcional

$$J_k(v) = \frac{\Delta t}{p}\int_\Omega \mid \nabla v \mid^p dx + \frac{1}{2}\int_\Omega v^2 dx - \int_\Omega u^k v dx, \tag{11.8.7}$$

en el espacio $W_0^{1,\,p}(\Omega)$. La unicidad de la solución se demuestra por técnicas análogas a las empleadas en la sección 11.6.

Además, el método de energía proporciona estimaciones inmediatas. En efecto, multiplicando en (11.8.6) por u^{k+1} e integrando en Ω obtenemos

$$\begin{aligned} \int_\Omega \left|u^{k+1}\right|^2 dx + \Delta t \int_\Omega \left|\nabla u^{k+1}\right|^p dx &= \int_\Omega u^{k+1}u^k dx \\ &\leq \left(\int_\Omega \left|u^{k+1}\right|^2\right)^{1/2}\left(\int_\Omega \left|u^k\right|^2 dx\right)^{1/2}, \end{aligned} \tag{11.8.8}$$

de donde se deduce que, en particular,

$$\left|\left|u^{k+1}\right|\right|_{L^2(\Omega)} \leq \left|\left|u^k\right|\right|_{L^2(\Omega)}. \tag{11.8.9}$$

Iterando esta desigualdad llegamos con facilidad a la cota

$$\max_k \left|\left|u^k\right|\right|_{L^2(\Omega)} \leq ||u_0||_{L^2(\Omega)}\,, \tag{11.8.10}$$

independiente de Δt.

Pero de (11.8.8) se deduce también que

$$\begin{aligned}\Delta t \sum_k \int_\Omega \left|\nabla u^{k+1}\right|^p dx &\leq \sum_k \left|\left|u^{k+1}\right|\right|_{L^2(\Omega)} \left(\left|\left|u^k\right|\right|_{L^2(\Omega)} - \left|\left|u^{k+1}\right|\right|_{L^2(\Omega)}\right) \\ &\leq ||u_0||_{L^2(\Omega)} \sum_k \left(\left|\left|u^k\right|\right| - \left|\left|u^{k+1}\right|\right|_{L^2(\Omega)}\right) \\ &\leq ||u_0||^2_{L^2(\Omega)} \,. \end{aligned} \tag{11.8.11}$$

Las estimaciones (11.8.10) y (11.8.11) son obviamente los análogos discretos de la estimación de energía.

A partir de la solución discreta $\{u^k\}_{k\geq 0}$ podemos construir una función continua $u_{\Delta t}(x,\, t)$ que en cada intervalo temporal de la forma $[k\Delta t,\, (k+1)\Delta t]$ tome el valor de $u^k(x)$. Obtenemos así una familia $\{u_{\Delta t}\}_{\Delta t>0}$ de funciones continuas, acotadas en el espacio $L^\infty(0,\, \infty;\, L^2(\Omega)) \cap L^p\left(0,\, \infty;\, W_0^{1,\,p}(\Omega)\right)$. Extrayendo subsucesiones podemos pasar al límite débilmente en la sucesión $u_{\Delta t}$ cuando $\Delta t \longrightarrow 0$ y obtener una función límite $u(x,\, t)$. El problema que persiste es probar que este límite es solución de la ecuación no-lineal (11.6.1) o, si se prefiere, que verifica la formulación débil (11.7.4). Nuevamente, a causa del carácter no-lineal de los problemas considerados, esto no es del todo inmediato.

Con el objeto de poder garantizar que el límite u es solución de (11.6.1) o (11.7.4) necesitamos probar la convergencia fuerte de $\nabla u_{\Delta t}$ en $L^1_{loc}\left(0,\, \infty;\, L^{p-1}(\Omega)\right)$. Para ello, el ingrediente clave es la obtención de estimaciones de orden superior. Con el objeto de entencer cómo se pueden obtener dichas estimaciones volvemos al problema continuo. Multiplicando en (11.6.1) por u_t e integrando en Ω obtenemos

$$\int_\Omega u_t^2 dx + \int_\Omega \mid \nabla u \mid^{p-2} \nabla u \cdot \nabla u_t dx = 0.$$

Por otra parte

$$\int_\Omega \mid \nabla u \mid^{p-2} \nabla u \cdot \nabla u_t dx = \frac{1}{p}\frac{d}{dt}\int_\Omega \mid \nabla u \mid^2 dx.$$

Obtenemos por tanto la identidad

$$\frac{d}{dt}\left[\frac{1}{p}\int_\Omega \mid \nabla u \mid^p dx\right] = -\int_\Omega u_t^2 dx = -\int_\Omega \left|\mathrm{div}\left(\mid \nabla u \mid^{p-2} \nabla u\right)\right|^2 dx \tag{11.8.12}$$

que, integrada en t, proporciona

$$\frac{1}{p}\int_\Omega \mid \nabla u(x,\, t) \mid^p dx + \int_0^t \int_\Omega \left|\mathrm{div}\left(\mid \nabla u \mid^{p-2} \nabla u\right)\right|^2 dxds = \frac{1}{p}\int_\Omega |\nabla u_0(x)|^p dx, \tag{11.8.13}$$

siempre y cuando $u_0 \in W_0^{1,\,p}(\Omega)$. Deducimos así una estimación para la solución continua en $L^\infty\left(0,\,\infty;\,W_0^{1,\,p}(\Omega)\right)$, además de una cota de $\operatorname{div}\left(\mid \nabla u \mid^{p-2} \nabla u\right)$ en $L^2(\Omega \times (0,\,\infty))$.

Esta estimación puede reproducirse en el marco discreto. Multiplicando en (11.8.6) por $\left(u^{k+1} - u^k\right)/\Delta t$ deducimos que

$$\int_\Omega \left|\frac{u^{k+1} - u^k}{\Delta t}\right|^2 dx + \frac{1}{\Delta t}\int_\Omega \operatorname{div}\left(\mid \nabla u^{k+1} \mid^{p-2} \nabla u^{k+1}\right)\cdot\left(u^{k+1} - u^k\right) dx = 0,$$

que, tras la integración por partes, proporciona

$$\int_\Omega \left|\frac{u^{k+1} - u^k}{\Delta t}\right|^2 dx + \frac{1}{\Delta t}\int_\Omega \mid \nabla u^{k+1} \mid^p dx - \frac{1}{\Delta t}\int_\Omega \mid \nabla u^{k+1} \mid^{p-2} \nabla u^{k+1}\cdot\nabla u^k dx = 0. \tag{11.8.14}$$

De esta expresión es fácil deducir que

$$\int_\Omega \mid \nabla u^{k+1} \mid^p dx \leq \int_\Omega \mid \nabla u^k \mid^p dx, \tag{11.8.15}$$

que, iterándola, proporciona la cota

$$\parallel \nabla u^k \parallel_{L^p(\Omega)} \leq \parallel \nabla u_0 \parallel_{L^p(\Omega)},\ \forall k \geq 0, \tag{11.8.16}$$

independiente de Δt.

Una vez que ya disponemos de esta cota es fácil concluir que

$$\Delta t \sum_k \int_\Omega \left|\frac{u^{k+1} - u^k}{\Delta t}\right|^2 dx \leq C \parallel \nabla u_0 \parallel^p_{L^p(\Omega)}. \tag{11.8.17}$$

Obtenemos así un análogo discreto de la estimación de energía de segundo orden (11.8.13).

Esta estimación permite obtener la convergencia fuerte de $\nabla u_{\Delta t}$ necesaria para probar que el límite u es solución de la ecuación (11.6.1) si bien en esta sección omitiremos los detalles.

11.9. Ecuaciones parabólicas: Comportamiento asintótico

Consideramos ahora la ecuación del calor lineal

$$\begin{cases} u_t - \Delta u = f(x) & \text{en} \quad \Omega,\ t > 0 \\ u = 0 & \text{en} \quad \partial\Omega,\ t > 0 \\ u(x,\,0) = u_0(x) & \text{en} \quad \Omega, \end{cases} \tag{11.9.1}$$

en un dominio acotado y regular Ω de $\mathbb{R}^N$ y $f = f(x)$ es una fuente externa independiente de t.

Mediante los métodos habituales (Galerkin, semigrupos,...) es fácil comprobar que si $u_0 \in L^2(\Omega)$ y $f \in H^{-1}(\Omega)$, existe una única solución de (11.9.1) en la clase $u \in C\big([0,\,\infty);\, L^2(\Omega)\big) \cap L^2_{loc}\big(0,\,\infty;\, H^1_0(\Omega)\big)$.

Por otra parte, la identidad de energía proporciona

$$\frac{1}{2}\frac{d}{dt}\int_\Omega u^2 dx + \int_\Omega |\,\nabla u\,|^2\,dx = \int_\Omega f u dx \tag{11.9.2}$$

de donde se deduce que

$$\frac{1}{2}\frac{d}{dt}\int_\Omega u^2 dx + \int_\Omega |\,\nabla u\,|^2\,dx \le \frac{\varepsilon}{2}\int_\Omega |\,\nabla u\,|^2\,dx + \frac{1}{2\varepsilon}\,\|\,f\,\|^2_{H^{-1}(\Omega)},$$

para todo $\varepsilon > 0$. Tomando por ejemplo $\varepsilon = 1$ se deduce que

$$\frac{d}{dt}\int_\Omega u^2 + \int_\Omega |\,\nabla u\,|^2 \le \|\,f\,\|^2_{H^{-1}(\Omega)}\,. \tag{11.9.3}$$

Aplicando la desigualdad de Poincaré, que garantiza la existencia de $c(\Omega) > 0$ tal que

$$\int_\Omega |\,\nabla\varphi\,|^2\,dx \ge c(\Omega)\int_\Omega \varphi^2 dx,\ \forall \varphi \in H^1_0(\Omega), \tag{11.9.4}$$

deducimos la desigualdad de energía

$$\frac{d}{dt}\int_\Omega u^2 dx + c(\Omega)\int_\Omega u^2 dx \le \|\,f\,\|^2_{H^{-1}(\Omega)}, \tag{11.9.5}$$

cuya resolución explícita da lugar a la cota

$$\int_\Omega u^2(x,\,t)dx \le e^{-c(\Omega)t}\int_\Omega u_0^2(x)dx + \frac{(1-e^{-c(\Omega)t})}{c(\Omega)}\,\|\,f\,\|^2_{H^{-1}(\Omega)}\,. \tag{11.9.6}$$

Deducimos así que

$$u \in L^\infty(0,\,\infty;\,L^2(\Omega)), \tag{11.9.7}$$

o, lo que es lo mismo, que la trayectoria $\{u(t)\}_{t\ge 0}$ está acotada en $L^2(\Omega)$.

Una vez que sabemos que la trayectoria está acotada, cabe plantearse la cuestión de su comportamiento asintótico cuando $t \to \infty$. En caso de que el límite de $u(t)$ cuando $t \to \infty$ existe, es previsible que ésta sea una solución estacionaria, independiente de t, de la ecuación que satisfará entonces la ecuación elíptica

$$\begin{cases} -\Delta u = f & \text{en} \quad \Omega \\ u\Big|_{\partial\Omega} = 0. \end{cases} \tag{11.9.8}$$

La teoría variacional clásica garantiza la existencia y unicidad de la solución $u^* \in H_0^1(\Omega)$ de (11.9.8).

Veamos ahora que

$$u(t) \longrightarrow u^*, \text{ exponencialmente en } L^2(\Omega),\ t \longrightarrow \infty. \tag{11.9.9}$$

En efecto, definimos

$$v(t) = u(t) - u^* \tag{11.9.10}$$

que es la solución de la ecuación del calor homogénea

$$\begin{cases} v_t - \Delta v = 0 & \text{en} \quad \Omega,\ t > 0 \\ v = 0 & \text{en} \quad \partial\Omega,\ t > 0 \\ v(x,\, 0) = v_0(x) = u_0(x) - u^*(x) & \text{en} \quad \Omega. \end{cases} \tag{11.9.11}$$

La identidad de energía asgura en este caso que

$$\frac{1}{2}\frac{d}{dt}\int_\Omega v^2 dx + \int_\Omega \mid \nabla v \mid^2 dx = 0. \tag{11.9.12}$$

Aplicando la desigualdad de Poincaré, como en el argumento anterior, deducimos que

$$\parallel v(t) \parallel_{L^2(\Omega)} \leq e^{-c(\Omega)t} \parallel u_0 - u^* \parallel_{L^2(\Omega)}, \tag{11.9.13}$$

lo cual proporciona la velocidad exponencial de convergencia anunciada.

De hecho (11.9.13) proporciona una tasa de decaimiento exponencial explícita, del orden de la constante de Poincaré $c(\Omega)$.

El método de la energía que acabamos de presentar permite obtener tasas de convergencia al equilibrio para una amplia clase de ecuaciones parabólicas lineales y no-lineales y también incluso, para ecuaciones de ondas disipativas. De este modo puede por ejemplo abordarse el caso de la ecuación parabólica p-Laplaciana:

$$u_t - \operatorname{div}\left(\mid \nabla u \mid^{p-2} \nabla u\right) = 0. \tag{11.9.14}$$

Conviene sin embargo tener en cuenta que no siempre se obtiene decaimiento exponencial, pudiendo ser este polinomial.

