

---

# Interpolación polinómica de alto orden Métodos espectrales

## Aplicación a problemas de contorno y de condiciones iniciales

*Juan A. Hernández Ramos*  
*Mario A. Zamecnik Barros*

*Departamento de Matemática Aplicada a la Ingeniería Aeroespacial*  
*Escuela Técnica Superior de Ingeniería Aeronáutica y del Espacio*  
*Universidad Politécnica de Madrid*

---

Portada:

Conquistar la cima de tu propósito, requiere confianza y perseverancia.

“Ascensión al Pico del Lobo” (febrero de 2019).

Fotografía y diseño de cubierta por Belén Moreno Santamaría.

Queda prohibida la reproducción de cualquier parte del texto por cualquier medio, incluido fotocopia, sin permiso escrito del autor.

2019 Primera edición

2020 Segundo edición. Revisión y complementos

2022 Tercera edición. Revisión y complementos.

© Juan A. Hernández Ramos, Mario A. Zamecnik Barros.

ISBN 978-1076625595

# Índice general

<b>Prefacio</b>	<b>III</b>
<b>1 Problemas objeto de estudio</b>	<b>1</b>
1.1 Introducción . . . . .	1
<b>2 Aproximación de funciones</b>	<b>5</b>
2.1 Introducción . . . . .	5
2.2 Serie truncada y precisión espectral . . . . .	7
2.3 Serie discreta o interpolante . . . . .	8
2.4 Serie discreta y serie truncada . . . . .	10
2.5 Fórmulas de cuadratura de Gauss . . . . .	12
2.6 Fórmulas de cuadratura de Gauss–Lobatto . . . . .	13
2.7 Puntos de interpolación y bases ortogonales . . . . .	14
<b>3 Interpolación polinómica</b>	<b>17</b>
3.1 Introducción . . . . .	17
3.2 Matriz de Vandermonde . . . . .	18
3.3 Interpolación de Lagrange . . . . .	20
3.4 Método recursivo . . . . .	22
3.5 Forma de Newton . . . . .	26
3.6 Error de interpolación . . . . .	32
3.6.1 Error de truncamiento . . . . .	33
3.6.2 Error de redondeo . . . . .	36
3.6.3 Acotación del error de truncamiento y redondeo . . . . .	37
3.7 Distribuciones equiespaciadas de puntos . . . . .	39
3.7.1 Error de truncamiento . . . . .	39
3.7.2 Error de redondeo . . . . .	44
3.8 Distribuciones de puntos no equiespaciados . . . . .	46
<b>4 Interpolantes de Fourier y Chebyshev</b>	<b>47</b>
4.1 Interpolante de Fourier . . . . .	48
4.2 Derivada del interpolante de Fourier . . . . .	49
4.3 Transformada rápida de Fourier . . . . .	50
4.4 Interpolante de Fourier en dos dimensiones . . . . .	52
4.5 Interpolantes de Chebyshev . . . . .	53

4.5.1	Ceros de Chebyshev . . . . .	54
4.5.2	Extremos de Chebyshev . . . . .	54
4.6	Derivada del interpolante de Chebyshev . . . . .	56
4.7	Transformada rápida coseno mediante la FFT . . . . .	57
4.8	Error y puntos de Chebyshev . . . . .	59
<b>5</b>	<b>Interpolación polinómica continua a trozos</b>	<b>61</b>
5.1	Fórmulas para la derivada primera y segunda . . . . .	63
5.2	Error de las derivadas primera y segunda . . . . .	65
5.3	Fórmulas para derivadas con tres puntos . . . . .	66
5.4	Error de interpolación con tres puntos . . . . .	67
5.4.1	Error de truncamiento . . . . .	67
5.4.2	Error de redondeo . . . . .	69
5.4.3	Error de interpolación: truncamiento y redondeo . . . . .	70
<b>6</b>	<b>Problema de contorno</b>	<b>73</b>
6.1	Ecuaciones diferenciales ordinarias . . . . .	73
6.2	Dominios bidimensionales . . . . .	77
<b>7</b>	<b>Problema de Cauchy en EDOS</b>	<b>79</b>
7.1	Métodos multipaso. Métodos Adams . . . . .	80
7.2	Métodos unipaso. Métodos Runge Kutta . . . . .	81
7.3	Error global y error local de truncamiento . . . . .	82
7.4	Acotación del error y estabilidad numérica . . . . .	84
7.5	Ejemplo: Integración del oscilador armónico . . . . .	87
<b>8</b>	<b>Problema de condiciones iniciales y de contorno</b>	<b>91</b>
8.1	Introducción . . . . .	91
8.2	Discretización espacial y temporal . . . . .	92
8.3	Error espacial y error temporal . . . . .	98
8.4	Error de la semidiscretización espacial . . . . .	99
8.5	Error en la ecuación del calor . . . . .	101
8.5.1	Acotación del error espacial . . . . .	102
8.5.2	Acotación del error temporal . . . . .	103
8.6	Extrapolación de Richardson . . . . .	105
<b>A</b>	<b>Nomenclatura</b>	<b>107</b>
	<b>Bibliografía</b>	<b>108</b>

# Prefacio

El uso de esquemas numéricos de alto orden para la simulación o integración de problemas físicos es hoy en día un tema de mucha importancia. Las dos técnicas mayormente empleadas son: métodos espectrales e interpolación polinómica de alto orden. Sin embargo, existen problemas asociados al mal condicionamiento numérico de la semidiscretización espacial cuando se utilizan interpolantes de alto orden.

El objetivo de este libro es poner de manifiesto la importancia de la interpolación en el resultado numérico obtenido por los diferentes esquemas numéricos o métodos. En concreto, se muestra que la concentración de la nube de puntos en los extremos del intervalo o dominio de integración es vital para el éxito de los métodos de interpolación de alto orden. Aunque todo el material que aparece en este libro es bien conocido, los autores han querido recoger de forma didáctica y ordenada algunos de los aspectos más relevantes para entender el origen de los problemas en la integración con los esquemas de alto orden.

En el capítulo 1 se hace una clasificación de los problemas objeto de estudio. En los problemas de evolución, la variable independiente temporal permite clasificar matemáticamente los problemas en: hiperbólicos o parabólicos. La evolución temporal de estos sistemas permite obtener la solución numérica aproximada en función de instantes anteriores. Es decir, a partir de la condición inicial, se puede determinar la solución aproximada en el siguiente paso de integración y así sucesivamente. Cuando los problemas no son de evolución, generalmente, son problemas elípticos en los que la información se propaga de manera instantánea y los algoritmos de cálculo son completamente diferentes. En concreto, una vez discretizado el problema espacialmente, la solución se debe obtener mediante la resolución de un sistema lineal o no lineal de ecuaciones. Aunque gran parte del contenido de este libro está dedicado a la interpolación, el principal objetivo de este capítulo es poner el foco en la simulación de los problemas gobernados por sistemas de ecuaciones con condiciones iniciales y de contorno.

En el capítulo 2 se hace una revisión somera de las técnicas de aproximación en espacios de dimensión finita e infinita. Se define el conjunto de funciones base del espacio junto con su producto interno. El uso del ordenador para la aproximación de funciones exige siempre el tratamiento en espacios de dimensión finita. Es por

esta razón que la idea principal de este capítulo es relacionar el error de aproximación de una serie discreta vinculada a un espacio de dimensión finita y una serie truncada vinculada a un espacio de dimensión infinita. La serie discreta se obtiene haciendo pasar el interpolante por una serie de puntos discretos o de colocación. El mayor atractivo de los métodos espectrales o desarrollos en serie en espacios de dimensión infinita reside en la precisión espectral. La precisión espectral está asociada a desarrollos de funciones infinitamente derivables mediante ciertas bases ortogonales. La idea principal de la precisión espectral es que mediante unos cuantos grados de libertad o términos del desarrollo en serie, el error de aproximación se hace extremadamente pequeño. Por esta razón, se busca que la aproximación mediante una serie discreta en un espacio de dimensión finita goce de propiedades similares a la precisión espectral. Para ello, se revisan las fórmulas de cuadratura de Gauss que permiten realizar una integral de manera exacta de ciertos polinomios mediante un conjunto de puntos de colocación. Si se eligen estos puntos de colocación de las cuadraturas de Gauss como puntos de interpolación en el espacio de dimensión finita, se demuestra que el producto interno coincide con el producto interno en dimensión infinita y la propiedad de precisión espectral se puede llevar a los espacios de dimensión finita. Es decir, siempre que la nube de puntos de interpolación o discretización se corresponda con los puntos gaussianos, la aproximación con una serie discreta y con una serie truncada se comportan asintóticamente con el mismo error.

En el capítulo 3 se hace un estudio en detalle de las diferentes técnicas de interpolación polinómica: método de Vandermonde, interpolación de Lagrange y fórmula de Newton. A continuación, se define el error de interpolación como la suma del error de truncamiento más el error de redondeo y se entra en el origen de ambos errores. Se demuestra que el error de truncamiento está gobernado por la función de error  $\pi(x)$  y se analiza su forma con la distribución de la nube de puntos. Para distribuciones con puntos equiespaciados, se observa la presencia de un máximo de la función  $\pi(x)$  próximo a los extremos del intervalo de aproximación. De igual forma, el error de redondeo se asocia a la función de Lebesgue que presenta un máximo cerca de los extremos del intervalo. Sin embargo, mientras que la función  $\pi(x)$  tiende a cero con el número de puntos  $N$  tendiendo a infinito, la función de Lebesgue no tiende a cero con  $N$ . Éste es el principal problema de la interpolación de alto orden en distribuciones de puntos arbitrarias. Para paliar este problema, se proponen distribuciones de puntos no equiespaciados que arrojan valores moderados para la el máximo de la función de Lebesgue o constante de Lebesgue.

El capítulo 4 está dedicado a los interpolantes de Chebyshev o la serie discreta de Chebyshev. La motivación de los puntos de Chebyshev pasa por buscar distribuciones de puntos que hagan uniforme los máximos o los extremos de la función  $\|\pi(x)\|$ . Se definen los polinomios de Chebyshev como  $T_k(x) = \cos(k\theta)$  con  $x = \cos(\theta)$ , y se demuestra que los ceros de Chebyshev o los puntos que hacen que  $T_{N+1}(x) = 0$  verifican que los máximos  $\|\pi(x)\|$  son todos iguales. Por otra parte,

---

se sabe que estos puntos coinciden con los puntos de la cuadratura de Gauss que hace que la serie discreta de Chebyshev y la serie truncada se comporten asintóticamente con el mismo error. Sin embargo, los ceros de Chebyshev no incluyen los extremos del intervalo de integración  $[-1, +1]$  y surge la definición de los extremos de Chebyshev. Aunque estos puntos no verifican que los máximos de la función  $\pi(x)$  sean uniformes, su comportamiento es mucho mejor que el comportamiento de los puntos equiespaciados. Además, la principal ventaja de los extremos de Chebyshev es que coinciden con los puntos Gauss-Lobatto que permiten realizar de forma exacta la integral para ciertos polinomios. Esta propiedad hace que el producto interno en el espacio de dimensión finita y dimensión infinita coincidan y que el error de la serie truncada y discreta tenga el mismo comportamiento. Por otra parte, en este capítulo se expone la posibilidad de representar la función mediante los coeficientes de la serie discreta o representación en el plano espectral o mediante los valores de la función en los nodos o puntos de Chebyshev o representación en el plano físico. El interés de ambas representaciones reside en la existencia de una transformada rápida entre el plano espectral y el plano físico que involucra  $O(N \log_2 N)$  operaciones. Mientras que el cálculo de una derivada espacial en el plano físico cuesta  $O(N^2)$  operaciones, el cálculo de la misma derivada en el plano espectral cuesta  $O(N)$  operaciones. Por lo tanto, aunque tengamos que hacer un par de transformadas del plano físico al espectral y del espectral al físico, el cálculo de la derivada espectral en el plano espectral es siempre mas atractivo desde el punto de vista computacional. Por el contrario, el cálculo de un término no lineal en el plano espectral cuesta  $O(N^2)$  operaciones mientras que en el plano físico cuesta tan solo  $O(N)$  operaciones. Por estas razones, los esquemas numéricos utilizan ambas representaciones física y espectral para realizar los cálculos con la mínima carga computacional.

En el capítulo 5 se define la interpolación continua a trozos. Es decir, se fija el grado del interpolante  $q$  y, en consecuencia, el número de puntos  $q + 1$  por los que pasa el interpolante. De esta forma, el interpolante del dominio de integración está definido por un conjunto de funciones polinómicas de grado  $q$ . Desde el punto de vista intuitivo, el comportamiento de la función en un punto depende de unos cuantos puntos vecinos a la derecha y a la izquierda de este punto. Se usa la teoría de interpolación desarrollada en el capítulo 3 para determinar el error de truncamiento y el error de redondeo. Cuando se interpola una función dada en una nube de  $N + 1$  puntos mediante interpolantes continuos a trozos de grado  $q$ , el error de truncamiento es  $O(\Delta x^{q+1})$  siendo  $\Delta x$  la distancia entre los puntos nodales. En este caso, se dice que el método de aproximación tiene una convergencia algebraica que es muy inferior a la convergencia espectral. En este capítulo se obtienen las fórmulas clásicas en diferencias finitas con tres puntos ( $q = 2$ ) para las derivadas primera y segunda y se analiza el error de truncamiento y redondeo. Finalmente, se estudia el paso  $\Delta x$  óptimo que minimiza el error total. Como el error de truncamiento tiende a cero cuando  $\Delta x$  tiende a cero y el error de redondeo tiende a infinito cuando  $\Delta x$  tiende a cero, existe un paso óptimo que minimiza el error. Si nos empeñáramos en disminuir el  $\Delta x$  por debajo de este valor óptimo, el

error en la derivada primera o en la derivada segunda aumentaría al aumentar el error de redondeo.

Una vez entendido en detalle el error de truncamiento y el error de redondeo en el cálculo de derivadas de interpolantes, se procede al uso de estos interpolantes globales o continuos a trozos para resolver problemas de contorno, problemas de condiciones iniciales y problemas de condiciones iniciales y de contorno.

En el capítulo 6 se aborda el problema de contorno en ecuaciones diferenciales ordinarias y ecuaciones en derivadas parciales. Se plantea el algoritmo numérico para resolver estos problemas. El resultado de la discretización espacial conduce a un conjunto de ecuaciones lineales o no lineales para la variable dependiente en el conjunto de puntos nodales o puntos de interpolación. La extensión a problemas bidimensionales se hace mediante el producto tensorial de dos interpolantes unidimensionales continuos a trozos o globales.

En el capítulo 7 se trata el problema de condiciones iniciales o de Cauchy en ecuaciones diferenciales ordinarias. Mediante interpolantes para la función  $F$  del problema de Cauchy, se obtienen los métodos de Adams cuyos coeficientes son integrales de la funciones de Lagrange. Se define el error global y el error local de truncamiento de un esquema temporal y se obtiene la ecuación en diferencias que gobierna el error. Cuando el problema de Cauchy es lineal y de coeficientes constantes, se puede resolver analíticamente la ecuación en diferencias del error. La acotación del error para un número de pasos de integración fijado permite definir el radio espectral de la matriz del sistema que determina la estabilidad de la solución del error. Se demuestra que si el radio espectral es menor que uno, el error global de la solución está acotado por el error local de truncamiento. Si el radio espectral es mayor que uno, el error tiende a infinito con el número de pasos tendiendo a infinito. Sin embargo, la determinación del radio espectral es complicada y mediante el teorema de transformación espectral y la definición de la región de estabilidad absoluta se puede determinar de una manera muy rápida y efectiva si el radio espectral es mayor o menor que la unidad. Si los autovalores de la matriz del sistema multiplicados por el paso temporal  $\Delta t$  están dentro de la región de estabilidad absoluta, entonces el radio espectral es menor que uno y podemos esperar que el error esté acotado por el error local de truncamiento. Si embargo, si los autovalores multiplicados por el  $\Delta t$  están fuera de la región de estabilidad absoluta, no tenemos ninguna esperanza de obtener una solución numérica.

En el capítulo 8 se aborda el problema de condiciones iniciales y de contorno. En este caso, la discretización de las variables espaciales se trata de forma independiente a la discretización de la variable temporal. Primero se realiza la discretización espacial y se reduce el problema a un conjunto de ecuaciones diferenciales con condiciones iniciales mediante las técnicas aprendidas en el capítulo 6. A este procedimiento se le conoce con el nombre del método de las líneas. Se define el error de la semidiscretización espacial y se plantea la ecuación diferencial ordinaria para

---



la evolución temporal del error espacial. Cuando la ecuación en derivadas parciales es lineal de coeficientes constantes, esta ecuación se puede resolver analíticamente y la acotación de la solución para un tiempo de integración dado permite definir la abscisa espectral de la matriz del sistema que determina la estabilidad de la solución. Si la abscisa espectral es mayor que cero, el problema está mal condicionado y el error espacial crece exponencialmente con el tiempo. Sin embargo, si la abscisa espectral es menor que cero, el error espacial está acotado por el error de truncamiento de la semidiscretización espacial. Generalmente, cuando se discretiza la ecuación de ondas mediante esquemas centrados de diferencias finitas de alto orden en mallas equiespaciadas, se obtienen autovalores que tienen parte real mayor que cero que dan origen a una abscisa espectral mayor que cero. Esta problemática está asociada al mal comportamiento de los interpolantes de alto orden en los contornos cuando la malla es equiespaciada. Este problema se suele resolver concentrando puntos en los contornos como lo hacen los puntos de Chebyshev.

En resumen, este libro trata de dar pistas y de entender el origen de los problemas mas comunes que nos podemos encontrar en el tratamiento con el ordenador de los problemas gobernados por ecuaciones diferenciales. Por otra parte, este libro trata de fusionar el mundo de la interpolación en mallas no equiespaciadas con los métodos espectrales o desarrollos en serie discretos. Desde el punto de vista de los autores, el uso del ordenador es imprescindible para entender en profundidad los retos aquí planteados. En esa línea de trabajo, el libro del mismo autor *“How to learn Applied Mathematics through modern FORTRAN”* trata de fusionar el mundo de la programación con el mundo del cálculo numérico a través de abstracciones de alto nivel utilizando un paradigma funcional. La experiencia en las aulas nos ha enseñado que la mejor manera de entender un operador diferencial o la influencia de una condición de contorno es en el ordenador. Esperemos que este libro permita al lector crear sus propios algoritmos basados en los pilares básicos de la interpolación polinómica y su estrecha relación con los métodos espectrales.

Juan A. Hernández  
Madrid  
Julio 2019

---



# Capítulo 1

## Problemas objeto de estudio

### 1.1. Introducción

De forma general, los problemas que provienen de la física y de la ingeniería se pueden clasificar en los siguientes grupos: *problemas de condiciones iniciales*, *problemas de condiciones iniciales y de contorno*, *problemas de contorno* y *problemas mixtos*. Desde un punto de vista matemático, los problemas de condiciones iniciales pueden dar lugar a ecuaciones hiperbólicas o parabólicas, los problemas de contorno derivan en ecuaciones de tipo elípticas y los problemas mixtos dan lugar a ecuaciones de tipo parabólico y elíptico. La obtención de soluciones numéricas para estos tres tipos de problemas implica realizar discretizaciones parciales, que se indican en el cuadro (1.1) y que determinan las necesidades o especificaciones siguientes:

En el problema de condiciones iniciales, el modelo matemático contiene derivadas parciales, de la función incógnita  $u : \mathbb{R}^3 \times \mathbb{R} \rightarrow \mathbb{R}^d$ , respecto de variables independientes espaciales y temporales. Para evaluar estas derivadas, se da un conjunto de puntos discretos en el dominio espacial y se aplica la *teoría de interpolación polinómica* y la *teoría de derivación numérica*, dando lugar a los métodos de discretización espacial, y que dan origen a los métodos de *diferencias finitas (FDM)*, *volúmenes finitos (FVM)* y *elementos finitos (FEM)*. Por otro lado, la *teoría de interpolación trigonométrica* permite obtener los denominados *métodos espectrales*. Este procedimiento se conoce como semi-discretización espacial y el resultado es un problema de valores iniciales para un sistema de ecuaciones diferenciales ordinarias en la función incógnita  $U : \mathbb{R} \rightarrow \mathbb{R}^N$  siendo  $N = dM$  con  $M$  el número de puntos de la semi-discretización espacial. Para evaluar las derivadas en la variable temporal,

se define un conjunto de puntos discretos en el dominio temporal y, nuevamente, se aplica la *teoría de interpolación polinómica* y la *teoría de derivación y cuadratura numérica*. En este caso, se obtienen los métodos numéricos *lineales multipaso* y *predictores–correctores*, que pueden ser tanto explícitos como implícitos. Otras técnicas de aproximación permiten obtener otro conjunto importante de métodos que son los denominados *Runge-Kutta* y que, de nuevo, pueden ser explícitos o implícitos. Los métodos implícitos, cuando se aplican a problemas diferenciales no lineales, dan lugar a sistemas de ecuaciones no lineales, que se resuelven aplicando los métodos numéricos correspondientes, como son el método de Newton-Raphson o el método de iteración de punto fijo. Esta aproximación se conoce como discretización temporal y el resultado es un problema de valores iniciales para un sistema de ecuaciones en diferencias. La primera columna del cuadro (1.1) resume el método de cálculo en problemas de valores iniciales explicado anteriormente.

En el problema de contorno, al igual que en el problema de valores iniciales, el modelo matemático contiene derivadas parciales, de la función incógnita  $u : \mathbb{R}^3 \rightarrow \mathbb{R}^d$ , respecto de las variables espaciales. Para evaluar estas derivadas, se emplean las técnicas de discretización espacial mencionadas en el problema de valores iniciales y el resultado es un sistema de ecuaciones algebraicas. Si el sistema es lineal, la solución se obtiene aplicando métodos numéricos para sistemas lineales como son el método de Gauss o factorización LU. Si el sistema es no lineal, la solución se obtiene aplicando métodos numéricos para sistemas no lineales como son el método de Newton o el método de iteración de punto fijo. La segunda columna del cuadro (1.1) resume el método de cálculo en problemas de contorno explicado anteriormente.

En un problema mixto, aparecen ecuaciones en derivadas parciales que evolucionan en el tiempo, por lo general de tipo parabólico, junto con ecuaciones de tipo elíptico. Para resolver los problemas mixtos se emplean las técnicas de discretización espacial y temporal junto con los métodos en ecuaciones algebraicas mencionadas en el tratamiento de los problemas de condiciones iniciales y de contorno explicados anteriormente.

---

Enunciado	Problema hiperbólico o parabólico.	Problema elíptico
Modelo matemático	Conjunto de ecuaciones en derivadas parciales $\frac{\partial u}{\partial t} = \mathcal{L}(u, t, x)$ + condiciones iniciales + condiciones de contorno	Conjunto de ecuaciones en derivadas parciales $\mathcal{L}(u, x) = 0$ + condiciones de contorno
Discretización espacial	Conjunto de ecuaciones diferenciales ordinarias $\frac{dU}{dt} = F(U, t)$ + condiciones iniciales	Conjunto de ecuaciones algebraicas $F(U) = 0$
Discretización temporal	Conjunto de ecuaciones en diferencias (DEs) $G(U^{n+1}, U^n, \dots) = 0$ + condiciones iniciales	
Solución discreta	$U^n$	$U$

Cuadro 1.1: Discretización espacial y temporal para problemas de condiciones iniciales y de contorno en ecuaciones en derivadas parciales.



# Capítulo 2

## Aproximación de funciones

### 2.1. Introducción

Un problema de gran interés en matemática aplicada es la evaluación de una función en un punto, cuando no se conoce la expresión analítica de la misma, o cuando, sí se dispone de tal expresión, pero es difícil de evaluar en el punto de estudio. Otro problema habitual, por ejemplo en cálculo numérico, es la obtención de derivadas de distinto orden e integrales de funciones de las que sólo se disponen de algunos valores puntuales. Para abordar el estudio de estos problemas, se dispone de un tema clásico de la matemática que es la *teoría de aproximación*.

Dada una función  $f$ , la teoría de aproximación estudia la obtención de otra función  $g$  más sencilla, que es *próxima* a  $f$  en un cierto sentido, y que debe precisarse. Dicha función sencilla  $g$  puede ser un polinomio, una función trigonométrica, una función racional, y en general, cualquier función elemental de las estudiadas en el cálculo infinitesimal. Particularmente, las funciones polinómicas y trigonométricas se han estudiado con detalle dando lugar al desarrollo de la teoría de aproximación polinómica y a la teoría de aproximación trigonométrica. La evolución en la teoría de aproximación se puede resumir a partir de los aportes fundamentales de Newton (1643), Euler (1707), Lagrange (1736), Weierstrass (1815), Chebyshev (1821) y Lebesgue (1875).

El teorema siguiente asegura la existencia de polinomios que convergen uniformemente a  $f$ .

*Teorema de aproximación de Weierstrass.*

Sea  $f$  una función continua definida en un intervalo  $[a, b]$ . Para todo  $\epsilon > 0$ , le corresponde un polinomio  $P$  tal que:

$$\|f - P\| < \epsilon.$$

Es decir que:

$$|f(x) - P(x)| < \epsilon, \quad \text{para todo } x \in [a, b].$$

Generalmente, una aproximación de una función se puede expresar como combinación lineal infinita de determinadas funciones elementales como polinomios o funciones trigonométricas, que forman la base de un cierto espacio vectorial. En lo que sigue, consideraremos espacios de Hilbert de dimensión finita o dimensión infinita. Los espacios de Hilbert son espacios vectoriales equipados con un producto interno que permite definir una distancia de forma que el espacio es un espacio métrico completo. La intuición geométrica juega un papel importante en los espacios de Hilbert. Un elemento de un espacio de Hilbert puede ser especificado de forma única por sus coordenadas con respecto a una base ortonormal en analogía a las coordenadas cartesianas en el plano.

Si  $f(x)$  es un elemento de un espacio de Hilbert de dimensión infinita y existe un conjunto numerable de funciones base  $\phi_k(x)$ , entonces la función  $f(x)$  se puede poner como:

$$f(x) = \sum_{k=0}^{\infty} \hat{c}_k \phi_k(x), \quad (2.1)$$

donde  $\hat{c}_k$  constituyen los coeficientes del desarrollo en serie. Siguiendo la analogía de un espacio vectorial, los coeficientes se obtienen proyectando la función  $f(x)$  en las funciones base  $\phi_k(x)$ . De manera más precisa, si definimos el producto interno siguiente con la función de peso  $\omega(x)$ :

$$\langle \phi_m, \phi_k \rangle_w = \int_a^b \phi_m(x) \phi_k(x) \omega(x) dx, \quad (2.2)$$

entonces la proyección de  $f(x)$  en  $\phi_m(x)$  queda:

$$\int_a^b f(x) \phi_m(x) \omega(x) dx = \sum_{k=0}^{\infty} \hat{c}_k \int_a^b \phi_k(x) \phi_m(x) \omega(x) dx \quad (2.3)$$

que permiten obtener los coeficientes  $\hat{c}_k$ . Generalmente, se eligen funciones ortogonales, es decir:

$$\langle \phi_m, \phi_k \rangle_w = \begin{cases} 0 & k \neq m \\ \gamma_m & k = m \end{cases}$$

En este caso, los coeficientes  $\hat{c}_k$  se calculan a partir de la siguiente expresión:

$$\hat{c}_m = \frac{1}{\gamma_m} \int_a^b f(x) \phi_m(x) \omega(x) dx. \quad (2.4)$$



## 2.2. Serie truncada y precisión espectral

Dado un desarrollo en serie de una función  $f(x)$ , se define la serie truncada como los  $N$  primeros términos del desarrollo

$$P_N(x) = \sum_{k=0}^N \hat{c}_k \phi_k(x). \quad (2.5)$$

El error de la serie truncada se define como el error de truncamiento y vale:

$$E_N(x) = f(x) - P_N(x) = \sum_{k=N+1}^{\infty} \hat{c}_k \phi_k(x).$$

Desde el punto de vista computacional interesa conocer cuántos términos debemos retener del desarrollo en serie para que el error de truncamiento sea pequeño. La regularidad de la función  $f(x)$  a desarrollar y las propiedades de las funciones base permiten conocer el número de términos que se deben retener para un error prefijado. La elección de las funciones base permite acelerar la convergencia del desarrollo en serie o minimizar el número de grados de libertad o coeficientes  $\hat{c}_k$  necesarios para representar de manera precisa la función  $f(x)$ .

Desde el punto de vista numérico, solo podemos manejar un número finito de modos  $k$  por lo que nos interesa que la convergencia de estos desarrollos en serie sea lo más rápida posible. De esta forma, la serie truncada aproxima bien a la función objeto de estudio. La aproximación más familiar es el desarrollo en serie de Fourier en donde las funciones ortogonales son senos y cosenos  $\phi_k(x) = e^{ikx}$ . Si la función es infinitamente suave, con todas sus derivadas periódicas en el contorno, el coeficiente  $k$ -ésimo del desarrollo en serie decae más rápidamente que cualquier potencia inversa de  $k$ . En la práctica se observa este comportamiento cuando tenemos suficientes términos del desarrollo en serie como para representar bien la estructura de la función  $f(x)$ . El rápido decaimiento de los coeficientes nos permite truncar las series con unos pocos términos y tener una excelente aproximación de la función. Esta característica básica de los métodos espectrales se denominada precisión espectral. La precisión espectral también se puede conseguir para funciones suaves pero que no sean periódicas utilizando funciones  $\phi_k(x)$  especiales. Para cualquier base ortogonal de funciones no se verifica, en principio, que los coeficientes del desarrollo en serie de una función suave en esta base tiendan a cero más rápidamente que cualquier potencia inversa de  $k$ . Usualmente, la precisión espectral se consigue solamente cuando la función tiene un comportamiento en el contorno muy especial. Las autofunciones del problema singular de Sturm–Liouville permiten precisión espectral en el desarrollo de cualquier función suave sin que sea necesario ninguna restricción adicional en el comportamiento en el contorno [2]. Un ejemplo de estas autofunciones que gozan de precisión espectral son los polinomios de Chebyshev y polinomios de Legendre. Por otra parte y bajo ciertas condiciones, los desarrollos en serie de senos y cosenos gozan de precisión espectral.

**Ejemplo:** Precisión espectral del desarrollo en serie de senos y cosenos de

$$f(x) = \sum_{m=-\infty}^{\infty} \hat{c}_m e^{imx}. \quad (2.6)$$

Para analizar el comportamiento de los coeficientes del desarrollo en serie de Fourier con el número de onda  $k$ , multiplicamos en (2.6) por  $e^{-ikx}$  e integramos por partes entre  $[0, 2\pi]$ . Para llevar a cabo esta operación tenemos que imponer que  $f(x)$  sea diferenciable en  $[0, 2\pi]$ , entonces para  $k \neq 0$

$$\hat{c}_k = \frac{-1}{ik} (f(2\pi) - f(0)) + \frac{1}{ik} \int_0^{2\pi} f'(x) e^{-ikx} dx. \quad (2.7)$$

Si no imponemos ninguna condición adicional, entonces se verifica que

$$|\hat{c}_k| = O\left(\frac{1}{k}\right).$$

Si además,  $f'(x)$  es diferenciable en  $[0, 2\pi]$  y  $f'(2\pi) = f'(0)$ , integrando nuevamente por partes en (2.7) queda

$$\hat{c}_k = \frac{1}{(ik)^2} (f'(2\pi) - f'(0)) + \frac{1}{(ik)^2} \int_0^{2\pi} f''(x) e^{-ikx} dx, \quad (2.8)$$

y los coeficientes

$$|\hat{c}_k| = O\left(\frac{1}{k^2}\right).$$

De esta forma, si  $f(x)$  es diferenciable  $m$  veces en  $[0, 2\pi]$  y todas sus derivadas hasta orden  $m - 1$  son periódicas, entonces

$$|\hat{c}_k| = O\left(\frac{1}{k^m}\right). \quad (2.9)$$

Como corolario, se puede concluir que el coeficiente  $\hat{c}_k$  del desarrollo en serie de Fourier de una función que es infinitamente diferenciable y todas sus derivadas periódicas en  $[0, 2\pi]$  decae más rápidamente que cualquier potencia negativa de  $k$ .

## 2.3. Serie discreta o interpolante

*Definición: Problema general de interpolación.*

Dada una función  $f$  y un conjunto de  $N + 1$  puntos discretos  $x_0, x_1, \dots, x_N$  del dominio de definición, el problema general de interpolación consiste en obtener otra función  $I$ , denominada función interpolante, de la forma:

$$I_N(x) = \tilde{c}_0 \phi_0(x) + \tilde{c}_1 \phi_1(x) + \dots + \tilde{c}_N \phi_N(x), \quad (2.10)$$

que es combinación lineal de funciones elementales  $\phi_0, \phi_1, \dots, \phi_N$ , y que debe verificar:

$$I_N(x_j) = f(x_j), \quad j = 0, 1, \dots, N. \quad (2.11)$$

Los puntos  $x_j$  se denominan *abscisas* o *nodos* de interpolación y los valores  $\tilde{c}_j$  son los *coeficientes* de interpolación. Resolver el problema de interpolación implica obtener los coeficientes  $\tilde{c}_j$ . Esta definición también puede enunciarse diciendo que dado un espacio vectorial  $V$ , de dimensión finita  $N$ , y dada una base  $\phi_0, \phi_1, \dots, \phi_N$ , cualquier elemento de  $V$  se puede expresar como una combinación lineal de los elementos de la base. Funciones base polinómicas dan lugar a problemas de *interpolación polinómica* y funciones base trigonométricas dan lugar a problemas de *interpolación trigonométrica*.