Por ejemplo, en el caso de la ecuación (11.9.14) con condiciones de contorno de Dirichlet, por el método de la energía deducimos la identidad

$$\frac{1}{2}\frac{d}{dt}\int_\Omega \mid u \mid^2 dx + \int_\Omega \mid \nabla u \mid^p dx = 0.$$

La desigualdad de Poincaré en este caso garantiza la existencia de una constante $c_p(\Omega) > 0$ tal que

$$\int \mid \nabla u \mid^p dx \geq c_p(\Omega) \int_\Omega \mid u \mid^p dx,$$

de donde obtenemos la estimación

$$\frac{1}{2}\frac{d}{dt}\int_\Omega u^2 dx + c_p(\Omega)\int_\Omega \mid u \mid^p dx \leq 0.$$

Suponiendo que $p > 2$ y aplicando la desigualdad de Hölder deducimos que

$$\int_\Omega u^2 dx \leq \left(\int_\Omega \mid u \mid^p dx\right)^{2/p} \mid \Omega \mid^{(p-2)/p}.$$

De estas dos desigualdades deducimos que

$$\frac{d}{dt}\int_\Omega u^2 dx + \frac{2c_p(\Omega)}{\mid \Omega \mid^{(p-2)/2}}\left(\int_\Omega \mid u \mid^2 dx\right)^{p/2} \leq 0.$$

Resolviendo esta desigualdad deducimos que

$$\int_\Omega u^2 dx \leq \left[\frac{(p-2)}{2}\left(\frac{2c_p(\Omega)}{\mid \Omega \mid^{(p-2)/2}}t + \frac{2\int_\Omega u_0^2 dx}{(p-2)}\right)^{(2-p)/2}\right]^{2/(p-2)} \leq c_p(\Omega)t^{-2/(p-2)},$$

es decir, un decaimiento polinomial con un exponente $2/(p-2)$. A medida que $p \nearrow \infty$ esta tasa de convergencia se debilita, lo cual refleja la creciente debilidad del efecto disipativo del p-Laplaciano. Esto es natural puesto que, en la medida que, $u \to 0$ cuando $t \to \infty$, al aumentar p la difusividad del p-Laplaciano disminuye. Sin embargo, a medida que $p \searrow 2$, la tasa de decaimiento polinomial $2/(p-2)$ tiende a infinito, lo cual refleja el hecho de que para el caso de la ecuación lineal ($p = 2$), se tenga una tasa de decaimiento exponencial.

Pero volvamos a la ecuación lineal (11.9.1). En (11.9.13) hemos obtenido una tasa de decaimiento exponencial con constante $C(\Omega)$, la constante de Poincaré. Con el objeto de analizar con más cuidado el comportamiento exponencial de las soluciones es conveniente utilizar desarrollos en serie de Fourier. Para ello consideramos una base ortogonal de $L^2(\Omega)$, $\{\varphi_j\}_{j\geq 1}$, constituida por autofunciones del Laplaciano:

$$\begin{cases} -\Delta\varphi_j = \lambda_j\varphi_j & \text{en} \quad \Omega \\ \varphi_j = 0 & \text{en} \quad \partial\Omega, \end{cases}$$

y asociada a la correspondiente sucesión de autovalores

$$0 < \lambda_1 < \lambda_2 \leq \ldots \leq \lambda_N \leq \ldots \longrightarrow \infty.$$

Desarrollamos en serie de Fourier los datos u_0 y f de (11.9.1).

Tenemos así

$$f = \sum_{j\geq 1} f_j\varphi_j;\ u_0 = \sum_{j\geq 1} u_{0,\,j}\varphi_j, \tag{11.9.15}$$

mientras que la solución habrá de ser de la forma

$$u(x,\, t) = \sum_{j\geq 1} u_j(t)\varphi_j. \tag{11.9.16}$$

La búsqueda de la solución se reduce a la determinación de sus coeficientes de Fourier. Estos vienen caracterizados por las ecuaciones

$$\begin{cases} u_j' + \lambda_j u_j = f_j, & t > 0 \\ u_j(0) = u_{0,\,j}, & j \geq 1 \end{cases} \tag{11.9.17}$$

que pueden calcularse explícitamente:

$$u_j(t) = u_{0,\,j} e^{-\lambda_j t} + \frac{(1 - e^{-\lambda_j t})}{\lambda_j} f_j,\, j \geq 1. \tag{11.9.18}$$

De esta expresión se deduce inmediatamente que

$$u_j(t) \longrightarrow \frac{f_j}{\lambda_j},\, t \longrightarrow \infty,\, \forall j \geq 1. \tag{11.9.19}$$

Un análisis un poco más cuidadoso permite comprobar que

$$u(\cdot,\, t) \rightarrow_{t\to\infty} \sum_{j\geq 1} \frac{f_j}{\lambda_j}\varphi_j(x), \text{ en } L^2(\Omega). \tag{11.9.20}$$

Esto coincide con el resultado obtenido mediante el método de la energía puesto que

$$\sum_{j\geq 1} \frac{f_j}{\lambda_j}\varphi_j(x), \tag{11.9.21}$$

es precisamente el desarrollo en serie de Fourier de la solución estacionaria u^* de (11.9.8).

De hecho este análisis confirma que

$$\| u(t) - u^* \|_{L^2(\Omega)} \leq e^{-\lambda_1 t} \| u_0 - u^* \|_{L^2(\Omega)}, \tag{11.9.22}$$

siendo $\lambda_1 > 0$ el primer autovalor del operador $-\Delta$ en $H_0^1(\Omega)$.

En realidad esta estimación coincide con la obtenida en (11.9.13) puesto que la constante de Poincaré $c(\Omega)$ coincide con λ_1, i.e.

$$c(\Omega) = \lambda_1.$$

Esto es así puesto que, por el principio mini-max, el primer autovalor λ_1 se caracteriza como el mínimo del cociente de Rayleigh,

$$\lambda_1 = \min_{\psi \in H_0^1(\Omega)} \frac{\displaystyle\int_\Omega |\, \nabla\psi \,|^2\, dx}{\displaystyle\int_\Omega \psi^2 dx}.$$

De modo que λ_1, el primer autovalor, es la mayor constante de la desigualdad

$$\lambda_1 \int_\Omega \psi^2 dx \leq \int_\Omega \mid \nabla\psi \mid^2 dx, \, \forall\psi \in H_0^1(\Omega).$$

En esta sección hemos probado que las soluciones de la ecuación del calor convergen a una solución estacionaria, cuando la fuente externa f es independiente del tiempo. El mismo tipo de argumentos permiten probar, por ejemplo, que si $f = f(x, t)$ depende de manera periódica (de período $\tau > 0$) en el tiempo t, entonces existe una única solución τ-periódica $u^* = u^*(x, t)$ de

$$\begin{cases} u_t - \Delta u = f & \text{en} \quad \Omega, \, 0 < t < \tau \\ u = 0 & \text{en} \quad \partial\Omega, 0 < t < \tau \\ u(0) = u(\tau) \end{cases}$$

de modo que, para cualquier solución del problema de valores iniciales

$$\begin{cases} u_t - \Delta u = f & \text{en} \quad \Omega, \quad t > 0 \\ u = 0 & \text{en} \quad \partial\Omega, \quad t > 0 \\ u(x, 0) = u_0(x) & \text{en} \quad \Omega, \end{cases}$$

se cumple que

$$\| \, u(t) - u^*(t) \, \|_{L^2(\Omega)} \to 0, \, t \to \infty$$

con velocidad exponencial. Dejamos los detalles de esta prueba al lector interesado.

En esta sección y en la anterior hemos mostrado algunos ejemplos en los que las soluciones de la ecuación de evolución, cuando $t \to \infty$, se simplifican asintóticamente al converger a un estado estacionario. Esto permite en algunos casos sustituir el modelo de evolución por el estacionario pero no sin el riesgo que supone haber realizado a priori un estudio cuantitativo de la distancia de ambas soluciones (evolución / estacionaria) o, lo que es lo mismo, de la velocidad de convergencia cuando $t \to \infty$.

En la práctica son muchos los casos en los que se adopta un modelo estacionario sin disponer de una prueba rigurosa de la convergencia asintótica de las soluciones de la ecuación de evolución cuando $t \to \infty$. Esto es una constante en el ámbito del análisis numérico y computacional en el que, con frecuencia, nos vemos obligados a emplear métodos sin poseer prueba rigurosa de convergencia más que en algunos casos simplificados.

11.10. Conclusión

En esta sección hemos visto cómo la discretización temporal implícita de Euler y la aproximación de Galerkin continua en tiempo son dos métodos de

aproximación para ecuaciones parabólicas no-lineales de la forma (11.6.1).

El hecho de que el esquema discreto en tiempo sea una manera natural de aproximar la EDP de evolución es lo que nos ha motivado en la sección anterior a introducir los esquemas de "splitting" o de descomposición en ecuaciones discretas en tiempo.

Los desarrollos de esta sección, en el marco de la ecuación del calor p-Laplaciana (11.6.1), pueden también hacerse, con argumentos análogos, para otros modelos como son las ecuaciones de Navier-Stokes $2-d$ o la ecuación de Burgers viscosa $1-d$.

Como el sistema (11.6.1) es autónomo y posee soluciones únicas, es el generador de un semigrupo que, por analogía con el caso lineal, denotamos $S(t)$. Así la solución u de (11.6.1) en el instante t se puede denotar como $u(t) = S(t)u_0$, donde $S(t)$ es la aplicación no-lineal solución. Esta familia de aplicaciones $\{S(t)\}_{t>0}$ satisface las dos relaciones

$$\begin{cases} S(0) = Id, \\ S(t) \circ S(s) = S(t+s), \end{cases}$$

por lo que se dice semigrupo.

Los desarrollos de esta sección están en la base de la teoría de semigrupos no-lineales en el marco de la cual (11.6.1) y los demás modelos mencionados no son más que ejemplos concretos que pueden abordarse mediante esta técnica sistemática.

Apéndice A

Aproximación de dominios en el problema de Dirichlet

Tal y como veíamos en la sección 9.2.5, la primera fuente de error en la aproximación del MEF P_1 en $2D$ es la que se deriva de la aproximación del dominio regular Ω mediante un dominio poligonal Ω_h. En este Apéndice mostramos brevemente que el error producido por esta aproximación es despreciable.

Consideramos en primer lugar el caso en que Ω_h es una aproximación interior de modo que $\Omega_h \subset \Omega$. En este caso, todo elemento del espacio $H_0^1(\Omega_h)$ puede también considerarse como elemento de $H_0^1(\Omega)$ puesto que la extensión por cero fuera de Ω_h es una aplicación lineal y continua de Ω_h en Ω.

La solución u en Ω y u_h en Ω_h satisfacen respectivamente,

$$\begin{cases} u \in H_0^1(\Omega) \\ \displaystyle\int_\Omega \nabla u \cdot \nabla v dx = \langle f,\, v\rangle_\Omega,\ \forall v \in H_0^1(\Omega) \end{cases} \tag{A.0.1}$$

y

$$\begin{cases} u_h \in H_0^1(\Omega_h) \\ \displaystyle\int_\Omega \nabla u_h \cdot \nabla v_h dx = \langle f,\, v_h\rangle_{\Omega_h},\ \forall v_h \in H_0^1(\Omega_h). \end{cases} \tag{A.0.2}$$

Como toda función $v_h \in H_0^1(\Omega_h)$ mediante el operador lineal de extensión por cero fuera del dominio Ω_h puede considerarse como un elemento de $H_0^1(\Omega)$ y de este modo $H_0^1(\Omega_h)$ es un subespacio cerrado de $H_0^1(\Omega)$, tenemos que

$$u_h = \pi_h u \tag{A.0.3}$$

donde $\pi_h : H_0^1(\Omega) \to H_0^1(\Omega_h)$ es el operador de proyección ortogonal.

Por tanto

$$\| u - u_h \|_{H_0^1(\Omega)} = \| u - \pi_h u \|_{H_0^1(\Omega)} = \min_{w_h \in H_0^1(\Omega_h)} \| u - w_h \|_{H_0^1(\Omega)} . \qquad (A.0.4)$$

Por consiguiente con el objeto de probar que $u_h \to u$ en $H_0^1(\Omega)$ o, incluso, de obtener una estimación sobre la velocidad de convergencia, es suficiente encontrar elementos $w_h \in H_0^1(\Omega_h)$ lo más próximos posible a $u \in H_0^1(\Omega)$.

Procedemos por un argumento de densidad, tal y como es habitual en el MEF.[1]

Sabemos que $C_c^\infty(\Omega)$ es denso en $H_0^1(\Omega)$. Por otra parte, si Ω_h aproxima a Ω de modo que Ω_h acabe conteniendo a cada compacto K de Ω la propiedad de aproximación necesariamente se satisface.

Deducimos por tanto que la distancia (A.0.4) tiende necesariamente a cero cuando Ω_h es una aproximación interna de Ω con la propiedad de que todo compacto contenido en Ω está contenido en Ω_h para todo $h \leqslant h_0$, con $h_0 > 0$ suficientemente pequeño.

Estas propiedades se cumplen eligiendo Ω_h como una aproximación interior de Ω, poligonal, con lados del orden de h, obtenidos uniendo una familia de puntos sobre $\partial\Omega$ y corrigiendola ligeramente en caso de que, por la falta de convexidad de Ω, el dominio Ω_h exceda parcialmente a Ω.

Dejamos para el lector el caso en que Ω_h es una aproximación que no está estrictamente contenida en Ω, así como el problema de la obtención de velocidades de convergencia cuando $u \in H^2 \cap H_0^1(\Omega)$.

Hemos probado la convergencia de orden en h del MEF $2D$ para las soluciones H^2 del problema de Dirichlet para el Laplaciano en dominios poligonales. Sin embargo, cuando se aborda el caso de un dominio curvo regular se comete un error adicional al aproximarlo por un dominio poligonal.

En este Apéndice describimos brévemente cómo la estimación del error puede mantenerse en ese caso.

Consideramos por tanto el problema

$$\begin{cases} -\Delta u = f & \text{en} \quad \Omega \\ u = 0 & \text{en} \quad \partial\Omega \end{cases} \qquad (A.0.5)$$

con $f \in L^2(\Omega)$ y Ω un abierto de clase C^2.

Consideramos una familia de triangulaciones regulares $\{\mathcal{C}_h\}_{h>0}$ de Ω tales que los dominios Ω_h poligonales que constituyen sean tales que $\Omega_h \subset \Omega$ para

[1] Este argumento ha sido usado en la prueba de la convergencia del método de elementos finitos en $1D$, por ejemplo, en la sección 9.2.3.

todo $h > 0$ y de modo que la distancia máxima entre los puntos $\partial\Omega$ y de $\partial\Omega_h$ sea del orden de h^2.

Esta aproximación es la que se obtiene de manera espontánea cuando Ω es un abierto convexo de clase C^2 y se toma una triangulación que, sobre el borde, induzca una aproximación lineal a trozos de la frontera con vértices distantes del orden de h. Véase la figura:

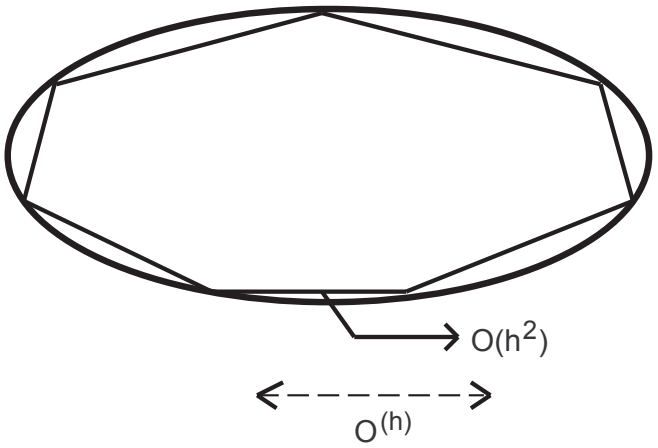


Figura A.1: Aproximación de un dominio regular mediante un polígono de lados del orden de $O(h)$ donde se observa que los sectores próximos a la frontera que quedan fuera del polígono son de un area del orden de $O(h^2)$.

Cuando el dominio Ω es no convexo la aproximación Ω_h así obtenida no tiene por qué ser interna. Es por eso que es necesario eventualmente corregirla. Pero esto puede realizarse manteniendo una distancia del orden $O(h^2)$ entre las fronteras

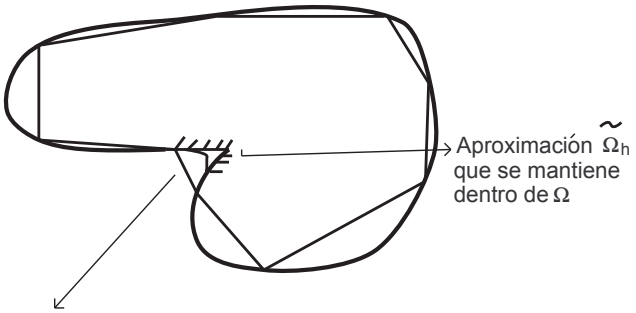


Figura A.2: Aproximación Ω_h que excede Ω que corregimos para que de lugar a una aproximación interior $\tilde{\Omega}_h$.

A partir de ahora suponemos por tanto que $\Omega_h \subset \Omega$ y que Ω_h proviene de una triangulación regular de tamaño relativo h y de modo que la distancia máxima entre $\partial\Omega$ y $\partial\Omega_h$ sea del orden de $O(h^2)$.

Sea $u_h \in V_h$ la solución aproximada obtenida por la aplicación del MEF en Ω_h. Extendemos u_h por cero fuera de Ω_h y obtenemos así $\tilde{u}_h \in H_0^1(\Omega)$. Nos proponemos probar la existencia de $C > 0$ tal que

$$\| u - \tilde{u}_h \|_{H_0^1(\Omega)} \leqslant Ch \| u \|_{H^2(\Omega)}, \tag{A.0.6}$$

para todo $h > 0$.

Descomponemos la norma a estimar en dos partes: El dominio poligonal interior Ω_h y la banda en torno a la frontera $B_h = \Omega \backslash \bar{\Omega}_h$.

Tenemos que

$$\| u - \tilde{u}_h \|^2_{H_0^1(\Omega)} = \int_{\Omega_h} | \nabla(u - u_h) |^2 \, dx + \int_{B_h} | \nabla u |^2 \, dx = I_1(h) + I_2(h). \tag{A.0.7}$$

Distinguimos ahora el tratamiento de cada una de las integrales.

• **Integral** $I_1(h)$

La función u es de clase H^2 en el dominio Ω_h en consideración. Por tanto, el análisis realizado para soluciones H^2 en un polígono se aplica sin ninguna alteración y obtenemos que

$$I_1(h) \leqslant C\, h^2. \tag{A.0.8}$$

Conviene observar que en este análisis no se utiliza para nada el hecho de que la función u se anula en el borde del dominio poligonal o no, cosa que no ocurre en este caso.

• **Integral** $I_2(h)$

Hay una primera estimación de esta integral que es fácil de obtener bajo la hipótesis adicional de que $\nabla u \in L^\infty(\Omega)$ es decir que $u \in W^{1,\infty}(\Omega)$.

En este caso tenemos

$$I_2(h) \leqslant \| \nabla u \|^2_{L^\infty(\Omega)} | B_h |, \tag{A.0.9}$$

donde $| B_h |$ denota la medida de B_h.

Ahora bien, teniendo en cuenta que B_h está contenido en un entorno de la frontera de Ω de anchura del orden de $O(h^2)$ deducimos que $| B_h | = O(h^2)$, de donde se deduce la estimación que buscábamos

$$I_2(h) \leqslant C\, h^2. \tag{A.0.10}$$

Pero hay que observar que en dos dimensiones espaciales el espacio $H^2(\Omega)$ no está contenido en $W^{1,\infty}(\Omega)$. En efecto, las funciones de $H^2(\Omega)$ pueden tener gradientes con singularidades logorítmicas. Por tanto, la hipótesis de que la solución $u \in W^{1,\infty}(\Omega)$ es una condición adicional. Hay resultados clásicos sobre regularidad para problemas elípticos que nos permiten dar condiciones suficientes para que esta condición se cumpla. Así, por ejemplo, en un dominio de clase C^2, si el segundo miembro f del problema de Dirichlet pertenece a $L^p(\Omega)$ con $p > 2$, la solución u pertence a $W^{2,p}(\Omega)$ que a su vez se inyecta con continuidad en $W^{1,\infty}(\Omega)$. Esto proporciona condiciones suficientes para que (A.0.10) se cumpla pero no garantiza el resultado bajo la condición óptima de que $u \in H^2(\Omega)$. Analicemos pues este caso que tiene naturaleza crítica.

El conjunto B_h es unión de elementos curvos de la forma mostrada en la figura que denotamos mediante ω_h:

Su diámetro es del orden de $O(h)$ pero su "anchura" es del orden de $O(h^2)$. La frontera de ω_h consta de dos partes, la exterior, $\partial\omega_{h,\,\mathrm{ext}}$ que está contenida en $\partial\Omega$ y la interior que denotamos por γ_h.

Multiplicando la ecuación $-\Delta u = f$ por u en ω_h, integrando por partes y usando que u se anula en la frontera exterior $\partial\omega_{h,\,\mathrm{ext}}$ deducimos que

$$\begin{aligned} \int_{\omega_h} | \nabla u |^2 \, dx &= \int_{\omega_h} f\, u\, dx + \int_{\gamma_h} \frac{\partial u}{\partial n} u \, d\sigma \\ &\leqslant \| f \|_{L^2(\omega_h)} \| u \|_{L^2(\omega_h)} + \left\| \frac{\partial u}{\partial n} \right\|_{L^2(\gamma_h)} \| u \|_{L^2(\gamma_h)} . \end{aligned} \tag{A.0.11}$$

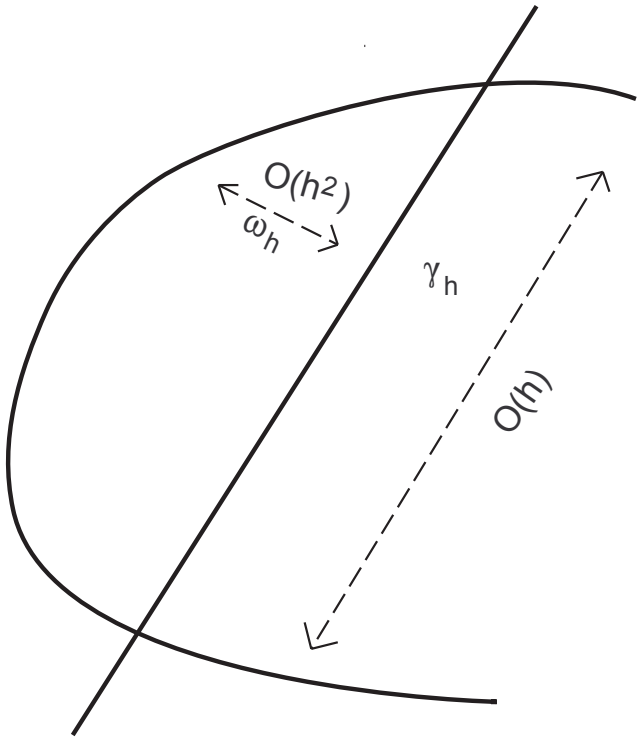


Figura A.3: Vista aumentada de la zona que queda fuera de la aproximación poligonal donde se observa que el diámetro es del orden de $O(h)$ mientras que el área es del orden de $O(h^2)$.

En la medida en que el dominio ω_h es de anchura $O(h^2)$, retomando la prueba de la desigualdad de Poincaré se obtiene que

$$\| u \|_{L^2(\omega_h)} \leqslant C\, h \, \| \nabla u \|_{L^2(\omega_h)} \, . \tag{A.0.12}$$

Esto es así puesto que u se anula en la frontera exterior de ω_h.

De (A.0.12) se deduce que

$$\| f \|_{L^2(\omega_h)} \| u \|_{L^2(\omega_h)} \leqslant C \, \| f \|_{L^2(\omega_h)} \, h \, \| \nabla u \|_{L^2(\omega_h)} \leqslant C \, \| f \|^2_{L^2(\omega_h)} \, h^2 + \frac{1}{2} \, \| \nabla u \|^2_{L^2(\omega_h)} \, . \tag{A.0.13}$$

Combinando (A.0.12) y (A.0.13) deducimos que

$$\int_{\omega_h} | \nabla u |^2 \, dx \leqslant C\, h^2 \, \| f \|^2_{L^2(\omega_h)} + \Big|\Big| \frac{\partial u}{\partial n} \Big|\Big|_{L^2(\gamma_h)} \, \| u \|_{L^2(\gamma_h)} \, . \tag{A.0.14}$$

Basta por tanto, estimar los últimos términos de esta desigualdad puesto que los primeros del miembro de la derecha no entrañan dificultad en el sentido que

$$\begin{aligned}\int_{B_h} |\nabla u|^2\, dx &= \sum_{\omega_h} |\nabla u|^2\, dx \\ &\leqslant C\,h^2 \sum_{\omega_h} \int_{\omega_h} f^2 dx + \sum_{\omega_h} \left|\left|\frac{\partial u}{\partial n}\right|\right|_{L^2(\gamma_h)} \|\, u\,\|_{L^2(\gamma_h)} \\ &\leqslant C\,h^2 \,\|\, f\,\|^2_{L^2(\Omega)} + \sum_{\omega_h} \left|\left|\frac{\partial u}{\partial n}\right|\right|_{L^2(\gamma_h)} \|\, u\,\|_{L^2(\gamma_h)} \end{aligned} \tag{A.0.15}$$

Por otra parte

$$\sum_{\omega_h} \left|\left|\frac{\partial u}{\partial n}\right|\right|_{L^2(\gamma_h)} \|\, u\,\|_{L^2(\gamma_h)} \leqslant \left(\sum_{\omega_h} \left|\left|\frac{\partial u}{\partial n}\right|\right|^2_{L^2(\gamma_h)}\right)^{1/2} \left(\sum_{\omega_h} \|\, u\,\|^2_{L^2(\gamma_h)}\right)^{1/2}. \tag{A.0.16}$$

y

$$\sum_{\omega_h} \left|\left|\frac{\partial u}{\partial n}\right|\right|^2_{L^2(\gamma_h)} = \left|\left|\frac{\partial u}{\partial n}\right|\right|^2_{L^2(\partial\Omega_h)}. \tag{A.0.17}$$

La demostración del Teorema de trazas garantiza la existencia de una constante C independiente de h tal que

$$\left|\left|\frac{\partial u}{\partial n}\right|\right|_{L^2(\partial\Omega_h)} \leqslant C\,\|\, u\,\|_{H^2(\Omega)}\,. \tag{A.0.18}$$

Basta por tanto probar que

$$\sum_{\omega_h} \|\, u\,\|^2_{L^2(\gamma_h)} = \|\, u\,\|^2_{L^2(\partial\Omega_h)} = O(h^2). \tag{A.0.19}$$

Esto es nuevamente fácil de comprobar en el caso en que $\nabla u \in L^\infty(\Omega)$. En este caso, como u se anula en el borde exterior de Ω y la "anchura" de la banda B_h es $O(h^2)$ tenemos que $\|\, u\,\|_{L^\infty(\partial\Omega_h)} = O(h^2)$. Como la longitud o perímetro de $\partial\Omega_h$ está uniformemente acotada deduciríamos que (A.0.19) se cumple. Pero, nuevamente, hemos de hacerlo bajo la hipótesis $u \in H^2(\Omega)$.

Con el objeto de probar (A.0.19), sin pérdida de generalidad, suponemos que ω_h es un conjunto del plano de la forma

$$\omega_h = \{(x,\, y) \in \mathbb{R}^2 :\, v \leqslant x \leqslant h,\, 0 \leqslant y \leqslant \varphi(x)\}$$

donde $\|\, \varphi\,\|_\infty = O(h^2)$. Entonces $\gamma_h = \{(x,\, y) :\, y = 0\}$, tal y como se indica en la siguiente figura.