El interpolante  $I_N(x)$  (2.10) también se denomina como la serie discreta. La diferencia fundamental de la serie discreta frente a la serie truncada (2.5) está en el método de obtención de los coeficientes. El sistema (2.11) constituye un sistema lineal de ecuaciones para la determinación de los coeficientes  $\tilde{c}_k$ . En general, salvo que la nube de puntos  $x_j$  sea óptima, la convergencia de la serie discreta (2.10) es mala. Existen nubes de puntos óptimas desde el punto de vista de interpolación. Es objeto de este libro entender el origen de esta nube de puntos y en donde reside su bondad.

Desde el punto de vista intuitivo se busca una nube de puntos que con un producto interno en un espacio de dimensión finita goce de las propiedades de ortogonalidad de los espacios de dimensión infinita. Dada una función  $f(x)$  definida en un intervalo  $[a, b]$ , se considera una partición de  $N + 1$  puntos discretos  $x_j$  con  $x_0 = a$  y  $x_N = b$ . Dada una función base  $\phi_m(x)$ , se pueden definir el vector base  $\Phi_m$  cuyas componentes son la evaluación de la función base en los puntos nodales:

$$\Phi_m = \{ \phi_m(x_0), \phi_m(x_1), \dots, \phi_m(x_j), \dots, \phi_m(x_N) \}.$$

A partir de estos vectores base y mediante los coeficientes de peso  $\alpha_j$  se define el producto interno en un espacio de dimensión finita mediante:

$$\langle \Phi_m, \Phi_k \rangle_N = \sum_{j=0}^N \phi_m(x_j) \phi_k(x_j) \alpha_j. \quad (2.12)$$

Consideremos una nube de puntos que verifique las siguientes ecuaciones:

$$\langle \Phi_m, \Phi_k \rangle_N = \langle \phi_m, \phi_k \rangle_w \quad (2.13)$$

Como disponemos de  $N + 1$  puntos nodales a determinar y  $N + 1$  coeficientes  $\alpha_j$  desconocidos, podemos forzar al cumplimiento de  $2N + 2$  ecuaciones de (2.13). Es decir, podemos forzar a que

$$\langle \Phi_k, \Phi_{k-1} \rangle_N = 0, \quad \langle \Phi_k, \Phi_k \rangle_N = \gamma_k, \quad k = 0, \dots, N. \quad (2.14)$$

De esta manera, si se elige adecuadamente la nube de puntos, entonces los vectores base del espacio de dimensión finita son ortogonales:

$$\langle \Phi_m, \Phi_k \rangle_N = \begin{cases} 0 & k \neq m, \\ \gamma_m & k = m. \end{cases} \quad (2.15)$$

Es importante hacer notar que la ortogonalidad se verifica exclusivamente para valores de  $k \leq N$  y  $m \leq N$ . Para estos puntos nodales, los coeficientes  $\tilde{c}_m$  del interpolante se obtienen proyectando el vector

$$\mathbf{f} = \{ f(x_0), f(x_1), \dots, f(x_j), \dots, f(x_N) \}$$

en los vectores de la base. Es decir,

$$\tilde{c}_m = \frac{1}{\gamma_m} \sum_{j=0}^N f(x_j) \phi_m(x_j) \alpha_j, \quad (2.16)$$

o mediante la definición del producto interno (2.12) como la proyección de  $\mathbf{f}$  sobre el vector de la base  $\Phi_m$

$$\tilde{c}_m = \frac{1}{\gamma_m} \langle \mathbf{f}, \Phi_m \rangle_N. \quad (2.17)$$

## 2.4. Serie discreta y serie truncada

En esta sección se estudia la relación que existe entre la aproximación por series y la aproximación por interpolación. En particular, se analiza la relación entre los coeficientes  $\hat{c}_k$  de la serie truncada y los coeficientes  $\tilde{c}_k$  de la serie discreta.

Para relacionar estos coeficientes, se particulariza el interpolante (2.10) en  $x_j$  y el desarrollo en serie dado por (2.1) y se igualan sus valores

$$\sum_{k=0}^N \tilde{c}_k \phi_k(x_j) = \sum_{k=0}^{\infty} \hat{c}_k \phi_k(x_j). \quad (2.18)$$

En esta expresión multiplicamos por  $\phi_m(x_j)\alpha_j$  y sumamos desde  $j = 0$  hasta  $j = N$

$$\sum_{k=0}^N \tilde{c}_k \sum_{j=0}^N \phi_m(x_j) \phi_k(x_j) \alpha_j = \sum_{k=0}^{\infty} \hat{c}_k \sum_{j=0}^N \phi_m(x_j) \phi_k(x_j) \alpha_j. \quad (2.19)$$

Para la nube de puntos  $x_j$  que verifica las ecuaciones (2.13) y utilizando la ortogonalidad dada por (2.15), los coeficientes de la serie discreta se expresan como:

$$\tilde{c}_m = \hat{c}_m + \frac{1}{\gamma_m} \sum_{k=N+1}^{\infty} \hat{c}_k \sum_{j=0}^N \phi_m(x_j) \phi_k(x_j) \alpha_j. \quad (2.20)$$

Esa misma expresión se puede poner mediante la definición del producto interno discreto (2.12) como:

$$\tilde{c}_m = \hat{c}_m + \frac{1}{\gamma_m} \sum_{k=N+1}^{\infty} \langle \Phi_m, \Phi_k \rangle_N \hat{c}_k. \quad (2.21)$$

La diferencia entre los coeficientes de la serie discreta y los coeficientes de la serie truncada se denomina error de aliasing. Es importante hacer notar que si el desarrollo en serie converge con precisión espectral, las colas de desarrollo en serie a partir del valor  $k = N + 1$  pueden ser muy pequeñas y la diferencia entre los coeficientes de la serie discreta y los coeficientes de la serie truncada ser despreciable.

Desde el punto de vista computacional, las conclusiones mas importantes relacionadas con los desarrollos en serie y los interpolantes o series discretas son las siguientes:

1. El tratamiento con ordenador de una determinada función obliga a aproximar la misma con una serie de grados de libertad. Estos grados de libertad pueden ser los coeficientes de un desarrollo en serie o los valores de la función en un conjunto de puntos nodales.
2. Si la función es infinitamente derivable, el desarrollo en serie de funciones base puede tener convergencia espectral. Es decir, que el módulo de los coeficientes de la serie tiendan a cero mucho más rápidamente que cualquier potencia negativa del índice del coeficiente. Esta propiedad es especialmente importante para reducir de forma drástica el número de grados de libertad cuando se aproxima una función en el ordenador y conduce a lo que se denomina esquemas con precisión espectral.
3. Existen puntos nodales óptimos que permiten interpolar la función con un error similar a los de los desarrollos en serie espectrales salvo un pequeño error de aliasing. Generalmente, la simulación o tratamiento con ordenador de los problemas de matemática aplicada se implementan de manera mucho mas sencilla mediante mallas o nubes de puntos que mediante desarrollos en serie. La confianza en los métodos de interpolación como métodos de precisión espectral hace muy interesante su uso en la aplicación práctica.

En las siguientes secciones se discuten las fórmulas de cuadratura de Gauss que permiten determinar los puntos de colocación para que exista equivalencia entre el producto interno en un espacio de dimensión finita y un espacio de dimensión infinita y, en consecuencia, se pueda extrapolar el concepto de precisión espectral a una distribución de puntos nodales.

## 2.5. Fórmulas de cuadratura de Gauss

En esta sección trataremos de obtener las nubes de puntos óptimas de las que hablábamos en la sección anterior. Para entender el origen de esta nube de puntos, tenemos que profundizar en las fórmulas de cuadratura de Gauss. Una cuadratura de Gauss es una fórmula que selecciona  $N + 1$  puntos nodales  $x_j$  y  $N + 1$  pesos  $\alpha_j$  de manera óptima para dar un resultado exacto para polinomios  $g(x)$  de grado  $2N + 1$  con una función de ponderación  $\omega(x)$  no negativa en intervalo  $[a, b]$ . Es decir,

$$\int_a^b g(x) \omega(x) dx = \sum_{j=0}^N \alpha_j g(x_j). \quad (2.22)$$

Para llevar a cabo la demostración, expresamos  $g(x)$  mediante

$$g(x) = c(x) \phi_{N+1}(x) + r(x) \quad (2.23)$$

donde  $c(x)$  es el polinomio cociente y  $r(x)$  es el polinomio resto. Como  $g(x)$  es un polinomio de grado  $2N + 1$ , el polinomio cociente en serie de las funciones base es

$$c(x) = \beta_0 \phi_0(x) + \dots + \beta_N \phi_N(x). \quad (2.24)$$

Si se lleva este desarrollo a (2.22) y debido a la ortogonalidad de  $\phi_k$  y  $\phi_{N+1}$  para todo  $k \leq N$ , se comprueba

$$\int_a^b g(x) \omega(x) dx = \int_a^b r(x) \omega(x) dx. \quad (2.25)$$

Como las integrales de (2.25) se pueden calcular mediante (2.22),  $r(x_j)$  debe coincidir con  $g(x_j)$ . Teniendo en cuenta (2.23) y para que  $g(x_j) = r(x_j)$ ,  $x_j$  deben ser los ceros de  $\phi_{N+1}(x)$ .

$$\int_a^b r(x) \omega(x) dx = \sum_{j=0}^N \alpha_j g(x_j). \quad (2.26)$$

Por último, falta por determinar los pesos  $\alpha_j$ . Como  $r(x)$  es un polinomio de grado  $N$ ,  $r(x)$  se puede expresar

$$r(x) = a_0 + a_1 x + \dots + a_N x^N. \quad (2.27)$$

donde  $a_k$  son coeficientes arbitrarios. Si llevamos este desarrollo a (2.26), se obtiene

$$\begin{aligned} & a_0 \int_a^b \omega(x) dx + a_1 \int_a^b x \omega(x) dx + \dots + a_N \int_a^b x^N \omega(x) dx = \\ & a_0 \sum_{j=0}^N \alpha_j + a_1 \sum_{j=0}^N x_j \alpha_j + \dots + a_N \sum_{j=0}^N x_j^N \alpha_j. \end{aligned}$$


---

para cualquier valor de  $a_k$ . De esta forma y dada una función de peso  $\omega(x)$ , el sistema lineal de ecuaciones

$$\int_a^b x^k \omega(x) dx = \sum_{j=0}^N x_j^k \alpha_j, \quad k = 0, \dots, N \quad (2.28)$$

permite obtener los pesos  $\alpha_j$ .

En conclusión, si los puntos nodales  $\{x_j, j = 0, \dots, N\}$  son los ceros de  $\phi_{N+1}(x)$  y los pesos  $\alpha_j$  están determinados por el sistema lineal (2.28), la fórmula de cuadratura de Gauss (2.22) es exacta para polinomios de grado  $2N + 1$ .

Con estos pesos  $\alpha_j$  y estos puntos nodales  $x_j$ , la fórmula de cuadratura de Gauss permite calcular de forma exacta el producto interno de un espacio de dimensión infinita mediante el producto interno de un espacio de dimensión finita

$$\int_a^b \phi_m(x) \phi_k(x) \omega(x) dx = \sum_{j=0}^N \phi_k(x_j) \phi_m(x_j) \alpha_j \quad (2.29)$$

para todo  $k \leq N$  y todo  $m \leq N$ .

En esta fórmula de integración gaussiana los puntos nodales  $x_j$  no incluyen los contornos  $x = a$  y  $x = b$ . Cuando el problema que queramos resolver incluya condiciones de contorno en los extremos del intervalo, necesitamos generalizar las fórmulas de integración gaussiana que incluyan estos puntos. Éste es el objeto de la siguiente sección.

## 2.6. Fórmulas de cuadratura de Gauss–Lobatto

Para obtener las fórmulas de integración gaussiana que incluyan los extremos del intervalo, se considera el siguiente polinomio:

$$q_{N+1}(x) = \phi_{N+1}(x) + c_1 \phi_N(x) + c_2 \phi_{N-1}(x), \quad (2.30)$$

donde  $c_1$  y  $c_2$  se eligen para que  $q_{N+1}(a) = q_{N+1}(b) = 0$ . De esta forma, las  $N + 1$  raíces de  $q_{N+1}(x)$  incluyen  $x_0 = a$  y  $x_N = b$ .

Una cuadratura de Gauss–Lobatto es una fórmula que selecciona  $N - 1$  puntos nodales  $x_1, \dots, x_{N-1}$  y  $N + 1$  pesos  $\alpha_j$  de manera óptima para dar un resultado exacto para polinomios  $g(x)$  de grado  $2N - 1$ . Es decir,

$$\int_a^b g(x) \omega(x) dx = \sum_{j=0}^N \alpha_j g(x_j). \quad (2.31)$$

La demostración se hace de manera similar a la cuadratura de Gauss. Consideremos que  $g(x)$  es un polinomio de grado  $2N - 1$  que lo expresamos de la siguiente forma:

$$g(x) = c(x) q_{N+1}(x) + r(x), \quad (2.32)$$

donde  $r(x)$  es el polinomio resto de grado  $N$  y  $c(x)$  es el polinomio cociente de grado  $N - 2$  siguiente:

$$c(x) = \beta_0 \phi_0(x) + \dots \beta_{N-2} \phi_{N-2}(x). \quad (2.33)$$

Es importante hacer notar que como (2.30) incluye  $\phi_{N-1}(x)$ ,  $c(x)$  puede llegar a grado  $N - 2$  para que

$$\int_a^b c(x) q_{N+1}(x) \omega(x) dx = 0, \quad (2.34)$$

y podamos calcular la integral mediante (2.31) con los pesos  $\alpha_j$  dados por (2.28).

En resumen, si los puntos nodales  $\{x_j, j = 0, \dots, N\}$  son los ceros de  $q_{N+1}(x)$  que incluyen los extremos del intervalo y los pesos  $\alpha_j$  están determinados por el sistema lineal (2.28), la formula de cuadratura de Gauss (2.31) es exacta para polinomios de grado  $2N - 1$ .

Con estos pesos  $\alpha_j$  y estos puntos nodales  $x_j$ , la fórmula de cuadratura de Gauss permite calcular de forma exacta el producto interno de un espacio de dimensión infinita mediante el producto interno de un espacio de dimensión finita

$$\int_a^b \phi_m(x) \phi_k(x) \omega(x) dx = \sum_{j=0}^N \phi_k(x_j) \phi_m(x_j) \alpha_j \quad (2.35)$$

para todo  $k \leq N$  y todo  $m \leq N - 1$ .

## 2.7. Puntos de interpolación y bases ortogonales

Cuando se trata aproximar una función  $f(x)$  mediante un conjunto finito de funciones base  $\phi_k(x)$ , el error se expresa:

$$E(x) = f(x) - \sum_{k=0}^N \tilde{c}_k \phi_k(x) \quad (2.36)$$

y su norma  $\|E\|$  se minimiza cuando los coeficientes  $\tilde{c}_k$  de la serie discreta se obtienen mediante la proyección de la  $f(x)$  en las funciones base:

$$\tilde{c}_k = \frac{1}{\gamma_k} \sum_{j=0}^{j=N} f(x_j) \phi_k(x_j) \alpha_j. \quad (2.37)$$



Como se ha discutido en las secciones 2.4-2.6, si los puntos de interpolación se eligen de forma adecuada, la diferencia entre la serie discreta y la serie truncada es el error de aliasing que es, generalmente, insignificante.

Es importante hacer notar que existe una equivalencia entre una distribución de puntos de interpolación y una base de polinomios ortogonales:

$$x_0, x_1, \dots, x_N \iff \phi_0(x), \phi_1(x), \dots, \phi_N(x). \quad (2.38)$$

Para poner de manifiesto este hecho, se expresa la relación de recurrencia a tres términos que verifica una sucesión de polinomios ortogonales a partir de un producto interno:

$$\phi_{k+1}(x) = \left( x - \frac{\langle x \phi_k, \phi_k \rangle}{\langle \phi_k, \phi_k \rangle} \right) \phi_k(x) - \frac{\langle \phi_k, \phi_k \rangle}{\langle \phi_{k-1}, \phi_{k-1} \rangle} \phi_{k-1}(x), \quad (2.39)$$

donde  $\langle \cdot, \cdot \rangle$  puede ser el producto interno  $\langle \cdot, \cdot \rangle_N$  de un espacio de dimensión finita (2.12) o el producto interno  $\langle \cdot, \cdot \rangle_w$  de un espacio dimensión infinita (2.2). La relación de recurrencia a tres términos permite construir una secuencia de polinomios ortogonales con respecto a un producto interno a partir de los valores iniciales  $\phi_0(x) = 1$  y  $\phi_1(x) = x$ . El producto interno del espacio de dimensión infinita se define en el intervalo  $[a, b]$  y la función de peso dada  $w(x)$ . Por otra parte, el producto interno del espacio de dimensión finita se define mediante la distribución de puntos  $x_j$  y los pesos  $\alpha_j$ .

De manera recíproca, dada una sucesión de polinomios ortogonales, las fórmulas de cuadratura de Gauss o Gauss-Lobatto permiten obtener los puntos de interpolación para que la serie discreta o el interpolante difiera de la serie truncada en el error de aliasing. Los polinomios de Chebyshev y los polinomios de Legendre son dos ejemplos de sucesiones de polinomios ortogonales que se usan con mucha frecuencia en problemas numéricos debido a su precisión espectral [2].

Con esta equivalencia entre los puntos de interpolación y la base de polinomios ortogonales, la resolución de un problema se puede hacer en el plano físico o en el plano espectral. Es decir, las incógnitas de un determinado problema pueden ser los valores que toma la función a determinar en los puntos nodales  $f(x_j)$  o los coeficientes  $\tilde{c}_k$  del desarrollo de la serie discreta. Dependiendo de la linealidad del problema y la existencia de una transformada rápida entre el plano físico y el plano espectral, el tratamiento como serie discreta puede ser computacionalmente más eficiente.

La pregunta que ahora se nos plantea es obtener la nube de puntos de interpolación o la base de polinomios ortogonales que minimizan el error de aproximación. Como se discutió en la sección 2.2, no todos los polinomios ortogonales gozan de precisión espectral. Nuestro interés se centra en aquellas distribuciones de puntos que se obtienen como ceros de las funciones bases con precisión espectral.

**Ejemplo:** Determinar la base ortogonal asociada a los puntos de interpolación

$$\{x_0, x_1, x_2, x_3, x_4\} = \{-1, -\sqrt{2}/2, 0, -\sqrt{2}/2, +1\}, \quad (2.40)$$

cuando el producto interno está definido por la expresión (2.12) con los coeficientes de peso

$$\{\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4\} = \left\{ \frac{\pi}{2N}, \frac{\pi}{N}, \frac{\pi}{N}, \frac{\pi}{N}, \frac{\pi}{2N} \right\}. \quad (2.41)$$

Utilizando estos puntos nodales y mediante la relación de recurrencia (2.39) se obtienen los siguientes polinomios una vez normalizados con el valor del polinomio en  $x = 1$ :

$$\{T_0, T_1, T_2, T_3, T_4\} = \{1, x, 2x^2 - 1, 4x^3 - 3x, 8x^4 - 8x^2 + 1\}. \quad (2.42)$$

Estos polinomios coinciden con los polinomios de Chebyshev. Con este ejemplo se pone de manifiesto que el interpolante con ciertos puntos de colocación es equivalente al desarrollo de la función en serie de polinomios de Chebyshev.

De forma recíproca, dada una base de polinomios y mediante la expresión (2.30) se puede obtener los puntos de interpolación o puntos de Gauss–Lobatto que contienen los extremos  $x = \pm 1$ . En el caso de los polinomios de Chebyshev, la expresión (2.30) queda:

$$q_5(x) = T_5(x) - T_4(x) = 16x^5 - 20x^3 + 5x - 4x^3 + 3x. \quad (2.43)$$

Los ceros de  $q_5(x)$  o puntos de Gauss–Lobatto son los puntos de interpolación dados en (2.40).

A modo de resumen se esquematizan las siguientes conclusiones:

1. Si la base de polinomios tiene precisión espectral, el interpolante basado en los puntos de Gauss o Gauss–Lobatto y la serie truncada se comportan de manera similar con respecto al error.
2. Un conjunto de puntos de colocación junto con un producto interno definen una base de polinomios mediante la relación de recurrencia a tres términos (2.39).
3. El atractivo de los métodos de alto orden reside en la precisión espectral. Las autofunciones del problema singular de Sturm–Liouville tienen precisión espectral. Existen infinidad de funciones base polinómicas que pueden gozar de la precisión espectral.
4. Una distribución de puntos de interpolación se puede comportar de manera análoga a las bases polinómicas con precisión espectral.

El objetivo del siguiente capítulo es analizar el error de interpolación y el error de las derivadas del interpolante para poder decidir de manera óptima el conjunto de puntos nodales o puntos de interpolación que tienen comportamientos análogos a las bases polinómicas con precisión espectral.

---

# Capítulo 3

## Interpolación polinómica

### 3.1. Introducción

La base no ortogonal formada por los monomios  $1, x, \dots, x^k, \dots$  permite aproximar una función  $f(x)$  mediante la serie truncada siguiente:

$$P_N(x) = \sum_{k=0}^N \hat{a}_k x^k. \quad (3.1)$$

La proyección de  $f(x)$  sobre los elementos de la base permite obtener los coeficientes  $\hat{a}_k$ . Por otra parte, la interpolación polinómica definida en (2.10)

$$I_N(x) = \sum_{k=0}^N a_k x^k. \quad (3.2)$$

permite obtener un polinomio interpolante que pase por un conjunto de puntos nodales  $x_j$ . La matriz del sistema lineal de ecuaciones que permite obtener  $a_k$  es la matriz de Vandermonde que es una matriz mal condicionada (sección 3.2). Para evitar el mal condicionamiento en la obtención de los coeficientes  $a_k$ , se formula la interpolación de Lagrange (sección 3.3) y la forma de Newton (sección 3.5).

### 3.2. Matriz de Vandermonde

Se consideran  $N + 1$  puntos  $x_0, x_1, \dots, x_N$  del dominio de definición de una función  $f$ , en los que se conocen los valores  $f(x_j) = y_j$ ,  $j = 0, 1, \dots, N$ . El problema de interpolación consiste en obtener una función interpolante  $I_N$  de grado  $\leq N$ , como una combinación lineal de monomios de la forma siguiente:

$$I_N(x) = a_0 + a_1 x + \dots + a_N x^N.$$

Por definición de interpolación, el interpolante  $I_N$  debe verificar que:

$$I_N(x_j) = f(x_j), \quad j = 0, 1, \dots, N.$$

En la expresión anterior,  $a_0, a_1, \dots, a_N$  son los coeficientes del polinomio y la obtención de los mismos, resuelve el problema de interpolación. La condición anterior conduce al sistema lineal de ecuaciones siguiente:

$$\begin{aligned} a_0 + a_1 x_0 + a_2 x_0^2 + \dots + a_N x_0^N &= f(x_0), \\ a_0 + a_1 x_1 + a_2 x_1^2 + \dots + a_N x_1^N &= f(x_1), \\ &\vdots \\ a_0 + a_1 x_N + a_2 x_N^2 + \dots + a_N x_N^N &= f(x_N), \end{aligned}$$

que, en forma matricial, puede expresarse como sigue

$$\begin{pmatrix} 1 & x_0 & x_0^2 & \dots & x_0^N \\ 1 & x_1 & x_1^2 & \dots & x_1^N \\ \vdots & & & & \vdots \\ 1 & x_N & x_N^2 & \dots & x_N^N \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_N \end{pmatrix} = \begin{pmatrix} f(x_0) \\ f(x_1) \\ \vdots \\ f(x_N) \end{pmatrix}. \quad (3.3)$$

La matriz de la expresión anterior, se conoce como matriz de Vandermonde. Si los puntos de interpolación son distintos, la matriz de Vandermonde no es singular y el sistema (3.3) tiene solución única. La solución del sistema lineal (3.3) se puede obtener mediante cualquier método directo de resolución de sistemas lineales como, por ejemplo, el método de eliminación Gaussiana o de factorización LU. El problema de interpolación queda finalmente resuelto con el polinomio siguiente:

$$I_N(x) = a_0 + a_1 x + \dots + a_N x^N,$$

expresado como una combinación lineal de monomios. Una dificultad que presenta esta forma de resolver el problema de interpolación es que la matriz de Vandermonde puede presentar un número de condición grande, dando lugar a sistemas

mal condicionados. A continuación se presenta un ejemplo de interpolación en monomios con tres puntos de interpolación.

**Ejemplo:**

Dada una función  $f$  y 3 puntos discretos  $x_0, x_1, x_2$  del dominio de definición, se propone obtener el polinomio interpolante  $I_2$  de grado  $\leq 2$ , expresado como combinación lineal de monomios. Por definición del problema de interpolación,  $I_2$  debe verificar que:

$$I_2(x_0) = f(x_0) = y_0, \quad I_2(x_1) = f(x_1) = y_1, \quad I_2(x_2) = f(x_2) = y_2.$$

Tal como se observa en las expresiones anteriores, se reemplaza  $f(x_j)$  por  $y_j$  para simplificar las expresiones finales.

Para resolver este problema se plantea el polinomio de grado  $\leq 2$  siguiente:

$$I_2(x) = a_0 + a_1 x + a_2 x^2,$$

donde  $a_0, a_1$  y  $a_2$  son coeficientes que deben determinarse. Para ello, se fuerza a que el polinomio  $I_2$  coincida con la función  $f$  en los 3 puntos discretos:

$$I_2(x_0) = y_0, \quad I_2(x_1) = y_1, \quad I_2(x_2) = y_2,$$

que conduce a un sistema de tres ecuaciones algebraicas lineales, cuya matriz de Vandermonde (matriz ampliada) es:

$$\left( \begin{array}{ccc|c} 1 & x_0 & x_0^2 & y_0 \\ 1 & x_1 & x_1^2 & y_1 \\ 1 & x_2 & x_2^2 & y_2 \end{array} \right).$$

El sistema se resuelve aplicando el método de Gauss. Mediante operaciones elementales de filas se obtienen los sistemas equivalentes siguientes:

$$\left( \begin{array}{ccc|c} 1 & x_0 & x_0^2 & y_0 \\ 0 & x_1 - x_0 & x_1^2 - x_0^2 & y_1 - y_0 \\ 0 & x_2 - x_0 & x_2^2 - x_0^2 & y_2 - y_0 \end{array} \right),$$

$$\left( \begin{array}{ccc|c} 1 & x_0 & x_0^2 & y_0 \\ 0 & 1 & x_1 + x_0 & \frac{y_1 - y_0}{x_1 - x_0} \\ 0 & 1 & x_2 + x_0 & \frac{y_2 - y_1}{x_2 - x_1} \end{array} \right),$$

$$\left( \begin{array}{ccc|c} 1 & x_0 & x_0^2 & y_0 \\ 0 & 1 & x_1 + x_0 & \frac{y_1 - y_0}{x_1 - x_0} \\ 0 & 0 & x_2 - x_0 & \left( \frac{y_2 - y_1}{x_2 - x_1} - \frac{y_1 - y_0}{x_1 - x_0} \right) \end{array} \right).$$

Aplicando sustitución inversa al sistema anterior, se calculan los coeficientes del polinomio:

$$\begin{aligned} a_2 &= \frac{1}{x_2 - x_0} \left( \frac{y_2 - y_1}{x_2 - x_1} - \frac{y_1 - y_0}{x_1 - x_0} \right), \\ a_1 &= \left( \frac{y_1 - y_0}{x_1 - x_0} \right) - \frac{x_1 + x_0}{x_2 - x_0} \left( \frac{y_2 - y_1}{x_2 - x_1} - \frac{y_1 - y_0}{x_1 - x_0} \right), \\ a_0 &= y_0 - x_0 a_1 - x_0^2 a_2. \end{aligned}$$

La forma de solución del problema de interpolación, expresada como una combinación lineal de monomios, no es la habitualmente empleada, dado que requiere resolver el sistema lineal (3.3) cada vez que se modifica el conjunto de nodos de interpolación. Esto hace que este método resulte ser muy laborioso. Por otro lado, desde un punto de vista computacional, el mal condicionamiento de la matriz de Vandermonde hace que el método no sea el adecuado. A continuación, se presenta otra forma de resolver el problema de interpolación que resulta ser mucho más práctica al momento de generar esquemas numéricos para la solución de ecuaciones diferenciales.

### 3.3. Interpolación de Lagrange

En este apartado, se presenta otra forma de resolver el problema de interpolación polinómica, que resulta ser más eficiente desde el punto de vista de la obtención de esquemas numéricos para la solución de ecuaciones diferenciales. Se consideran  $N + 1$  puntos  $x_0, x_1, \dots, x_N$  del dominio de definición de una función  $f$ , en los que se conocen los valores  $f(x_j)$ ,  $j = 0, 1, \dots, N$ . A continuación, se expresa el polinomio  $I_N$  de grado  $N$  en la forma siguiente:

$$I_N(x) = b_0 \ell_0(x) + b_1 \ell_1(x) + \dots + b_N \ell_N(x), \quad (3.4)$$

donde  $b_0, b_1, \dots, b_N$  son coeficientes a calcular y  $\ell_j(x)$  son polinomios de grado  $N$  que tienen la expresión siguiente:

$$\ell_j(x) = \frac{(x - x_0)(x - x_1) \dots (x - x_{j-1})(x - x_{j+1}) \dots (x - x_N)}{(x_j - x_0)(x_j - x_1) \dots (x_j - x_{j-1})(x_j - x_{j+1}) \dots (x_j - x_N)}, \quad (3.5)$$

o en su forma mas compacta

$$\ell_j(x) = \prod_{\substack{i=0 \\ i \neq j}}^N \frac{(x - x_i)}{(x_j - x_i)}. \quad (3.6)$$

La forma (3.4) se conoce como interpolación de Lagrange y los polinomios (3.5) son los polinomios de Lagrange. Como los polinomios (3.5) toman los valores siguientes:

$$\begin{cases} \ell_j(x) = 1, & x = x_j, \\ \ell_j(x) = 0, & x = x_i, \quad i = 0, 1, \dots, j-1, j+1, \dots, N, \end{cases} \quad (3.7)$$

entonces la interpolación de Lagrange se puede expresar como se indica a continuación:

$$I_N(x) = \sum_{j=0}^N f(x_j) \ell_j(x). \quad (3.8)$$

A continuación se presenta un ejemplo de interpolación en la forma de Lagrange, con tres puntos de interpolación.

### Ejemplo:

Dada una función  $f$  y 3 puntos de interpolación  $x_0, x_1, x_2$  del dominio de definición, se propone obtener el polinomio interpolante  $I_2$ , de grado  $\leq 2$ , mediante la interpolación de Lagrange, que verifique que

$$I_2(x_0) = f(x_0) = y_0, \quad I_2(x_1) = f(x_1) = y_1, \quad I_2(x_2) = f(x_2) = y_2.$$

Aplicando la expresión (3.4), el interpolante de grado  $\leq 2$  es:

$$I_2(x) = b_0 \ell_0(x) + b_1 \ell_1(x) + b_2 \ell_2(x),$$

donde  $\ell_0, \ell_1, \ell_2$  son polinomios de grado  $\leq 2$  que, según (3.5), tienen las expresiones siguientes:

$$\ell_0(x) = \frac{(x - x_1)(x - x_2)}{(x_0 - x_1)(x_0 - x_2)},$$

$$\ell_1(x) = \frac{(x - x_0)(x - x_2)}{(x_1 - x_0)(x_1 - x_2)},$$

$$\ell_2(x) = \frac{(x - x_0)(x - x_1)}{(x_2 - x_0)(x_2 - x_1)}.$$


---

Obsérvese que los polinomios anteriores verifican los valores siguientes (ver 3.7):

$$\ell_0(x_0) = 1, \quad \ell_0(x_1) = 0, \quad \ell_0(x_2) = 0,$$

$$\ell_1(x_0) = 0, \quad \ell_1(x_1) = 1, \quad \ell_1(x_2) = 0,$$

$$\ell_2(x_0) = 0, \quad \ell_2(x_1) = 0, \quad \ell_2(x_2) = 1.$$

Finalmente, los coeficientes del interpolante son:

$$b_0 = f(x_0), \quad b_1 = f(x_1), \quad b_2 = f(x_2),$$

que son datos del problema. El polinomio buscado es:

$$I_2(x) = f(x_0) \ell_0(x) + f(x_1) \ell_1(x) + f(x_2) \ell_2(x).$$

### 3.4. Método recursivo

En la resolución del problema de interpolación polinómica, resulta de gran utilidad contar con un procedimiento recursivo que permita obtener el polinomio interpolante de grado  $N$  a partir del polinomio interpolante de grado  $N-1$ . Dicho de otra forma, al incorporar un nuevo punto de interpolación y a partir del interpolante construido con  $N$  puntos, interesa obtener el nuevo interpolante construido con  $N+1$  puntos de interpolación. En este apartado se describe un método recursivo para obtener los polinomios de Lagrange cuando, a partir de un conjunto de puntos dados, se considera un nuevo punto de interpolación. Con el fin de desarrollar este tema, a modo de notación, se introduce un subíndice adicional a los polinomios de Lagrange. En lo que sigue, se indicará con  $\ell_{ji}$  al polinomio de Lagrange que verifica el punto de interpolación  $j$  construido a partir de  $i$  puntos de interpolación, de grado  $i-1$ . Por tanto, a partir de la expresión (3.5), el polinomio de Lagrange obtenido a partir de  $i$  puntos, puede expresarse de la forma siguiente:

$$\ell_{ji}(x) = \frac{(x-x_0)(x-x_1)\dots(x-x_{j-1})(x-x_{j+1})\dots(x-x_{i-1})}{(x_j-x_0)(x_j-x_1)\dots(x_j-x_{j-1})(x_j-x_{j+1})\dots(x_j-x_{i-1})},$$

$$j = 0, \dots, N, \quad i = 1, \dots, N+1.$$

La expresión anterior puede escribirse como:

$$\ell_{ji}(x) = \frac{(x-x_0)\dots(x-x_{i-2})}{(x_j-x_0)\dots(x_j-x_{i-2})} \left( \frac{x-x_{i-1}}{x_j-x_{i-1}} \right), \quad j = 0, \dots, N, \quad i = 1, \dots, N+1.$$

De esta forma,  $\ell_{ji}(x)$  se puede expresar mediante el polinomio de Lagrange  $\ell_{ji-1}$  obtenido a partir de  $i-1$  puntos. Finalmente, se obtiene la recursión

$$\ell_{ji}(x) = \ell_{ji-1}(x) \left( \frac{x-x_{i-1}}{x_j-x_{i-1}} \right), \quad j = 0, \dots, N, \quad i = 1, \dots, N+1,$$



que permite obtener el polinomio de Lagrange de grado  $i - 1$  a partir del polinomio de Lagrange de grado  $i - 2$ , cuando se incorpora el punto de interpolación  $x_i$ . La condición inicial del método recursivo propuesto es  $\ell_{j1} = 1$  que corresponde al polinomio de Lagrange, de grado 0, construido con un único punto de interpolación  $x_j$ . A continuación, se presenta un ejemplo donde se aplica el procedimiento recursivo explicado.