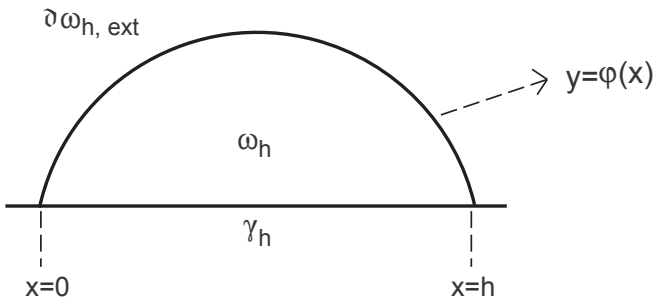


Figura A.4: Representación del dominio ω_h de referencia.

Se trata entonces de estimar

$$\int_{\gamma_h} u^2 d\sigma = \int_0^h u^2(x,\, 0)dx.$$

Como $u(x,\,\varphi(x))=0$, puesto que $(x,\,\varphi(x))\in\partial\Omega$, tenemos que

$$\begin{aligned}
&\int_{\gamma_h} u^2 d\sigma = \int_0^h u^2(x,\,0)dx = \int_0^h \left[\int_0^{\varphi(x)} u_y(x,\,s)ds\right]^2 dx \\
= &\int_0^h \left[\int_0^{\varphi(x)} \left(\int_0^s u_{yy}(x,\,\sigma)+u_y(x,\,0)\right) ds\right]^2 dx \\
\leqslant &\int_0^h \int_0^{\varphi(x)} \left(\int_0^s u_{yy}(x,\,\sigma)d\sigma + u_y(x,\,0)\right)^2 ds\varphi(x)dx \\
\leqslant &\; C\,h^2 \int_0^h \int_0^{\varphi(x)} \left[\left(\int_0^s u_{yy}(x,\,\sigma)d\sigma\right)^2 + u_y^2(x,\,0)\right] dsdx \\
\leqslant &\; C\,h^2 \int_0^h \int_0^{\varphi(x)} \left[\int_0^s u_{yy}(x,\,\sigma)d\sigma\right]^2 ds\,dx + C\,h^2 \int_0^h \varphi(x) u_y^2(x,\,0)dx \\
\leqslant &\; C\,h^2 \int_0^h \int_0^{\varphi(x)} \int_0^s u_{yy}^2(x,\,\sigma)d\sigma s ds\,dx + C\,h^4 \int_0^h u_y^2(x,\,0)dx \\
\leqslant &\; C\,h^4 \left[\int_0^h \int_0^{\varphi(x)} \int_0^s u_{yy}^2(x,\,\sigma)d\sigma\,ds dx + \int_0^h u_y^2(x,\,0)dx\right] \\
\leqslant &\; C\,h^4 \left[\int_0^h \int_0^{\varphi(x)} u_{yy}^2(x,\,s)ds\varphi(x)dx + \int_0^h u_y^2(x,\,0)dx\right] \\
\leqslant &\; C\,h^4 \left[h^2 \int_0^h \int_0^{\varphi(x)} u_{yy}^2(x,\,s)ds + \int_0^h u_y^2(x,\,0)dx\right] \\
\leqslant &\; C\,h^4 \left[h^2 \parallel u \parallel^2_{H^2(\omega_h)} + \int_{\partial\omega_h,\,\mathrm{ext}} \left|\frac{\partial u}{\partial n}\right|^2 d\sigma\right].
\end{aligned}$$

Por tanto

$$\sum_{\omega_h} \parallel u \parallel^2_{L^2(\gamma_h)} \leqslant C\,h^6 \parallel u \parallel^2_{H^2(\Omega)} + C\,h^4 \int_{\partial\Omega} \left|\frac{\partial u}{\partial n}\right|^2 d\sigma.$$

Pero, los resultados de trazas garantizan que

$$\int_{\partial\Omega} \left|\frac{\partial u}{\partial n}\right|^2 d\sigma \leqslant C \parallel u \parallel^2_{H^2(\Omega)}.$$

Llegamos por tanto a la conclusión que

$$\sum_{\omega_n} \parallel u \parallel^2_{L^2(\gamma_h)} \leqslant C\,h^4 \parallel u \parallel_{H^2(\Omega)}.$$

Se satisface por tanto con creces la desigualdad (A.0.19) que necesitabamos probar.

Apéndice B

Aceleración del MDD para datos rápidamente oscilantes

En los experimentos numéricos realizados en clase hemos constatado que, en algunas ocasiones, el hecho de que los datos del problema oscilen rápidamente a lo largo de la interfase hace que el MDD converja más rápidamente. El objeto de esta sección es dar una prueba rigurosa de este hecho.

Consideramos para ello el caso sencillo de un cuadrado $\Omega = (0,\, 1) \times (0,\, 1)$ que descomponemos en dos subdominios $\Omega_1 = (0,\, 2/3) \times (0,\, 1)$ y $\Omega_2 = (1/3,\, 1) \times (0,\, 1)$, siendo el dominio de solapamiento la banda vertical $\Omega_1 \cap \Omega_2 = (1/3,\, 2/3) \times (0,\, 1)$.

En los experimentos numéricos constatamos que las oscilaciones de las soluciones a lo largo de las dos interfases verticales $x = 1/3$ y $x = 2/3$ parecía acelerar la convergencia del método.

Veamos que ésto es efectivamente así. Consideramos para ello soluciones de la forma $u(x,\, y) = \varphi(x)\,\mathrm{sen}(m\pi y)$. Se trata entonces de soluciones que en la dirección vertical y paralela a las interfases oscilan más rápido a medida que m aumenta.

El operador de Laplace para estas soluciones admite entonces la forma

$$-\Delta u = \left(-d_x^2 + m^2\pi^2\right) \varphi(x)\,\mathrm{sen}(m\pi y). \tag{B.0.1}$$

Como la descomposición de dominios elegida no perturba la variable vertical y que en cada subdominio y pertenece al intervalo $(0,\, 1)$ las sucesivas iteraciones del método proporcionan siempre funciones que preservan la misma estructura en variables separadas, siendo la dependencia en y siempre la misma, es decir, de la forma $\mathrm{sen}(m\pi y)$.

El análisis habitual de la convergencia del MDD $2-D$ en este caso indica que la ecuación $1-D$ que gobierna el proceso es

$$-d_x^2\varphi + m^2\varphi = 0. \tag{B.0.2}$$

En otras palabras, al resolver esta ecuación en el intervalo de x correspondiente al dominio Ω_1 que es $x \in (0,\, 2/3)$ con condiciones $\varphi(0) = 0$ y $\varphi(2/3) = 1$, la constante que nos da el orden de convergencia es el valor de esta solución en el punto $1/3$. Tal y como veremos este valor tiende a cero cuando m aumenta.

Para realizar este cálculo conviene situarnos en el intervalo unidad $(0,\, 1)$. La ecuación a resolver es entonces

$$\begin{cases} -\varphi'' + m^2\varphi = 0, & 0 < x < 1 \\ \varphi(0) = 0, & \varphi(1) = 1 \end{cases} \tag{B.0.3}$$

cuya solución es

$$\varphi_m(x) = \frac{e^m}{e^{2m}-1}\left(e^{mx} - e^{-mx}\right). \tag{B.0.4}$$

Consideremos ahora un punto intermedio $\alpha \in (0,\, 1)$. Entonces

$$\varphi_m(\alpha) = \frac{e^m}{e^{2m}-1}\left(e^{m\alpha} - e^{-m\alpha}\right). \tag{B.0.5}$$

Cuando $m = 0$ tenemos

$$\varphi_0(x) = x \tag{B.0.6}$$

y por tanto

$$\varphi_0(\alpha) = \alpha. \tag{B.0.7}$$

A medida que m aumenta vemos que

$$\varphi_m(\alpha) \to 0,\ m \to \infty. \tag{B.0.8}$$

De hecho

$$\varphi_m(\alpha) \sim e^{2m(\alpha-1)}. \tag{B.0.9}$$

La constante $\varphi_m(\alpha)$ establece la contractividad de cada paso de aplicación del MDD en este caso. Es por eso que el método converge más rápido a medida que m aumenta. De la expresión (B.0.9) se deduce que la aceleración de la convergencia es exponencial a medida que m aumenta.

Apéndice C

Ejercicios

Ejercicio C.0.1 En el marco continuo, en una dimensión espacial, hemos probado la convergencia del método de descomposición de dominios tomando en los extremos de cada subdominio Ω_- y Ω_+ condiciones de contorno de Dirichlet.

Analizar la convergencia del método consistente en utilizar la ecuación de Neumann en los puntos interiores de la interfase x_- y x_+. En este caso el esquema se escribe:

$$\begin{cases} -\left(u_-^k\right)'' = 0,\ -1 < x < x_- \\ u_-^k(-1) = \alpha_-,\ u_{-,x}^k(x_-) = u_{+,x}^{k-1}(x_-) \end{cases}$$

$$\begin{cases} -\left(u_+^k\right)'' = 0,\ x_+ < x < 1 \\ u_{+,x}^k(x_+) = u_{-,x}^k(x_-),\ u_+^k(1) = \alpha_+. \end{cases}$$

Ejercicio C.0.2 Estudiar el mismo problema cuando en una de las interfases tomamos condiciones de contorno de Dirichlet y en la otra de Neumann.

Ejercicio C.0.3 Probar la convergencia del método de descomposición de dominios en el marco continuo para el problema de Dirichlet en el caso en que el operador $-d_x^2$ se sustituye por $-d_x(a(x)d_x)$ siendo $a = a(x)$ un coeficiente medible acotado y elíptico, i.e.

$$\exists a_0 > 0 :\ a(x) \geqslant a_0 > 0 \text{ p.c.t. } 1 < x < 1.$$

Ejercicio C.0.4 Resolvemos la ecuación de Laplace

$$-\Delta u = f \text{ en } \mathbb{R}^n$$

donde f es una función de soporte compacto.

(a) Escribir la representación de u por convolución con la solución fundamental.

(b) Suponiendo además que $f \in L^r(\mathbb{R}^n)$ determinar, en función de la dimensión espacial n, el rango de espacios L^p_{loc} al que pertenece la solución.

> **Indicación:** En la medida en que f es de soporte compacto y sólo nos interesan las propiedades locales de la solución podemos truncar el soporte de la solución fundamental y aplicar después la desigualdad de Young.

Ejercicio C.0.5 (a) Demostrar que la desigualdad de Poincaré no se verifica en $\mathbb{R}^n$.

(b) Deducir que el espacio $\dot{H}^1(\mathbb{R}^n)$, completado de $H^1(\mathbb{R}^n)$ con respecto a la norma

$$\| \varphi \|_{\dot{H}^1(\mathbb{R}^n)} = \left(\int_{\mathbb{R}^n} | \nabla \varphi |^2 \, dx \right)^{1/2}$$

es un espacio más grande que $H^1(\mathbb{R}^n)$.

(c) Demuestra sin embargo que, cuando $n \geqslant 2$, por la desigualdad de Sobolev y de Trudinger, la única función constante que pertenece a $\dot{H}^1(\mathbb{R}^n)$ es la trivial $u \equiv 0$.

(d) >Qué ocurre cuando $n = 1$?

Ejercicio C.0.6 (a) Demostrar que si Ω es un dominio acotado se cumple la siguiente variante de la desigualdad de Poincaré:

$$\int_{\mathbb{R}^n} \varphi^2 dx \leqslant C \left[\int_{\mathbb{R}^n} | \nabla \varphi |^2 \, dx + \int_{\Omega^c} \varphi^2 dx \right], \forall \varphi \in H^1(\mathbb{R}^n).$$

(b) Probarlo tanto por un argumento de contradicción por compacidad como por la demostración constructiva habitual de la desigualdad de Poincaré en un dominio acotado

(c) >Se puede relajar la condición de que Ω sea acotado?

Ejercicio C.0.7 (a) Suponiendo que $n \geqslant 2$ demostrar que existen funciones de $H^1(\mathbb{R}^n)$ que no son continuas en $x = 0$.

(b) Usando un argumento de superposición y la existencia de conjuntos densos y numerables en $\mathbb{R}^n$ deducir la existencia de funciones de $H^1(\mathbb{R}^n)$ que no son continuas en ningún punto.

Ejercicio C.0.8 Sea Ω un dominio acotado de $\mathbb{R}^n$. Probar que si $f \in H^1(\Omega)$ y $g \in C^1(\bar{\Omega})$, entonces $fg \in H^1(\Omega)$.

Ejercicio C.0.9 Probar que la desigualdad de Sobolev

$$\left(\int_{\mathbb{R}^n} \mid \varphi \mid^{2n/(n-2)} \, dx \right)^{(n-2)/2n} \leqslant C_n \left(\int_{\mathbb{R}^n} \mid \nabla \varphi \mid^2 \, dx \right)^{1/2}$$

que se cumple en dimensiones $n \geqslant 3$, es invariante por cambios de escala $\varphi_R(x) = \varphi(Rx)$.

Ejercicio C.0.10 Consideramos la transformada de Fourier parcial en la variable $x' \in \mathbb{R}^{n-1}$ de $x = (x',\, x_n) \in \mathbb{R}^n$:

$$\hat{f}\,(\xi',\, x_n) = \int_{\mathbb{R}^{n-1}} e^{-ix' \cdot \xi'} f\,(x',\, x_n)\, dx'.$$

Probar que

$$\int_{\mathbb{R}^n} \mid f\,(x',\, x_n) \mid^2 \, dx' dx_n = \int_{\mathbb{R}^n} \mid \hat{f}\,(\xi',\, x_n) \mid^2 \, d\xi' dx_n;$$

$$\int_{\mathbb{R}^n} \mid \nabla f\,(x',\, x_n) \mid^2 \, dx' dx_n = \int_{\mathbb{R}^n} \left[\mid \xi' \mid^2 \left| \hat{f}\,(\xi', x_n) \right|^2 + \left| \partial_n \hat{f}\,(\xi',\, x_n) \right|^2 \right] d\xi' dx_n.$$

Ejercicio C.0.11 Suponiendo que A es simétrica definida positiva comparar la velocidad de convergencia obtenida para los métodos de máximo descenso y de gradiente conjugado.