**Ejemplo:**

Dada una función  $f$  se propone obtener el polinomio interpolante  $I_2$ , de grado  $\leq 2$ , partiendo del polinomio de Lagrange de grado 0, que verifique el punto  $x_0$ , posteriormente incorporar el punto  $x_1$  para obtener el polinomio de grado 1, y por último, agregar el punto de interpolación  $x_2$  y obtener el polinomio de grado 2. En la figura (3.1) se representa gráficamente el algoritmo recursivo que se explica a continuación.

Se comienza con el punto de interpolación  $x_0$  y el polinomio de Lagrange de grado 0 es:

$$\ell_{01} = 1,$$

siendo esta la condición inicial del método recursivo. A continuación, se considera el punto  $x_1$  y se calcula el polinomio de Lagrange de grado 1, empleando la expresión (3.4)

$$\ell_{02}(x) = \ell_{01} \frac{x - x_1}{x_0 - x_1} = \frac{x - x_1}{x_0 - x_1}.$$

Por último, se considera el punto  $x_2$  y se obtiene el polinomio de Lagrange de grado 2, aplicando nuevamente la expresión (3.4)

$$\ell_{03}(x) = \ell_{02}(x) \frac{x - x_2}{x_0 - x_2} = \frac{(x - x_1)(x - x_2)}{(x_0 - x_1)(x_0 - x_2)}.$$

Los polinomios de Lagrange  $\ell_{13}(x)$  y  $\ell_{23}(x)$  se obtienen de igual forma que  $\ell_{03}(x)$ . Finalmente, el interpolante de grado  $\leq 2$  de la función  $f$  es:

$$I_2(x) = f(x_0) \ell_{03}(x) + f(x_1) \ell_{13}(x) + f(x_2) \ell_{23}(x),$$

que se representa en la figura (3.2).

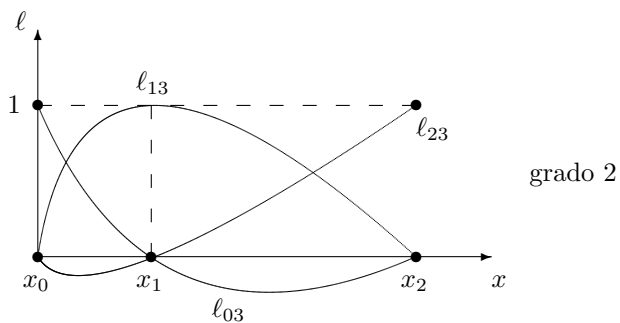


Figura 3.1: Algoritmo recursivo para la obtención de los polinomios de Lagrange

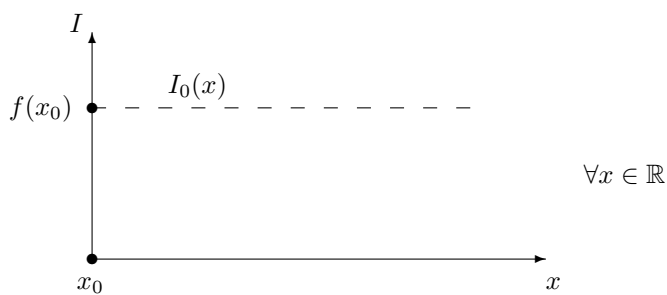
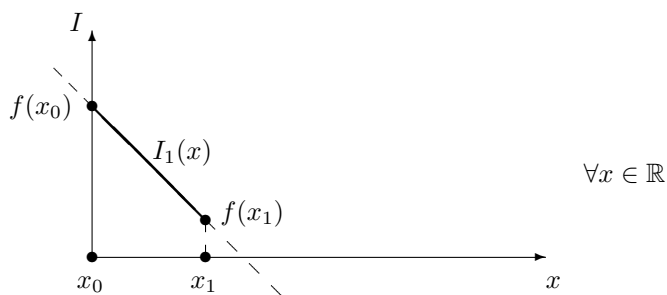
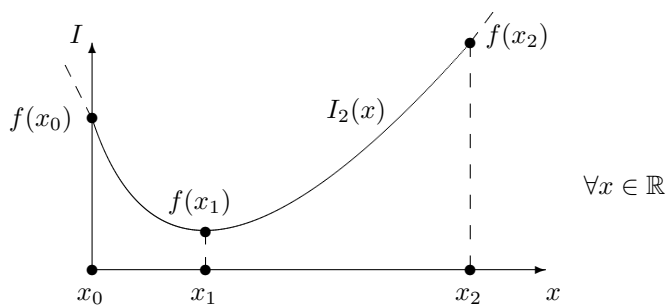


Figura 3.2: Interpolante de Lagrange de grado  $\leq 2$  de una función  $f$

### 3.5. Forma de Newton

En esta sección, se presenta otra forma de resolver el problema de interpolación polinómica que se conoce como forma de Newton. Al igual que en la interpolación de Lagrange, este método aporta un procedimiento recursivo para construir polinomios de grado superior, a medida que se añaden puntos de interpolación. Además, este método permite introducir el concepto de cociente incremental, también conocido como de diferencia dividida, que permite definir otro método recursivo para determinar los coeficientes de la forma de Newton.

Se consideran  $N + 1$  puntos distintos  $x_0, x_1, \dots, x_N$  del dominio de definición de una función  $f$ , en los que se conocen los valores  $f(x_j)$ ,  $j = 0, 1, \dots, N$ . A continuación, se expresa el polinomio  $I_N$  de grado  $\leq N$ , en la forma siguiente:

$$I_N(x) = c_0 + c_1(x - x_0) + c_2(x - x_0)(x - x_1) + \dots + c_N(x - x_0)(x - x_1)(x - x_2) \dots (x - x_{N-2})(x - x_{N-1}), \quad (3.9)$$

donde  $c_0, c_1, \dots, c_N$  son los coeficientes de interpolación a calcular. La expresión (3.9) se conoce como forma de Newton del problema de interpolación polinómica y, de forma compacta, se puede expresar como:

$$I_N(x) = \sum_{j=0}^N c_j \prod_{i=0}^{j-1} (x - x_i). \quad (3.10)$$

Para calcular los coeficientes de interpolación, se recurre al procedimiento recursivo que se describe a continuación. Como condición inicial, se consideran los datos siguientes: el punto de interpolación  $x_0$  y la imagen de la función  $f(x_0)$ . A partir del punto  $x_0$ , se obtiene el interpolante  $I_0(x)$  de grado 0 que verifica la función  $f$  en el nodo de interpolación  $x_0$ . De esta forma, se obtiene:

$$I_0(x) = f(x_0) = c_0.$$

A partir de  $I_0(x)$ , se añade un segundo punto  $x_1$  y se calcula el interpolante  $I_1(x)$  de grado  $\leq 1$  de la forma siguiente:

$$I_1(x) = I_0(x) + \varphi_1(x).$$

Para obtener la función  $\varphi_1(x)$  se imponen las dos condiciones siguientes: debe tratarse de un polinomio de grado  $\leq 1$  dado que  $I_0$  es de grado 0, y debe anularse en el punto  $x_0$  para que se cumpla que  $I_1(x_0) = I_0(x_0)$ . Por tanto, la función  $\varphi_1(x)$  debe tener la expresión siguiente:

$$\varphi_1(x) = c_1(x - x_0),$$

donde  $c_1$  es un coeficiente que se determina a partir de la condición  $I_1(x_1) = f(x_1)$ . Una vez calculado el polinomio  $I_1$ , y a partir del mismo, se incorpora otro punto

de interpolación  $x_2$  y se busca el interpolante  $I_2(x)$  de grado  $\leq 2$ , que tenga la expresión general

$$I_2(x) = I_1(x) + \varphi_2(x).$$

La función  $\varphi_2(x)$  debe tratarse de un polinomio de grado  $\leq 2$  dado que  $I_1$  es de grado  $\leq 1$ , y debe anularse en los puntos  $x_1$  para que se cumpla que  $I_2(x_1) = I_1(x_1)$ , y  $x_0$  para que  $I_2(x_0) = I_0(x_0)$ . A partir de estas condiciones,  $\varphi_2(x)$  tendrá la forma:

$$\varphi_2(x) = c_2(x - x_0)(x - x_1),$$

donde el coeficiente  $c_2$  se determina imponiendo la condición  $I_2(x_2) = f(x_2)$ . Finalmente, el interpolante  $I_N(x)$  de grado  $\leq N$  se obtiene a partir del polinomio  $I_{N-1}$ , añadiendo el punto  $x_N$ . La expresión general para este paso es:

$$I_N(x) = I_{N-1}(x) + \varphi_N(x),$$

donde  $\varphi_N(x)$  debe ser un polinomio de grado  $\leq N$  y debe anularse en los puntos  $x_{N-1}, x_{N-2}, \dots, x_1, x_0$ . La función  $\varphi_N(x)$  que cumple las condiciones citadas anteriormente es:

$$\varphi_N(x) = c_N(x - x_0)(x - x_1) \dots (x - x_{N-2})(x - x_{N-1}).$$

El coeficiente  $c_N$  se determina imponiendo la condición  $I_N(x_N) = f(x_N)$ . Finalmente, se puede expresar la recursión siguiente:

$$I_i(x) = I_{i-1}(x) + \varphi_i(x), \quad (3.11)$$

siendo

$$\varphi_i(x) = c_i(x - x_0)(x - x_1) \dots (x - x_{i-1}) = c_i \prod_{j=0}^{i-1} (x - x_j),$$

donde los coeficientes  $c_i$  se obtienen imponiendo

$$I_i(x_i) = f(x_i), \quad i = 0, \dots, N.$$

Esta recursión permite construir el interpolante de grado  $i$ , en la forma de Newton, a partir del polinomio de Newton de grado  $i - 1$ , cuando se incorpora el punto de interpolación  $x_i$ .

### Ejemplo:

Dada una función  $f$  y 3 puntos de interpolación  $x_0, x_1, x_2$  del dominio de definición, se propone obtener el polinomio interpolante  $I_2$ , de grado  $\leq 2$ , en la forma de Newton, que verifique que

$$I_2(x_0) = f(x_0) = y_0, \quad I_2(x_1) = f(x_1) = y_1, \quad I_2(x_2) = f(x_2) = y_2.$$

Se parte de los datos iniciales  $x_0$  y  $f(x_0)$  y se obtiene el polinomio de grado 0 siguiente:

$$I_0(x) = f(x_0) = c_0.$$

A partir del interpolante anterior, se añade el punto de interpolación  $x_1$  y se calcula el interpolante de grado  $\leq 1$ , aplicando la recursión (3.11). La expresión resultante es:

$$I_1(x) = I_0(x) + c_1(x - x_0),$$

donde  $c_1$  se calcula forzando a que  $I_1(x_1) = f(x_1)$ , de la forma siguiente:

$$\begin{aligned} I_1(x_1) &= f(x_1), \\ I_0(x_1) + c_1(x_1 - x_0) &= f(x_1), \\ c_1 &= \frac{f(x_1) - I_0(x_1)}{x_1 - x_0}, \\ c_1 &= \frac{f(x_1) - f(x_0)}{x_1 - x_0}. \end{aligned}$$

El interpolante de grado  $\leq 1$  resulta:

$$I_1(x) = f(x_0) + \frac{f(x_1) - f(x_0)}{x_1 - x_0} (x - x_0).$$

A partir de  $I_1$ , se incorpora el punto  $x_2$  y, aplicando la recursión (3.11), se construye  $I_2(x)$  de grado  $\leq 2$ :

$$I_2(x) = I_1(x) + c_2(x - x_0)(x - x_1),$$

donde  $c_2$  se calcula forzando a que  $I_2(x_2) = f(x_2)$ , de la forma siguiente:

$$\begin{aligned} I_2(x_2) &= f(x_2), \\ I_1(x_2) + c_2(x_2 - x_0)(x_2 - x_1) &= f(x_2), \\ c_2 &= \frac{f(x_2) - I_1(x_2)}{(x_2 - x_0)(x_2 - x_1)}, \\ c_2 &= \frac{f(x_2) - f(x_0) - \frac{f(x_1) - f(x_0)}{x_1 - x_0} (x_2 - x_0)}{(x_2 - x_0)(x_2 - x_1)}, \\ c_2 &= \frac{\frac{f(x_2) - f(x_1)}{x_2 - x_1} - \frac{f(x_1) - f(x_0)}{x_1 - x_0}}{(x_2 - x_0)}. \end{aligned}$$


---

El interpolante de grado  $\leq 2$  resulta:

$$I_2(x) = f(x_0) + \left[ \frac{f(x_1) - f(x_0)}{x_1 - x_0} \right] (x - x_0) + \left[ \frac{\frac{f(x_2) - f(x_1)}{x_2 - x_1} - \frac{f(x_1) - f(x_0)}{x_1 - x_0}}{(x_2 - x_0)} \right] (x - x_0)(x - x_1).$$

La forma de Newton permite introducir un nuevo concepto que se conoce como *cociente de incrementos* o *diferencias divididas*, que aporta un algoritmo recursivo para la determinación de los coeficientes de interpolación.

*Definición: Diferencia dividida.*

*Dada una función  $f$  y un conjunto de  $N + 1$  puntos distintos  $x_0, x_1, \dots, x_N$  del dominio de definición de  $f$ , se define como diferencia dividida de la función  $f$  en los  $N + 1$  puntos a los coeficientes del interpolante polinómico en la forma de Newton.*

La diferencia dividida de orden 0 de la función  $f$  con respecto al punto  $x_0$  es el valor de la función en dicho punto y se indica de la forma siguiente:

$$f[x_0] = f(x_0).$$

La diferencia dividida de orden 1 de la función  $f$  con respecto a los puntos  $x_0$  y  $x_1$  tiene la expresión siguiente:

$$f[x_0, x_1] = \frac{f(x_1) - f(x_0)}{x_1 - x_0},$$

que es, por definición, el coeficiente del monomio  $x$  del interpolante de Newton (ver ejemplo anterior).

Si se añade el punto siguiente  $x_2$ , la diferencia dividida de orden 2 de la función  $f$  con respecto a los puntos  $x_0, x_1$  y  $x_2$ , resulta:

$$f[x_0, x_1, x_2] = \frac{\frac{f(x_2) - f(x_1)}{x_2 - x_1} - \frac{f(x_1) - f(x_0)}{x_1 - x_0}}{(x_2 - x_0)},$$

que es, por definición, el coeficiente del monomio  $x^2$  del interpolante de Newton (ver ejemplo anterior). Se puede observar que la diferencia dividida de segundo orden, se puede expresar en función de las diferencias divididas  $f[x_0, x_1]$  y  $f[x_1, x_2]$ , de

primer orden, tal como se indica a continuación:

$$f[x_0, x_1, x_2] = \frac{f[x_1, x_2] - f[x_0, x_1]}{(x_2 - x_0)}.$$

De forma recursiva, se pueden ir añadiendo puntos e ir obteniendo diferencias divididas de orden superior, o lo que es equivalente, se pueden ir obteniendo los coeficientes de los monomios de mayor potencia del interpolante correspondiente. Por tanto, a partir de las diferencias divididas de orden  $i - 1$ , se puede construir la diferencia dividida de orden  $i$  mediante la expresión siguiente:

$$f[x_0, x_1, \dots, x_i] = \frac{f[x_1, x_2, \dots, x_i] - f[x_0, x_1, \dots, x_{i-1}]}{x_i - x_0}. \quad (3.12)$$

Obsérvese que si se permutan los argumentos  $x_0$  y  $x_1$  en la diferencia dividida de orden 1, el resultado no cambia. Es decir que:

$$f[x_0, x_1] = f[x_1, x_0].$$

Por otro lado, si se permutan los argumentos  $x_0$ ,  $x_1$  y  $x_2$  en la diferencia dividida de orden 2, el resultado de las  $3! = 6$  permutaciones posibles dan el mismo resultado. Es decir que:

$$\begin{aligned} f[x_0, x_1, x_2] &= f[x_0, x_2, x_1] = f[x_1, x_0, x_2] = \\ f[x_1, x_2, x_0] &= f[x_2, x_0, x_1] = f[x_2, x_1, x_0]. \end{aligned}$$

Se puede demostrar que cualquier permutación de los argumentos no modifica el valor de la diferencia dividida. Es decir que para la diferencia dividida de orden  $N$ , se puede expresar que:

$$f[x_0, x_1, \dots, x_{N-1}, x_N] = f[x_{\alpha_0}, x_{\alpha_1}, \dots, x_{\alpha_N}],$$

donde la serie finita de enteros  $\alpha_0, \alpha_1, \dots, \alpha_N$  es cualquier permutación de  $0, 1, \dots, N$ .

Para demostrar las anteriores propiedades se considera la fórmula  $P_1(x)$  de Newton del polinomio interpolante de grado  $N$  que pasa los puntos:  $x_0, x_1, \dots, x_N$

$$P_1(x) = f[x_0] + f[x_0, x_1](x - x_0) \dots + f[x_0, \dots, x_N](x - x_0) \dots (x - x_{N-1}).$$

Ahora consideramos la siguiente permutación de los valores nodales:  $x_N, x_{N-1}, \dots, x_0$  y construimos el polinomio interpolante  $P_2(x)$

$$P_2(x) = f[x_N] + f[x_N, x_{N-1}](x - x_N) \dots + f[x_N, \dots, x_0](x - x_N) \dots (x - x_1).$$

Por la unicidad del polinomio interpolante  $P_1(x)$  y  $P_2(x)$  deben ser el mismo. Identificando los coeficientes de orden máximo  $x^N$  se obtiene:

$$f[x_0, \dots, x_N] = f[x_N, \dots, x_0]. \quad (3.13)$$



Es decir, la diferencia dividida de orden  $N$  es la misma para la secuencia de puntos  $x_0, \dots, x_N$  y para la secuencia  $x_N, \dots, x_0$  y, en general, para cualquier otra permutación.

Si identificamos ahora el coeficiente de orden  $x^{N-1}$ , se obtiene:

$$\begin{aligned} f[x_0, x_1, \dots, x_{N-1}] - f[x_0, x_1, \dots, x_N](x_0 + x_1 + \dots + x_{N-1}) = \\ f[x_N, x_{N-1}, \dots, x_1] - f[x_N, x_{N-1}, \dots, x_0](x_N + x_{N-1} + \dots + x_1). \end{aligned}$$

Si utilizamos la propiedad (3.13) y despejamos la diferencia dividida de orden  $N$  en función de la diferencia dividida de orden  $N - 1$ , se obtiene

$$f[x_0, \dots, x_N] = \frac{f[x_1, \dots, x_N] - f[x_0, \dots, x_{N-1}]}{x_N - x_0}. \quad (3.14)$$

Por último, dados  $N + 1$  puntos distintos  $x_0, x_1, \dots, x_N$  del dominio de definición de una función  $f(x)$  y si se definen las diferencias divididas de orden cero como  $f[x_k] = f(x_k)$ ,  $k = 0, 1, \dots, N$ , empleando los conceptos explicados anteriormente, se puede construir la tabla de diferencias divididas siguiente:

$x_0$	$f[x_0]$				
		$f[x_0, x_1]$			
$x_1$	$f[x_1]$		$f[x_0, x_1, x_2]$		
		$f[x_1, x_2]$		$\ddots$	
$x_2$	$f[x_2]$		$f[x_1, x_2, x_3]$		
		$f[x_2, x_3]$		$\ddots$	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\dots$	$f[x_0, x_1, \dots, x_N]$
		$f[x_{N-3}, x_{N-2}]$			
$x_{N-2}$	$f[x_{N-2}]$		$f[x_{N-3}, x_{N-2}, x_{N-1}]$		
		$f[x_{N-2}, x_{N-1}]$		$\ddots$	
$x_{N-1}$	$f[x_{N-1}]$		$f[x_{N-2}, x_{N-1}, x_N]$		
		$f[x_{N-1}, x_N]$		$\ddots$	
$x_N$	$f[x_N]$				

(3.15)

La columna primera indica el conjunto de  $N + 1$  puntos distintos del dominio de definición de  $f$ , y cada columna, a partir de la segunda, representa las diferencias divididas de orden 0, 1, 2 hasta orden  $N$ . Moviéndose a través de la diagonal de la tabla de diferencias divididas se puede obtener la diferencia dividida de orden  $i$  a partir de las diferencias divididas de orden  $i - 1$ .

Finalmente, el interpolante de grado  $\leq N$ , en la forma de Newton, se puede expresar en la forma siguiente:

$$I_N(x) = f[x_0] + f[x_0, x_1] (x - x_0) + f[x_0, x_1, x_2] (x - x_0)(x - x_1) + \cdots + f[x_0, x_1, \dots, x_N] (x - x_0)(x - x_1)(x - x_2) \dots (x - x_{N-2})(x - x_{N-1}),$$

donde los coeficientes han sido reemplazados por las diferencias divididas del orden correspondiente que obtienen de forma recursiva mediante la expresión (3.14).

Es importante hacer notar que con la fórmula de Newton, el polinomio de grado  $N$  se construye añadiendo un punto  $(x_N, f(x_N))$  al polinomio de grado  $N - 1$ . Es decir,

$$I_N(x) = I_{N-1}(x) + f[x_0, x_1, \dots, x_N] (x - x_0) \dots (x - x_{N-1}). \quad (3.16)$$

### 3.6. Error de interpolación

Una vez analizadas las tres posibles formas de un polinomio interpolante, resulta de gran importancia conocer la diferencia que existe entre la función  $f(x)$  a interpolar y sus derivadas y el interpolante  $I_N(x)$  en el dominio de definición de la función.

De forma general, el error total al aproximar una función proviene de dos fuentes o componentes de error que se explican a continuación. Por un lado, existe un error asociado al método de aproximación empleado que se conoce como *error de truncamiento*. Este error se puede analizar cuestionando la igualdad siguiente:

$$\lim_{N \rightarrow \infty} I_N(x) = f(x). \quad (3.17)$$

Por otro lado, existe un error asociado a la sensibilidad del método de aproximación a perturbaciones que se conoce como condicionamiento del método de aproximación. Al implementar un método de aproximación en el ordenador, aparecen perturbaciones asociadas a la aritmética de precisión finita por lo que este error se lo conoce como *error de redondeo*. Por lo tanto, el error total de aproximación se expresa de la forma siguiente:

$$E_{\text{total}} = R_N(x) + R_L(x), \quad (3.18)$$

donde  $R_N(x)$  es el error de truncamiento y  $R_L(x)$  es el error de redondeo asociado a las perturbaciones de la función  $f(x)$  a interpolar.

En las secciones siguientes se estudian con detalle el error de truncamiento a partir del teorema del error de interpolación y el error de redondeo a partir de la función de Lebesgue.

### 3.6.1. Error de truncamiento

Se pretende obtener una expresión analítica para el error de truncamiento que permita analizar las diferentes fuentes de error. El siguiente teorema nos relaciona la nube de puntos y la regularidad de la función con el error de truncamiento.

*Teorema: Error de truncamiento.*

Sea  $f$  una función de clase  $C^{N+1}$ , definida en un intervalo  $[a, b]$  y sean  $x_0, x_1, \dots, x_N$ ,  $N+1$  puntos de interpolación distintos, pertenecientes al intervalo  $[a, b]$ , de los que se obtiene el polinomio interpolante  $I_N$  de grado  $\leq N$ . Entonces, a cada punto  $x$  perteneciente al intervalo  $[a, b]$ , le corresponde un punto  $\xi$ , que corresponde al intervalo:

$$\min(x_0, x_1, \dots, x_N, x) < \xi < \max(x_0, x_1, \dots, x_N, x), \quad (3.19)$$

para el que se cumple la expresión siguiente que corresponde al error de interpolación y que se indica con  $R_N(x)$ :

$$R_N(x) = f(x) - I_N(x) = \pi_{N+1}(x) \frac{f^{(N+1)}(\xi)}{(N+1)!}, \quad (3.20)$$

donde  $\pi_{N+1}$  es la función de error dada por :

$$\pi_{N+1}(x) = (x - x_0)(x - x_1) \dots (x - x_N). \quad (3.21)$$

Es importante destacar que, a priori, el punto  $\xi$  se desconoce y que es función de  $x$ , es decir que  $\xi = \xi(x)$ .

*Demostración:*

Para demostrar el teorema del error de interpolación, se plantean dos pasos previos, que se indican a continuación. En un primer paso se demostrará que el error de interpolación de grado  $N$  se puede expresar a partir de las diferencias divididas mediante la expresión siguiente:

$$R_N(x) = \pi_{N+1}(x) f[x, x_0, x_1, \dots, x_N], \quad (3.22)$$

y en un segundo paso, se demostrará que:

$$f[x, x_0, x_1, \dots, x_N] = \frac{f^{(N+1)}(\xi)}{(N+1)!}. \quad (3.23)$$

Para comprobar (3.22) se demuestra mediante la relación de recurrencia (3.16) que el error de interpolación  $R_N(x)$  es:

$$R_N(x) = \pi_{N+1}(x) f[x, x_0, x_1, \dots, x_N]. \quad (3.24)$$

Se parte de  $N + 1$  puntos de interpolación  $x_0, \dots, x_N$  y se construye el interpolante de grado  $N$ . A continuación, se añade el punto genérico  $(t, f(t))$  y se construye el interpolante mediante la relación de recurrencia (3.16),

$$I(x) = I_N(x) + f[x_0, x_1, \dots, x_N, t] (x - x_0) \dots (x - x_{N-1})(x - x_N). \quad (3.25)$$

Como el punto añadido es genérico el interpolante con  $t = x$  coincide con la función a interpolar ( $I(x) = f(x)$ ). Así, la diferencia entre la función  $f(x)$  y el interpolante  $I_N(x)$  es el error de truncamiento

$$R_N(x) = f[x_0, x_1, \dots, x_N, x] (x - x_0) \dots (x - x_{N-1})(x - x_N),$$

que es la expresión que queríamos demostrar.

En un segundo paso, se pretende demostrar (3.23), es decir que la diferencia dividida del paso anterior es:

$$f[x, x_0, x_1, \dots, x_N] = \frac{f^{N+1}(\xi)}{(N+1)!}. \quad (3.26)$$

En principio se sabe que:

$$R_N(x) = \pi_{N+1}(x) f[x, x_0, x_1, \dots, x_N] = f(x) - I_N(x). \quad (3.27)$$

Para esta demostración, se construye una función  $Q(x)$  a partir de la expresión siguiente:

$$Q(x) = f(x) - I_N(x) - \pi_{N+1}(x) G(t), \quad (3.28)$$

donde  $G(t)$  es:

$$G(t) = f[t, x_0, \dots, x_N]. \quad (3.29)$$

A continuación se analizan los ceros de la función  $Q$  y de sus derivadas sucesivas.

La función  $Q$  tiene  $N + 2$  ceros que se describen a continuación:

$$\begin{cases} N + 1 \text{ ceros en: } & x_j, \quad j = 0, \dots, N, \\ 1 \text{ cero en :} & x = t. \end{cases} \quad (3.30)$$

Los  $N + 1$  primeros ceros provienen de los  $N + 1$  puntos de interpolación

$$x_j, \quad j = 0, \dots, N.$$

En ellos, por definición de interpolación, se cumple que:

$$f(x_j) - I_N(x_j) = 0. \quad (3.31)$$

Por otro lado, la función  $\pi_{N+1}(x_j) = 0$ . Finalmente, la función  $Q$  se anula en los  $N + 1$  puntos de interpolación:

$$Q(x_j) = f(x_j) - I_N(x_j) - \pi_{N+1}(x_j) G(t) = 0, \quad j = 0, \dots, N. \quad (3.32)$$

Además, la función  $Q$  tiene un cero en  $x = t$ , por definición de interpolación:

$$Q(t) = f(t) - I_N(t) - \pi_{N+1}(t) G(t) = 0. \quad (3.33)$$

Como la función  $Q(x)$  tiene  $N + 2$  ceros, el teorema generalizado de Rolle permite asegurar que la derivada  $N + 1$  de  $Q(x)$  tiene un cero en un punto intermedio.

Derivando consecutivamente con respecto a  $x$ , se observa que  $Q'$  tiene  $N + 1$  ceros,  $Q''$  tiene  $N$  ceros,  $Q'''$  tiene  $N - 1$  ceros y así hasta la derivada de orden  $N + 1$  de  $Q$  que tiene un cero con la expresión siguiente:

$$Q^{(N+1)}(x) = f^{(N+1)}(x) - I_N^{(N+1)}(x) - (N + 1)! G(t). \quad (3.34)$$

Como  $I_N^{(N+1)} = 0$ , la función  $Q^{(N+1)}$  queda:

$$Q^{(N+1)}(x) = f^{(N+1)}(x) - (N + 1)! f[t, x_0, x_1, \dots, x_N]. \quad (3.35)$$

La expresión anterior tiene un cero en algún valor  $x = \xi$ , que permite expresar que:

$$f^{(N+1)}(\xi) = (N + 1)! G(t) = (N + 1)! f[t, x_0, x_1, \dots, x_N], \quad (3.36)$$

donde  $\xi$  es función de  $t, x_0, x_1, \dots, x_N$ . De esta forma queda demostrado (3.23)

$$f[x, x_0, x_1, \dots, x_N] = \frac{f^{(N+1)}(\xi)}{(N + 1)!}. \quad (3.37)$$

A partir de los resultados (3.22) y (3.23) se demuestra el teorema del error de interpolación:

$$R_N(x) = \pi_{N+1}(x) \frac{f^{(N+1)}(\xi)}{(N + 1)!}, \quad (3.38)$$

donde  $\xi$  pertenece al intervalo:

$$\min(x_0, x_1, \dots, x_N, x) < \xi < \max(x_0, x_1, \dots, x_N, x).$$

De la expresión del error de truncamiento (3.38) se observa que el error está relacionado con dos cuestiones, una de ellas es la regularidad de la función a interpolar y la otra son los puntos de interpolación. Con respecto a la regularidad de la función, se observa que la misma debe ser derivable hasta el orden  $N + 1$ . Independientemente de su regularidad, si  $f^{N+1}$  es grande o pequeña, el error de interpolación será grande o pequeño. En particular, si  $f^{N+1}$  no está acotada con  $N$  en el dominio de definición, el error de interpolación resultará no acotado. Con respecto a los puntos de interpolación, se observa que afectan al error de interpolación en el número de puntos a través de  $(N + 1)!$  y en la distribución de puntos de interpolación a través de la función  $\pi_{N+1}(x)$ .

### 3.6.2. Error de redondeo

Siempre que interpolamos una función aparecen errores de redondeo asociados a la precisión del ordenador. Si trabajamos con precisión simple estos errores son del  $O(10^{-7})$  y con precisión doble de  $O(10^{-15})$ . Este error que introduce el ordenador por la precisión finita del mismo se le conoce como error de redondeo o *round-off* y se le denomina por la letra  $\epsilon$ . De esta manera, al introducir en el ordenador para el cálculo del interpolante la evaluación de la función  $f(x)$  en el punto nodal  $x_j$  aparece un error de redondeo que denominamos por la variable  $\epsilon_j$ . Es decir,  $\epsilon_j$  es la indeterminación que tenemos en el ordenador al evaluar la función en el nodo  $x_j$ . Sin embargo, estos valores se pueden acotar por el valor  $\epsilon$  asociado a la precisión con la que trabajemos.

De esta manera, en el interpolante sustituimos  $f(x_j)$  por  $f(x_j) + \epsilon_j$  para tener en cuenta precisión finita del ordenador. Esta indeterminación la denominamos por error de redondeo y tiene la expresión:

$$R_L(x) = \epsilon_0 \ell_0(x) + \epsilon_1 \ell_1(x) + \dots + \epsilon_N \ell_N(x). \quad (3.39)$$

Aunque no podamos conocer el error de interpolación, nos interesa acotar este error para saber si el problema de interpolación está bien condicionado. Si tomamos el valor absoluto en (3.39), se obtiene la siguiente expresión mediante la desigualdad triangular:

$$|R_L(x)| \leq |\epsilon_0| |\ell_0(x)| + \dots + |\epsilon_N| |\ell_N(x)|. \quad (3.40)$$

Si consideramos que todos los errores de redondeo  $\epsilon_j$  de cada nodo  $x_j$  están acotados por  $\epsilon$ , entonces

$$|R_L(x)| \leq \epsilon \lambda_N(x), \quad (3.41)$$

donde  $\lambda_N(x)$  es la función de Lebesgue definida como:

$$\lambda_N(x) = \sum_{j=0}^N |\ell_j(x)|. \quad (3.42)$$

Si ahora calculamos el máximo del error de interpolación en todo el dominio de interpolación, la cota superior viene determinada por la constante de Lebesgue.

$$\max_{x \in [a,b]} |R_L(x)| \leq \epsilon \Lambda_N, \quad (3.43)$$

donde  $\Lambda_N$  es la constante de Lebesgue dada por:

$$\Lambda_N = \max_{x \in [a,b]} \lambda_N(x). \quad (3.44)$$

La constante de Lebesgue también está relacionada con la mejor aproximación que podemos hacer mediante un interpolante.

---

*Teorema: Error de interpolación con respecto al mejor interpolante.*

Sea  $\Lambda_N$  la constante de Lebesgue dada por (3.44),  $f$  una función definida en  $[a, b]$ ,  $I_N$  el polinomio interpolante correspondiente y  $I_N^*$  el mejor polinomio interpolante. Entonces,

$$\|f - I_N\| \leq (\Lambda_N + 1) \|f - I_N^*\|. \quad (3.45)$$

Si la constante de Lebesgue de nuestro conjunto de puntos  $x_j, j = 0, \dots, N$  es grande, el de error interpolación con respecto al error del interpolante óptimo es grande.

*Demostración:*

En primer lugar tenemos que relacionar la norma del interpolante con la constante de Lebesgue mediante su definición. Si  $I_N$  es un interpolante de  $f$  de grado  $N$ , mediante la definición (3.44) entonces  $\|I_N\| \leq \|f\| \Lambda_N$ . Si expresamos  $f - I_N$  como  $f - I_N^* - I_N + I_N^*$  y utilizamos la desigualdad triangular, entonces

$$\|f - I_N\| \leq \|f - I_N^*\| + \|I_N - I_N^*\|.$$

El polinomio  $Q_N = I_N - I_N^*$  es un interpolante  $f - I_N^*$  porque

$$Q_N(x_j) = I_N(x_j) - I_N^*(x_j) = f(x_j) - I_N^*(x_j).$$

De esta manera y mediante la definición de la constante de Lebesgue

$$\|I_N - I_N^*\| \leq \Lambda_N \|f - I_N^*\|,$$

y

$$\|f - I_N\| \leq (\Lambda_N + 1) \|f - I_N^*\|, \quad (3.46)$$

con lo que el teorema queda demostrado.

### 3.6.3. Acotación del error de truncamiento y redondeo

Una vez analizado el origen y las expresiones analíticas para el error de interpolación, pasamos a estudiar su comportamiento con las distribuciones nodales de puntos y con el número de puntos  $N$ . La expresión para el error total de interpolación es:

$$E_N(x) = \pi_{N+1}(x) \frac{f^{(N+1)}(\xi)}{(N+1)!} + \sum_{j=0}^N \epsilon_j \ell_j(x), \quad (3.47)$$

donde el primer término es el error de truncamiento  $R_N(x)$  y el segundo el error de redondeo  $R_L(x)$ . Tomando la norma del supremo en esta expresión y utilizando la desigualdad triangular, se obtiene:

$$\|E_N\| \leq \|\pi_{N+1}\| \frac{\|f^{(N+1)}(\xi)\|}{(N+1)!} + \|\epsilon\| \Lambda_N. \quad (3.48)$$

Antes de estudiar el comportamiento de los términos anteriores hacemos un cambio de variable  $x$  independiente para transformar el dominio  $[a, b]$  en el dominio  $[-1, 1]$  mediante la siguiente expresión:

$$x = \frac{a+b}{2} + \frac{b-a}{2}t, \quad (3.49)$$

donde  $t \in [-1, 1]$ . Tanto en los polinomios de Lagrange como en la función de error  $\pi_{N+1}$  aparecen los factores  $x - x_j$  que al hacer el cambio de variable quedan:

$$x - x_j = \frac{b-a}{2}(t - t_j), \quad (3.50)$$

siendo  $t_j$  una partición del intervalo  $[-1, 1]$ . Así, la función de error  $\pi_{N+1}$  queda

$$\pi_{N+1}(t) = (t - t_0)(t - t_1) \dots (t - t_N) \left( \frac{b-a}{2} \right)^{N+1}, \quad (3.51)$$

y los polinomios de Lagrange

$$\ell_j(t) = \frac{(t - t_0) \dots (t - t_{j-1})(t - t_{j+1}) \dots (t - t_N)}{(t_j - t_0) \dots (t_j - t_{j-1})(t_j - t_{j+1}) \dots (t_j - t_N)} \quad (3.52)$$

Por otra parte, la derivada  $N + 1$  veces de  $f$  con respecto a  $t$  queda:

$$\frac{d^{N+1}f}{dt^{N+1}} = \frac{d^{N+1}f}{dx^{N+1}} \left( \frac{2}{b-a} \right)^{N+1}. \quad (3.53)$$

Al llevar estas expresiones al error de truncamiento, el segundo factor de (3.53) se cancela con el segundo factor de (3.51) quedando las mismas expresiones que en la variable independiente  $x$ . Por lo tanto, sin ninguna pérdida de generalidad se puede suponer que en todo el análisis que sigue el dominio de interpolación es el  $[-1, 1]$ .

A la vista de la expresión (3.48) se plantean las siguientes cuestiones:

1. Es necesario conocer el comportamiento de  $\|\pi_{N+1}\|$  con  $N \rightarrow \infty$  y la distribución de la nube de puntos.
2. Cuando se interpola con alto orden  $N$ , se debe exigir la existencia al menos de la derivada  $N + 1$  de la función a interpolar.
3. Es necesario conocer el comportamiento de la constante de Lebesgue  $\Lambda_N$  con  $N$  y con la distribución de puntos.

Si los puntos de interpolación están equiespaciados a una distancia  $\Delta x$  en el dominio  $[-1, 1]$ , veremos en las siguientes secciones que la función de error  $\pi_{N+1}$  está acotada por

$$\|\pi_{N+1}\| \leq \frac{\Delta x^{N+1}}{4} N! = \left( \frac{2}{e} \right)^N \sqrt{\frac{\pi}{2N}},$$



que tiende a cero con  $N \rightarrow \infty$ . Sin embargo, la constante de Lebesgue se puede acotar inferior y superiormente por

$$\frac{2^{N-2}}{N^2} < \Lambda_N < \frac{2^{N+3}}{N},$$

y se observa que no tiende a cero con  $N \rightarrow \infty$ . Esta característica asociada a las distribuciones de puntos equiespaciadas invalida su uso en las interpolaciones de alto orden.

En la sección siguiente se analizará el comportamiento del error con diferentes distribuciones de puntos y se llegará a la conclusión de que para paliar el efecto anteriormente descrito se deben concentrar puntos en los extremos del intervalo de interpolación.

## 3.7. Distribuciones equiespaciadas de puntos

En este caso, se considera una función  $f$  y  $N + 1$  puntos del dominio de definición de  $f$  a partir de los cuales se construye el interpolante  $I_N$ . Se propone estudiar el comportamiento del error de interpolación a través de  $\pi_{N+1}(x)$  y de  $\Lambda_N$  con  $N \rightarrow \infty$  y una distribución equiespaciada de puntos.

### 3.7.1. Error de truncamiento

En este apartado se estudia el comportamiento del error de interpolación cuando el número de puntos tiende a infinito, con una distribución equiespaciada de los mismos. En particular, se explica por qué el error de interpolación es mayor en los extremos del intervalo de interpolación y se determina cuál es el valor máximo de este error. Se plantea una distribución equiespaciada de  $N + 1$  puntos de interpolación donde la distancia entre puntos consecutivos es  $\Delta x$ . Para dos puntos consecutivos genéricos  $(x_j, x_{j+1})$  la distancia tiene la expresión siguiente:

$$\Delta x = x_{j+1} - x_j.$$

Esta distribución se muestra en la figura (3.3). Del teorema del error de interpolación se tiene que  $\pi_{N+1}(x)$  es:

$$\pi_{N+1}(x) = (x - x_0)(x - x_1) \cdots (x - x_N). \quad (3.54)$$

Se consideran dos puntos de interpolación consecutivos genéricos  $x_j$  y  $x_{j+1}$ . Suponiendo que  $\pi_{N+1}(x)$  alcanza un extremo en un valor de  $x$  que pertenece al intervalo  $(x_j, x_{j+1})$ , se propone demostrar que:

$$|(x - x_j)(x - x_{j+1})| \leq \frac{\Delta x^2}{4}. \quad (3.55)$$

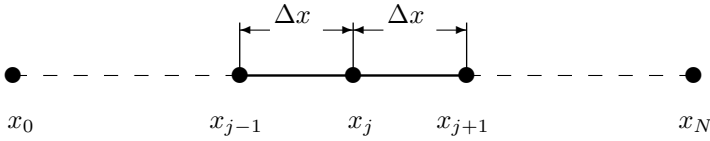


Figura 3.3: Distribución equiespaciada de puntos de interpolación

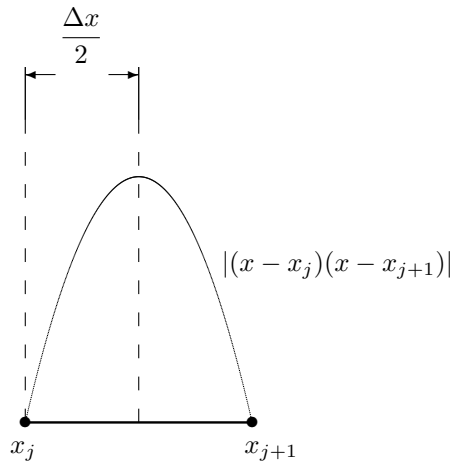


Figura 3.4: Representación de  $|\pi_{N+1}(x)|$  entre los puntos de interpolación  $x_j$  y  $x_{j+1}$

Para ello, se busca el extremo de la función  $y = (x - x_j)(x - x_{j+1})$ :

$$\begin{aligned} y &= (x - x_j)(x - x_{j+1}), \\ y' &= x - x_{j+1} + x - x_j = 0, \end{aligned}$$

$$\begin{aligned} x_{\text{extremo}} &= \frac{1}{2}(x_j + x_{j+1}), \\ y_{\text{extremo}} &= \frac{1}{2}(x_{j+1} - x_j) \frac{1}{2}(x_j - x_{j+1}) = -\frac{\Delta x^2}{4}. \end{aligned}$$

Se observa que el extremo de  $\pi_{N+1}(x)$ , en caso de obtenerse en el intervalo  $(x_j, x_{j+1})$ , se alcanza en el punto medio de intervalo y su valor es  $\Delta x^2/4$  (ver figura 3.4). Aplicando módulos a la expresión anterior, se demuestra (3.55).

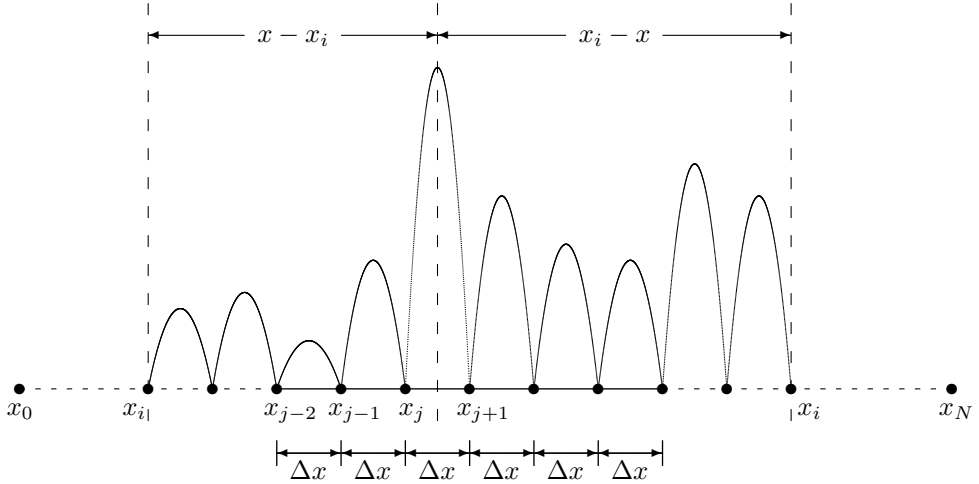


Figura 3.5: Extremos de  $|\pi_{N+1}(x)|$  en un caso general de una interpolación global equiespaciada

A continuación, se plantea el caso general de una interpolación global tal como se representa en la figura (3.5). Observando la misma se puede comprobar que:

$$\begin{aligned} |x - x_i| &\leq (j - i + 1)\Delta x, & \text{para } i < j, \\ |x - x_i| &\leq (i - j)\Delta x, & \text{para } i > j + 1. \end{aligned} \quad (3.56)$$

En la figura (3.5) el índice  $i$  indica puntos de interpolación genéricos por delante y por detrás del intervalo  $(x_j, x_{j+1})$  donde se supone que existe el máximo de  $|\pi_{N+1}(x)|$ . Aplicando módulos a la expresión (3.54), se tiene que:

$$|\pi_{N+1}(x)| \leq |(x - x_0)| \cdots |(x - x_{j-1})| \frac{\Delta x^2}{4} |(x - x_{j+2})| \cdots |(x - x_{N-1})| |(x - x_N)|,$$

y haciendo uso de las desigualdades (3.56), se llega a que:

$$|\pi_{N+1}(x)| \leq (j + 1) j (j - 1) \cdots 2 \Delta x^j \frac{\Delta x^2}{4} 2 \cdot 3 \cdots (N - j) \Delta x^{N-j-1},$$

$$|\pi_{N+1}(x)| \leq (j + 1)! (N - j)! \frac{\Delta x^{N+1}}{4}.$$

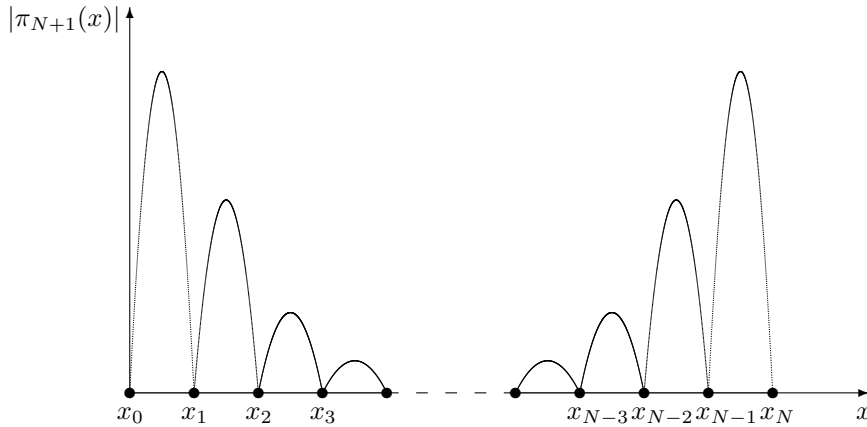


Figura 3.6:  $|\pi_{N+1}(x)|$  en una interpolación global equiespaciada

La expresión anterior da una cota de  $|\pi_{N+1}(x)|$ , si se considera que el máximo está en el intervalo  $(x_j, x_{j+1})$ . Si ahora se mueve el índice  $j$  desde  $j = 0, 1, \dots, N-1$ , se puede comprobar que los valores máximos se presentan en los extremos del dominio de interpolación, correspondiendo a los índices  $j = 0$  y  $j = N-1$ . Es decir, los máximos de  $|\pi_{N+1}(x)|$  se alcanzan en los intervalos  $(x_0, x_1)$  y  $(x_{N-1}, x_N)$  y están acotados por:

$$\max_{x \in [a, b]} |\pi_{N+1}(x)| \leq N! \frac{\Delta x^{N+1}}{4}. \quad (3.57)$$

Utilizando la fórmula de Stirling para el factorial de  $N$  y sustituyendo  $\Delta x$  por  $2/N$ , se obtiene:

$$|\pi_{N+1}(x)| \leq \left(\frac{2}{e}\right)^N \sqrt{\frac{\pi}{2N}}. \quad (3.58)$$

El resultado (3.58) tiene una gran importancia ya que permite conocer una cota del error de interpolación en función del número de puntos elegido. La cota superior de la función de error  $\pi_{N+1}$  tiende a cero con  $N \rightarrow \infty$ . De forma esquemática, en la figura (3.6) se representa  $|\pi_{N+1}(x)|$  para una interpolación equiespaciada genérica donde puede observarse que los errores son importantes en los extremos del dominio de interpolación y son pequeños en las proximidades del punto medio de dicho dominio.

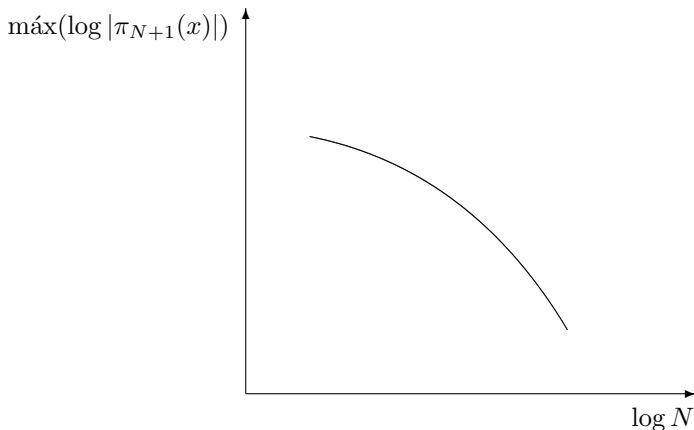


Figura 3.7: Convergencia de  $|\pi_{N+1}(x)|$  en función de  $N$  para una interpolación global equiespaciada

Tomando logaritmos en la expresión (3.58) se obtiene:

$$\log \|\pi_{N+1}(x)\| \leq -\frac{1}{2} \log N - N \log \left( \frac{e}{2} \right) + \frac{\log(2\pi)}{2}.$$

De forma esquemática, la expresión anterior tiene la representación gráfica que se indica en la figura (3.7) y que representa la convergencia de  $\|\pi_{N+1}(x)\|$  con el número de puntos  $N$  en una interpolación equiespaciada.

Una vez estudiado el comportamiento de  $\pi_{N+1}(x)$ , se estudia una cota del error de interpolación  $R_N(x)$ . Mediante la expresión (3.48) y la acotación (3.57), se obtiene la desigualdad siguiente:

$$\|R_N(x)\| \leq \frac{\Delta x^{N+1}}{4} \frac{|f^{(N+1)}(\xi)|}{N+1}. \quad (3.59)$$

Si la función  $f$  a interpolar tiene derivadas de cualquier orden acotadas, como por ejemplo es la función  $f(x) = \sin(x)$ , y se estudia el límite del módulo del error con  $N \rightarrow \infty$  se tiene que:

$$\lim_{N \rightarrow \infty} |R_N(x)| \leq \lim_{N \rightarrow \infty} \left[ \frac{\Delta x^{N+1}}{4} \frac{1}{(N+1)} \right] = 0.$$

Es decir, el error de truncamiento tiende a cero cuando  $N \rightarrow \infty$ . Si por el contrario, se interpola una función que tiene derivadas no acotadas como por ejemplo es la

función de Runge, que tiene la expresión siguiente:

$$f(x) = \frac{1}{1 + 25x^2}, \quad x \in [-1, 1],$$

se tiene

$$\lim_{N \rightarrow \infty} |R_N(x)| = \infty,$$

porque las derivadas  $f^{(N+1)}$  para la función de Runge no están acotadas cuando  $N \rightarrow \infty$ .

### 3.7.2. Error de redondeo

Como hemos comentado con anterioridad, el error de redondeo  $R_L(x)$  está asociado a la precisión finita del ordenador y se puede medir a través de la constante de Lebesgue

$$\Lambda_N = \max_{x \in [a, b]} \sum_{j=0}^N |\ell_j(x)|.$$

Cuando seleccionamos la nube de puntos  $x_0, \dots, x_N$  nos interesa: (i) que la constante de Lebesgue sea lo menor posible para que la aproximación esté próxima a la óptima (3.46) y (ii) que el problema esté bien condicionado. Es decir, que una perturbación pequeña  $\epsilon$  de los valores de la función a interpolar no introduzcan grandes variaciones en el interpolante.

Procedemos a buscar una cota inferior para la constante de Lebesgue. Como la función de Lebesgue  $\lambda_N(x)$  es una suma de  $N + 1$  términos en valor absoluto, podemos acotarla inferiormente mediante el término central del sumatorio. Es decir,

$$\lambda_N(x) = |\ell_0(x)| + \dots + |\ell_{N/2}(x)| + \dots + |\ell_N(x)| \geq |\ell_{N/2}(x)|. \quad (3.60)$$

Por otra parte, si la constante de Lebesgue  $\Lambda_N$  es el máximo de la función de Lebesgue  $\lambda_N(x)$  en  $x \in [a, b]$ , podemos calcular el máximo en la expresión anterior y acotar particularizando por el valor medio  $x = a + \Delta x/2$

$$\max_{x \in [a, b]} \lambda_N(x) \geq \max_{x \in [a, b]} |\ell_{N/2}(x)| \geq |\ell_{N/2}(a + \frac{\Delta x}{2})|. \quad (3.61)$$

Consideramos la expresión del polinomio de Lagrange

$$\ell_j(x) = \frac{(x - x_0) \dots (x - x_{j-1})(x - x_{j+1}) \dots (x - x_N)}{(x_j - x_0) \dots (x_j - x_{j-1})(x_j - x_{j+1}) \dots (x_j - x_N)}, \quad (3.62)$$

con una distribución equiespaciada de puntos

$$x_j = a + \Delta x \, j, \quad j = 0, \dots, N,$$

con  $\Delta x = (b - a)/N$ . Si definimos  $y = (x - x_0)/\Delta x$ , la expresión (3.62) para una distribución equiespaciada queda:

$$\ell_j(y) = \frac{y(y-1) \dots (y-j+1)(y-j-1) \dots (y-N)}{j(j-1) \dots 1 \, 1 \dots (j-N)}, \quad (3.63)$$

con  $y \in [0, N]$ . Si en (3.63) hacemos  $y = 1/2$ , que es equivalente a  $x = a + \Delta x/2$  en la ecuación (3.62), con  $j = N/2$ , se obtiene:

$$\ell_{N/2}(1/2) = \frac{\frac{1}{2}(-\frac{1}{2})(-\frac{3}{2})(-\frac{5}{2}) \dots (-\frac{2N-1}{2})}{(\frac{N}{2})! (\frac{N}{2})!}. \quad (3.64)$$

Es importante hacer notar que en el numerador de la expresión anterior no está el factor  $(N-1)/2$ . De esta forma, el valor absoluto de la expresión (3.65) queda como:

$$|\ell_{N/2}(1/2)| = \frac{2 \, (2N-1)(2N-3) \dots 1}{2^N (N-1) (\frac{N}{2})! (\frac{N}{2})!}. \quad (3.65)$$

El numerador está formado por el producto de los  $N-1$  factores impares que se puede poner como

$$(2N-1)(2N-3)(2N-5) \dots 1 = \frac{(2N-1)(2N-2)(2N-3)(2N-4) \dots 1}{(2N-2)(2N-4) \dots 2}. \quad (3.66)$$

O lo que es lo mismo

$$(2N-1)(2N-3)(2N-5) \dots 1 = \frac{(2N-1)!}{2^{N-1}(N-1)!} = \frac{(2N)!}{2^N N!}. \quad (3.67)$$

Utilizando la expresión (3.67) y la fórmula de Stirling  $N! \sim \sqrt{2\pi N}(N/e)^N$ , la cota (3.65) queda:

$$|\ell_{N/2}(1/2)| = \frac{2^N \sqrt{2}}{\pi N(N-1)}. \quad (3.68)$$

De esta manera la constante de Lebesgue es

$$\Lambda_N \geq \frac{2^N \sqrt{2}}{\pi N(N-1)}. \quad (3.69)$$

Incluso para valores moderados de  $N$ , la constante de Lebesgue puede ser muy grande haciendo que la interpolación en una malla equiespaciada presente errores muy grandes. En conclusión, este valor de la constante de Lebesgue para mallas equiespaciadas invalida el uso de interpolantes de alto orden en estas distribuciones de puntos.

### 3.8. Distribuciones de puntos no equiespaciados

Una distribución equiespaciada de puntos de interpolación da lugar a una constante de Lebesgue muy grande (3.69) que hace que el problema esté mal condicionado. Es decir, un pequeño error en las evaluaciones de la función a interpolar origina un error de redondeo muy grande de interpolación. Tanto la función de Lebesgue como la función de error  $\pi_{N+1}(x)$  presentan sus máximos en las proximidades del contorno. A diferencia de la constante de Lebesgue que tiende a infinito con  $N$  tendiendo a infinito, el máximo de la función de error  $\pi_{N+1}(x)$  tiende a cero con  $N$  tendiendo a infinito. La pregunta que surge ahora es si es posible encontrar distribuciones de puntos que tengan valores de la función de Lebesgue próximos a la unidad para todo el dominio. La respuesta a esta pregunta es afirmativa y pasa por concentrar puntos en los extremos del intervalo. Existen tres formas de obtener puntos de interpolación que consiguen el comportamiento deseado:

1. Puntos de Gauss o Gauss–Lobatto asociados a bases polinómicas que gocen de precisión espectral (secciones 2.5–2.6). Los puntos de Gauss son los ceros de la función polinómica de la base  $\Phi_{N+1}(x)$  y los puntos de Gauss–Lobatto que incluyen los extremos  $x = \pm 1$  son los ceros de una combinación lineal de  $\Phi_{N+1}(x)$ ,  $\Phi_N(x)$  y  $\Phi_{N-1}(x)$ . Los puntos de Chebyshev–Gauss o Chebyshev–Gauss–Lobatto son ejemplos de estos puntos de interpolación basados en los polinomios base de Chebyshev.
2. Puntos de interpolación que minimicen la función de error  $\pi_{N+1}(x)$ . En el capítulo siguiente se muestra que los puntos de interpolación que minimizan la función de error  $\pi_{N+1}(x)$  son los puntos de Chebyshev–Gauss.
3. Puntos de interpolación que minimicen la función de Lebesgue  $\lambda_N(x)$ . El conjunto de puntos de interpolación que minimiza la constante de Lebesgue se dice que es óptimo y se puede determinar numéricamente. Este conjunto tiene una constante de Lebesgue ligeramente inferior a los puntos de Chebyshev y su comportamiento desde el punto de vista de error es muy similar. En su contra, los puntos de Chebyshev tienen una transformada rápida de Fourier que les hace mas atractivos.

En cualquiera de los tres procedimientos anteriores, la nube de puntos de interpolación verifica las siguientes propiedades:

1. Las distribuciones de puntos concentran puntos en los extremos  $x = \pm 1$ .
2. Los máximos de la función de error  $\pi_{N+1}(x)$  son similares en todo el dominio de interpolación  $x \in [-1, +1]$ .
3. Los máximos de la función de Lebesgue  $\lambda_N(x)$  son similares en todo el dominio y la constante de Lebesgue disminuye dando lugar a un problema de interpolación bien condicionado.



# Capítulo 4

## Interpolantes de Fourier y Chebyshev

En este capítulo se analizan las distribuciones de puntos nodales o de interpolación que permiten construir los interpolantes trigonométricos de Fourier y los interpolantes polinómicos de Chebyshev que gozan de una gran popularidad por su previsión espectral.

En el capítulo 2 se analizó la relación entre los desarrollos truncados con  $N$  términos en serie de funciones ortogonales  $\Phi_k(x)$  y desarrollos discretos en serie basados en una nube particular de puntos nodales. En la sección 2.2 se mostró que si la función a aproximar es suficiente diferenciable y las funciones de la base son las autofunciones del problema singular de Sturm–Liouville, la convergencia de la serie truncada es espectral. Además, si la nube de puntos se elige como los ceros de  $\Phi_{N+1}(x)$ , la aproximación mediante la serie discreta es prácticamente la misma que la de la serie discreta o el interpolante y la diferencia entre ambas series es pequeña y conocida como el error de *aliasing*. Por otra parte, en el capítulo 3 se analizó el origen del error de la interpolación polinómica. El error de truncamiento se vinculó a la función del error  $\pi_{N+1}(x)$  y el error de redondeo a la función de Lebesgue  $\lambda_N(x)$ .

En la sección 4.1 se define la serie truncada y la serie discreta o interpolante trigonométrico de Fourier y se definen los nodos o puntos de interpolación trigonométrica. Estos puntos nodales permitan establecer una equivalencia entre la serie discreta y la serie truncada de Fourier. En la sección 4.5 se define la serie truncada y discreta de Chebyshev y mediante las fórmulas de cuadratura de Gauss (secciones 2.5–2.6) se obtienen puntos nodales para calcular la serie discreta. Como el

interpolante de Chebyshev es polinómico, los resultados del capítulo 3 se usan en la sección 4.8 para calcular la nube de puntos de Chebyshev que se compararán con los obtenidos cuando se usan las fórmulas de cuadratura de Gauss.

Otra de las razones que hace atractivo el uso de la serie discreta de Chebyshev y la serie discreta de Fourier es la existencia de una transformada rápida entre el plano espectral y el plano físico. En este capítulo se relaciona la transformada rápida de Fourier con la transformada coseno o transformada de Chebyshev. Es importante hacer notar que cuando se conoce el interpolante de Chebyshev o de Fourier de una función dada, la derivación del interpolante se puede hacer de manera muy precisa en el plano espectral y, posteriormente, mediante la transformada rápida obtener los valores de la derivada en los puntos nodales o puntos de interpolación. De esta manera, se pueden determinar las derivadas de una función incógnita en los puntos nodales o puntos de colocación. El uso del interpolante o la serie discreta permite desarrollar métodos de colocación en problemas de contorno y de valores iniciales.

## 4.1. Interpolante de Fourier

Los senos y los cosenos son un sistema de funciones ortogonales en  $x \in [0, 2\pi]$ . La serie truncada asociada a esta base se expresa:

$$P_N(x) = \sum_{k=0}^{N-1} a_k \cos(kx) + b_k \sin(kx). \quad (4.1)$$

Este desarrollo en serie se puede expresar de forma compacta mediante:

$$P_N(x) = \sum_{k=-N/2}^{N/2-1} \hat{c}_k e^{ikx}, \quad (4.2)$$

donde los  $N - 1$  coeficientes  $\hat{c}_k$  son:

$$\hat{c}_k = \frac{1}{2}(a_k - ib_k), \quad \hat{c}_{-k} = \frac{1}{2}(a_k + ib_k). \quad (4.3)$$

Se define el interpolante o la serie discreta de Fourier mediante:

$$I_N(x) = \sum_{k=-N/2}^{N/2-1} \tilde{c}_k e^{ikx}. \quad (4.4)$$

Para determinar los  $N - 1$  coeficientes  $\tilde{c}_k$ , se fuerza a que el interpolante pase por un conjunto de puntos  $x_0, \dots, x_{N-1}$  que son los ceros de  $\Phi_N(x) = e^{iNx}$  (sección 2.6)

$$x_j = \frac{2\pi j}{N}, \quad j = 0, \dots, N - 1, \quad (4.5)$$

que son puntos equiespaciados entre 0 y  $2\pi$ . Forzando a que el interpolante verifique estos puntos se obtiene:

$$f(x_j) = \sum_{k=-N/2}^{N/2-1} \tilde{c}_k e^{ikx_j}, \quad j = 0, \dots, N-1. \quad (4.6)$$

La diferencia entre los coeficientes de la serie discreta y la serie truncada de Fourier es el error de *aliasing* dado por la expresión (2.21):

$$\tilde{c}_m = \hat{c}_m + \sum_{|p| \geq 1} \hat{c}_{m+pN}. \quad (4.7)$$

Es importante hacer notar que si la serie truncada  $\hat{c}_m$  tiene precisión espectral, la serie discreta  $\tilde{c}_m$  también tiene comportamiento espectral y por lo tanto el interpolante trigonométrico con los puntos (4.5) tiene un error tan pequeño como el error de la serie truncada.

## 4.2. Derivada del interpolante de Fourier

Conocidos los valores nodales  $(x_j, f(x_j))$ , se pueden obtener los coeficientes  $\tilde{c}_m$  resolviendo el sistema de  $N-1$  ecuaciones dado por (4.6). Como los vectores de la base  $[\Phi_k(x_0), \dots, \Phi_k(x_{N-1})]$  son ortogonales, los coeficientes  $\tilde{c}_k$  que se obtienen de (4.6) son:

$$\tilde{c}_k = \frac{1}{N} \sum_{j=0}^{N-1} f_j e^{-ikx_j}, \quad k = -\frac{N}{2}, \dots, \frac{N}{2} - 1. \quad (4.8)$$

Una vez calculados los coeficientes de la serie discreta, la derivada del interpolante de Fourier dado por (4.4) vale:

$$\frac{dI_N}{dx} = \sum_{k=-N/2}^{N/2-1} i k \tilde{c}_k e^{ikx}, \quad (4.9)$$

Es importante hacer notar que la obtención de los coeficientes  $\tilde{c}_k$  dados por la ecuación (4.8) exige realizar  $O(N^2)$  operaciones. Una de la principales ventajas del interpolante de Fourier es la existencia de transformada rápida. Esto permite transformar los valores  $f_j$  entre el plano físico y los valores  $\tilde{c}_m$  del plano espectral en  $O(N \log N)$  operaciones. El paso entre el plano físico  $f_j$  y el plano espectral  $\tilde{c}_m$  mediante la transformada rápida de Fourier se discute en la siguiente sección.

### 4.3. Transformada rápida de Fourier

La transformada rápida de Fourier es un algoritmo recursivo para evaluar la transformada discreta de Fourier (4.8) y su inversa (4.6). La transformada rápida de Fourier más simple requiere que  $N$  sea una potencia de 2. La transformada directa e inversa de Fourier se expresan mediante:

$$\tilde{c}_k = \frac{1}{N} \sum_{j=0}^{N-1} f_j e^{-ikx_j}, \quad k = -\frac{N}{2}, \dots, \frac{N}{2} - 1 \quad (4.10)$$

$$f_j = \sum_{k=-N/2}^{N/2-1} \tilde{c}_k e^{ikx_j}, \quad j = 0, \dots, N-1 \quad (4.11)$$

El algoritmo de Cooley–Turkey (1965) permite evaluar las sumas de (4.10) en  $5 N \log_2 N$  operaciones reales (cuando  $N$  es una potencia de 2) en lugar de  $8N^2$  operaciones reales requeridas por la suma directa. Además, el cálculo de (4.10) por la transformada rápida de Fourier produce menos error debido a la falta de precisión de la suma directa.