Ejercicio C.0.12 Consideramos el espacio de las sucesiones de cuadrado sumable ℓ^2. Consideramos asimismo el subespacio

$$h^1 = \left\{ \{a_j\}_{j \in \mathbb{N}} \in \ell^2 : \sum_{j=1}^{\infty} j^2 a_j^2 < \infty \right\}.$$

Consideramos por último el subespacio ℓ^2_N de los elementos de ℓ^2 tales que los términos de la sucesión correspondiente a índices $j \geqslant N$ se anulen.

Probar que:

- h^1 es denso en ℓ^2,
- Para cada $\vec{a} \in \ell^2$ y $N \in \mathbf{N}$ existe $\vec{a}_N \in \ell^2_N$ tal que $\vec{a}_N \to \vec{a}$ en ℓ^2 cuando $N \to \infty$.
- No existe ninguna función $\varepsilon(N) \to 0$ cuando $N \to \infty$ tal que para cada $\vec{a}_N \in \ell^2_N$ tal que

$$\| \vec{a} - \vec{a}_N \|_{\ell^2} \leqslant \varepsilon(N) \| \vec{a} \|_{\ell^2}.$$

- Probar que existe $\varepsilon(N)$ y calcularlo explícitamente de modo que

$$\| \vec{a} - \vec{a}_N \|_{\ell^2} \leqslant \varepsilon(N) \| \vec{a} \|_{h^1}, \forall \vec{a} \in h^1,$$

donde $\| \cdot \|_{h^1}$ denota la norma canónica de h^1, i.e.

$$\| \vec{a} \|_{h^1} = \left[\sum_{j=1}^{\infty} j^2 a_j^2 \right]^{1/2}.$$

Ejercicio C.0.13 Demuestra que se mantiene el orden $O(h)$ de convergencia del método de elementos finitos P_1 en una dimensión espacial sin necesidad de suponer que el mallado sea uniforme, es decir, bajo la hipótesis más general que

$$\max_{j=1\cdots N_h} | x_{j+1} - x_j | = O(h)$$

siendo $\{x_j\}_{j=1,\cdots,N_h}$ los nodos del mallado.

Ejercicio C.0.14 Obtén una estimación del orden de convergencia para el método de elementos finitos P_1 en una dimensión espacial bajo la hipótesis de que $u \in W^{2,p}$ con $1 \leqslant p \leqslant 2$.

Ejercicio C.0.15 Sea Ω un dominio acotado de $\mathbb{R}^n$. Suponemos que u y $v \in L^p(\Omega)$ con $1 < p < \infty$.

Para cada $h > 0$ consideramos la función:

$$J(h) = \frac{1}{h} \int_{\Omega} [| u + hv |^p - | u |^p] \, dx.$$

(a) Comprobar que $J : \mathbb{R}^+ \to \mathbb{R}$ está bien definida.

(b) Aplicando el Teorema de la Convergencia Dominada comprueba que J es continua en cada punto $h > 0$.

(c) Calcula

$$\lim_{h \to 0} J(h).$$

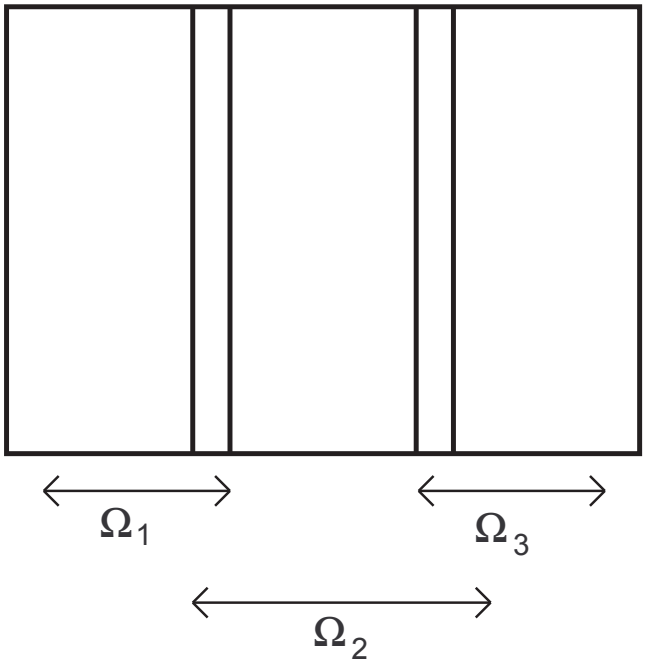
(d) Responde a las mismas cuestiones en el caso del funcional

$$\hat{J}(h) = \frac{1}{h} \int_{\Omega} \left[| u + hv |^{p-1} (u + hv) - | u |^{p-1} u \right] dx.$$

Ejercicio C.0.16 Consideramos la ecuación de Laplace en un dominio cuadrado Ω de $\mathbb{R}^2$:

$$\text{(1)} \qquad \begin{cases} -\Delta u = f & \Omega \\ u\Big|_{\partial\Omega} = 0. \end{cases}$$

Descomponemos el dominio en tres bandas verticales con solapamiento:



(a) Construir un método de descomposición de dominios en el que se itere entre los subdominios Ω_1, Ω_2 y Ω_3.

(b) Demostrar la convergencia del método.

(c) Hacer lo mismo en el caso de una discretización de la ecuación por diferencias finitas.

Ejercicio C.0.17 Sea X un espacio de Banach y $L \in \mathcal{L}(X, X)$ un operador lineal acotado. Probar que la función $x_0 \in X \to e^{tL}x_0$ es analítica real.

Ejercicio C.0.18 Consideramos un operador no acotado lineal A en un espacio de Hilbert H. Probar que las dos siguientes caracterizaciones de la propiedad de disipatividad son equivalentes:

- $\langle Ax, x\rangle \leqslant 0, \quad \forall x \in D(A)$
- $\| x - \lambda Ax \| \geqslant \| x \|, \quad \forall x \in D(A), \quad \forall \lambda > 0.$

Ejercicio C.0.19 Consideramos la ecuación del calor

$$\begin{cases} u_t - \Delta u = 0 & \text{en} \quad \Omega \times (0, \infty) \\ u = 0 & \text{en} \quad \partial\Omega \times (0, \infty) \\ u(0) = u_0 & \text{en} \quad \Omega \end{cases}$$

con dato inicial $u_0 \in L^2(\Omega)$.

Analizar la regularidad e integrabilidad de las funciones $t \to \int_\Omega u\varphi dx$ y $t \to \int_\Omega \nabla u \cdot \nabla \varphi dx$ tanto cuando $\varphi \in L^2(\Omega)$ como cuando $\varphi \in H_0^1(\Omega)$ distinguiendo el caso en que en Ω se cumple la desigualdad de Poincaré del que no.

Ejercicio C.0.20 Sea Ω un abierto no vacío de $\mathbb{R}^n$ y $T > 0$. Demostrar que el espacio de las combinaciones lineales finales de funciones test en variables separadas de la forma $\varphi(x)\psi(t)$ con $\varphi \in \mathcal{D}(\Omega)$ y $\varphi \in \mathcal{D}(0, T)$ es denso en $\mathcal{D}(\Omega \times (0, T))$.

Ejercicio C.0.21 Consideramos la ecuación elíptica con condiciones de contorno de Neumann:

$$-\Delta u + u = f \quad \text{en} \quad \Omega, \quad \partial u/\partial \nu = g \quad \text{en} \quad \partial\Omega.$$

Mediante ν denotamos la normal exterior unitaria. Suponemos que Ω es un abierto acotado y regular y que $f \in L^2(\Omega)$, $g \in L^2(\partial\Omega)$.

a) Comprobar que las soluciones fuertes satisfacen también la siguiente versión variacional del problema:

$$u \in H^1(\Omega); \int_\Omega [\nabla u \cdot \nabla \varphi + u\varphi]dx = \int_\Omega f\varphi dx + \int_{\partial\Omega} g\varphi d\sigma, \forall \varphi \in H^1(\Omega),$$

donde $d\sigma$ denota la medida superficial.

b) Comprueba que todas las integrales que intervienen en la formulación variacional son finitas.

c) Comprueba que el problema está en las condiciones del lema de Lax-Milgram y deduce la existencia y unicidad de una solución débil.

d) Construye el funcional a minimizar que corresponde a este problema variacional y verifica que está en las condiciones de aplicar el Método Directo del Cálculo de Variaciones.

e) En el caso de una dimensión espacial introduce un esquema iterativo de aproximación por descomposición de dominios y prueba su convergencia. Calcula la velocidad de convergencia.

Suponemos ahora que el dominio es poligonal.

f) Introduce una fomulación aproximada de Galerkin con elementos finitos P1. Verifica la existencia y unicidad de la aproximación.

g) Escribe el problema en la forma matricial

$$RU = M_1 F + M_2 G$$

comentando el siginificado de cada uno de los elementos que aparecen en la misma (vector columna de incógnitas U, de datos F y G, matrices de masa y de rigidez...)

Comenta las diferencias principales con respecto al problema de Dirichlet.

h) Prueba que la sucesión de soluciones aproximadas está acotada en $H^1(\Omega)$ e indica los pasos principales de la prueba de la convergencia en $H^1(\Omega)$ para soluciones débiles y convergencia con tasa de orden uno para las soluciones fuertes en $H^2(\Omega)$.

Ejercicio C.0.22 Consideramos ahora el problema de evolución con condiciones de contorno de Neumann:

$$u_t - \Delta u = 0 \text{ en } \Omega \times (0,\infty),\ \partial u/\partial \nu = 0 \text{ en } \partial\Omega \times (0,T);\ u(x,0) = u_0(x) \text{ en } \Omega.$$

Suponemos que Ω es un dominio acotado de clase C^2.

a) Escribe el problema en la forma de un problema de Cauchy abstracto en $L^2(\Omega)$

$$U_t = AU, \quad t > 0; U(0) = U_0.$$

b) Demuestra que el operador A involucrado es m-disipativo.

c) Identifica el dominio del operador utilizando que las soluciones del problema elíptico del problema anterior están en $H^2(\Omega)$ cuando $g \equiv 0$ y que $f \in L^2(\Omega)$.

d) Deduce un resultado de existencia y unicidad de soluciones débiles y otro de soluciones fuertes.

Suponemos ahora nuevamente que el dominio es poligonal.

e) Introduce una aproximación Galerkin por elementos finitos $P1$. Escríbela en la forma algebráica

$$MU_t = RU.$$

f) Deduce la existencia y unicidad de las soluciones aproximadas.

g) Utilizando el método de la energía obtén una cota de las soluciones aproximadas en $C([0,T]; L^2(\Omega)) \cap L^2(0,T; H^1(\Omega))$.

h) Indica los pasos principales de la prueba de la convergencia del método.

i) Integrando la ecuación en derivadas parciales original prueba que la integral $\int_\Omega u(x,t)dx$ se conserva a lo largo de trayectorias.

j) >Se conserva esta integral para las soluciones aproximadas obtenidas por elementos finitos? Justifica la respuesta.

Ejercicio C.0.23 Consideramos la ecuación elíptica con convección y condiciones de contorno de Dirichlet:

$$-\Delta u + \vec{a} \cdot \nabla u = f + \operatorname{div}(\vec{g}) \quad \text{en} \quad \Omega, \quad u = 0 \quad \text{en} \quad \partial\Omega.$$

En este sistema $\vec{a}$ es un vector fijo arbitrario de $\mathbf{R}^n$, y $\cdot$ denota el producto escalar euclideo.

Suponemos que Ω es un abierto acotado y regular y que $f \in L^2(\Omega)$, $g \in (L^2(\Omega))^n$.

a) Enuncia la formulación variacional del problema para soluciones débiles $u \in H_0^1(\Omega)$.

b) Comprueba que todas las integrales que intervienen en la formulación variacional son finitas.

c) Comprueba que el problema está en las condiciones del lema de Lax-Milgram y deduce la existencia y unicidad de una solución débil.

d) >El problema puede abordarse mediante la minimización de un funcional? Razona la respuesta.

e) Demuestra una cota de la solución que sea independiente del valor del vector de convección $\vec{a}$.

f) En el caso de una dimensión espacial introduce un esquema iterativo de aproximación por descomposición de dominios y prueba su convergencia. Calcula la velocidad de convergencia.

Suponemos ahora que el dominio es poligonal.

g) Introduce una fomulación aproximada de Galerkin con elementos finitos P1. Verifica la existencia y unicidad de la aproximación.

h) Escribe el problema en la forma matricial

$$RU = M_1F + M_2G$$

comentando el significado de cada uno de los elementos que aparecen en la misma (vector columna de incógnitas U, de datos F y G, matrices de masa y de rigidez...)

Comenta las diferencias principales con respecto al caso más habitual en el que $\vec{a} = 0$ y $g \equiv 0$.

i) Prueba que la sucesión de soluciones aproximadas está acotada en $H^1(\Omega)$ e indica los pasos principales de la prueba de la convergencia en $H^1(\Omega)$ para soluciones débiles y convergencia con tasa de orden uno para las soluciones fuertes en $H^2(\Omega)$.

Ejercicio C.0.24 Consideramos ahora la ecuación del calor

$$u_t - \Delta u = 0 \text{ en } \Omega \times (0,\infty),\ u = 0 \text{ en } \partial\Omega \times (0,T);\ u(x,0) = u_0(x) \text{ en } \Omega.$$

Suponemos que Ω es un dominio acotado de clase C^2.

a) Escribe el problema en la forma de un problema de Cauchy abstracto en $L^2(\Omega)$

$$U_t = AU, \quad t > 0; U(0) = U_0.$$

b) Demuestra que el operador A involucrado es m-disipativo.

c) Identifica el dominio del operador.

d) Deduce un resultado de existencia y unicidad de soluciones débiles y otro de soluciones fuertes.

e) Comprueba que las soluciones verifican la identidad de energía:

$$\frac{1}{2}\frac{d}{dt}\int_\Omega u^2 dx = -\int_\Omega |\nabla u|^2 dx.$$

Suponemos ahora nuevamente que el dominio es poligonal.

f) Introduce una aproximación Galerkin por elementos finitos $P1$. Escríbela en la forma algebráica

$$MU_t = RU.$$

g) Deduce la existencia y unicidad de las soluciones aproximadas.

h) Utilizando el método de la energía obtén una cota de las soluciones aproximadas en $C([0,T]; L^2(\Omega)) \cap L^2(0,T; H^1(\Omega))$.

i) Indica los pasos principales de la prueba de la convergencia del método.

k) En el caso de una dimensión espacial indica cuál debería ser el límite de las soluciones del problema continuo cuando $a \to 0$ y cuando $a \to \infty$.