Si  $N = 2^m$ , entonces la transformada de Fourier se puede calcular de forma rápida. La transformada directa de Fourier se puede poner como la transformada de Fourier para los términos pares mas la transformada de Fourier para los términos impares.

$$\tilde{c}_k = \frac{1}{N} \sum_{j=0}^{N/2} f_{2j} e^{-ikx_{2j}} + \frac{1}{N} \sum_{j=0}^{N/2-1} f_{2j+1} e^{-ikx_{2j+1}}. \quad (4.12)$$

Como los puntos de colocación están equiespaciados, entonces

$$x_{2j+1} = x_{2j} + \frac{2\pi}{N} \quad (4.13)$$

y la expresión (4.12) se puede poner como:

$$\tilde{c}_k = \frac{1}{N} \sum_{j=0}^{N/2} f_{2j} e^{-ikx_{2j}} + w^k \frac{1}{N} \sum_{j=0}^{N/2-1} f_{2j+1} e^{-ikx_{2j}}, \quad w = e^{-2\pi i/N} \quad (4.14)$$

La transformada de Fourier para calcular  $\tilde{c}_k$  se puede calcular a partir de la transformada de Fourier para los términos pares  $f_{2j}$  mas la transformada de Fourier para los términos impares  $f_{2j+1}$ . La expresión (4.14) se calcula para valores de  $k = -N/2, \dots, N/2 - 1$ . Sin embargo, debido a la expresión que tienen los coeficientes  $\tilde{c}_{k-N/2}$ ,

$$\tilde{c}_{k-N/2} = \frac{1}{N} \sum_{j=0}^{N/2} f_{2j} e^{-ikx_{2j}} - w^k \frac{1}{N} \sum_{j=0}^{N/2-1} f_{2j+1} e^{-ikx_{2j}}, \quad (4.15)$$

solo es necesario calcular (4.14) desde  $k = 0$  hasta  $k = N/2$  y mediante la expresión (4.14) calcular  $\tilde{c}_{-N/2}$  hasta  $\tilde{c}_{-1}$  con las sumas de los términos pares y los términos impares obtenidas en (4.15).

Este algoritmo recursivo nos permite crear un árbol binario subdividiendo consecutivamente los puntos de colocación o puntos nodales en pares e impares. En la figura 4.1 se muestra un ejemplo de árbol binario que resulta de aplicar la transformada rápida de Fourier con 8 puntos interpolación.

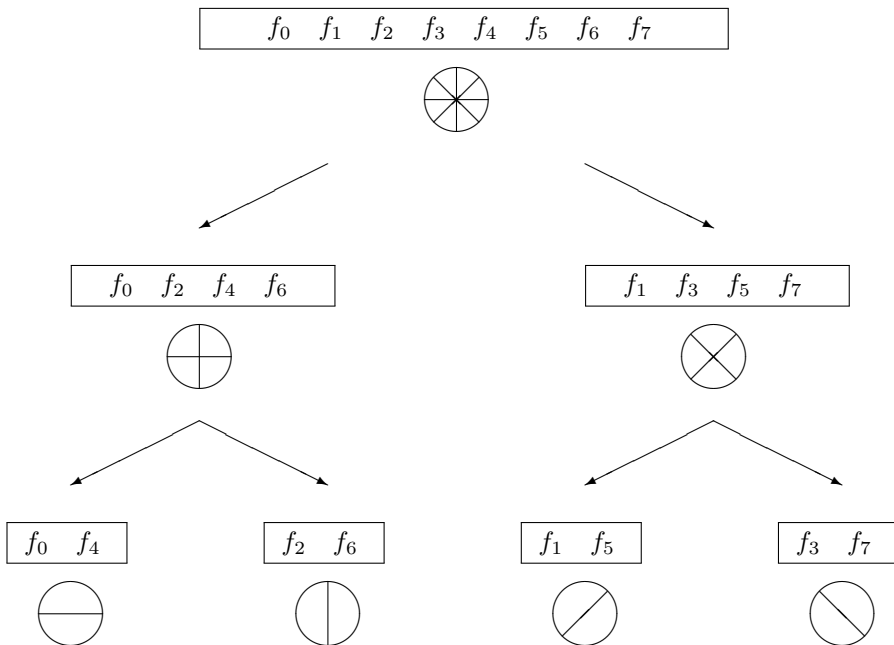


Figura 4.1: Transformada rápida de Fourier de ocho puntos mediante la regla de recursión de puntos pares e impares.

Generalmente, necesitaremos la transformada de Fourier de una función real. En este caso, y considerando que en la expresión (4.10)  $f_j$  es un número complejo, podríamos calcular dos transformadas de Fourier reales almacenando un conjunto de valores en la parte real de  $f_j$  y otro conjunto de valores en la parte imaginaria de  $f_j$ . Por ejemplo, si se necesitan transformadas de Fourier multidimensionales, las transformadas se pueden obtener por parejas. Supongamos que  $f_j^1$  y  $f_j^2, j = 0, \dots, N-1$  son dos conjuntos de datos reales. Entonces se puede definir,

$$f_j = f_j^1 + i f_j^2 \quad (4.16)$$

y calcular los  $\tilde{c}_k$  mediante (4.10). Entonces, los coeficientes de Fourier  $\tilde{c}_k^1, \tilde{c}_k^2$  se

pueden obtener con las siguientes expresiones,

$$\tilde{c}_k^1 = \frac{1}{2}(\tilde{c}_k + \overline{\tilde{c}_{-k}}), \quad (4.17)$$

$$\tilde{c}_k^2 = -\frac{i}{2}(\tilde{c}_k - \overline{\tilde{c}_{-k}}). \quad (4.18)$$

## 4.4. Interpolante de Fourier en dos dimensiones

Se define el interpolante o la serie discreta de Fourier en  $\mathbb{R}^2$  mediante:

$$I_N(x, y) = \sum_{k=-N/2}^{N/2-1} \sum_{m=-M/2}^{M/2-1} \tilde{c}_{km} e^{ikx+imy}. \quad (4.19)$$

Para determinar los  $(N-1)(M-1)$  coeficientes  $\tilde{c}_{km}$ , se fuerza a que el interpolante pase por un conjunto de puntos  $(x_i, y_j)$  con  $i = 0, \dots, N-1$  y  $j = 0, \dots, M-1$  que son puntos equiespaciados en  $[0, 2\pi] \times [0, 2\pi]$ . Forzando a que el interpolante verifique estos puntos se obtiene:

$$f(x_i, y_j) = \sum_{k=-N/2}^{N/2-1} \sum_{m=-M/2}^{M/2-1} \tilde{c}_{km} e^{ikx_i+imy_j}, \quad i = 0, \dots, N-1, \quad j = 0, \dots, M-1. \quad (4.20)$$

El sistema anterior se invierte para dar:

$$\tilde{c}_{km} = \frac{1}{NM} \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} f_{ij} e^{-ikx_i-imy_j}, \quad k = -\frac{N}{2}, \dots, \frac{N}{2}-1, \quad m = -\frac{M}{2}, \dots, \frac{M}{2}-1. \quad (4.21)$$

La expresión anterior se puede poner de la siguiente manera:

$$\tilde{c}_{km} = \frac{1}{NM} \sum_{j=0}^{M-1} e^{-imy_j} \left[ \sum_{i=0}^{N-1} f_{ij} e^{-ikx_i} \right]. \quad (4.22)$$

Esta forma nos permite calcular la transformada bidimensional de Fourier mediante transformadas unidimensionales de Fourier. El corchete de la expresión (4.22) representa un conjunto de transformadas unidimensionales de Fourier en la dirección  $x$  con  $y_j$  constante. Con el resultado de estas transformadas se realiza un conjunto de transformadas en la dirección  $y$  con  $x_i$  constante y se obtienen los coeficientes  $\tilde{c}_{km}$ .

## 4.5. Interpolantes de Chebyshev

Los polinomios de Chebyshev  $T_k(x)$  constituyen una base de polinomios ortogonales en  $x \in [-1, 1]$  con respecto al producto interno

$$\langle u, v \rangle_w = \int_{-1}^{+1} T_m(x) T_k(x) \frac{1}{\sqrt{1-x^2}} dx. \quad (4.23)$$

Se definen los polinomios de la base de la siguiente forma:

$$T_k(x) = \cos(k\theta), \quad (4.24)$$

donde  $x = \cos \theta$ . A continuación se demuestra que la expresión anterior es un polinomio de grado  $k$ . Mediante la fórmula trigonométrica

$$\cos(k+1)\theta + \cos(k-1)\theta = 2 \cos k\theta \cos \theta, \quad (4.25)$$

se puede escribir la relación de recurrencia para  $T_k(x)$  de la siguiente manera:

$$T_{k+1}(x) = 2x T_k(x) - T_{k-1}(x) \quad (4.26)$$

junto con  $T_0(x) = 1$  y  $T_1(x) = x$ . La recurrencia (4.26) permite verificar que  $T_k(x)$  es un polinomio de grado  $k$  y que el coeficiente que multiplica a  $x^k$  vale  $2^{k-1}$ .

La serie truncada asociada a esta base se define como:

$$P_N(x) = \sum_{k=0}^N \hat{c}_k T_k(x). \quad (4.27)$$

Asociado a una nube de puntos nodales o puntos de interpolación se construye el interpolante o la serie discreta

$$I_N(x) = \sum_{k=0}^N \tilde{c}_k T_k(x). \quad (4.28)$$

La pregunta que surge ahora es determinar la bondad del ajuste del interpolante de Lagrange basado en las distribuciones de puntos nodales y su relación con la serie truncada. Como vimos en las secciones 2.5 y 2.6, existen distribuciones de puntos que nos permiten igualar el producto interno  $\langle u, v \rangle_N$  del espacio vectorial de dimensión finita determinado por los puntos nodales (2.15) con el producto interno  $\langle u, v \rangle_w$  del espacio de dimensión infinita (2.3). Es decir,

$$\langle u, v \rangle_N = \langle u, v \rangle_w \quad (4.29)$$

para  $u, v$  polinomios de grado máximo  $2N+1$  para las fórmulas de cuadratura de Gauss o polinomios de grado máximo  $2N-1$  para las fórmulas de cuadratura de Gauss-Lobatto.

### 4.5.1. Ceros de Chebyshev

Se considera el interpolante de Chebyshev o la serie discreta siguiente:

$$I_N(x) = \sum_{k=0}^N \tilde{c}_k T_k(x). \quad (4.30)$$

Imponiendo que el interpolante pase por los ceros de  $T_{N+1}(x) = \cos(N+1)\theta$ , se determinan los coeficientes de esta serie discreta. Los ceros de  $T_{N+1}(x)$  son:

$$(N+1)\theta_j = \frac{\pi}{2} + \pi j, \quad j = 0, \dots, N. \quad (4.31)$$

que se corresponden con los nodos de interpolación

$$x_j = \cos \left( \frac{\frac{\pi}{2} + \pi j}{N+1} \right), \quad j = 0, \dots, N. \quad (4.32)$$

Con estos ceros, el producto interno del espacio de dimensión finita  $\langle u, v \rangle_N$  coincide con el producto interno del espacio de dimensión infinita  $\langle u, v \rangle_w$  (sección 2.5) con los coeficientes de peso  $\alpha_j = \pi/(N+1)$ . Así, el producto interno queda:

$$\langle u, v \rangle_N = \frac{\pi}{N+1} \sum_{j=0}^N u(x_j) v(x_j). \quad (4.33)$$

Los coeficientes de normalización  $\gamma_m$  definidos en (2.15) quedan:

$$\gamma_m = \frac{\pi}{N+1} \sum_{j=0}^N \cos^2(m\theta_j) = \frac{\pi}{2}. \quad (4.34)$$

De esta forma, los coeficientes de la serie discreta se obtienen mediante

$$\tilde{c}_m = \frac{2}{(N+1)} \sum_{j=0}^N f(x_j) \cos m \theta_j, \quad m = 1, \dots, N. \quad (4.35)$$

Cuando el interpolante se usa para integrar un problema en donde es necesario imponer condiciones de contorno, los ceros de Chebyshev no tienen mucha utilidad porque no incluyen los contornos  $x = -1$  y  $x = +1$ . En la siguiente sección se definen los extremos de Chebyshev que incluyen los contornos y son los que tienen una mayor utilidad.

### 4.5.2. Extremos de Chebyshev

En esta sección se discuten los puntos de interpolación que incluyen los contornos  $x = -1$  y  $x = +1$  y que se denominan extremos de Chebyshev.



Las fórmulas de cuadratura de Gauss–Lobatto (sección 2.6) se obtienen cuando los puntos nodales son los  $N + 1$  ceros de la combinación de polinomios de la base:

$$q_{N+1}(x) = T_{N+1}(x) + c_1 T_N(x) + c_2 T_{N-1}(x), \quad (4.36)$$

donde  $c_1$  y  $c_2$  se eligen para que  $q_{N+1}(-1) = q_{N+1}(+1) = 0$ . Resolviendo el sistema anterior se obtiene  $c_1 = 0$  y  $c_2 = -1$ . Es decir,

$$q_{N+1}(x) = T_{N+1}(x) - T_{N-1}(x) = \cos \theta \operatorname{sen} N\theta. \quad (4.37)$$

Los ceros de  $q_{N+1}(x)$  son los puntos nodales:

$$x_j = \cos\left(\frac{\pi j}{N}\right), \quad j = 0, \dots, N. \quad (4.38)$$

Imponiendo que el interpolante pase por los extremos de Chebyshev o ceros de  $q_{N+1}(x)$  se determinan los coeficientes de esta serie discreta

$$I_N(x) = \sum_{k=0}^N \tilde{c}_k T_k(x). \quad (4.39)$$

Con estos puntos nodales, la expresión de los coeficientes de peso (2.28) permite obtener:

$$\{\alpha_0, \alpha_1, \dots, \alpha_{N-1}, \alpha_N\} = \left\{ \frac{\pi}{2N}, \frac{\pi}{N}, \dots, \frac{\pi}{N}, \frac{\pi}{2N} \right\},$$

y el producto interno queda:

$$\langle u, v \rangle_N = \frac{\pi}{2N} (u_0 v_0 + u_N v_N) + \frac{\pi}{N} \sum_{j=1}^{N-1} u_j v_j. \quad (4.40)$$

Los coeficientes de normalización  $\gamma_m$  son:

$$\gamma_m = \frac{\pi}{2N} (\cos^2(m\theta_0) + \cos^2(m\theta_N)) + \frac{\pi}{N} \sum_{j=1}^{N-1} \cos^2(m\theta_j). \quad (4.41)$$

Como los extremos de Chebyshev incluyen  $\theta_N = \pi$  y  $\theta_0 = 0$ , la expresión anterior se simplifica para dar:

$$\gamma_m = \begin{cases} \frac{\pi}{2} & m < N, \\ \pi & m = N. \end{cases} \quad (4.42)$$

Así, los coeficientes de la serie discreta se obtienen mediante

$$\tilde{c}_m = \frac{1}{2N} (f(x_0) + (-1)^m f(x_N)) + \frac{1}{N} \sum_{j=1}^{N-1} f(x_j) \cos\left(\frac{m \pi j}{N}\right). \quad (4.43)$$

El error de aliasing (2.20) entre la serie discreta y la serie truncada cuando se se consideran los extremos de Chebyshev tiene una expresión muy simple:

$$\tilde{c}_m = \hat{c}_m + \sum_{|p| \geq 1} \hat{c}_{m+pN}. \quad (4.44)$$

Es importante hacer notar que si la serie truncada  $\hat{c}_k$  tiene precisión espectral, la serie discreta también tiene comportamiento espectral y por lo tanto el interpolante de Lagrange con los puntos (4.63) tiene un error tan pequeño como el error de la serie truncada. Por otra parte, el paso entre el plano físico  $f(x_j)$  y el plano espectral  $\tilde{c}_m$  se puede realizar con una transformada rápida de Fourier o transformada coseno dada por (4.43).

## 4.6. Derivada del interpolante de Chebyshev

Una de la principales ventajas del interpolante de Chebyshev es la existencia de transformada rápida (sección 4.7). Esto permite transformar entre el plano físico y el plano espectral en  $O(N \log N)$  operaciones. Cualquier problema de simulación involucra el interpolante de Chebyshev y sus derivadas. Mientras que la derivación del interpolante en los  $N+1$  nodales involucra  $O(N^2)$  operaciones, la derivación en el plano espectral y su posterior transformación al plano físico involucra  $O(N \log N)$  operaciones. Es por esta razón por la que nos interesa conocer una expresión de la derivada del polinomio de Chebyshev en el plano espectral.

La derivada del interpolante de Chebyshev dado por (4.28) vale:

$$\frac{dI_N}{dx} = \sum_{k=0}^N \tilde{c}_k \frac{dT_k}{dx}, \quad (4.45)$$

donde

$$\frac{dT_k}{dx} = k \frac{\sin k\theta}{\sin \theta}. \quad (4.46)$$

Mediante la siguiente igualdad trigonométrica

$$\frac{\sin(k+1)\theta}{\sin \theta} - \frac{\sin(k-1)\theta}{\sin \theta} = 2 \cos \theta, \quad (4.47)$$

expresamos las derivadas  $T'_k(x)$  en función de los polinomios de Chebyshev,

$$\frac{T'_{k+1}}{k+1} - \frac{T'_{k-1}}{k-1} = 2T_k, \quad k = 0, \dots, N-1. \quad (4.48)$$

El sistema anterior en forma matricial se puede expresar

$$\begin{pmatrix} 2 & 0 & 0 & 0 & \dots & \dots & 0 \\ 0 & 0 & 1/2 & 0 & \dots & \dots & 0 \\ 0 & -1 & 0 & 1/3 & \dots & \dots & 0 \\ 0 & 0 & -1/2 & 0 & 1/4 & \dots & 0 \\ \vdots & & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & \dots & 0 & -1/(N-2) & 0 & 1/N \end{pmatrix} \begin{pmatrix} T'_1 \\ T'_2 \\ T'_3 \\ T'_4 \\ \vdots \\ T'_N \end{pmatrix} = 2 \begin{pmatrix} T_0 \\ T_1 \\ T_2 \\ T_3 \\ \vdots \\ T_{N-1} \end{pmatrix}. \quad (4.49)$$

Si definimos mediante  $\tilde{c}_k^{(1)}$  los coeficientes del desarrollo en serie de Chebyshev de la derivada del interpolante

$$\frac{dI_N}{dx} = \sum_{k=0}^N \tilde{c}_k^{(1)} T'_k, \quad (4.50)$$

entonces los coeficientes  $\tilde{c}_k^{(1)}$  se relacionan con los coeficientes  $\tilde{c}_k$  mediante,

$$2(\tilde{c}_0, \dots, \tilde{c}_{N-1}) = (\tilde{c}_1^{(1)}, \dots, \tilde{c}_N^{(1)}) D, \quad (4.51)$$

donde  $D$  es la matriz del sistema (4.49). La componente genérica  $k$  del sistema (4.51) queda:

$$\tilde{c}_{k-1}^{(1)} = \tilde{c}_{k+1}^{(1)} + 2k \tilde{c}_k, \quad k = N-1, \dots, 1. \quad (4.52)$$

Esta es una ecuación en diferencias de segundo orden para los valores  $\tilde{c}_k^{(1)}$  que se integra con  $\tilde{c}_N^{(1)} = 0$  y  $\tilde{c}_{N-1}^{(1)} = 2N \tilde{c}_N$ . El número de operaciones necesario para obtener los coeficientes  $\tilde{c}_k^{(1)}$  es  $O(N)$ . Posteriormente, para obtener la derivada del interpolante en los puntos nodales se hace una transformada coseno que involucra  $O(N \log N)$  operaciones.

## 4.7. Transformada rápida coseno mediante la FFT

La transformada de Fourier (4.10) con  $M$  puntos se expresa:

$$\tilde{f}_k = \frac{1}{M} \sum_{j=0}^{M-1} f_j e^{-ik \theta_j}, \quad k = -\frac{M}{2}, \dots, \frac{M}{2} - 1 \quad (4.53)$$

con  $\theta_j = 2\pi j/M$ . Por otra parte, la transformada coseno (4.43)

$$\tilde{c}_k = \frac{1}{2N} (f_0 + (-1)^k f_N) + \frac{1}{N} \sum_{j=1}^{N-1} f_j \cos\left(\frac{k \pi j}{N}\right). \quad (4.54)$$

Para relacionar estas dos transformadas hacemos  $M = 2N$  y prolongamos periódicamente  $f_j$  con  $N - 1$  puntos extra haciendo

$$\begin{aligned} f_{N+1} &= f_{N-1}, \\ f_{N+2} &= f_{N-2}, \\ &\vdots \\ f_{2N-1} &= f_1. \end{aligned}$$

Si usamos estos puntos prolongados en la expresión (4.53) y dividimos la suma en los  $N + 1$  primeros puntos y en los  $N - 1$  siguiente puntos, se obtiene:

$$M\tilde{f}_k = \sum_{j=0}^N f_j e^{-ik\theta_j} + \sum_{j=N+1}^{2N-1} f_j e^{-ik\theta_j}. \quad (4.55)$$

Del primer sumatorio de la expresión anterior separamos los términos correspondientes a  $j = 0$  y  $j = N$

$$M\tilde{f}_k = f_0 + f_N(-1)^N + \sum_{j=1}^{N-1} f_j e^{-ik\theta_j} + \sum_{j=N+1}^{2N-1} f_j e^{-ik\theta_j}. \quad (4.56)$$

Finalmente, los dos sumatorios de esta expresión los agrupamos por parejas  $f_j$  y  $f_{2N-j}$

$$M\tilde{f}_k = f_0 + f_N(-1)^N + \sum_{j=1}^{N-1} f_j e^{-ik\theta_j} + f_{2N-j} e^{-ik\theta_{2N-j}}. \quad (4.57)$$

Como  $\theta_{2N-j} = 2\pi - \theta_j$  y  $f_{2N-j} = f_j$  la expresión anterior queda

$$M\tilde{f}_k = f_0 + f_N(-1)^N + 2 \sum_{j=1}^{N-1} f_j \cos k\theta_j. \quad (4.58)$$

Es decir, la transformada coseno (4.54) se calcula mediante la transformada rápida de Fourier (4.53) haciendo

$$\tilde{c}_k = \frac{\tilde{f}_k}{2}. \quad (4.59)$$

## 4.8. Error y puntos de Chebyshev

En la sección 4.5.2 se discutió el error de *aliasing* entre la serie truncada y la serie discreta cuando se usan los extremos de Chebyshev como puntos de interpolación.

En esta sección se analiza el error de truncamiento asociado a la función de error  $\pi_{N+1}(x)$  y su relación con los puntos de interpolación. La minimización de la función  $\pi_{N+1}(x)$  pasa por buscar distribuciones de puntos que hagan que los extremos relativos de  $\pi_{N+1}(x)$  sean iguales. Un ejemplo de funciones con extremos iguales son los senos y los cosenos. Definimos la función de error  $\pi_{N+1}(x)$  mediante un coseno que tenga  $N + 1$  ceros en nuestro intervalo de integración de la forma siguiente:

$$\pi_{N+1}(x) = K \cos(N + 1)\theta, \quad (4.60)$$

con  $x = \cos \theta$ .

La función  $\pi_{N+1}(x)$  así definida es un polinomio de grado  $N + 1$  como vimos en la sección 4.5. Además, a través de la definición de  $\pi_{N+1}(x)$  (ecuación 3.21) el coeficiente que multiplica a  $x^{N+1}$  es uno y como el coeficiente de mayor orden de  $T_{N+1}(x)$  es  $2^N$  (sección 4.5), se obtiene el valor de la constante  $K$  para dar:

$$\pi_{N+1}(x) = \frac{1}{2^N} \cos(N + 1)\theta \quad (4.61)$$

Es importante resaltar la siguientes observaciones:

1. El error de truncamiento involucra la función  $\pi_{N+1}(x)$  que tiende a cero mucho más rápidamente que cualquier potencia negativa de  $N$ . Esa condición está asociada a la precisión espectral. Es decir, muy pocos grados de libertad  $N$  permiten tener un error insignificante. Es importante hacer notar que la cota (3.58) para puntos de interpolación equiespaciados es mucho mayor que  $1/2^N$  (4.61) y no permite hablar de precisión espectral.
2. Los puntos de interpolación coinciden con los ceros de Chebyshev (4.32) que se obtienen a través de un razonamiento completamente diferente basado en las fórmulas de cuadratura de Gauss.
3. Estos puntos no contienen los extremos  $x = \pm 1$  como vimos en la sección 4.5.1. A modo de ejemplo, la figura (4.2) presenta los ceros  $x_j$  y extremos  $y_j$  de  $|\pi_{N+1}(x)|$  para 9 puntos. distribuidos a partir de la expresión (4.32).
4. La cota de la constante de Lebesgue definida por (3.44) es mucho más pequeña que para distribuciones de puntos equiespaciados y queda:

$$\Lambda_N \leq \frac{2}{\pi} \log(N + 1) + 1. \quad (4.62)$$

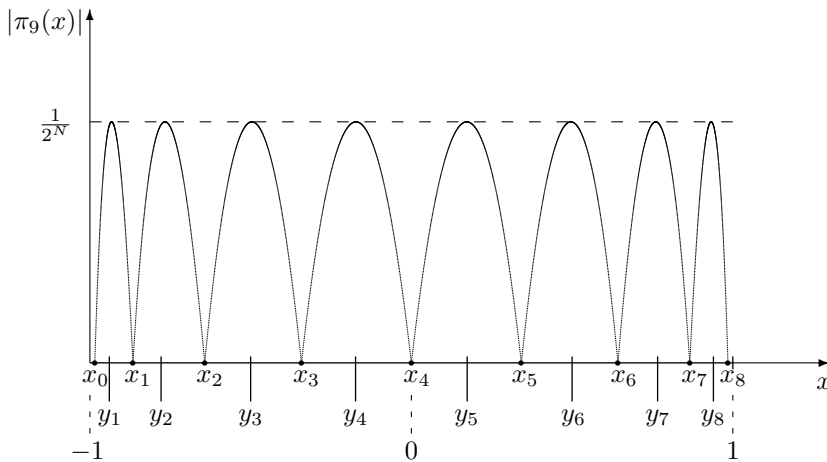


Figura 4.2: Ceros  $x_j$  y extremos  $y_j$  de  $|\pi_{N+1}(x)|$  para un conjunto de 9 puntos ( $N = 8$ ) no equiespaciados obtenidos mediante la expresión (4.32)

Tal como puede observarse en la figura y como puede comprobarse en las expresiones correspondientes, un inconveniente que presentan los ceros  $x_j$  (4.32) es que no contienen los puntos extremos del intervalo de interpolación. Los puntos nodales (4.32) se los conoce como puntos de Chebyshev de primera clase o ceros de Chebyshev.

Si se deriva la expresión  $\pi_{N+1}(x)$  con respecto a  $x$  y se iguala a cero, se obtienen los puntos los extremos de  $\pi_{N+1}(x)$ :

$$y_j = \cos\left(\frac{\pi j}{N}\right), \quad j = 1, \dots, N-1. \quad (4.63)$$

Cuando estos extremos se complementan con los puntos  $y_0 = -1$  y  $y_N = +1$ , se obtienen los puntos dados por la expresión (4.63) que se les conoce como puntos de Chebyshev de segunda clase o extremos de Chebyshev. Es importante hacer notar que esta distribución de puntos nodales hace que  $\pi_{N+1}(x)$  y  $\lambda_N(x)$  no sean uniformes pero se comporten mucho mejor que para distribuciones equiespaciadas de puntos. Generalmente, los extremos de Chebyshev son los puntos de mayor uso porque permiten imponer condiciones de contorno en los extremos del intervalo. Además, la existencia de una transformada rápida hace que su uso sea muy extendido.

# Capítulo 5

## Interpolación polinómica continua a trozos

Los métodos para aproximar funciones mediante interpolantes son la *interpolación global* y la *interpolación continua a trozos*. La interpolación global consiste en utilizar todos los puntos de interpolación dentro del dominio de definición, para construir un interpolante que será continuo en dicho dominio. La interpolación continua a trozos consiste en utilizar una cantidad, por lo general pequeña, de puntos de interpolación para construir un interpolante continuo a trozos. En este caso, se habla de  $q + 1$  puntos nodales de interpolación y el interpolante resultante  $I_q$  es un polinomio de grado  $q$ . El método de interpolación continua a trozos es el más empleado cuando, a partir de la teoría de interpolación, se pretende obtener métodos numéricos para la solución de ecuaciones diferenciales. La estrategia de construcción de un interpolante continuo a trozos se explicará en secciones siguientes de este capítulo.

En el capítulo anterior se analizó el comportamiento del error cuando se emplean interpolaciones globales equiespaciadas y no equiespaciadas. En este capítulo se estudia la *interpolación continua a trozos* con distribuciones equiespaciadas y no equiespaciadas de puntos. Ésta es la forma más empleada al momento de generar esquemas numéricos para la solución de ecuaciones diferenciales. Para este tipo de interpolación se estudia el comportamiento del error de interpolación a través de la función de error  $\pi$  y de la función de Lebesgue  $\lambda_N(x)$ .

Para describir la interpolación equiespaciada continua a trozos se consideran  $N + 1$  puntos  $x_0, x_1, \dots, x_N$  del dominio de definición de una función  $f$ . A continuación, se consideran  $q/2$  puntos de interpolación a la derecha y a la izquierda de

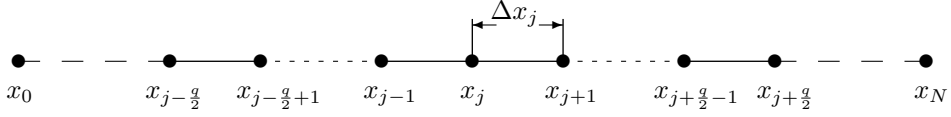


Figura 5.1: Molécula computacional centrada y equiespaciada

un punto genérico  $x_j$ , donde  $q$  es un número natural par. De esta forma se tienen  $q + 1$  puntos de interpolación donde  $x_j$  es el punto medio. A esta configuración de puntos se la suele denominar molécula computacional centrada y se representa en la figura (5.1). Para el conjunto de  $q + 1$  puntos de interpolación se puede construir un interpolante  $I_q$  de grado  $\leq q$  y si se repite esta construcción a todo el dominio de interpolación, se tiene una interpolación continua a trozos. Es importante hacer notar que para interpolar en las proximidades de los extremos  $x_0$  y  $x_N$  la molécula computacional se debe descentrar manteniendo el mismo grado del interpolante.

El error correspondiente a la interpolación polinómica a partir de los  $q + 1$  puntos equiespaciados tiene la expresión siguiente:

$$R_q(x) = \pi_{q+1}(x) \frac{f^{(q+1)}(\xi)}{(q+1)!}, \quad (5.1)$$

donde  $\pi_{q+1}(x)$  es:

$$\begin{aligned} \pi_{q+1}(x) &= (x - x_{j-\frac{q}{2}})(x - x_{j-\frac{q}{2}+1}) \cdots (x - x_{j-1})(x - x_j)(x - x_{j+1}) \cdots \\ &\quad \cdots (x - x_{j+\frac{q}{2}-1})(x - x_{j+\frac{q}{2}}). \end{aligned} \quad (5.2)$$

Al igual que se hizo en la interpolación global, a continuación se busca una cota del error de interpolación continua a trozos a través de la función (5.2). Si la distribución de puntos nodales es equiespaciada, la expresión (3.58) para la cota del error de truncamiento resulta:

$$\|\pi_{q+1}(x)\| \leq q! \frac{\Delta x^{q+1}}{4}. \quad (5.3)$$

En el caso de una nube de puntos equiespaciada,  $\Delta x = 2/N$  y la expresión (5.3) queda:

$$\|\pi_{q+1}(x)\| \leq \frac{q!}{4} \left( \frac{2}{N} \right)^{q+1}. \quad (5.4)$$



Tomando logaritmos en la expresión (5.3), se tiene que:

$$\log \|\pi_{q+1}(x)\| \leq \log \left( \frac{q!}{4} \right) + (q+1) [\log 2 - \log N].$$

De forma esquemática, la figura (5.2) representa la expresión anterior e indica la convergencia de  $|\pi_{q+1}(x)|$  con el número de puntos  $N$  en una interpolación continua a trozos equiespaciada.

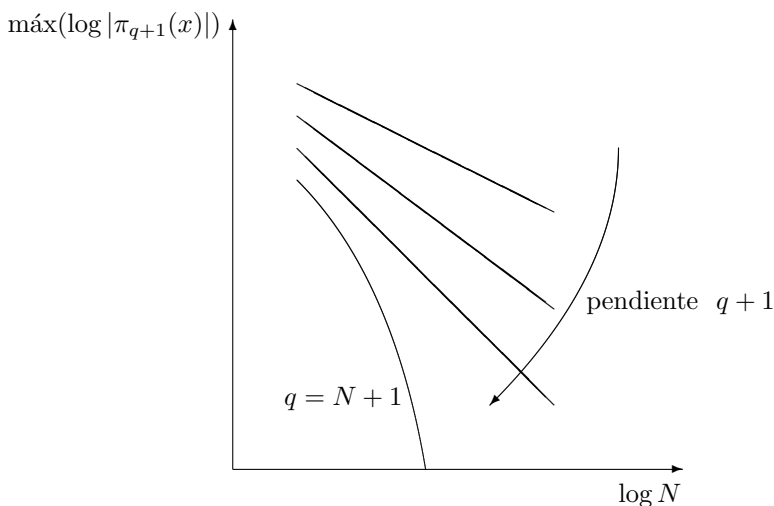


Figura 5.2: Convergencia de  $|\pi_{q+1}(x)|$  en función de  $N$  para una interpolación continua a trozos equiespaciada con  $q+1$  puntos

Se observa en la figura (5.2) que mientras que en la interpolación continua a trozos las expresiones son rectas con pendiente  $q+1$ , el error de truncamiento en la interpolación global disminuye con  $N$  más rápidamente que una recta de pendiente  $N+1$ .

## 5.1. Fórmulas para la derivada primera y segunda

En este apartado, se obtienen las fórmulas de las derivadas primera y segunda para interpolantes a trozos construidos a partir de  $q+1$  puntos. En la figura (5.3)

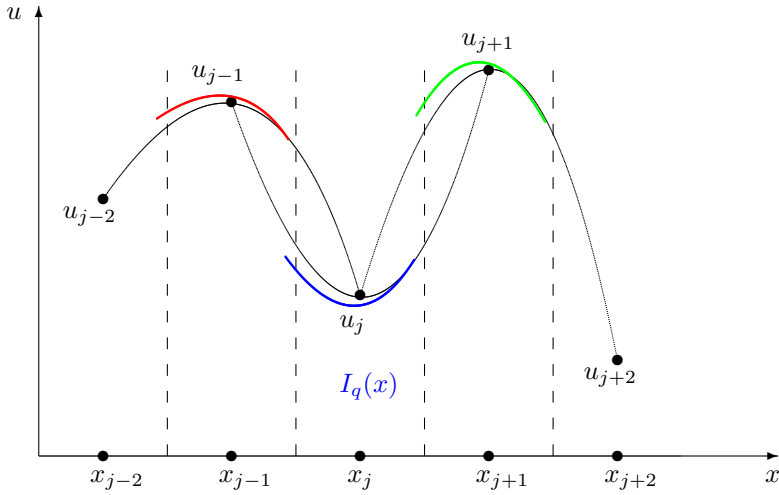


Figura 5.3: Interpolación continua a trozos y dominio de definición del inteprolante  $I_q(x)$

se representan los interpolantes de grado  $q$  y su dominio de validez representados con diferentes tramos. Es importante hacer notar que el dominio de validez del interpolante  $I_q$  es:

$$\forall x \in \left( x_j - \frac{\Delta x_{j-1}}{2}, x_j + \frac{\Delta x_{j+1}}{2} \right). \quad (5.5)$$

De esta forma, las expresiones de la derivada primera y segunda de  $I_q$  particularizadas en el punto nodal  $x_k$  perteneciente a su dominio de definición quedan:

$$\left( \frac{dI_q}{dx} \right)_{x_k} = \sum_{j=0}^q f_j \ell'_j(x_k), \quad (5.6)$$

$$\left( \frac{d^2I_q}{dx^2} \right)_{x_k} = \sum_{j=0}^q f_j \ell''_j(x_k). \quad (5.7)$$

Obsérvese que tanto  $\ell'_j(x_k)$  como  $\ell''_j(x_k)$  son números y las expresiones (5.6) y (5.7) se conocen como las fórmulas de diferencias finitas para las derivadas primera y segunda.

## 5.2. Error de las derivadas primera y segunda

El error de truncamiento y redondeo del interpolante  $I_q$  es:

$$E_q(x) = \pi_{q+1}(x) \frac{f^{(q+1)}(\xi)}{(q+1)!} + \sum_{j=0}^q \epsilon_j \ell_j(x). \quad (5.8)$$

La expresión (5.8) es válida para su dominio de definición correspondiente. Si la molécula computacional está centrada en  $x_k$ , el dominio de validez de este interpolante es un entorno de  $x_k$  dado por la expresión (5.5). Obsérvese que la interpolación no es continua en  $x = x_k - \Delta x_{k-1}/2$  y en  $x_k + \Delta x_k/2$ . En el caso de los contornos, el dominio de validez es mayor.

Derivando la expresión (5.8) con respecto a  $x$ , se obtiene el error de la derivada primera:

$$E'_q(x) = \pi'_{q+1}(x) \frac{f^{(q+1)}(\xi)}{(q+1)!} + \pi_{q+1}(x) \frac{f^{(q+2)}(\xi)}{(q+1)!} \left( \frac{d\xi}{dx} \right) + \sum_{j=0}^q \epsilon_j \ell'_j(x). \quad (5.9)$$

y derivando nuevamente la expresión anterior con respecto a  $x$ , se obtiene el error de la derivada segunda:

$$\begin{aligned} E''_q(x) = & \pi''_{q+1}(x) \frac{f^{(q+1)}(\xi)}{(q+1)!} + 2 \pi'_{q+1}(x) \frac{f^{(q+2)}(\xi)}{(q+1)!} \left( \frac{d\xi}{dx} \right) + \\ & \pi_{q+1}(x) \left( \frac{f^{(q+2)}(\xi)}{(q+1)!} \left( \frac{d\xi}{dx} \right) \right)' + \sum_{j=0}^q \epsilon_j \ell''_j(x). \end{aligned} \quad (5.10)$$

En particular, las expresiones (5.8), (5.9) y (5.10) para los puntos *nodales*  $x_k$ , respectivamente quedan:

$$E_q(x_k) = \epsilon_k, \quad (5.11)$$

$$E'_q(x_k) = \pi'_{q+1}(x_k) \frac{f^{(q+1)}(\xi)}{(q+1)!} + \sum_{j=0}^q \epsilon_j \ell'_j(x_k), \quad (5.12)$$

$$E''_q(x_k) = \pi''_{q+1}(x_k) \frac{f^{(q+1)}(\xi)}{(q+1)!} + 2\pi'_{q+1}(x_k) \frac{f^{(q+2)}(\xi)}{(q+1)!} \left( \frac{d\xi}{dx} \right) + \sum_{j=0}^q \epsilon_j \ell''_j(x_k). \quad (5.13)$$

Si  $q$  es par, el número de puntos de interpolación es  $q+1$  impar y las moléculas computacionales son centradas. Si  $q$  es impar, el número de puntos de interpolación es  $q+1$  par y la molécula computacional no es centrada.

Es importante hacer notar que si la malla es equiespaciada y  $q$  es par,  $\pi''_{q+1}(x)$  se anula en el punto central y el factor  $\pi''_{q+1}(x)$  en (5.13) se hace cero, quedando la derivada primera y la derivada segunda con el mismo orden de error. Por esta razón, siempre se elige  $q$  par cuando se interpola con polinomios continuos a trozos.

En los siguientes apartados se discutirá el error de la derivada primera y segunda para una interpolación continua a trozos basada en tres puntos ( $q = 2$ ).

### 5.3. Fórmulas para derivadas con tres puntos

Se consideran tres puntos de interpolación no equiespaciados tal como se representa en la figura (5.4).

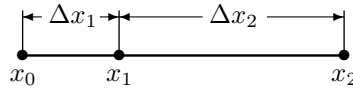


Figura 5.4: Tres puntos de interpolación no equiespaciados

Con estos tres puntos se define un interpolante de segundo orden  $I_2(x)$

$$I_2(x) = f_0 \ell_0(x) + f_1 \ell_1(x) + f_2 \ell_2(x) \quad (5.14)$$

con los polinomios de Lagrange asociados:

$$\ell_0(x) = \frac{(x - x_1)(x - x_2)}{(x_0 - x_1)(x_0 - x_2)}, \quad (5.15)$$

$$\ell_1(x) = \frac{(x - x_0)(x - x_2)}{(x_1 - x_0)(x_1 - x_2)}, \quad (5.16)$$

$$\ell_2(x) = \frac{(x - x_0)(x - x_1)}{(x_2 - x_0)(x_2 - x_1)}. \quad (5.17)$$

Para calcular las fórmulas de las derivadas primera y segunda se deriva el interpolante y se particulariza en el punto deseado. Estas expresiones para la derivada primera y segunda particularizadas en  $x = x_1$  quedan:

$$I'_2(x_1) = -\frac{\Delta x_2 f_0}{(\Delta x_2 + \Delta x_1) \Delta x_1} + \frac{(\Delta x_2 - \Delta x_1) f_1}{\Delta x_1 \Delta x_2} + \frac{\Delta x_1 f_2}{(\Delta x_2 + \Delta x_1) \Delta x_2}. \quad (5.18)$$

$$I_2''(x_1) = \frac{2 f_0}{(\Delta x_2 + \Delta x_1) \Delta x_1} - \frac{2 f_1}{\Delta x_1 \Delta x_2} + \frac{2 f_2}{(\Delta x_2 + \Delta x_1) \Delta x_2}. \quad (5.19)$$

Cuando la derivada primera se particulariza en el extremo  $x = x_0$ , se obtiene:

$$I_2'(x_0) = -\frac{(\Delta x_2 + 2\Delta x_1)f_0}{(\Delta x_2 + \Delta x_1) \Delta x_1} + \frac{(\Delta x_2 + \Delta x_1)f_1}{\Delta x_1 \Delta x_2} - \frac{\Delta x_1 f_2}{(\Delta x_2 + \Delta x_1) \Delta x_2}. \quad (5.20)$$

Las fórmulas anteriores permiten aproximar las derivadas primera y segunda en un punto central o en un punto del contorno. Si la malla es equiespaciada, las expresiones anteriores se reducen a las conocidas fórmulas de diferencias finitas para las derivadas primera y segunda

$$I_2'(x_1) = \frac{1}{2 \Delta x} (f_2 - f_0), \quad (5.21)$$

$$I_2''(x_1) = \frac{1}{\Delta x^2} (f_2 - 2f_1 + f_0), \quad (5.22)$$

$$I_2'(x_0) = \frac{1}{2\Delta x} (-3f_0 + 4f_1 - f_2), \quad (5.23)$$

$$I_2'(x_2) = \frac{1}{2\Delta x} (+3f_2 - 4f_1 + f_0). \quad (5.24)$$

## 5.4. Error de interpolación con tres puntos

Las expresiones para el error de truncamiento y el error de redondeo de la derivada primera (5.12) y de la derivada segunda (5.13) particularizadas para  $q = 2$  quedan:

$$R_2'(x_k) = \pi_3'(x_k) \frac{f^{(3)}(\xi)}{3!} + \sum_{j=0}^2 \epsilon_j \ell_j'(x_k), \quad (5.25)$$

$$R_2''(x_k) = \pi_3''(x_k) \frac{f^{(3)}(\xi)}{3!} + 2\pi_3'(x_k) \frac{f^{(4)}(\xi)}{3!} \left( \frac{d\xi}{dx} \right) + \sum_{j=0}^2 \epsilon_j \ell_j''(x_k). \quad (5.26)$$

### 5.4.1. Error de truncamiento

La influencia de la distribución de puntos en el error de truncamiento está caracterizada por  $\pi_3'(x_k)$  y por  $\pi_3''(x_k)$ . La función  $\pi_3$  para este caso es:

$$\pi_3(x) = (x - x_0)(x - x_1)(x - x_2), \quad (5.27)$$

y sus derivadas primera y segunda son:

$$\begin{aligned}\pi'_3(x) &= (x - x_0)(x - x_1) + (x - x_0)(x - x_2) + (x - x_1)(x - x_2), \\ \pi''_3(x) &= 6x - 2(x_0 + x_1 + x_2).\end{aligned}\tag{5.28}$$

Particularizamos la derivada primera y segunda de  $\pi_3(x)$  en el punto nodal medio  $x_1$  de la figura (5.4)

$$\begin{aligned}\pi'_3(x_1) &= (x_1 - x_0)(x_1 - x_2), \\ \pi''_3(x_1) &= 6x_1 - 2x_0 - 2x_1 - 2x_2 = 2(x_1 - x_0 - x_2) = 4 \left[ x_1 - \frac{1}{2}(x_0 + x_2) \right],\end{aligned}$$

y sustituyendo las distancias entre los puntos nodales por su espaciado correspondiente  $\Delta x_1 = x_1 - x_0$  y  $\Delta x_2 = x_2 - x_1$ , las derivadas primera y segunda de  $\pi$  en  $x_1$  resultan:

$$\begin{aligned}\pi'_3(x_1) &= -\Delta x_1 \Delta x_2, \\ \pi''_3(x_1) &= 2(\Delta x_1 - \Delta x_2).\end{aligned}$$

Con estas expresiones, los errores de truncamiento para la derivada primera y segunda en el punto  $x_1$  quedan:

$$R'_2(x_1) = -\Delta x_1 \Delta x_2 \frac{f^{(3)}(\xi)}{3!},\tag{5.29}$$

$$R''_2(x_1) = -2\Delta x_1 \Delta x_2 \frac{f^{(4)}(\xi)}{3!} \left( \frac{d\xi}{dx} \right) + 2(\Delta x_1 - \Delta x_2) \frac{f^{(3)}(\xi)}{3!}.\tag{5.30}$$

A la vista de las anteriores expresiones (5.29) y (5.30) podemos extraer las siguientes conclusiones:

1. Si consideramos  $N + 1$  puntos nodales equiespaciados en el compacto  $[-1, 1]$ , entonces,  $\Delta x = 2/N$  y de la expresión (5.29) vemos que el error truncamiento es inversamente proporcional al número de puntos al cuadrado. O lo que es lo mismo, si integramos numéricamente con  $2N$  puntos, entonces el error de truncamiento se divide por  $2^2$ . En general, se dice que un determinado esquema numérico es de orden  $q$  si al duplicar el número de puntos manteniendo la distribución de los puntos nodales constante, el error se divide por  $2^q$ .

2. Por otra parte, no podemos asegurar nada acerca del comportamiento del error cuando se cambia el número de puntos y además se cambia su distribución. En este caso, si el cambio de la distribución es favorable a la solución, el error disminuirá aunque no sabemos con que ley. Sin embargo, si la nueva distribución es desfavorable a la solución, el error podrá aumentar aún incluso habiendo aumentado el número de puntos.
3. Mientras que el primer término del desarrollo (5.30) es el error local de truncamiento propio del esquema centrado de tres puntos y aparece incluso en mallas con espaciamiento uniforme, el segundo término es propio de las mallas no uniformes. Si la malla es uniforme ( $\Delta x_1 = \Delta x_2$ ), este segundo término es cero. En distribuciones de puntos en las que la variación del paso es del mismo orden que el paso, este término llega a ser  $O(\Delta x)$ . Para estas distribuciones nodales, el error de truncamiento de la derivada segunda es  $O(\Delta x)$  en lugar de  $O(\Delta x^2)$  como lo es el error de truncamiento de la derivada primera. Por lo tanto, si queremos que el error de truncamiento de la derivada primera y de la derivada segunda sean del mismo orden, deberemos exigir condiciones de regularidad a la distribución de puntos nodales. Si

$$\Delta x_i = \Delta x_{i-1} + O(\Delta x^2),$$

entonces todos los términos de (5.30) son del mismo orden y el error de truncamiento es  $O(\Delta x^2)$ .

4. En zonas con gradientes fuertes, la distancia entre los puntos nodales se puede reducir de forma que el error de truncamiento se mantenga constante en todo el dominio. De esta manera, la distribución de puntos nodales se puede adaptar a la forma de la solución para minimizar el error de truncamiento.

### 5.4.2. Error de redondeo

Las expresiones para el error de redondeo de la derivada primera y segunda para las fórmulas con tres puntos particularizadas para el punto  $x_k$  son:

$$R'_L(x_k) = \sum_{j=0}^2 \epsilon_j \ell'_j(x_k), \quad (5.31)$$

$$R''_L(x_k) = \sum_{j=0}^2 \epsilon_j \ell''_j(x_k). \quad (5.32)$$

Estas expresiones coinciden con las fórmulas (5.18) y (5.19) de la derivada primera y de la derivada segunda cuando se sustituye  $(f_0, f_1, f_2)$  por  $(\epsilon_0, \epsilon_1, \epsilon_2)$ .

En el caso de una malla equiespaciada, las expresiones (5.31) y (5.32) particularizadas para  $x_k = x_1$  quedan:

$$R'_L(x_1) = \frac{\epsilon_2 - \epsilon_1}{2 \Delta x}, \quad R''_L(x_1) = \frac{\epsilon_2 - 2\epsilon_1 + \epsilon_0}{\Delta x^2}. \quad (5.33)$$

### 5.4.3. Error de interpolación: truncamiento y redondeo

Una vez hemos obtenido el error de truncamiento y el error de redondeo pasamos a analizar el tamaño o la importancia de estos dos términos en distribuciones equiespaciadas de puntos en función del número de puntos nodales  $N$ . Las expresiones para el error de truncamiento (5.29) (5.30) junto con las del error de redondeo (5.33) permiten escribir el error total de interpolación para la derivada primera y segunda

$$E'_2(x_1) = -\Delta x^2 \frac{f^{(3)}(\xi)}{3!} + \frac{\epsilon_2 - \epsilon_1}{2 \Delta x}, \quad (5.34)$$

$$E''_2(x_1) = -2\Delta x^2 \frac{f^{(4)}(\xi)}{4!} \left( \frac{d\xi}{dx} \right) + \frac{\epsilon_2 - 2\epsilon_1 + \epsilon_0}{\Delta x^2}. \quad (5.35)$$

Acotando las expresiones anteriores se tiene:

$$|E'_2(x_1)| \leq \frac{\Delta x^2}{6} K_1 + \frac{2 \epsilon}{2 \Delta x}, \quad (5.36)$$

$$|E''_2(x_1)| \leq \frac{\Delta x^2}{3} K_2 + \frac{4 \epsilon}{\Delta x^2}, \quad (5.37)$$

donde  $|f^{(3)}(\xi)| \leq K_1$ ,  $\left| f^{(4)}(\xi) \left( \frac{d\xi}{dx} \right) \right| \leq K_2$  y  $\epsilon = \max(\epsilon_0, \epsilon_1, \epsilon_2)$ . Las expresiones (5.36) y (5.37) se representan respectivamente en las figuras (5.5) y (5.6) en función de  $\Delta x$ .

Para los dos casos se observa la presencia de un mínimo de error para un  $\Delta x$  óptimo. La razón reside en la importancia del error de redondeo frente al error de truncamiento.



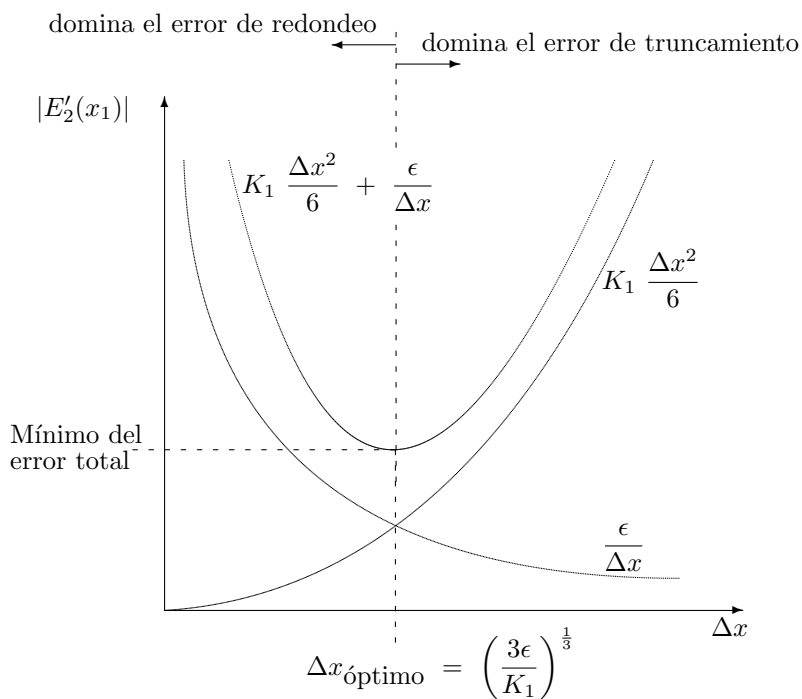


Figura 5.5: Error total de la fórmula para la derivada primera centrada con tres puntos en una malla equiespaciada

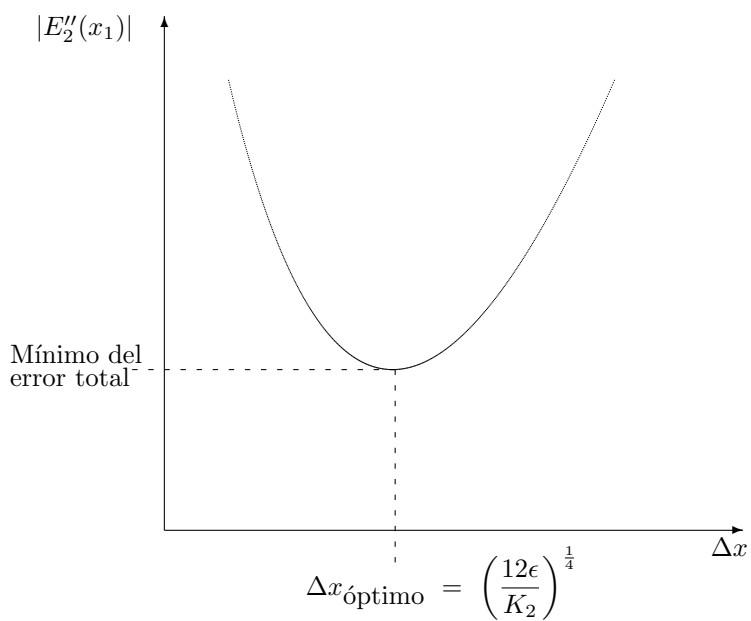


Figura 5.6: Error total de la fórmula para la derivada segunda centrada con tres puntos en una malla equiespaciada

# Capítulo 6

## Problema de contorno

En este capítulo se analizará por separado el problema de contorno en ecuaciones diferenciales ordinarias y el problema de contorno en ecuaciones en derivadas parciales.

Un problema de contorno se define por un operador diferencial  $\mathcal{L}(u)$  y sus condiciones de contorno  $BC(u) = 0$  siguientes:

$$\begin{aligned}\mathcal{L}(u) &= 0, & \forall x \in \Omega, & \quad \Omega \subset \mathbb{R}^d \\ BC(u)|_{\partial\Omega} &= 0,\end{aligned}\tag{6.1}$$

donde  $\mathcal{L}(u)$  constituye una relación entre los valores de la función y sus derivadas. El orden de un problema diferencial es el mayor orden de las derivadas que involucra el operador  $\mathcal{L}(u)$ . Esta relación es válida en un dominio  $\Omega$ . El problema de contorno se caracteriza por dar las condiciones que debe cumplir la solución general de (6.1) en el contorno del dominio  $\Omega$ .

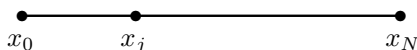
### 6.1. Ecuaciones diferenciales ordinarias

En particular, si el problema fuera unidimensional y la expresión  $\mathcal{L}(u)$  fuera un operador escalar de segundo orden, se deben dar dos condiciones para determinar la solución, una en  $x = x_0$  y la otra en  $x = x_f$ . Si el problema fuera de orden superior a dos, el problema se puede reducir a un problema vectorial de segundo orden mediante la inclusión de una variable auxiliar. Generalmente, las leyes de

conservación de la física en formulación primitiva constituyen sistemas de ecuaciones en derivadas parciales que involucran a lo sumo derivadas segundas de las magnitudes físicas.

A continuación, se expone el algoritmo para resolver un problema de contorno escalar en un dominio unidimensional:

1. Partición del dominio de integración mediante una malla equiespaciada o no equiespaciada.



2. Elegir el interpolante: global o continuo a trozos.  
Si es global, estaremos frente a métodos espectrales y si es continuo a trozos frente a diferencias finitas.
3. Calcular las expresiones para las derivadas primera y segunda. En general, las condiciones de contorno se representan por el operador  $BC(u) = 0$  que involucra a la función y a su derivada primera. Estas condiciones o relaciones diferenciales se deben cumplir en el contorno de  $\Omega$ . En el caso unidimensional el contorno de  $\Omega$  es  $x = x_0$  y  $x = x_f$ . En el caso de diferencias finitas se utilizan las fórmulas centradas o descentradas obtenidas en los capítulos anteriores. En el caso de interpolación global, se pueden obtener las fórmulas de las derivadas en el plano espectral o en el plano físico. Si el problema se resuelve en el plano físico, las incógnitas son los valores de la función en los puntos nodales mientras que si el problema se resuelve en el plano espectral, las incógnitas del problema son las amplitudes de los armónicos de la serie discreta.
4. Imponemos el cumplimiento del operador diferencial  $\mathcal{L}(u) = 0$  en cada uno de los puntos interiores  $x_j$ ,  $j = 1, \dots, N - 1$ .
5. Imponemos las condiciones de contorno  $BC(u) = 0$  en  $x = x_0$  y en  $x = x_N$  mediante las fórmulas para las derivadas obtenidas en el paso 3.
6. Si el problema es escalar, el sistema anterior constituye un sistema de  $N + 1$  ecuaciones con  $N + 1$  incógnitas. Si la resolución se hace en el plano físico, las incógnitas son los valores de la función en los puntos nodales  $u_j$ ,  $j = 0, \dots, N$ . Por el contrario, si la resolución se hace en el plano espectral, las incógnitas son las amplitudes de los armónicos del desarrollo en

serie:

$$I_N(x) = \sum_{k=0}^N \tilde{c}_k \phi_k(x), \quad \{\tilde{c}_k, \quad k = 0, \dots, N\}.$$

Dependiendo de la linealidad del operador diferencial y de la linealidad de las condiciones de contorno, el sistema resultante de las ecuaciones algebraicas puede ser lineal o no lineal.

7. El último punto pasa por resolver el sistema algebraico del paso 6.

La figura (6.1) resume el algoritmo explicado anteriormente.

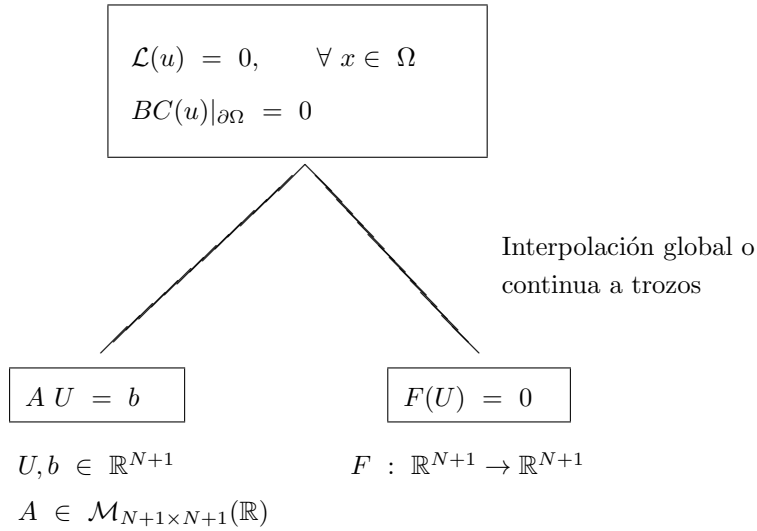


Figura 6.1: Esquema del algoritmo para la solución de un problema de contorno

A continuación, a modo de ejemplo, se presenta el siguiente problema de contorno:

$$\begin{aligned} u'' + \pi^2 u &= 0, \\ u(-1) &= 1, \\ u(+1) &= 1, \end{aligned} \tag{6.2}$$

donde la solución analítica es:

$$u(x) = -\cos(\pi x). \tag{6.3}$$

Se plantea una partición espacial tal como se indica en la figura (6.2). Imponiendo que se verifique el operador diferencial en los puntos interiores y las condiciones de

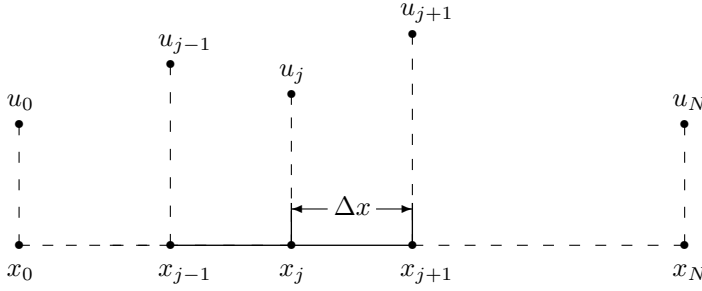


Figura 6.2: Partición de la malla espacial

contorno se obtienen el siguiente sistema de  $N + 1$  ecuaciones con  $N + 1$  incógnitas  $u_j$  o valores nodales.

$$\begin{aligned} \frac{1}{\Delta x^2} (u_{j+1} - 2u_j + u_{j-1}) + \pi^2 u_j &= 0, \quad j = 1, \dots, N-1, \\ u_0 &= 1, \quad u_N = 1. \end{aligned} \quad (6.4)$$

El sistema anterior expresado en forma vectorial queda:

$$A U = b, \quad (6.5)$$

donde la matriz del sistema es:

$$A = \begin{pmatrix} 1 & 0 & & \dots & 0 \\ 1 & -2 + \Delta x^2 \pi^2 & & 1 & 0 \\ 0 & \ddots & \ddots & & \\ 0 & 1 & -2 + \Delta x^2 \pi^2 & & 1 \\ 0 & \dots & & 0 & 1 \end{pmatrix}, \quad (6.6)$$

con el vector de incógnitas  $U$  y el término independiente  $b$

$$U = \begin{pmatrix} u_0 \\ u_1 \\ \vdots \\ u_{N-1} \\ u_N \end{pmatrix}, \quad b = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}. \quad (6.7)$$

Generalmente, las discretizaciones de problemas lineales de contorno llevan asociados matrices que tienen formas mucho más complicadas que la del ejemplo anterior.

Para facilitar la construcción de la matriz del sistema, se define una función vectorial  $L(U)$  cuyo argumento es un vector de incógnitas o puntos nodales  $U$  y la imagen es el conjunto de ecuaciones (6.4). Si la función depende linealmente de  $U$ , entonces  $F(U) = A U + b$  donde  $A$  es la matriz del sistema y  $b$  es el término independiente.

Mediante sucesivas evaluaciones de la función vectorial  $L(U)$  obtenemos tanto el término independiente como la matriz del sistema. En concreto,  $F(0)$  coincide con el término independiente  $b$ . Además, si hacemos todas las componentes de  $U$  cero menos la componente  $j$  que valga 1, la imagen de la función vectorial  $L$  para este valor de  $U$  es la columna  $j$  de la matriz más el término independiente  $b$ . De esta forma, barriendo todos los posibles valores de  $U$  desde  $j = 0, \dots, N$ , obtenemos todas las columnas de la matriz  $A$ .

## 6.2. Dominios bidimensionales

Cuando el dominio de integración  $\Omega$  es bidimensional o tridimensional, el operador diferencial  $\mathcal{L}(u) = 0$  puede involucrar derivadas parciales en cada una de las direcciones del espacio. Desde el punto de vista conceptual, el problema de contorno es el mismo que en un dominio unidimensional salvo que en este caso es necesario calcular derivadas a lo largo de direcciones coordenadas diferentes. En el caso de que las direcciones coordenadas sean ortogonales, el problema se puede reducir a un interpolante unidimensional de forma muy simple. Se considera un interpolante bidimensional a trozos

$$I(x, y) = \sum_{k=0}^q \sum_{m=0}^q f_{km} \ell_k(x) \ell_m(y), \quad (6.8)$$

donde,

$$\ell_k(x) = \prod_{\substack{n=i-q/2 \\ n \neq k}}^{i+q/2} \frac{(x - x_n)}{(x_k - x_n)}, \quad \ell_m(y) = \prod_{\substack{n=j-q/2 \\ n \neq m}}^{j+q/2} \frac{(y - y_n)}{(y_m - y_n)}. \quad (6.9)$$

La expresión (6.8) es un interpolante que satisface los puntos nodales  $(x_k, y_m)$  con los valores de la función  $f(x_k, y_m)$  que se representan por  $f_{km}$  para los siguientes puntos de la molécula computacional:

$$\begin{aligned} &\{x_{i-\frac{q}{2}}, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_{i+\frac{q}{2}}\}, \\ &\{y_{j-\frac{q}{2}}, \dots, y_{j-1}, y_j, y_{j+1}, \dots, y_{j+\frac{q}{2}}\}. \end{aligned} \quad (6.10)$$

De esta forma se construye una partición equiespaciada o no equiespaciada del dominio  $\Omega$  con  $N_x + 1$  puntos en la dirección  $x$  y  $N_y + 1$  puntos en la dirección  $y$ .

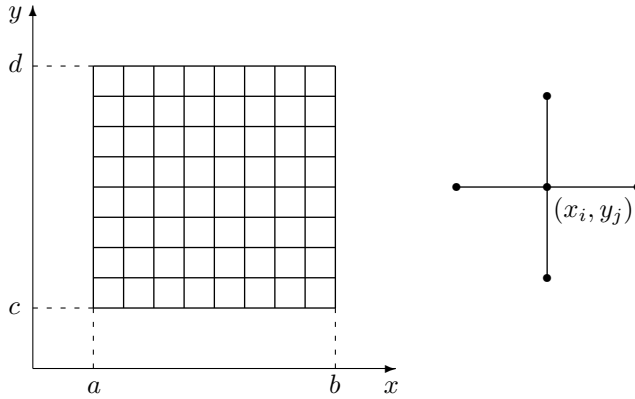


Figura 6.3: Partición equiespaciada en un dominio bidimensional

De igual forma que en el caso 1D, se fuerza al cumplimiento del operador diferencial en los puntos interiores o puntos de colocación:

$$\mathcal{L}(u)|_{x_i, y_j} = 0, \quad (6.11)$$

que constituyen un sistema de  $(N_x - 1) \times (N_y - 1)$  ecuaciones que junto con las condiciones de contorno:

$$BC(u)|_{\partial\Omega} = 0, \quad (6.12)$$

que son  $2(N_x - 1) + 2(N_y + 1) = 2N_x + 2N_y$  ecuaciones permiten obtener las  $(N_x + 1)(N_y + 1)$  incógnitas de la función en los puntos nodales.