Ejercicio C.0.25 Consideramos la ecuación elíptica con condiciones de contorno de Neumann:

$$-\Delta u = f \quad \text{en} \quad \Omega, \quad \partial u/\partial \nu + u = g \quad \text{en} \quad \partial\Omega.$$

Mediante ν denotamos la normal exterior unitaria y mediante $\partial \cdot /\partial \nu$ la derivada en esa dirección.

Suponemos que Ω es un abierto acotado y regular y que $f \in L^2(\Omega)$, $g \in L^2(\partial\Omega)$.

a) Comprobar que las soluciones fuertes satisfacen también la siguiente versión variacional del problema:

$$u \in H^1(\Omega); \int_\Omega \nabla u \cdot \nabla \varphi dx + \int_{\partial\Omega} u\varphi d\sigma = \int_\Omega f\varphi dx + \int_{\partial\Omega} g\varphi d\sigma, \forall \varphi \in H^1(\Omega),$$

donde $d\sigma$ denota la medida superficial.

b) Comprueba que todas las integrales que intervienen en la formulación variacional son finitas.

c) Prueba la siguiente variante de la desigualdad de Poincaré:

$$\int_\Omega u^2 dx \leq C[\int_\Omega |\nabla u|^2 dx + \int_{\partial\Omega} u^2 d\sigma].$$

d) Utilizando esta desigualdad, comprueba que el problema está en las condiciones del lema de Lax-Milgram y deduce la existencia y unicidad de una solución débil.

e) Construye el funcional a minimizar que corresponde a este problema variacional y verifica que está en las condiciones de aplicar el Método Directo del Cálculo de Variaciones.

f) En el caso de una dimensión espacial introduce un esquema iterativo de aproximación por descomposición de dominios y prueba su convergencia. Calcula la velocidad de convergencia.

Suponemos ahora que el dominio es poligonal.

g) Introduce una fomulación aproximada de Galerkin con elementos finitos P1. Verifica la existencia y unicidad de la aproximación.

h) Escribe el problema en la forma matricial

$$RU = M_1 F + M_2 G$$

comentando el siginificado de cada uno de los elementos que aparecen en la misma (vector columna de incógnitas U, de datos F y G, matrices de masa y de rigidez...)

Comenta las diferencias principales con respecto al problema de Dirichlet.

i) Prueba que la sucesión de soluciones aproximadas está acotada en $H^1(\Omega)$ e indica los pasos principales de la prueba de la convergencia en $H^1(\Omega)$ para soluciones débiles y convergencia con tasa de orden uno para las soluciones fuertes en $H^2(\Omega)$.

Ejercicio C.0.26 Consideramos ahora el problema de evolución con condiciones de contorno de Robin:

$$u_t - \Delta u = 0 \text{ en } \Omega\times(0,\infty),\ \partial u/\partial\nu + u = 0 \text{ en } \partial\Omega\times(0,T);\ u(x,0) = u_0(x) \text{ en } \Omega.$$

Suponemos que Ω es un dominio acotado de clase C^2.

a) Escribe el problema en la forma de un problema de Cauchy abstracto en $L^2(\Omega)$

$$U_t = AU, \quad t > 0; U(0) = U_0.$$

b) Demuestra que el operador A involucrado es m-disipativo.

c) Identifica el dominio del operador utilizando que las soluciones del problema elíptico del problema anterior están en $H^2(\Omega)$ cuando $g \equiv 0$ y que $f \in L^2(\Omega)$.

d) Deduce un resultado de existencia y unicidad de soluciones débiles y otro de soluciones fuertes.

Suponemos ahora nuevamente que el dominio es poligonal.

e) Introduce una aproximación Galerkin por elementos finitos $P1$. Escríbela en la forma algebráica

$$MU_t = RU.$$

f) Deduce la existencia y unicidad de las soluciones aproximadas.

g) Utilizando el método de la energía obtén una cota de las soluciones aproximadas en $C([0,T]; L^2(\Omega)) \cap L^2(0,T; H^1(\Omega))$.

h) Indica los pasos principales de la prueba de la convergencia del método.

i) Integrando la ecuación en derivadas parciales original prueba que la integral $\int_\Omega u(x,t)dx + \int_{\partial\Omega} u(x,t)d\sigma$ se conserva a lo largo de trayectorias.

j) >Se conserva esta integral para las soluciones aproximadas obtenidas por elementos finitos? Justifica la respuesta.

Ejercicio C.0.27 Consideramos la ecuación elíptica con condiciones de contorno de Dirichlet:

$$-\Delta u = f \quad \text{en} \quad \Omega, \quad u = 0 \quad \text{en} \quad \partial\Omega. \tag{C.0.1}$$

Suponemos que Ω es un abierto acotado y regular y que $f \in L^2(\Omega)$.

Consideramos asimismo una aproximación poligonal Ω_h de Ω y una triangulación $\mathcal{T}_h$ que satisface las condiciones de regularidad necesarias para garantizar la convergencia del método de elementos finitos $P1$ con un orden h en $H_0^1(\Omega)$.

Consideramos asimismo el problema de autovalores

$$-\Delta\varphi = \lambda\varphi \quad \text{en} \quad \Omega, \quad \varphi = 0 \quad \text{en} \quad \partial\Omega. \tag{C.0.2}$$

Por la teoría de descomposición espectral de operadores autoadjuntos compactos sabemos que (C.0.2) admite una sucesión λ_j de autovalores positivos que tiende a infinito y una sucesión de autofunciones $\{\varphi_j\}_{j\geq 1}$ que constituye una base ortonormal de $L^2(\Omega)$, a la vez que $\{\varphi_j/\sqrt{\lambda_j}\}_{j\geq 1}$ es una base ortonormal de $H_0^1(\Omega)$.

a) Escribe el problema de autovalores que corresponde a la aproximación por elementos finitos $P1$ de (C.0.2).

b) Prueba que ésta admite N autovalores y autovectores, siendo N el número de nodos interiores de la triangulación. Demuestra que existe una base ortonormal de $\mathbf{R}^N$ constituida por los autovectores. Prueba que la ortogonalidad de los autovectores se verifica tanto en los productos escalares inducidos por las matrices de masa M_h como de rigidez R_h.

A partir de ahora denotamos por λ_j^h y φ_j^h los autovalores y autovectores del problema discreto.

c) Escribe el desarrollo en serie de la solución de (C.0.1) y de su versión discreta utilizando la base espectral.

Comprueba en particular que de este modo se puede recuperar el resultado clásico que garantiza que la solución u de (C.0.1) está en $H_0^1(\Omega)$ cuando f está en $L^2(\Omega)$.

d) Comprueba que la convergencia espectral, i. e. el hecho de que cada autovalor y autovector converja cuando $h \to 0$, permite recuperar la convergencia de las soluciones de la aproximación por elementos finitos de (C.0.1).

e) Demuestra que el pimer autovalor λ_1 se puede caracterizar mediante el cociente de Rayleigh como:

$$\lambda_1 = \min_{\varphi \in H_0^1(\Omega)} \frac{\int_\Omega |\nabla \varphi|^2 dx}{\int_\Omega |\varphi|^2 dx.} \tag{C.0.3}$$

f) Utilizando una caracterización semejante para λ_1^h, prueba que el primer autovalor λ_1^h del problema discreto converge, cuando $h \to 0$, al autovalor λ_1 de la ecuación de Laplace (C.0.1).

g) Pasando al límite en la ecuación que satisface φ_1^h prueba su convergencia fuerte en $H_0^1(\Omega)$ a φ_1. Para ello deduce primero la convergencia débil y después la convergencia de las normas a partir de las ecuaciones correspondientes.

Ejercicio C.0.28 Consideramos ahora el problema de diseño óptimo

$$-\Delta u + \varepsilon u = f \quad \text{en} \quad \Omega, \quad u = 0 \quad \text{en} \quad \partial\Omega, \tag{C.0.4}$$

en el que f es un elemento dado de $L^2(\Omega)$ y ε es un paramétro no-negativo cuyo valor óptimo hemos de determinar de modo que la solución u_ε de (C.0.4) se asemeje lo más posible a una función u_d dada de $H_0^1(\Omega)$.

Para ello introducimos el funcional

$$J(\varepsilon) = \frac{1}{2} \int_\Omega |u - u_d|^2 dx + \frac{\varepsilon^2}{2}.$$

a) Prueba que para cada $\varepsilon \geq 0$, (C.0.4) admite una única solución $u_\varepsilon \in H_0^1(\Omega)$.

b) Prueba que el funcional $J : (0, \infty) \to \mathbf{R}$ es continuo.

c) >Se puede asegurar que J sea convexo?

d) Demuestra que J es coercivo y deduce, aplicando el MDCV, que el funcional alcanza su mínimo en un punto ε^*.

e) >Puede asegurarse que $\varepsilon^* = 0$?

f) Formula la versión elementos finitos $P1$ de este problema de diseño. Demuestra que las soluciones óptimas ε_h^* del problema discreto convergen cuando $h \to 0$ a la solución óptima continua ε^*.

g) Comprueba que la condición $J'(\varepsilon^*) = 0$ que necesariamente se cumple en el punto de mínimo se puede escribir en la forma

$$\int_\Omega (u^* - u_d) v dx + \varepsilon^* = 0,$$

siendo v la solución de

$$-\Delta v + \varepsilon^* v + u^* = 0 \quad \text{en} \quad \Omega, \quad v = 0 \quad \text{en} \quad \partial\Omega. \tag{C.0.5}$$

h) Prueba la existencia y unicidad de la solución v de (C.0.5).

Apéndice D

Soluciones a problemas

21.

a) Si Ω es de clase C^2, con datos f y g suficientemente regulares ($f \in L^2(\Omega)$ y $g \in H^{1/2}(\partial\Omega)$) la solución del problema elíptico u pertenece a $H^2(\Omega)$ como consecuencia de los resultados clásicos de regularidad.

Por consiguiente, utilizando la fórmula de Green de integración por partes obtenemos que, como

$$-\Delta u + u = f \text{ p.c.t. } x \in \Omega; \quad \int_\Omega (-\Delta u + u)\varphi dx = \int_\Omega f\varphi dx.$$

Ahora bien

$$-\int_\Omega \Delta u\varphi dx = \int_\Omega \nabla u \cdot \nabla\varphi dx - \int_{\partial\Omega} \frac{\partial u}{\partial \nu}\varphi d\sigma.$$

La condición de Neumann garantiza que $\partial u/\partial\nu = g$ en $\partial\Omega$, de donde se deduce que

$$-\int_\Omega \Delta u\varphi dx = \int_\Omega \nabla u \cdot \nabla\varphi dx - \int_{\partial\Omega} g\varphi d\sigma.$$

Obtenemos así la formulación débil buscada:

$$\int_\Omega \nabla u \cdot \nabla\varphi dx + \int_\Omega u\varphi dx = \int_\Omega f\varphi dx + \int_{\partial\Omega} g\varphi d\sigma.$$

b) Tenemos, por Cauchy,

$$\begin{aligned}\left|\int_\Omega \nabla u\cdot\nabla\varphi dx\right| &\leqslant \| \nabla u \|_{L^2(\Omega)}\| \nabla\varphi \|_{L^2(\Omega)},\\ \left|\int_\Omega u\varphi dx\right| &\leqslant \| u \|_{L^2(\Omega)}\| \varphi \|_{L^2(\Omega)},\\ \left|\int_\Omega f\varphi dx\right| &\leqslant \| f \|_{L^2(\Omega)}\| \varphi \|_{L^2(\Omega)},\\ \left|\int_{\partial\Omega} g\varphi d\sigma\right| &\leqslant \| g \|_{L^2(\partial\Omega)}\| \varphi \|_{L^2(\partial\Omega)}\ .\end{aligned}$$

Con el objeto de comprobar que la última integral converge basta utilizar el resultado clásico de trazas que garantiza que si $\varphi\in H^1(\Omega)$, entonces $\varphi\Big|_{\partial\Omega}\in H^{1/2}(\partial\Omega)$ y por lo tanto, en particular, $\varphi\Big|_{\partial\Omega}\in L^2(\partial\Omega)$.

c) El problema variacional

$$u\in H^1(\Omega);\quad \int_\Omega\left[\nabla u\cdot\nabla\varphi+u\varphi\right]dx=\int_\Omega f\varphi dx+\int_{\partial\Omega} g\varphi d\sigma,\quad \forall\varphi\in H^1(\Omega),$$

está, en efecto, en las condiciones del Lema de Lax-Milgram.

Basta observar que $H=H^1(\Omega)$ es un espacio de Hilbert, que la forma bilineal

$$a(u,\,v)=\int_\Omega\left[\nabla u\cdot\nabla v+uv\right]dx$$

es en realidad el producto escalar canónico de $H^1(\Omega)$, y que la forma lineal

$$\varphi\in H^1(\Omega)\to\int_\Omega f\varphi dx+\int_{\partial\Omega} g\varphi d\sigma$$

es continua cuando $f\in L^2(\Omega)$ y $g\in L^2(\partial\Omega)$, por las cotas establecidas en el apartado b).

d) El funcional a minimizar para obtener una solución es:

$$\begin{gathered}J\ :\ H^1(\Omega)\to\mathbb{R}\\ J(u)=\frac{1}{2}\int_\Omega\left[|\ \nabla u\ |^2+u^2\right]dx-\int_\Omega fudx-\int_{\partial\Omega} gud\sigma.\end{gathered}$$

En virtud de las estimaciones del apartado anterior sobre las formas bilineales y lineales implicadas en la formulación variacional adaptada al Lema de Lax-Milgram es fácil comprobar que $J\ :\ H^1(\Omega)\to\mathbb{R}$ es una función continua, convexa y coerciva.

De la aplicación directa del Método Directo del Cálculo de Variaciones (MDCV) se deduce que el funcional J alcanza el mínimo en un punto $u \in H^1(\Omega)$.

Además, como J es estrictamente convexo, el punto de mínimo es único.

Por último, en el punto de mínimo se deduce que

$$DJ(u) = 0, \text{ en } \left(H^1(\Omega)\right)'$$

o, lo que es lo mismo,

$$\langle DJ(u),\, \varphi\rangle = 0,\, \forall \varphi \in H^1(\Omega),$$

donde $\langle \cdot,\, \cdot \rangle$ denota el producto de dualidad entre $\left(H^1(\Omega)\right)'$ y $H^1(\Omega)$.

Es fácil comprobar que

$$\langle DJ(u),\, \varphi\rangle = \int_\Omega \left[\nabla u \cdot \nabla \varphi + u\varphi\right] dx - \int_\Omega f\varphi dx - \int_{\partial\Omega} g\varphi d\sigma.$$

Se observa por tanto que $u \in H^1(\Omega)$ es una solución débil.

e) Descomponemos el dominio $\Omega = (0,\, 1)$ en dos subdominios $\Omega_- = (0,\, x_+)$ y $\Omega_+ = (x_-,\, 1)$ donde

$$0 < x_- < x_+ < 1.$$

Utilizando esta descomposición y usando en las interfasses $x = x_-$ y $x = x_+$ la condición de contorno de Dirichlet para transmitir la información de un dominio a otro, podemos construir la siguiente iteración, para $k \geqslant 1$,

$$\begin{cases} -u_-^{k,''} + u_-^k = 0, & 0 < x < x_+ \\ u_-^{k,'}(0) = \alpha \\ u_-^k(x_+) = u_-^{k-1}(x_+) \end{cases}$$

$$\begin{cases} -u_+^{k,''} + u_+^k = 0, & x_- < x < 1 \\ u_+^k(x_-) = u_-^k(x_+) \\ u_+^{k,'}(1) = \beta. \end{cases}$$

Esta iteración puede inicializarse mediante la elección de una función u_0^+ de arrancada arbitraria, que podría se por ejemplo una función parabólica que toma en los extremos del intervalo los valores de Neumann exigidos.

La demostración de la convergencia del método iterativo es entonces idéntica a la del problema de Dirichlet. Basta comprobar que las soluciones de

los problemas de referencia

$$\begin{cases} -W_-^{''} + W_- = 0, \quad 0 < x < x_+ \\ W_-^{'}(0) = 0 \\ W_-(x_+) = 1 \end{cases}$$

$$\begin{cases} -W_+^{''} + W_+ = 0, \quad x_- < x < 1 \\ W_+(x_-) = 1 \\ W_+^{'}(1) = 0 \end{cases}$$

son tales que

$$\text{máx}\Big[\mid W_-(x_-) \mid, \mid W_+(x_+) \mid \Big] < 1.$$

Para comprobar esta última propiedad basta hacerlo para cualquiera de las ecuaciones pues la de la otra es muy similar.

Consideremos el primer caso. Tenemos que

$$W_-(x) = ae^x + be^{-x}.$$

De las condiciones de contorno se deduce que, primero, $a = b$ y, en segundo lugar,

$$W_-(x_+) = a\Big(e^{x_+} + e^{-x_+}\Big) = 1.$$

Por tanto

$$a = \Big(e^{x_+} + e^{-x_+}\Big)^{-1}.$$

Vemos pues que

$$\mid W^-(x_-) \mid = \frac{e^{x_-} + e^{-x_-}}{e^{x_+} + e^{-x_+}} < 1.$$

Esto concluye la convergencia, con velocidad exponencial, del método de descomposición de dominios. Como es habitual, el método acelera su convergencia a medida que la banda intermedia $(x_-,\, x_+)$ de solapamiento de los subdominios aumenta.

f) Dada una familia de subespacios $\Big(V_h\Big)_{h>0}$ de dimensión finita de $H^1(\Omega)$ que satisface la "condición de llenado":

$$\forall v \in H^1(\Omega),\, \exists v_h \in V_h \,:\, v_h \to v \text{ en } H^1(\Omega),\, h \to 0,$$

la aproximación de Galerkin del problema es la siguiente:

$$\begin{cases} u_h \in V_h, \\ \displaystyle\int_\Omega \Big[\nabla u_h \cdot \nabla \varphi_h + u_h \varphi_h\Big] dx + \int_\Omega [f\varphi_h] dx - \int_{\partial\Omega} g\varphi_h d\sigma,\, \forall \varphi_h \in V_h. \end{cases}$$

Por los Teoremas generales probados en clase, a partir de la "condición de llenado" se deduce la convergencia del método.

Con el objeto de describir el método de elementos finitos $P1$ en este caso conviene descomponer la implementación del método del siguiente modo:

- Aproximamos el dominio Ω mediante dominios poligonales Ω_h.

  Una vez realizada esta aproximación en el resto de la construcción suponemos que el dominio Ω es poligonal.

- Introducimos una triangulación $(\mathcal{T}_h)_{h>0}$ del dominio Ω bajo las condiciones habituales que garantizan la convergencia de orden 1 del método en el problema de Dirichlet.

- Fijado h numeramos los nodos del mallado mediante el índice $j = 1, \cdots, N_h$. Introducimos entonces las funciones de base polinomiales $\{\phi_j\}_{j=1,\cdots,N_h}$ de modo que

$$\phi_j(x_k) = \delta_{jk}.$$

- Definimos entonces

$$V_h = \text{span}\,\{\phi_j\}_{j=1,\cdots,N_h}.$$

  El método de elementos finitos $P1$ consiste pues en aplicar el método de Galerkin con esta familia de subespacios.

La diferencia más relevante con respecto al problema de Dirichlet radica que, en la construcción del espacio V_h se tienen en cuenta las funciones de base correspondientes a los nodos del mallado situados en la frontera.

La convergencia del método se prueba de manera idéntica al caso del problema de Dirichlet. Se deduce así que:

- Cuando $u \in H^1(\Omega)$ las soluciones aproximadas $(u_h)_{h>0} \in V_h$ convergen en $H^1(\Omega)$ a la solución u.

- Cuando la solución u es más regular y, más concretamente, $u \in H^2(\Omega)$, la velocidad de convergencia en $H^1(\Omega)$ es de orden uno, $O(h)$.

g) Suponiendo que

$$f(x) = \sum_{j=1}^{N_h} f_j \phi_j(x), \quad g(x) = \sum_{j=1}^{N_h} g_j \phi_j(x),$$

el problema aproximado de Galerkin puede escribirse en la forma

$$RU = M_1 F + M_2 G,$$

con

$$R = \Big(r_{jk}\Big)_{1\leqslant j,\, k\leqslant N_h};\ r_{jk} = \int_\Omega \Big[\nabla\phi_j \cdot \nabla\phi_k + \phi_j\phi_k\Big]dx$$
$$M_1 = \Big(m_{1,\,jk}\Big)_{1\leqslant j,\, k\leqslant N_h};\ m_{1,\,jk} = \int_\Omega \phi_j\phi_k dx,$$
$$M_2 = \Big(m_{2,\,jk}\Big)_{1\leqslant j,\, k\leqslant N_h};\ m_{2,\,jk} = \int_{\partial\Omega} \phi_j\phi_k d\sigma.$$

Las diferencias principales con respecto al problema de Dirichlet son:

- En la base de V_h se introducen los elementos que se apoyan en los nodos que se ubican en la frontera del dominio.
- La matriz R de rigidez es en realidad la suma de las que en el contexto del problema de Dirichlet se denominan matrices de rigidez y de masa.
- La matriz de masa M_1 es la habitual en el método de elementos finitos. La novedad en este caso es la aparición de una segunda matriz de masa M_2 derivada de los términos frontera. Esta matriz M_2 es hueca en el sentido, por ejemplo, en una dimensión espacial, de que sólo dos elementos de esa matriz son no nulos

h) Tal y como hemos indicado anteriormente la prueba es análoga a la del caso de Dirichlet.

22.

a) El espacio de Hilbert ambiente es $H = L^2(\Omega)$. El operador $A = \Delta$ es lineal y no acotado con dominio

$$D(A) = \left\{v \in L^2(\Omega) :\ \Delta v \in L^2(\Omega),\ \partial v/\partial\nu = 0 \text{ en } \partial\Omega\right\}.$$

En estas condiciones el problema se escribe como un problema de Cauchy abstracto

$$\begin{cases} U_t = AU, & t > 0 \\ U(0) = U_0. \end{cases}$$

A este respecto conviene precisar el sentido de la condición de traza nula de la derivada normal.

Cuando Ω es regular de clase C^2, $D(A)$ coincide con

$$D(A) = \left\{ v \in H^2(\Omega) : \partial v \Big/ \partial \nu = 0 \text{ en } \partial\Omega \right\}.$$

En este caso, no hay ninguna ambigüedad en el sentido de la condición de traza nula de la derivada normal puesto que si $v \in H^2(\Omega)$, $\nabla v \in \left(H^1(\Omega)\right)^n$ y por lo tanto $\nabla v\Big|_{\partial\Omega} \in \left(H^{1/2}(\partial\Omega)\right)^n$.
Cuando Ω es menos regular hay que elaborar un poco más el sentido de la condición de traza nula de la derivada normal. Cuando el dominio Ω es de clase C^1 el campo normal $\nu = \nu(x)$ está bien definido y se trata de una función continua definida sobre $\partial\Omega$ y a valores en la esfera unidad de $\mathbb{R}^n$. Por otra parte, cuando Ω es de clase C^1 se verifica la fórmula de Green de modo que, formalmente, se tiene la identidad:

$$\int_\Omega \Delta v \varphi dx = \int_\Omega v \Delta\varphi dx + \int_{\partial\Omega} \left[\frac{\partial v}{\partial \nu}\varphi - v\frac{\partial \varphi}{\partial \nu} \right] d\sigma.$$

En particular, si $\varphi \in C^2(\bar{\Omega})$ y $\partial\varphi/\partial\nu = 0$ sobre $\partial\Omega$ tenemos que

$$\int_\Omega \Delta v \varphi dx = \int_\Omega v \Delta\varphi dx + \int_{\partial\Omega} \frac{\partial v}{\partial \nu}\varphi d\sigma,$$

o, lo que es lo mismo,

$$\int_{\partial\Omega} \frac{\partial v}{\partial \nu}\varphi d\sigma = \int_\Omega \Delta v \varphi dx - \int_\Omega v \Delta\varphi dx$$

de modo que

$$\left| \int_{\partial\Omega} \frac{\partial v}{\partial \nu}\varphi d\sigma \right| \leqslant \| \Delta v \|_{L^2(\Omega)} \| \varphi \|_{L^2(\Omega)} + \| v \|_{L^2(\Omega)} \| \Delta\varphi \|_{L^2(\Omega)}$$

Utilizamos ahora resultados cásicos de trazas que garantizan que si $\rho_1 \in H^{3/2}(\partial\Omega)$ y $\rho_2 \in H^{1/2}(\partial\Omega)$, existe $\varphi \in H^2(\Omega)$ tal que $\varphi\Big|_{\partial\Omega} = \rho_1$ y $\partial\varphi/\partial\nu\Big|_{\partial\Omega} = \rho_2$. En particular podemos tomar $\rho_2 \equiv 0$ y tenemos entonces

$$\| \varphi \|_{H^2(\Omega)} \leqslant C \| \rho_1 \|_{H^{3/2}(\partial\Omega)}.$$

De la identidad anterior deducimos entonces que si $v \in L^2(\Omega)$ y $\Delta v \in L^2(\Omega)$, $\partial v \Big/ \partial\nu$ define una forma lineal continua sobre $H^{3/2}(\partial\Omega)$. La traza de $\partial v \Big/ \partial\nu$ sobre la frontera es por tanto un elemento de $H^{-3/2}(\partial\Omega)$. Esto da sentido a la traza de la derivada normal.

b) • A es disipativo.

Basta observar que si $v \in D(A)$,

$$(Av,\, v)_{L^2(\Omega)} = \int_\Omega \Delta v v dx = -\int_\Omega \mid \nabla v \mid^2 dx + \int_{\partial\Omega} \frac{\partial v}{\partial \nu} v d\sigma = -\int_\Omega \mid \nabla v \mid^2 dx \leqslant 0$$

puesto que $\partial v \Big/ \partial \nu = 0$ en $\partial\Omega$

• A es maximal.

Dado $f \in L^2(\Omega)$ hemos de probar la existencia de una única solución de la ecuación

$$(I - A)v = f,\, v \in D(A).$$

Este es precisamente el problema considerado en el ejercicio anterior

$$\left\{ \begin{array}{ll} -\Delta v + v = f & \text{en} \quad \Omega \\ \partial v \Big/ \partial \nu = 0 & \text{en} \quad \partial\Omega. \end{array} \right.$$

Aplicando el Lema de Lax-Milgram deducimos la existencia de una única solución débil $v \in H^1(\Omega)$. Es fácil comprobar que $v \in D(A)$.

c) Tal y como hemos visto anteriormente, cuando el dominio Ω es de clase C^2, el dominio del operador $D(A)$ coincide con:

$$D(A) = \left\{ v \in H^2(\Omega);\, \partial v \Big/ \partial \nu = 0 \text{ en } \partial\Omega \right\}.$$

d) Aplicando el Teorema de Hille-Yosida deducimos inmediatamente los dos siguientes resultados:

• Soluciones débiles.

Si $u_0 \in L^2(\Omega)$, existe una única solución débil $u \in C\Big([0,\, \infty);\, L^2(\Omega)\Big)$. Además $\nabla u \in L^2(\Omega \times (0,\, \infty))$.