# Capítulo 7

## Problema de Cauchy en EDOS

Un problema de Cauchy en EDOS está formado por un conjunto de ecuaciones diferenciales de evolución junto con una condición inicial  $U(t_0)$ .

$$\begin{aligned} \frac{dU}{dt} &= F(U; t), \quad U \in \mathbb{R}^N, \quad F : \mathbb{R}^N \times \mathbb{R} \rightarrow \mathbb{R}^N, \\ U(t_0) &= U^0, \quad \forall t \in [t_0, +\infty). \end{aligned} \quad (7.1)$$

En los problemas de Cauchy, la variable independiente suele ser el tiempo. La nomenclatura adoptada en esta sección se muestra en la Figura 7.1. Con el superíndice  $n$  se indica la aproximación en el instante temporal  $t_n$ .

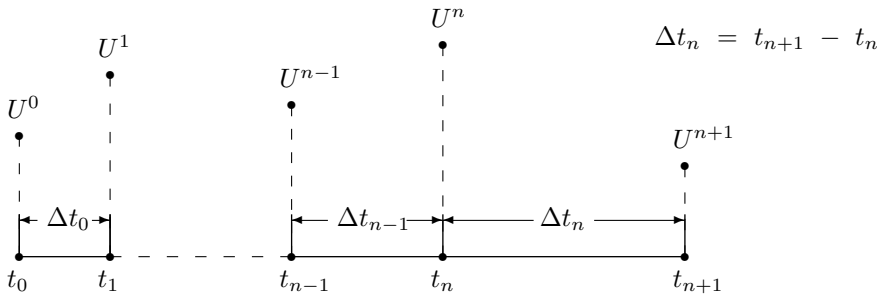


Figura 7.1: Partición de la malla temporal

En los problemas de Cauchy o problemas de evolución se trata de obtener una aproximación para  $U^{n+1}$  en función de los valores anteriores  $U^j$  y  $F^j$ . Los métodos numéricos para la discretización temporal se pueden dividir en dos grandes grupos: (i) métodos unipaso y (ii) métodos multipaso. Mientras que los métodos multipaso utilizan los valores  $U^j$  y  $F^j$  en pasos anteriores, los métodos unipaso utilizan la evaluación de  $F(U; t)$  en etapas intermedias entre  $t_n$  y  $t_{n+1}$ .

## 7.1. Métodos multipaso. Métodos Adams

Para obtener los esquemas multipaso o esquemas Adams, se integra la ecuación diferencial de (7.1) entre  $t_n$  y  $t_{n+1}$

$$U(t_{n+1}) = U(t_n) + \int_{t_n}^{t_{n+1}} F(U; t) dt. \quad (7.2)$$

para aproximar la integral de (7.2) se construye un interpolante para  $F$  de grado  $p$  que involucra los  $p+1$  siguientes:

$$F^{n+1}, F^n, F^{n-1}, \dots, F^{n+1-p}.$$

De esta forma, el interpolante de grado  $p$  se escribe:

$$I_p(t) = \sum_{j=0}^p F^{n+1-j} \ell_j(t), \quad \ell_j(t) = \prod_{\substack{k=n+1-p \\ k \neq j}}^{k=n+1} \frac{(t - t_k)}{(t_j - t_k)}. \quad (7.3)$$

Cuando se lleva esta expresión del interpolante a la ecuación (7.2) y se integra, los esquemas resultantes o esquemas de Adams quedan:

$$U^{n+1} = U^n + \Delta t_n \sum_{j=0}^p \beta_j F^{n+1-j}, \quad n = p-1, \dots \quad (7.4)$$

donde los valores  $\beta_j$  son números que se obtienen al integrar los polinomios de Lagrange

$$\beta_j = \frac{1}{\Delta t_n} \int_{t_n}^{t_{n+1}} \ell_j(t) dt. \quad (7.5)$$

Si el interpolante (7.3) involucra  $F^{n+1}$  entonces, el esquema (7.4) no permite obtener de forma explícita  $U^{n+1}$  en función de los puntos anteriores y los esquemas resultantes son implícitos y se denominan esquemas Adams–Moulton. Por el contrario, cuando el interpolante (7.3) no involucra  $F^{n+1}$ , entonces los esquemas (7.4) son explícitos y se denominan esquemas Adams–Bashforth. La expresión (7.4) permite obtener la solución  $U^{n+1}$  en el instante  $t_{n+1}$  conocida la solución en  $p$  pasos

anteriores. Es decir, la ecuación en diferencias (7.4) requiere  $p$  condiciones iniciales para poder ser utilizado. Como solo disponemos de una condición inicial, las  $p - 1$  condiciones iniciales o de arranque de los métodos multipaso se deben calcular mediante otros esquemas unipaso o esquemas que involucren menos pasos. Una vez determinadas las  $p$  condiciones iniciales, la expresión (7.4) permite obtener de manera directa  $U^{n+1}$  si el esquema es explícito o  $\beta_0 = 0$

$$U^{n+1} = U^n + \Delta t_n (\beta_1 F^n + \dots + \beta_p F^{n+1-p}), \quad n = p - 1. \quad (7.6)$$

Sin embargo, cuando el esquema es implícito o  $\beta_0$  es distinto de cero, la ecuación resultante (7.4) queda:

$$U^{n+1} = C + \Delta t_n \beta_0 F(U^{n+1}; t_{n+1}), \quad (7.7)$$

donde  $C$  se puede determinar a partir de los pasos anteriores y el sistema resultante (7.7) constituye un sistema vectorial no lineal de  $N$  ecuaciones con  $N$  incógnitas o componentes de  $U^{n+1}$ . Mientras que los esquemas explícitos permiten obtener  $U^{n+1}$  en función de los anteriores, los esquemas implícitos requieren resolver un sistema no lineal de ecuaciones cada vez que damos un paso temporal. La única razón para utilizar esquemas implícitos frente a los esquemas explícitos es que se comportan, en general, mejor en términos de estabilidad.

## 7.2. Métodos unipaso. Métodos Runge Kutta

Los métodos Runge-Kutta se basan en evaluar etapas intermedias entre  $t_n$  y  $t_{n+1}$ . Aunque son más robustos que los métodos multipaso exigen, en general, más carga computacional. La expresión general de un esquema Runge-Kutta de  $e$  etapas o número de evaluaciones intermedias entre  $t_n$  y  $t_{n+1}$  es:

$$U^{n+1} = U^n + \Delta t_n \sum_{i=1}^e b_i k_i, \quad (7.8)$$

donde

$$k_i = F \left( U^n + \Delta t \sum_{j=1}^e a_{ij} k_j; t_n + c_i \Delta t_n \right), \quad i = 1, \dots, e. \quad (7.9)$$

Los coeficientes  $b_i, c_i$  y  $a_{ij}$  están dados y son constantes propias del esquema. En general, los esquemas son implícitos y requieren resolver un sistema no lineal (7.9) de  $e N$  ecuaciones por cada paso de integración. Sin embargo, si la matriz  $a_{ij}$  es triangular inferior con ceros en la diagonal, los esquemas son explícitos.

**Ejemplo:**

Se considera el siguiente esquema Runge–Kutta de dos etapas:

$$U^{n+1} = U^n + \Delta t_n (b_1 k_1 + b_2 k_2), \quad (7.10)$$

donde

$$k_1 = F(U^n + \Delta t (a_{11}k_1 + a_{12}k_2); t_n + c_1 \Delta t_n), \quad (7.11)$$

$$k_2 = F(U^n + \Delta t (a_{21}k_1 + a_{22}k_2); t_n + c_2 \Delta t_n). \quad (7.12)$$

Como  $F$  es una función vectorial de dimensión  $N$ , las ecuaciones vectoriales (7.11)–(7.12) son de dimensión  $N$  e involucran las incógnitas  $k_1$  y  $k_2$ . Conocida la condición inicial  $U^0$ , la solución  $U^1$  se determina resolviendo el sistema no lineal de dimensión  $2N$  y llevando los valores obtenidos  $k_1$   $k_2$  a la expresión (7.10).

Si  $a_{11} = 0$ ,  $a_{12} = 0$  y  $a_{22} = 0$ , el esquema es explícito. La pendiente  $k_1$  se puede obtener a partir del paso anterior mediante:

$$k_1 = F(U^n; t_n + c_1 \Delta t_n), \quad (7.13)$$

y conocido  $k_1$  determinar la evaluación intermedia  $k_2$  mediante:

$$k_2 = F(U^n + \Delta t a_{21}k_1; t_n + c_2 \Delta t_n). \quad (7.14)$$

Finalmente, llevando los valores de  $k_1$  y  $k_2$  a la expresión (7.10) se puede obtener el valor en el paso siguiente.

### 7.3. Error global y error local de truncamiento

En esta sección se propone estudiar la relación que existe entre el error de la solución numérica y el error de truncamiento o aproximación asociado a un esquema numérico dado. Para poder obtener esta relación, se define el error global de la solución numérica y el error local de truncamiento del esquema.

*Definición:* El error global de la solución numérica en un instante  $t_n$  genérico se define como la diferencia entre la solución exacta evaluada en ese instante menos la solución numérica

$$E^n = U(t_n) - U^n. \quad (7.15)$$

*Definición:* Se dice que un esquema es de orden  $q$  cuando su error local de truncación es  $O(\Delta t^{q+1})$ .

Debido al proceso de acumulación del error local, se demuestra que el error global es orden  $q$  o lo que es lo mismo su error es  $O(\Delta t^q)$ . De esta forma, para que

el error global sea al menos de primer orden o que el error tienda a cero cuando  $\Delta t$  tienda a cero, el error local debe ser al menos  $O(\Delta t^2)$ .

*Definición:* El error local de truncamiento de un esquema numérico en un instante  $t_n$  genérico se define como el residuo que deja la solución exacta cuando se introduce en el esquema numérico.

Para los métodos lineales multipaso, esta expresión queda:

$$T^{n+1} = U(t_{n+1}) - U(t_n) - \Delta t_n \sum_{j=0}^p \beta_j F(U(t_{n+1-j}); t_{n+1-j}). \quad (7.16)$$

Para los métodos Runge–Kutta, esta expresión queda:

$$T^{n+1} = U(t_{n+1}) - U(t_n) - \Delta t_n \sum_{i=1}^e b_i k_i(U), \quad (7.17)$$

donde los valores  $k_i$  se determinan introduciendo la solución exacta en la ecuaciones (7.9). Una vez definido con precisión el error local y el error global, se procede a obtener la relación entre ambos.

Para los métodos Runge–Kutta, la relación se obtiene restando las ecuación (7.17) menos la ecuación (7.8)

$$T^{n+1} = E^{n+1} - E^n - \Delta t_n \sum_{i=1}^e b_i (k_i(U) - k_i). \quad (7.18)$$

Reordenando esta ecuación y considerando que  $k_i$  es una función suficientemente diferenciable de  $U$ ,

$$E^{n+1} = E^n + \Delta t_n \sum_{i=1}^e b_i \frac{\partial k_i}{\partial U} E^n + T^{n+1}. \quad (7.19)$$

Esta ecuación en diferencias de primer orden o un paso expresa la relación entre el error local y el error global. El error local aparece como un término fuente en la acumulación del error global.

Para los métodos lineales multipaso, la relación se obtiene de forma similar restando la ecuación (7.16) menos la ecuación (7.4). Reordenando esta ecuación y considerando que  $F$  es una función suficientemente diferenciable de  $U$ , la ecuación del error global queda:

$$E^{n+1} = E^n + \Delta t_n \sum_{j=0}^p \beta_j \frac{\partial F}{\partial U} E^{n+1-j} + T^{n+1}. \quad (7.20)$$

El error global para los métodos lineales multipaso está gobernado por una ecuación en diferencias de  $p$  pasos en la que el error local aparece como un término fuente en la acumulación del error global.

En la siguiente sección, se analizará el comportamiento del error global a lo largo de la integración para esquemas Runge–Kutta y esquemas Adams. La teoría general se formula para sistemas de ecuaciones en diferencias de un paso para el error global. Por esta razón, la ecuación en diferencias (7.20) de  $p$  pasos se lleva a una ecuación vectorial de un paso mediante el cambio de variable independiente siguiente:  $\mathbf{V}^n = (U^n, U^{n-1}, \dots, U^{n+1-p})^T$ .

## 7.4. Acotación del error y estabilidad numérica

Se considera que el error global está gobernado por una ecuación en diferencias de un paso del tipo (7.19) o de  $p$  pasos del tipo (7.20). Mediante la reducción a sistema de primer orden o un paso, la ecuación que gobierna el error global de un esquema numérico se escribe:

$$E^{n+1} = B E^n + T^{n+1}, \quad n = 0, 1, \dots \quad (7.21)$$

donde  $B$  es una matriz cuyos coeficientes dependen de la solución exacta  $U(t)$  particularizada en los diferentes pasos de integración. Conocido el error en la condición inicial  $E^0$ , se puede determinar el error en instantes sucesivos.

En esta sección se analiza el comportamiento del error global cuando el sistema (7.1) es lineal de coeficientes constantes y se integra un paso de tiempo constante  $\Delta t$ . En este caso, el sistema se expresa:

$$\frac{dU}{dt} = A U + b, \quad U \in \mathbb{R}^N, \quad F : \mathbb{R}^N \times \mathbb{R} \rightarrow \mathbb{R}^N, \quad (7.22)$$

$$U(t_0) = U^0, \quad \forall t \in [t_0, +\infty),$$

donde  $A$  es la matriz del sistema de coeficientes constantes y  $b$  es el término independiente de coeficientes constantes. En este caso, la matriz  $B$  del sistema (7.21) es de coeficientes constantes que son funciones de la matriz  $A$  y del paso de tiempo  $\Delta t$  y el sistema (7.21) se resuelve para dar:

$$E^n = B^n E^0 + \sum_{k=1}^n B^{n-k} T^k, \quad n = 1, 2, \dots \quad (7.23)$$

La norma del error global  $\|E^n\|$  mediante la desigualdad triangular se expresa:

$$\|E^n\| \leq \|B^n\| \|E^0\| + \sup_{k \in [1, n]} \|T^k\| \sum_{k=1}^n \|B^{n-k}\|. \quad (7.24)$$

Si  $B$  es normal,  $B B^T = B^T B$ , existe un conjunto ortonormal de autovalores que diagonaliza  $B$  y la norma:

$$\|B^n\| = \rho^n, \quad \text{con} \quad \rho = \sup_{z \in \Lambda(B)} |z|. \quad (7.25)$$

En esta expresión  $\rho$  representa el radio espectral de  $B$  o el valor máximo del módulo de todos los autovalores de  $B$  o espectro  $\Lambda(B)$  de  $B$ . De esta forma, la expresión (7.24) queda:

$$\|E^n\| \leq \rho^n \|E^0\| + \sup_{k \in [1, n]} \|T^k\| \sum_{k=1}^n \rho^{n-k}. \quad (7.26)$$

El sumatorio de la expresión anterior involucra una serie geométrica de razón  $\rho$  que se suma para dar:

$$\|E^n\| \leq \rho^n \|E^0\| + \sup_{k \in [1, n]} \|T^k\| \left| \frac{1 - \rho^{n+1}}{1 - \rho} \right|. \quad (7.27)$$

Cuando se integran órbitas o sistemas Hamiltonianos,  $\rho$  puede hacerse igual a la unidad y la expresión anterior (7.27) no es válida. Llevando el valor  $\rho = 1$  a la cota (7.26) y considerando que el tiempo total de integración  $\tau = \Delta t n$ , el error de las condiciones iniciales se mantiene acotado durante la integración y el error global permanece acotado por el error local de truncamiento dividido por  $\Delta t$ .

$$\|E^n\| \leq \|E^0\| + \frac{\tau}{\Delta t} \sup_{k \in [1, n]} \|T^k\|. \quad (7.28)$$

Si  $\rho < 1$  y  $n \rightarrow \infty$ , la propagación del error de las condiciones iniciales tiende a cero y el error global está acotado por el error local de truncamiento

$$\|E^n\| \leq \frac{1}{1 - \rho} \sup_{k \in [1, n]} \|T^k\|. \quad (7.29)$$

Si  $B$  es no normal, la ecuación (7.25) no es válida pero mediante la forma canónica de Jordan  $H = Q^{-1} B Q$  podemos acotar,

$$\|B^n\| \leq \|Q^{-1}\| \|Q\| \|H^n\|. \quad (7.30)$$

Una vez más si  $\rho < 1$ , entonces  $\|H^n\|$  tiende a cero con  $n \rightarrow \infty$ , y el error temporal de la solución numérica tiende a cero con  $n \rightarrow \infty$ . Si  $\rho > 1$ , el error crece sin límite para matrices normales o no normales. El error de las condiciones iniciales y el error local de truncamiento se amplifican invalidando la solución numérica.

La pregunta que surge ahora es determinar en qué situaciones  $\rho$  es menor que uno. Para poder analizar los autovalores de  $B$  en función de los autovalores de  $A$  y poder calcular el radio espectral, se utiliza el teorema de transformación espectral. El teorema de transformación espectral involucra el polinomio característico de estabilidad que se define como:

*Definición:* El polinomio característico de estabilidad de un método lineal multipaso se define mediante:

$$\pi(r, \omega) = \sum_{j=0}^p (\alpha_j - \omega \beta_j) r^{p-j}, \quad (7.31)$$

donde  $\omega$  es un parámetro complejo del que depende el polinomio.

*Definición:* El polinomio característico de estabilidad de un método lineal unipaso se define mediante:

$$\pi(r, \omega) = r - 1 - p(\omega), \quad (7.32)$$

donde  $p(\omega)$  es un polinomio cuyo grado es el número de etapas del esquema unipaso. El teorema de transformación espectral permite obtener los autovalores de  $B$  como las raíces del polinomio característico de estabilidad cuando hacemos  $\omega = \lambda \Delta t$  donde  $\lambda$  son los autovalores de la matriz  $A$ .

Finalmente, la definición de región de estabilidad absoluta permite asegurar que el radio espectral es menor que la unidad siempre que los autovalores de  $A$  por  $\Delta t$  pertenezcan a una región del plano complejo o región de estabilidad.

*Definición:* La región de estabilidad absoluta  $\mathcal{R}_A$  de un esquema numérico es el conjunto de números  $\omega \in \mathbb{C}$  para los cuales todas las raíces del polinomio característico de estabilidad satisfacen que su módulo es menor o igual a la unidad y aquellas que tienen módulo igual a la unidad son simples.

Si denominamos por  $\lambda \Delta t$  los autovalores de  $A \Delta t$  y utilizamos estas definiciones, el teorema de transformación espectral nos asegura que si todos los autovalores  $\lambda \Delta t$  están dentro de la región de estabilidad absoluta  $\mathcal{R}_A$ , entonces todos los autovalores  $r$  de la matriz  $B$  tienen módulo menor que la unidad, el radio espectral es menor que la unidad y, en consecuencia, el error global está acotado.

Es importante hacer notar que el signo de la parte real de  $\lambda$  determina la estabilidad de las soluciones (7.22) y que el módulo de los autovalores de  $r$  de  $B$  determinan la estabilidad de las soluciones numéricas de un sistema lineal de coeficientes constantes.

La región de estabilidad absoluta incluye siempre parte del semiplano complejo con parte real negativa. Dependiendo de la forma de la región de estabilidad absoluta autovalores  $\lambda$  de la matriz  $A$  con parte real negativa se pueden meter dentro de la región de estabilidad eligiendo un paso temporal  $\Delta t$  pequeño. Otras veces o para otros esquemas esto se hace imposible y se invalida un determinado esquema numérico para un determinado problema.



## 7.5. Ejemplo: Integración del oscilador armónico

En esta sección se integra el oscilador armónico con sus condiciones iniciales:

$$\ddot{x} + x = 0, \quad x(0) = 1, \quad \dot{x}(0) = 0, \quad (7.33)$$

mediante los esquemas Euler, Euler inverso, Crank-Nicolson y esquema Leap-Frog siguientes:

$$U^{n+1} = U^n + \Delta t F^n, \quad E_1(7.34)$$

$$U^{n+1} = U^n + \Delta t F^{n+1}, \quad E_2(7.35)$$

$$U^{n+1} = U^n + \Delta t/2 (F^n + F^{n+1}), \quad E_3(7.36)$$

$$U^{n+1} = U^{n-1} + 2\Delta t F^n. \quad E_4(7.37)$$

La idea de este ejemplo es fijar los conceptos desarrollados en las secciones anteriores.

## Polinomio característico de estabilidad y radio espectral

Los polinomios característicos de estabilidad los esquemas  $E_1$ - $E_4$  son:

$$\pi_1 = r - 1 - \omega, \quad E_1(7.38)$$

$$\pi_2 = (1 - \omega)r - 1, \quad E_2(7.39)$$

$$\pi_3 = (1 - \omega/2)r - 1 - \omega/2, \quad E_3(7.40)$$

$$\pi_4 = r^2 - 2r\omega - 1. \quad E_4(7.41)$$

El teorema de transformación espectral nos permite obtener los autovalores de la matriz del problema en diferencias (7.21) en función de los autovalores de la matriz del problema diferencial (7.22). La ecuación de segundo orden (7.33) se transforma en el sistema de primer orden siguiente:

$$\frac{d}{dt} \begin{pmatrix} x \\ \dot{x} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} x \\ \dot{x} \end{pmatrix}. \quad (7.42)$$

Los autovalores de la matriz del sistema son  $\lambda_1 = i$  y  $\lambda_2 = -i$ . Las raíces del polinomio característico de estabilidad con  $\omega = \lambda \Delta t$  son los autovalores  $r$  de la matriz (7.21). De esta manera, se obtiene el radio espectral (7.25) para los diferentes esquemas,

$$\rho_1 = \sqrt{1 + \Delta t^2}, \quad E_1(7.43)$$

$$\rho_2 = 1 / \sqrt{1 + \Delta t^2}, \quad E_2(7.44)$$

$$\rho_3 = 1, \quad E_3(7.45)$$

$$\rho_4 = |i\Delta t + \sqrt{\Delta t^2 - 1}|. \quad E_4(7.46)$$

Si  $\Delta t < 1$ ,  $\rho_4 = 1$  y si  $\Delta t > 1$ , entonces  $\rho_4 > 1$ .

## Error local de truncamiento y acotación del error global

El error de truncamiento de los esquemas anteriores se obtiene a través del residuo que deja una función test  $z(t)$  suficientemente diferenciable cuando se introduce en el esquema numérico y se hace un desarrollo en serie de Taylor alrededor del origen  $t_n$ . De esta forma, el error de truncamiento para los diferentes esquemas queda:

$$T_1^{n+1} = \Delta t^2 \ddot{z}(t_n) / 2 + O(\Delta t^3) \quad (7.47)$$

$$T_2^{n+1} = -\Delta t^2 \ddot{z}(t_n) / 2 + O(\Delta t^3), \quad (7.48)$$

$$T_3^{n+1} = -\Delta t^3 \ddot{z}(t_n) / 12 + O(\Delta t^4), \quad (7.49)$$

$$T_4^{n+1} = -\Delta t^3 \ddot{z}(t_n) / 3 + O(\Delta t^5). \quad (7.50)$$

La expresión (7.27) permite acotar el error global mediante el error local de truncamiento y el radio espectral. Cuando se integra el oscilador armónico con un esquema Euler, el radio espectral  $\rho_1 > 1$  (ecuación 7.44) y, en consecuencia, el error global tiende a infinito con  $n \rightarrow \infty$ . Para el esquema Euler inverso, el radio espectral  $\rho_2 < 1$  (ecuación 7.45) y el error global para valores de  $n$  grandes se acota por:

$$E_2^n \leq \frac{\Delta t^2}{2(1 - \rho_2)} \approx 1. \quad (7.51)$$

Al ser  $E_2^n$  una cota superior con valor próximo a la unidad, no podemos asegurar que el error global de la solución numérica tienda a cero con  $\Delta t \rightarrow \infty$ . Cuando el esquema utilizado es el esquema Leap-Frog y  $\Delta t < 1$ , entonces el radio espectral es  $\rho_3 = 1$  (ecuación 7.46). Si  $\Delta t > 1$ , el radio espectral  $\rho_3 > 1$  y  $E_3^n$  tiende a infinito con  $n \rightarrow \infty$ . Cuando el esquema utilizado es el Crank-Nicolson, el radio espectral  $\rho_4 = 1$  (ecuación 7.46). Las cotas (7.26) para el error global quedan:

$$E_3^n \leq \|E^0\| + \tau \Delta t / 12, \quad (7.52)$$

$$E_4^n \leq \|E^0\| + \tau \Delta t / 3, \quad \Delta t < 1. \quad (7.53)$$

$$(7.54)$$

## Región de estabilidad absoluta

Las regiones de estabilidad absoluta junto los autovalores del problema diferencial permiten llegar a conclusiones similares. En la figura 7.2 se representan las regiones de estabilidad (marcadas en gris) de los esquemas numéricos junto con los autovalores del problema diferencial multiplicados por  $\Delta t$ . En el caso del esquema Euler figura 7.2a, los autovalores multiplicados por  $\Delta t$  ( $\pm \Delta t$ ) caen fuera de la región de estabilidad absoluta originando una solución numérica es inestable para

todo  $\Delta t > 0$ . En el caso del Euler inverso figura 7.2b, la situación es la inversa. Para todo  $\Delta t > 0$ ,  $\pm i\Delta t$  caen dentro de la región de estabilidad absoluta y hacen que la solución numérica sea estable. La región de estabilidad absoluta del esquema Crank–Nicolson es todo el semiplano complejo con parte real negativa (figura 7.2c) y los autovalores caen en la frontera de la región de estabilidad para todo valor de  $\Delta t > 0$  haciendo que la solución numérica sea marginalmente estable. La región de estabilidad absoluta del esquema Leap-Frog es el segmento imaginario  $[-1, +1]$  (figura 7.2d). En este caso, cuando el  $\Delta t < 1$ , los autovalores multiplicados por  $\Delta t$  caen en la frontera de la región haciendo que la solución numérica sea marginalmente estable. Si  $\Delta t > 1$ , entonces los autovalores caen fuera de la región y la solución numérica se vuelve inestable.

## Soluciones de las ecuaciones en diferencias

Por último y debido a que las ecuaciones en diferencias son lineales y de coeficientes constantes, se pueden calcular las expresiones analíticas de las soluciones numéricas para poder compararlas con la solución exacta

$$x(t_n) = \cos(n\Delta t) \quad (7.55)$$

del problema de Cauchy (7.33) Las soluciones de las ecuaciones en diferencias correspondientes a los diferentes esquemas son:

$$x_1^n = \rho_1^n \cos(n\theta), \quad \rho_1 = \sqrt{1 + \Delta t^2}, \quad \tan \theta = -\Delta t, \quad (7.56)$$

$$x_2^n = \rho_2^n \cos(n\theta), \quad \rho_2 = 1/\sqrt{1 + \Delta t^2}, \quad \tan \theta = \Delta t, \quad (7.57)$$

$$x_3^n = \cos(n\theta), \quad \tan(\theta/2) = \Delta t/2, \quad (7.58)$$

$$x_4^n = \cos(n\theta)(c_1 + c_2(-1)^n), \quad \tan(\theta) = \sqrt{\Delta t^2 - 1}/\Delta t, \quad \Delta t < 1. \quad (7.59)$$

Notar que  $x_4^n$  está expresado para los valores  $\Delta t < 1$ . Las discusiones basadas en la región de estabilidad de los esquemas numéricos permitieron ver que: (i) la solución  $x_1^n$  tiende a infinito con  $n \rightarrow \infty$ , (ii) la solución  $x_2^n$  tiende a cero con  $n \rightarrow \infty$  y (iii) la solución  $x_3^n$  se mantiene acotada cuando  $\Delta t < 1$  y (iv) la solución  $x_4^n$  se mantiene acotada para todo  $\Delta t$  y  $n \rightarrow \infty$ . En concreto, se puede comprobar que  $x_3^n$  y  $x_4^n$  no tienen error de amplitud pero tienen un error de fase que es  $O(\Delta t)$ . Es importante hacer notar que debido a que el esquema Leap-Frog es un esquema de dos pasos, existen dos soluciones fundamentales correspondientes a las constantes  $c_1$  y  $c_2$ . La condición inicial  $x(0) = 1$  junto con la condición inicial extra que se obtiene del arranque del esquema Leap-Frog mediante un esquema de un paso, permiten obtener  $c_1$  y  $c_2$ . Generalmente  $c_1 = 1 + O(\Delta t^2)$  y  $c_2 = O(\Delta t^2)$ .

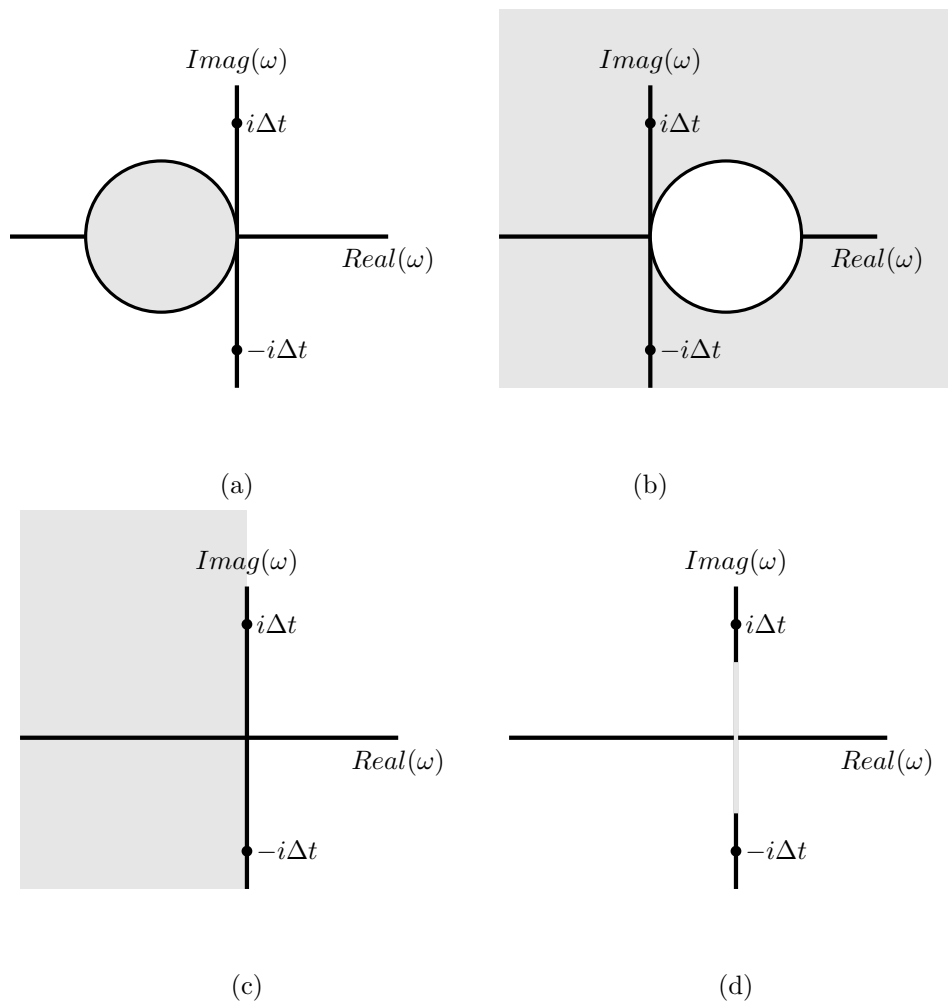


Figura 7.2: Regiones de estabilidad de esquemas numéricos junto con los autovalores del oscilador armónico. (a) Esquema Euler, (b) Esquema Euler inverso, (c) Esquema Crank-Nicolson y (d) Esquema Leap-frog.

# Capítulo 8

## Problema de condiciones iniciales y de contorno

### 8.1. Introducción

Un problema de condiciones iniciales y de contorno está caracterizado por un conjunto de ecuaciones de evolución que se deben verificar en un dominio espacial junto con condiciones de contorno para ese dominio. Generalmente, estas ecuaciones de evolución provienen de principios de conservación de energía, cantidad de movimiento y masa. Como ejemplos representativos de este tipo de problemas se encuentra la ecuación del calor que tiene carácter parabólico y la ecuación de ondas que tiene carácter hiperbólico. La ecuación del calor representa la conservación de energía térmica en un dominio genérico. Es decir, la diferencia de energía térmica que entra menos la energía térmica que sale de un determinado dominio se invierte en calentar o incrementar la temperatura de ese dominio. Las condiciones de contorno en este caso hacen referencia al flujo de energía térmica que se impone en el contorno del dominio. La condición de contorno más sencilla es imponer la temperatura del contorno. Sin embargo, en los problemas reales una condición de contorno muy usual es considerar que el flujo de energía que sale de un contorno es proporcional a la diferencia de energía entre el contorno y el espacio exterior. En este caso, como el flujo de calor se puede calcular mediante la ley de Fourier e involucra la gradientes normales al contorno, la condición de contorno es una relación entre las derivadas normales en el contorno y los valores de la temperatura exterior y del contorno.

Sea el siguiente problema de condiciones iniciales y de contorno:

$$\begin{aligned} \frac{\partial u}{\partial t} &= \mathcal{L}(u; t), & u : \Omega \times [t_0, +\infty) &\rightarrow \mathbb{R}, & \Omega &\subset \mathbb{R}^3, \\ BC(u)|_{\partial\Omega} &= 0, \\ u(x, t_0) &= f(x), \end{aligned} \tag{8.1}$$

donde  $\mathcal{L}(u; t)$  representa el operador diferencial espacial que involucra derivadas a lo largo de direcciones de  $\Omega$  y el primer término representa la variación frente al tiempo de la magnitud o magnitudes consideradas por el hecho de que existe un balance neto distinto de cero entre la entrada y la salida. Estas ecuaciones de evolución junto con las condiciones de contorno o relaciones  $BC(u)$  impuestas en el contorno para todo instante y la condición inicial  $u(x, t_0) = f(x)$  permiten determinar o predecir el comportamiento de  $u$  a lo largo del tiempo.