• Soluciones fuertes.

Cuando $u_0 \in D(A)$ existe una única solución $u \in C([0,\, \infty);\, D(A)) \cap C^1\Big([0,\, \infty);\, L^2(\Omega)\Big)$. En particular, si Ω es de clase C^2, por la caracterización del dominio del operador $D(A)$, tenemos que si $u_0 \in H^2(\Omega)$ y $\partial u_0 / \partial \nu = 0$ en $\partial\Omega$, entonces existe una única solución

$$u \in C\Big([0,\, \infty);\, H^2(\Omega)\Big) \cap C^1\Big([0,\, \infty);\, L^2(\Omega)\Big).$$

e) Tal y como estudiamos en el curso, las soluciones débiles pueden caracterizarse del siguiente modo:

$$\begin{cases} u \in C\Big([0,\,\infty);\, L^2(\Omega)\Big);\ \nabla u \in L^2(\Omega \times (0,\,\infty)) \\ \dfrac{d}{dt}\displaystyle\int_\Omega u(x,\,t)\varphi(x)dx = \int_\Omega \nabla u(x,\,t)\cdot\nabla\varphi(x)dx, \qquad \text{p.c.t. } t>0, \\ \displaystyle\int_\Omega u(x,\,0)\varphi(x)dx = \int_\Omega u_0(x)\varphi(x)dx. \end{cases}$$

Con el objeto de introducir la aproximación por elementos finitos $P1$, utilizamos la familia de espacios aproximantes $(V_h)_{h>0}$ del problema anterior. La aproximación de Galerkin del problema es entonces la siguiente:

$$\begin{cases} u_h \in C([0,\,\infty);\, V_h);\ \nabla u_h \in L^2(\Omega \times (0,\,\infty)) \\ \dfrac{d}{dt}\displaystyle\int_\Omega u_h(x,\,t)\varphi_h(x)dx = \int_\Omega \nabla u_h(x,\,t)\cdot\nabla\varphi_h(x)dx,\ \forall\varphi_h \in V_h,\ \forall t>0, \\ \displaystyle\int_\Omega u_h(x,\,0)\varphi_h(x)dx = \int_\Omega u_0(x)\varphi_h(x)dx. \end{cases}$$

Es fácil comprobar que el problema puede escribirse en la forma matricial:

$$MU_{h,\,t} = -RU_h,$$

donde R es la matriz de rigidez introducida en el problema anterior y $M = M_1$, siendo M_1 la primera matriz de masa del problema anterior.

f) Como R y M son matrices simétricas definidas positivas, el problema aproximado tiene una única solución que viene dada por la representación exponencial

$$U_h(t) = e^{-M^{-1}Rt}U_0.$$

Obviamente

$$U_h \in C^\omega([0,\,\infty);\, V_h).$$

g) Utilizando el método de la energía, consistente en justificar la utilización de la función test $\varphi_h = u_h(\cdot,\,t)$ en la formulación de Galerkin del problema aproximado obtenemos que

$$\frac{1}{2}\frac{d}{dt}\int_\Omega u_h^2(x,\,t) = -\int_\Omega |\ \nabla u_h(x,\,t)\ |^2\,dx,$$

de donde deducimos que

$$\frac{1}{2}\parallel u_h \parallel^2_{L^\infty(0,\,\infty;\,L^2(\Omega)} + \parallel \nabla u_h \parallel^2_{L^2(\Omega\times(0,\,\infty))} \leqslant \frac{1}{2}\parallel u_0 \parallel^2_{L^2(\Omega)},$$

lo cual proporciona las estimaciones uniformes buscadas.

Conviene en este punto recordar que las soluciones del problema de Galerkin son de la forma

$$u_h(x,\,t) = \sum_{j=1}^{N_h} u_j(t)\phi_j(x)$$

donde las componentes $\{u_j(t)\}_{1\leqslant j\leqslant N_j}$ son precisamente las de la incógnita vectorial $U_h(t)$ que interviene en la representación matricial del problema finito-dimensional.

h) Los pasos principales de la prueba de la convergencia son las mismas que en el caso del problema de Dirichlet.

• Extrayendo subsucesiones, que podemos seguir denotando mediante el parámetro h, tenemos

$$\begin{cases} u_h \rightharpoonup v & \text{en} \quad L^\infty\Big(0,\,\infty;\,L^2(\Omega)\Big) \text{ débil} - * \\ \nabla u_h \rightharpoonup \nabla v & \text{en} \quad L^2(\Omega\times(0,\,\infty)). \end{cases}$$

• El problema se reduce a probar que v es la solución débil del problema continuo. En efecto, en ese caso tendríamos que $v = u$ y toda la sucesión u_h convergería a u en el sentido anterior.

• Con el objeto de comprobar que v es la solución débil del problema continuo basta pasar al límite en la formulación de Galerkin del problema aproximado. En este punto la hipótesis de que los subespacios $(V_h)_{h>0}$ llenan el espacio $H^1(\Omega)$ juega un papel esencial.

Dado $\varphi \in H^1(\Omega)$ tenemos una sucesión $\varphi_h \in V_h$ tal que

$$\varphi_h \to \varphi \text{ en } H^1(\Omega).$$

Esta convergencia, junto con la convergencia de las funciones u_h garantiza que

$$\int_\Omega u_h(x,\,t)\varphi_h(x)dx \to \int_\Omega v(x,\,t)\varphi(x)dx \text{ en } L^\infty(0,\,\infty) - \text{débil} - *$$

y

$$\int_\Omega \nabla u_h(x,\,t)\cdot\nabla\varphi_h(x)dx \to \int_\Omega \nabla v(x,\,t)\cdot\nabla\varphi(x)dx \text{ débilmente en } L^2(0,\,\infty).$$

Tenemos así que el límite $v = v(x,\,t)$ satisface

$$\frac{d}{dt}\int_\Omega v(x,\,t)\varphi(x)dx + \int_\Omega \nabla u(x,\,t)\cdot\nabla\varphi(x)dx = 0, \text{ en } \mathcal{D}'(0,\,\infty),\ \forall\varphi \in H^1(\Omega).$$

Con el objeto de garantizar que v es la solución débil hemos de probar que:

- $v \in C([0,\, \infty);\, L^2(\Omega))$
- $v(0) = u_0$ en Ω.

La primera propiedad de continuidad en el tiempo no se obtiene en el proceso de paso al límite que sólo asegura que $v \in L^\infty(0,\, \infty;\, L^2(\Omega))$. Para probar la continuidad en tiempo basta proceder como en el curso. Esto es efectivamente consecuencia del hecho que

$$v \in L^2(0,\, T;\, H^1(\Omega))$$

y que

$$v_t \in L^2\Big(0,\, T;\, (H^1(\Omega))'\Big).$$

Esta última consecuencia de la primera $\Big(v \in L^2(0,\, T;\, H^1(\Omega))\Big)$ y de que v sea solución de la ecuación del calor en el sentido de las distribuciones, i.e.

$$v_t = \Delta v \text{ en } \mathcal{D}'(\Omega \times (0,\, \infty))$$

cosa que se deduce del hecho que v satisfaga la formulación débil de Galerkin del problema continuo.

Una vez que sabemos que $v \in C([0,\, \infty);\, L^2(\Omega))$ tiene sentido la traza de la solución en el instante inicial $t = 0$, $v(x,\, 0)$. Para ver que

$$v(x,\, 0) = u_0(x) \text{ p.c.t. } x \in \Omega$$

podemos proceder como en el curso, usando una formulación variacional de la solución mediante funciones test que dependan de x y de t y que incorpore el dato inicial del problema.

(i) Integrando la ecuación continua tenemos

$$\frac{d}{dt}\int_\Omega u(x,\, t)dx = \int_\Omega \Delta u(x,\, t)dx = \int_{\partial\Omega} \frac{\partial u}{\partial \nu}(x,\, t)d\sigma = 0$$

de donde se deduce que la integral $\displaystyle\int_\Omega u(x,\, t)dx$ de la solución se conserva en el tiempo.

En el caso del problema aproximado, para que esta propiedad se cumpla, es imprescindible que la función test $\varphi_h \equiv 1$ pertenezca al espacio aproximante V_h. Esto ocurre en el caso de los elementos finitos $P1$ considerados

puesto que la función constante $\varphi \equiv 1$ pertenece al espacio V_h pues se trata de una función continua y constante a trozos sobre los triángulos del mallado.

Bibliografía

[1] Bender, C.M. and Orszag, S.A. (1978). *Advanced Mathematical Methods for Scientists and Engineers*, McGraw-Hill.

[2] Brezis, H. (1983). *Analyse Fonctionnelle, Théorie et Applications*, Masson, Paris.

[3] Cazenave, T. and Haraux A. (1989), *Introduction aux problèmes d'évolution semilinéaires*, Mathématiques & Applications, Ellipses, Paris.

[4] Chorin, A. & J.; Marsden, J. E. (1993), *A mathematical introduction to fluid mechanics*. Third edition. Texts in Applied Mathematics, 4. Springer-Verlag, New York.

[5] Cohen, G. (2001). *Higher-order numerical methods for transient wave equations*. Scientific Computation, Springer.

[6] Courant, R. and Hilbert, D. (1989). *Methods of Mathematical Physics*, John Wiley & Sons.

[7] Eastham, M.S.P. (1973). *The Spectral Theory of Periodic Differential Equations*, Scottish Academic Press, Edinburgh.

[8] Evans, L. C. (1998). *Partial Differential Equations*, Graduate Studies in Mathematics, Vol.19, AMS.

[9] Glowinski, R. (1992). "Ensuring well-posedness by analogy; Stokes problem and boundary control of the wave equation". *J. Compt. Phys.*, **103**(2), 189–221.

[10] Glowinski, R., Li, C. H. and Lions, J.-L. (1990). "A numerical approach to the exact boundary controllability of the wave equation (I). Dirichlet controls: Description of the numerical methods". *Japan J. Appl. Math.*, **7**, 1–76.

[11] E. Godlewski y P. A. Raviart , (1987). "Hyperbolic Systems of Conservation Laws", Ellipses, Paris.

[12] Infante, J.A. and Zuazua, E. (1999). "Boundary observability for the space-discretizations of the one-dimensional wave equation", *Mathematical Modelling and Numerical Analysis,* **33**, 407–438.

[13] Isaacson, E. and Keller, H.B. (1966). *Analysis of Numerical Methods*, John Wiley & Sons.

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

[15] John, F. (1982) *Partial differential Equations,* (4. ed), Springer.

[16] LeVeque, R. J. (1992). *Numerical methods for conservation laws*. Second edition. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel.

[17] J.L. Lions, *Quelques méthodes de résolution des problèmes aux limites non linéaires*. Dunod; Gauthier-Villars, Paris, 1969.

[18] Negreanu, M. and Zuazua, E. (2003). "Uniform boundary controllability of a discrete 1D wave equation. *Systems and Control Letters*, **48** (3-4) , 261-280.

[19] Quarteroni A. y Valli, A. (1998). *Numerical approximation of Partial differential Equations*, Springer, Springer Series in Computational Mathematics, 23.

[20] Quarteroni A. y Valli, A. (1999). *Domain Decomposition Methods for Partial differential Equations*, Oxford Science Publications.

[Q] F. Quirós, *Notas de Cálculo Numérico.* 2002.

[21] Reed M. y Simon B. (1978) *Modern Methods of Mathematical Physics. Vol. 1. Functional Analysis*, Academic Press.

[22] Sethian, J. A. (2002), *Level set methods and fast marching methods*. Cambridge Monographs in Applied and Computational Mathematics, Cambridge University Press.

[23] Strikwerda, J. C. (1989) *Finite Difference Schemes and Partial Differential Equations*, Wadsworth & Brooks, California.

[24] Rauch, J. (1991) *Partial Differential Equations*, Graduate Texts in Mathematics, Springer Verlag.

[25] Sanz-Serna, J. (1985). "Stability and convergence in numerical analysis. I: linear problems—a simple, comprehensive account". *Res. Notes in Math.*, **132**, Pitman, Boston, MA, pp. 64–113.

[26] Temam, R. (1977). *Theory and Numerical Analysis of the Navier-Stokes equations.* North-Holland.

[27] Trefethen, L. N. (1982). "Group velocity in finite difference schemes", *SIAM Rev.,* **24** (2), pp. 113–136.

[28] Vázquez, J. L. (2003) *Fundamentos Matemáticos de las Mecánica de Fluidos.*

(http://www.uam.es/juanluis.vazquez)

[29] Vichnevetsky, R. and Bowles, J.B. (1982). *Fourier Analysis of Numerical Approximations of Hyperbolic Equations.* SIAM Studies in Applied Mathematics, **5**, SIAM, Philadelphia.

[30] Whitham, G. B. (1999), *Linear and nonlinear waves.* John Wiley & Sons, Inc., New York.

[31] Young, R. M. (1980). *An Introduction to Nonharmonic Fourier Series*, Academic Press, New York.

[32] O. C. Zienkiewicz, Achievements and some unsolved problems of the finite element method, Int. J. Numer. Meth. Engng. **47** (2000), 9–28.

[33] Zuazua, E. (1999). "Boundary observability for the finite-difference space semi-discretizations of the 2D wave equation in the square", *J. Math. Pures et Appliquées*, **78**, 523–563.

[34] Zuazua, E. (2002). "Controllability of Partial Differential Equations and its Semi-Discrete Approximations". *Discrete and Continuous Dynamical Systems*, **8** (2), 469–513.

[35] Zuazua, E. (1999). "Observability of 1D waves in heterogeneous and semi-discrete media". *Advances in Structural Control.* J. Rodellar et al., eds., CIMNE, Barcelona, pp. 1–30.

[36] Zuazua, E. (2003). Ecuaciones en Derivadas Parciales,

`http://www.uam.es/personal-pdi/ciencias/ezuazua/docencia.html`.