La característica fundamental que diferencia los problemas de evolución de los problemas de contorno es que mientras que en los problemas de contorno la información viaja a velocidad infinita, en los problemas de evolución la información viaja a la velocidad que determine el mecanismo físico en cuestión y la solución de puede integrar desde la condición inicial mediante sucesivos pasos temporales suficientemente pequeños. Esta característica permite abordar los problemas de evolución en ecuaciones en derivadas parciales mediante el método de las líneas. El método de la líneas consiste en dos pasos sucesivos: (i) discretización espacial del problema y (ii) discretización temporal del problema. Mediante el primer paso aproximamos una función como puede ser la temperatura de un dominio mediante un conjunto de puntos nodales finito. Es decir, pasamos de un espacio de dimensión infinita a un espacio de dimensión finita. Una vez realizada esta discretización el conjunto de los valores nodales de la función o de la temperatura constituye un vector de estado que está gobernado por un conjunto de ecuaciones de evolución junto con una condición inicial. Este segundo problema se trata con cualquiera de los esquemas numéricos analizados para la integración de problemas de Cauchy en ecuaciones diferenciales ordinarias. En la siguiente sección se describe el método de las líneas como algoritmo de cálculo para la resolución de estos problemas.

## 8.2. Discretización espacial y temporal

El operador  $\mathcal{L}(u; t)$  es un operador diferencial que actúa sobre la variable  $u(x, t)$  e incluye todas las variaciones espaciales de  $u(x, t)$ . Para abordar la integración de este tipo de problemas se utilizará el denominado método de las líneas que consiste en discretizar espacialmente el problema (8.1) para obtener un conjunto de ecuaciones diferenciales ordinarias. Posteriormente, mediante la discretización

temporal o mediante el uso de los esquemas estudiados para la solución numérica del problema de Cauchy se integra en el tiempo un problema de ecuaciones diferenciales ordinarias.

La discretización espacial se basa en pasar de un espacio de dimensión infinita donde está definida la función solución  $u(x, t)$  a un espacio de dimensión finita que aproxima la función solución mediante los valores  $u_i(t)$  en un conjunto de puntos nodales  $x_i$  o grados de libertad. Mediante estos puntos nodales se construye un interpolante global o continuo a trozos y se aproximan las derivadas espaciales que incluye el operador  $\mathcal{L}(u; t)$ . De esta manera, la evolución del grado de libertad  $u_i(t)$  viene gobernando por una ecuación de la forma:

$$\frac{du_i}{dt} = L_i(u_0, u_1, \dots, u_N), \quad i = 0, \dots, N, \quad (8.2)$$

donde  $L$  es una función vectorial que, en general, depende de todos los puntos nodales  $u_i(t)$ . Si expresamos este sistema en forma vectorial,

$$\frac{dU}{dt} = L(U; t), \quad (8.3)$$

donde  $U$  representa el vector de estado del sistema cuyas componentes son los valores nodales  $u_i(t)$ . Es importante hacer notar las condiciones de contorno al ser discretizadas aparecen en las componentes de  $L(U; t)$ .

La discretización temporal se realiza de la misma manera. Las funciones  $u_i(t)$  se aproximan en un espacio de dimensión finita por un conjunto de pasos temporales  $u_i^n$ . Para los métodos Adams se construyen interpolantes y para los métodos Runge-Kutta se evalúan etapas intermedias. De cualquier forma, los esquemas temporales conducen a relaciones algebraicas que permiten determinar  $u_i^{n+1}$  en función de valores o pasos anteriores. Si los esquemas son explícitos, estas relaciones se expresan:

$$u_i^{n+1} = G_i(\{u_j^k\}), \quad j = 0, \dots, N, \quad k = n, \dots, n+1-p, \quad i = 0, \dots, N. \quad (8.4)$$

Si expresamos este sistema en forma vectorial,

$$U^{n+1} = G(U^n, U^{n-1}, \dots, U^{n-1+p}), \quad (8.5)$$

donde el vector de estado  $U^n = (u_0^n, u_1^n, \dots, u_N^n)^T$  representa el valor aproximado de  $U(t)$  en el instante  $t_n$  y  $G$  es una función vectorial que depende de los valores aproximados del vector de estado en instantes anteriores. De esta forma y mediante (8.5) el vector de estado aproximado  $U^{n+1}$  se determina en función de los valores aproximados de los vectores de estado en instantes anteriores.

En la figura (8.1) se representa el método de las líneas.

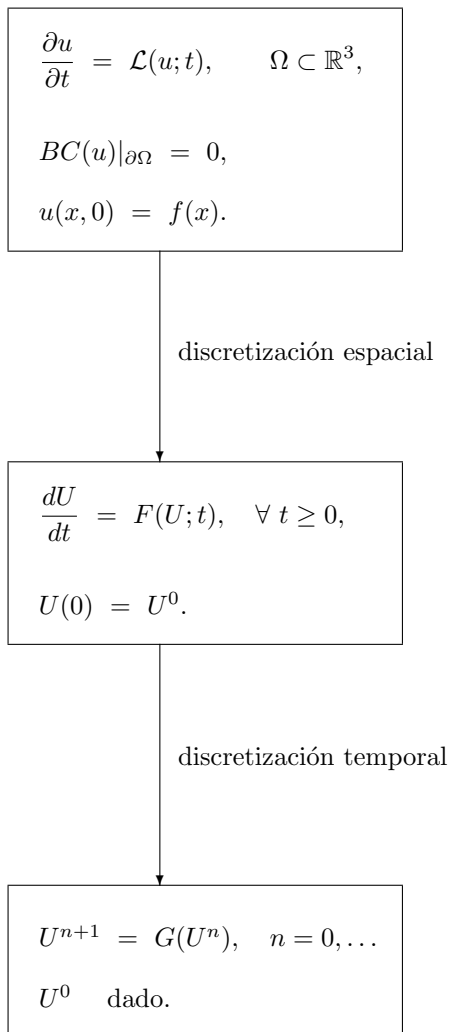


Figura 8.1: Método de las líneas para la discretización espacial y temporal de un problema de evolución



## EJEMPLO. Ecuación del calor en un dominio unidimensional

A continuación, se explica el método de las líneas para la ecuación del calor en un dominio unidimensional  $\forall x \in [0, 1]$  con condiciones de contorno adiabáticas o de flujo cero y considerando un perfil inicial de temperatura distinto de cero dado por  $f(x)$ .

$$\begin{aligned}\frac{\partial u}{\partial t} &= \frac{\partial^2 u}{\partial x^2}, \\ \frac{\partial u}{\partial x} \Big|_{(0,t)} &= 0, \\ \frac{\partial u}{\partial x} \Big|_{(1,t)} &= 0, \\ u(x, 0) &= f(x).\end{aligned}\tag{8.6}$$

El algoritmo que se basa en el método de las líneas es el siguiente:

1. Se procede a hacer una partición equiespaciada del intervalo  $[0, 1]$ .
2. Se impone la ecuación en derivadas parciales en el punto genérico  $x_j$ .

$$\left( \frac{\partial u}{\partial t} \right)_{x_j} = \left( \frac{\partial^2 u}{\partial x^2} \right)_{x_j}.\tag{8.7}$$

3. Se utiliza un interpolante continuo a trozos para  $u(x, t)$  basado en tres puntos y se calculan las fórmulas en diferencias finitas para las derivadas.
4. Se plantea el sistema de ecuaciones diferenciales ordinarias resultante una vez realizada la semidiscretización espacial para los puntos interiores.

$$\frac{du_j}{dt} = \frac{1}{\Delta x^2} (u_{j+1} - 2u_j + u_{j-1}), \quad j = 1, \dots, N-1.\tag{8.8}$$

5. Se discretizan las condiciones de contorno.

$$\frac{\partial u}{\partial x} \Big|_{x=0} = \frac{1}{2\Delta x} (-3u_0 + 4u_1 - u_2) = 0,\tag{8.9}$$

$$\frac{\partial u}{\partial x} \Big|_{x=1} = \frac{1}{2\Delta x} (3u_N - 4u_{N-1} + u_{N-2}) = 0,\tag{8.10}$$

y se obtienen los puntos del contorno en función de los puntos interiores:

$$u_0 = \frac{1}{3} (4u_1 - u_2),\tag{8.11}$$

$$u_N = \frac{1}{3} (4u_{N-1} - u_{N-2}).\tag{8.12}$$


---

6. Se realiza la discretización temporal del sistema (8.8) mediante un esquema temporal. En este caso se utilizará un esquema Euler.

$$\begin{aligned}
 u_j^{n+1} &= u_j^n + \frac{\Delta t}{\Delta x^2} (u_{j+1}^n - 2u_j^n + u_{j-1}^n), \quad j = 1, \dots, N-1, \quad n = 0, \dots \\
 u_0^n &= \frac{1}{3} (4u_1^n - u_2^n), \\
 u_N^n &= \frac{1}{3} (4u_{N-1}^n - u_{N-2}^n), \\
 u_j^0 &= f(x_j), \quad j = 0, \dots, N.
 \end{aligned} \tag{8.13}$$

A partir de la condición inicial  $u_j^0 = f(x_j)$  se determinan los puntos interiores del siguiente paso temporal  $u_j^1$ ,

$$u_j^1 = u_j^0 + \frac{\Delta t}{\Delta x^2} (u_{j+1}^0 - 2u_j^0 + u_{j-1}^0), \quad j = 1, \dots, N-1. \tag{8.14}$$

Las condiciones de contorno  $u_0^n$  y  $u_N^n$  aparecen cuando se terminan los puntos interiores próximos al contorno  $j = 1$  y  $j = N-1$ .

De esta forma, el problema se integra a lo largo del tiempo mediante dos bucles anidados. El bucle exterior se encarga de actualizar el instante temporal desde la condición inicial. Para cada instante temporal existe un bucle interno que permite determinar las variaciones de la temperatura con respecto al tiempo para cada punto nodal  $x_j$ . En cada instante temporal las condiciones de contorno  $u_0^n$  y  $u_N^n$  se imponen antes de terminar las variaciones de temperatura de los nodos interiores. Es decir, constituyen ligaduras o relaciones algebraicas que se deben satisfacer para todo instante  $t_n$  y se resuelven conocidos los valores de temperatura de los puntos interiores.

EJEMPLO. *Ecuación de ondas en un dominio unidimensional*

Consideramos las oscilaciones transversales de una cuerda unida entre dos puntos dados  $x = 0$  y  $x = 1$ . Se considera como condiciones iniciales que la cuerda está elongada y parte del reposo. Como condiciones de contorno se supone que la oscilación es cero en los puntos en los que está fijada la cuerda. Así, la ecuación de ondas junto con sus condiciones iniciales y de contorno que permite modelar esta física son:

$$\begin{aligned} \frac{\partial^2 u}{\partial t^2} &= \frac{\partial^2 u}{\partial x^2}, \\ u(0, t) &= 0, \quad u(1, t) = 0, \\ u(x, 0) &= f(x), \quad \frac{\partial u}{\partial t}(x, 0) = 0. \end{aligned} \tag{8.15}$$

En este caso, la integración numérica del problema pasa por reducir inicialmente la ecuación de ondas a una ecuación de primer orden en el tiempo mediante un cambio de variables. Se realiza el cambio de variable  $w = \partial u / \partial t$  y el sistema resultante de la ecuación de ondas queda:

$$\frac{\partial}{\partial t} \begin{pmatrix} u \\ w \end{pmatrix} = \begin{pmatrix} w \\ \frac{\partial^2 u}{\partial x^2} \end{pmatrix}. \tag{8.16}$$

La ecuación de ondas queda así reducida a un sistema de primer orden en el tiempo que puede ser integrado mediante el método de las líneas anteriormente descrito. Si se discretiza la derivada segunda mediante un esquema centrado de tres puntos, se obtiene el siguiente sistema de ecuaciones diferenciales ordinarias:

$$\begin{pmatrix} \frac{du_j}{dt} \\ \frac{dw_j}{dt} \end{pmatrix} = \begin{pmatrix} w_j \\ \frac{1}{\Delta x^2} (u_{j+1} - 2u_j + u_{j-1}) \end{pmatrix}, \tag{8.17}$$

donde se define el vector de estado  $U = (u_0, w_0, u_1, w_1, \dots, u_N, w_N)^T$ . Finalmente, el sistema de ecuaciones diferenciales ordinarias así planteado se integra con cualquier esquema temporal.

---

### 8.3. Error espacial y error temporal

Cuando un problema se discretiza temporal y espacialmente aparecen dos tipos de problemas de errores: (i) el asociado a la aproximación de las derivadas en el espacio y (ii) el asociado a la aproximación de las derivadas en el tiempo o la integración temporal.

La solución exacta del problema evaluada en el nodo  $x_i$  se representa por  $u(x_i, t)$ . Cuando se realiza la semidiscretización espacial,  $u_i(t)$  representa el valor aproximado en el nodo  $x_i$ . Finalmente, cuando se realiza la discretización temporal,  $u_i^n$  representa el valor numérico aproximado en el instante  $t_n$  y el nodo  $x_i$ .

**Definición.** *Error espacial*

Se define como la diferencia entre el valor exacto y el valor aproximado en cada nodo  $x_i$ ,

$$E_i = u(x_i, t) - u_i(t). \quad (8.18)$$

**Definición.** *Error temporal*

Se define en cada nodo  $x_i$  como la diferencia entre el valor  $u_i(t_n)$  y el valor aproximado  $u_i^n$  en el instante  $t_n$ ,

$$E_i^n = u_i(t_n) - u_i^n. \quad (8.19)$$

**Definición.** *Error total*

Se define el error total asociado a las dos discretizaciones como la diferencia entre el valor exacto  $u(x_i, t_n)$  evaluado en el nodo  $x_i$  y el instante  $t_n$  menos el valor aproximado  $u_i^n$ ,

$$E_i^T = u(x_i, t_n) - u_i^n. \quad (8.20)$$

Con estas definiciones, se puede comprobar que el error total es la suma del error espacial más el error temporal

$$E_{T_i} = \underbrace{u(x_i, t_n) - u_i(t_n)}_{E_i(t_n)} + \underbrace{u_i(t_n) - u_i^n}_{E_i^n}, \quad (8.21)$$

o en notación vectorial:

$$E_T = E + E^n. \quad (8.22)$$

A continuación, se pasa a analizar el comportamiento del error espacial  $E$  y del error temporal  $E^n$ .

## 8.4. Error de la semidiscretización espacial

Para obtener la ecuación que gobierna el error de la semidiscretización espacial, se resta la ecuación (8.1) en el nodo genérico  $x_i$  menos la ecuación (8.2),

$$\frac{d}{dt} [u(x_i, t) - u_i] = \mathcal{L}(u; t) |_{x_i} - L_i(u_0, u_1, \dots, u_N). \quad (8.23)$$

Se suma y resta el operador discreto aplicado a la solución exacta.

$$\begin{aligned} \frac{dE_i}{dt} &= \mathcal{L}(u; t) |_{x_i} - L_i(u_0, u_1, \dots, u_N) \\ &\quad + L_i(u(x_0; t), u(x_1; t), \dots, u(x_N; t)) \\ &\quad - L_i(u(x_0; t), u(x_1; t), \dots, u(x_N; t)). \end{aligned} \quad (8.24)$$

Se define el error local de truncamiento de la semidiscretización espacial en cada nodo  $x_i$  mediante la diferencia entre el operador exacto y el operador discreto u operador en diferencias aplicado a la solución exacta:

$$R_i = \mathcal{L}(u; t) |_{x_i} - L_i(u(x_0; t), u(x_1; t), \dots, u(x_N; t)). \quad (8.25)$$

De esta forma, la ecuación para el error  $E_i$  queda:

$$\frac{dE_i}{dt} = R_i + L_i(u(x_0; t), u(x_1; t), \dots, u(x_N; t)) - L_i(u_0, u_1, \dots, u_N). \quad (8.26)$$

Utilizando el teorema del valor medio para la función vectorial  $L_i$

$$\frac{dE_i}{dt} = R_i + \frac{\partial L_i}{\partial u_j}(\xi_0, \xi_1, \dots, \xi_N) (u(x_j; t) - u_j), \quad (8.27)$$

donde  $\partial L_i / \partial u_j$  representa la matriz jacobiana de la función vectorial  $L_i$  evaluada en un punto intermedio  $(\xi_0, \dots, \xi_N)$ . Si denominamos por  $A$  la matriz jacobiana particularizada, el sistema anterior en forma vectorial se expresa:

$$\frac{dE}{dt} = A E + R. \quad (8.28)$$

Si la matriz  $A$  fuera de coeficientes constante, esta ecuación junto con la condición inicial  $E(0)$  o error en las condiciones iniciales se puede integrar para dar:

$$E(t) = e^{At} E(0) + \int_0^t e^{A(t-\tau)} R(\tau) d\tau. \quad (8.29)$$

La acotación de la expresión anterior resulta:

$$\|E\| \leq \|e^{At}\| \|E(0)\| + \sup_{t \in [0, t_f]} \|R\| \left\| \int_0^t e^{A(t-\tau)} d\tau \right\|. \quad (8.30)$$

Si la matriz  $A$  es normal  $A A^T = A^T A$ , entonces existe un conjunto de autovectores ortogonales que diagonaliza la matriz  $A$  y

$$\|e^{At}\| = e^{\alpha(A)t}, \quad (8.31)$$

donde  $\alpha(A)$  es la abscisa espectral de  $A$  que se define como:

$$\alpha(A) = \sup_{z \in \Lambda(A)} \operatorname{Re}(z), \quad (8.32)$$

con  $\Lambda(A)$  como el espectro o conjunto de todos los autovalores de  $A$ . Es decir, la abscisa espectral es el valor máximo de la parte real de los autovalores de la matriz  $A$ . Mediante la definición de la abscisa espectral (8.32), la integral del segundo término de la cota de error (8.30) queda:

$$\int_0^t \|e^{A(t-\tau)}\| d\tau = \int_0^t e^{\alpha(t-\tau)} d\tau = \left. \frac{e^{\alpha(t-\tau)}}{-\alpha} \right|_0^t = \frac{1}{\alpha} (e^{\alpha t} - 1). \quad (8.33)$$

Finalmente, la acotación del error (8.30) de la semidiscretización espacial queda:

$$\|E\| \leq e^{\alpha t} \|E(0)\| + \frac{1}{\alpha} (e^{\alpha t} - 1) \sup_{t \in [0, t_f]} \|R\|. \quad (8.34)$$

Si  $\alpha < 0$ , el error espacial con  $t \rightarrow \infty$  está acotado por el error de truncamiento de la discretización espacial:

$$\|E\| \leq \frac{1}{\alpha} \sup_{t \in [0, t_f]} \|R\|. \quad (8.35)$$

Si  $\alpha > 0$ , el error en las condiciones iniciales o el error de truncamiento, por pequeño que sea, se amplifica exponencialmente con  $t \rightarrow \infty$ .

Por esta razón es importante asegurar que la discretización espacial no cambia el carácter de estabilidad del problema diferencial. Generalmente, cuando se discretiza la ecuación de ondas mediante esquemas centrados de diferencias finitas de alto orden en mallas equiespaciadas, se obtienen autovalores que tienen parte real mayor que cero. Estos esquemas o fórmulas de diferencias finitas no son válidas para realizar la discretización espacial. Esta problemática está asociada al mal comportamiento de los interpolantes de alto orden en los contornos cuando la malla es equiespaciada. Este problema se suele resolver concentrando puntos en los contornos como lo hacen los métodos de Chebyshev.

Si  $A$  es no normal, la ecuación (8.31) no es válida pero la forma canónica de Jordan  $J = P^{-1}AP$ , donde  $P$  es una matriz de  $N \times N$ , nos permite dar la siguiente cota para la norma de la matriz exponencial:

$$\|e^{At}\| \leq \|P^{-1}\| \|P\| \|e^{Jt}\|. \quad (8.36)$$

Una vez más si  $\alpha(A) < 0$ , entonces  $\|e^{Jt}\|$  tiende a cero con  $t \rightarrow \infty$  y el error de la solución numérica se comporta como el error de truncación para  $t \rightarrow \infty$ . Si la abscisa espectral es mayor que cero, el error crece sin límite para matrices normales o no normales.

## 8.5. Error en la ecuación del calor

El comportamiento del error espacial está gobernado por la ecuación (8.28) que incluye la matriz del sistema  $A$  y el error espacial de truncamiento  $R$ . De igual forma, el comportamiento del error temporal está gobernado por la ecuación (7.21) que incluye la matriz del sistema  $B$  y el error temporal de truncamiento  $T^{n+1}$ .

En esta sección y a modo de ejemplo, se vuelve a reproducir el procedimiento de obtención de la ecuación del error temporal y espacial para la integración de la ecuación del calor en un dominio unidimensional con condiciones de contorno homogéneas. La discretización espacial se lleva cabo mediante diferencias finitas centradas con tres puntos y la temporal mediante un esquema Euler. La semidiscretización espacial conduce al siguiente sistema de ecuaciones diferenciales ordinarias:

$$\frac{du_j}{dt} = \frac{1}{\Delta x^2} (u_{j+1} - 2u_j - u_{j-1}), \quad j = 1, \dots, N-1, \quad (8.37)$$

con  $u_0 = 0$  y  $u_N = 0$ . Restando la ecuación del calor (8.7) menos la ecuación anterior (8.37) en cada punto  $x_j$

$$\left( \frac{\partial u}{\partial t} \right)_{x_j} - \frac{du_j}{dt} = \left( \frac{\partial^2 u}{\partial x^2} \right)_{x_j} - \frac{1}{\Delta x^2} (u_{j+1} - 2u_j - u_{j-1}). \quad (8.38)$$

A continuación, se suma y se resta el operador discreto aplicado a la solución exacta

$$\begin{aligned} \frac{dE_j}{dt} &= \left( \frac{\partial^2 u}{\partial x^2} \right)_{x_j} - \frac{1}{\Delta x^2} (u_{j+1} - 2u_j - u_{j-1}) \\ &\quad + \frac{1}{\Delta x^2} (u(x_{j+1}) - 2u(x_j) - u(x_{j-1})) \\ &\quad - \frac{1}{\Delta x^2} (u(x_{j+1}) - 2u(x_j) - u(x_{j-1})). \end{aligned} \quad (8.39)$$

Mediante la definición del error de truncamiento de la ecuación del calor

$$R_j = \left( \frac{\partial^2 u}{\partial x^2} \right)_{x_j} - \frac{1}{\Delta x^2} (u(x_{j+1}) - 2u(x_j) - u(x_{j-1})), \quad (8.40)$$

la ecuación (8.39) queda:

$$\frac{dE_j}{dt} = R_j + \frac{1}{\Delta x^2} (E_{j+1} - 2E_j - E_{j-1}). \quad (8.41)$$

que es el ejemplo de la ecuación del calor para el sistema general (8.28).

Para obtener la ecuación del error temporal se procede de la siguiente forma. Se considera el sistema (8.37) expresado en forma vectorial con el vector columna  $U = (u_1, \dots, u_{N-1})^T$

$$\frac{dU}{dt} = AU, \quad (8.42)$$

donde  $A$  es la matriz del sistema que es tridiagonal y está definida por (8.37). A continuación, se discretiza en el tiempo el sistema anterior mediante un esquema Euler y un paso temporal constante  $\Delta t$

$$U^{n+1} = U^n + \Delta t A U^n. \quad (8.43)$$

La definición del error de truncamiento del esquema temporal (7.16)

$$T^{n+1} = U(t_{n+1}) - U(t_n) - \Delta t A U(t_n), \quad (8.44)$$

permite obtener la ecuación del error global de la semidiscretización espacial. Se resta la ecuación (8.44) menos la ecuación (8.43) y como  $E^n = U(t_n) - U^n$  se obtiene:

$$E^{n+1} = T^{n+1} + E^n + \Delta t A E^n. \quad (8.45)$$

Si se define la matriz  $B = I + \Delta t A$ , el sistema anterior queda:

$$E^{n+1} = B E^n + T^{n+1}, \quad (8.46)$$

que es el ejemplo del esquema Euler para la ecuación del error global (7.21).

### 8.5.1. Acotación del error espacial

Conocida la matriz del sistema  $A$  y conocido el error de truncamiento  $R$ , la acotación del error de la semidiscretización espacial pasa por determinar la abscisa espectral de  $A$ . En este caso, la matriz  $A$  es normal y la determinación de los autovalores de  $A$  se puede realizar de forma analítica mediante el siguiente sistema de ecuaciones

$$\frac{1}{\Delta x^2} (u_{j+1} - 2u_j - u_{j-1}) = \lambda u_j, \quad j = 1, \dots, N-1, \quad (8.47)$$

con condiciones de contorno  $u_0 = 0$  y  $u_N = 0$ . Es fácil comprobar que existen soluciones no homogéneas de la forma

$$u_j = \sin(k\pi x_j), \quad (8.48)$$

con  $k$  un número entero. Si se introducen estas soluciones o autovectores en (8.47), se obtienen los autovalores correspondientes:

$$\lambda_k = -\frac{2}{\Delta x^2} (1 - \cos(k\pi\Delta x)) = -\frac{4}{\Delta x^2} \sin^2\left(\frac{k\pi\Delta x}{2}\right). \quad (8.49)$$



Como la matriz  $A$  es normal y cuadrada de dimensión  $N - 1$ , deben existir  $N - 1$  autovectores y autovalores que se corresponden con los valores de  $k = 1, \dots, N - 1$ .

En este caso, la abscisa espectral vale:

$$\alpha = \sup_{\lambda \in \Lambda(A)} \operatorname{Re}(\lambda) = -\frac{4}{\Delta x^2} \operatorname{sen}^2 \left( \frac{\pi \Delta x}{2} \right) \approx -\frac{4}{\Delta x^2} \frac{\pi^2 \Delta x^2}{4} = -\pi^2, \quad (8.50)$$

y la ecuación para la acotación del error de semidiscretización espacial (8.34) se puede escribir como:

$$\|E(t)\| \leq e^{-\pi^2 t} \|E(0)\| + \frac{1}{\pi^2} \left( 1 - e^{-\pi^2 t} \right) \sup_{\forall \tau} \|R\|. \quad (8.51)$$

A continuación, se calcula el error de truncamiento  $R$  mediante (8.40) que coincide con el error de la derivada segunda de un interpolante con tres puntos equiespaciados

$$R_j = -2\Delta x^2 \frac{u^{(4)}(\xi)}{3!} \left( \frac{d\xi}{dx} \right), \quad (8.52)$$

donde  $\xi$  es un punto intermedio que depende de  $x_j$ . De esta forma,

$$\|E(t)\| \leq e^{-\pi^2 t} \|E(0)\| + \frac{1}{\pi^2} \left( 1 - e^{-\pi^2 t} \right) \Delta x^2 M, \quad (8.53)$$

donde  $M$  se obtiene mediante una cota de la derivada cuarta de la función. Se observa que el error en las condiciones iniciales decrece con el tiempo y el error espacial crece con el tiempo y permanece acotado por algo  $O(\Delta x^2)$ . Se dice que un esquema espacial es de orden  $q$  cuando su error de truncamiento es  $O(\Delta x^q)$ .

### 8.5.2. Acotación del error temporal

Conocida la matriz del sistema  $B$  y conocido el error de truncamiento  $T$ , la acotación del error de la semidiscretización temporal pasa por determinar el radio espectral de  $B$ . En este caso, la matriz  $B = I + \Delta t A$  y los autovalores de  $B$  son:

$$r_k = 1 + \Delta t \lambda_k = 1 - \frac{4\Delta t}{\Delta x^2} \operatorname{sen}^2 \left( \frac{k\pi \Delta x}{2} \right). \quad (8.54)$$

En este caso, el radio espectral vale:

$$\rho = \sup_{r \in \Lambda(B)} |r| = \left| 1 - \frac{4\Delta t}{\Delta x^2} \right|. \quad (8.55)$$

Para que el error de semidiscretización temporal (7.27) esté acotado, el radio espectral debe ser menor que uno,

$$\frac{\Delta t}{\Delta x^2} < \frac{1}{2}. \quad (8.56)$$

Por otra parte, el error local de truncamiento del esquema Euler (8.44) desarrollando en serie de Taylor alrededor de  $t_n$  se puede poner como:

$$T^{n+1} = \frac{\Delta t^2}{2} A^2 U^n + O(\Delta t^3). \quad (8.57)$$

De esta forma, la ecuación (7.27) queda:

$$\|E^n\| \leq \rho^n \|E^0\| + \left| \frac{1 - \rho^{n+1}}{1 - \rho} \right| C \Delta t^2, \quad (8.58)$$

donde  $C$  se obtiene mediante una cota de la derivada segunda de la función  $U(t)$ . Si  $\rho$  es menor que uno y  $n \rightarrow \infty$ , el error en las condiciones iniciales tiende a cero y el error de truncamiento está acotado por

$$\|E^n\| \leq C \Delta x^2 \Delta t. \quad (8.59)$$

Así, el error de la discretización temporal del esquema Euler es  $O(\Delta t)$ . Es importante hacer notar cómo el denominador de la expresión (8.58) reduce un orden el error de truncamiento. Se dice que un esquema temporal para la discretización temporal es de orden  $q$  cuando su error local de truncamiento es de  $O(\Delta t^{q+1})$ .

---

## 8.6. Extrapolación de Richardson

La extrapolación de Richardson (1927) es una técnica que nos permite estimar el error tanto espacial como temporal de una derivación o una integración numérica sin conocer la solución exacta del problema.

Supongamos que  $\Phi(h)$  es el resultado numérico conseguido con una discretización temporal o espacial caracterizada por el paso de malla  $h$ . Para poder aplicar la técnica de la extrapolación de Richardson, debe existir un desarrollo en serie de Taylor de la solución calculada en potencias  $h$ . En el caso de las discretizaciones espaciales realizadas mediante aproximaciones polinómicas este desarrollo existe sin ninguna limitación y  $h$  representa el inverso de los  $N$  puntos nodales o grados de libertad del interpolante. De igual forma, todos los esquemas temporales salvo los esquemas Adams multipaso admiten desarrollos en serie de Taylor con el paso temporal  $h$ .

Si  $\Phi^0$  es el valor exacto de la magnitud que queremos calcular y  $\Phi(h)$  admite un desarrollo en serie de Taylor en potencias de  $h$ , entonces

$$\Phi(h) = \Phi^0 + k_1 h^q + O(h^{q+1}). \quad (8.60)$$

Consideremos la solución  $\Phi^1$  obtenida con un esquema numérico y la malla espacial o temporal  $h_1$

$$\Phi^1 = \Phi^0 + k_1 h_1^q + O(h_1^{q+1}), \quad (8.61)$$

y la solución  $\Phi^2$  con la malla  $h_2$

$$\Phi^2 = \Phi^0 + k_1 h_2^q + O(h_2^{q+1}). \quad (8.62)$$

Restando las ecuaciones (8.61) y (8.62), se puede obtener el valor de  $k_1$  de la ecuación (8.60),

$$k_1 = \frac{\Phi^1 - \Phi^2}{h_1^q - h_2^q} + O(h_1). \quad (8.63)$$

Llevando esta aproximación de  $k_1$  a la ecuación (8.60), se puede obtener una aproximación del error  $E^1 = \Phi^1 - \Phi^0$  cometido en la determinación de  $\Phi_0$  mediante el cálculo numérico basado en la malla  $h_1$  y la malla  $h_2$ ,

$$E^1 = \frac{(\Phi^1 - \Phi^2)h_1^q}{h_1^q - h_2^q} + O(h_1^{q+1}). \quad (8.64)$$

Es importante hacer notar que  $\Phi$  puede ser un valor real si es el resultado de una integral o un valor final de una simulación. En otros casos,  $\Phi$  puede ser un vector columna cuyas componentes representan la aproximación de un problema en puntos nodales o en los instantes temporales de una discretización temporal. Por este motivo, y para poder restar los vectores columnas  $\Phi^1$  y  $\Phi^2$  se suele calcular la solución  $\Phi^2$  con una malla  $h_2 = h_1/2$ . De esta forma, se hacen coincidir en posición

espacial o temporal las componentes de  $\Phi^1$  con las componentes pares de  $\Phi^2$  para poder restarlas. Es decir, la solución  $\Phi^1$  se calcula con  $N + 1$  valores nodales y la solución  $\Phi^2$  con  $2N + 2$  valores nodales. De esta forma, el error en cada punto nodal de la solución  $\Phi^1$  queda:

$$E_j^1 = \frac{(\Phi_j^1 - \Phi_{2j}^2)}{1 - 1/2^q}, \quad j = 0, \dots, N. \quad (8.65)$$

Se exponen a continuación tres ejemplos basados en la extrapolación de Richardson para determinar los errores temporales y errores espaciales en problemas valores iniciales y de contorno.

1. Error temporal. Sea  $U^n$  con  $n = 0, \dots, N$  una solución numérica de un problema de Cauchy en ecuaciones diferenciales ordinarias (7.1) calculada mediante un esquema temporal de orden  $q$  y una partición temporal equiespaciada de paso  $\Delta t$ . Sea  $V^m$  con  $m = 0, \dots, 2N$  otra solución con paso  $\Delta t/2$ . De esta forma, el error  $E^n$  de la solución numérica  $U^n$  en cada instante temporal  $t_n$  se puede calcular mediante:

$$E^n = \frac{(U^n - V^{2n})}{1 - 1/2^q}, \quad n = 0, \dots, N. \quad (8.66)$$

2. Error de truncamiento de discretización de operadores espaciales. Sea  $R_j$  el error local de truncamiento de la semidiscretización espacial en cada nodo  $x_j$  definido mediante la ecuación (8.25) y sea  $L_j^1$  el operador espacial semidiscretizado espacialmente definido por (8.2) mediante  $M + 1$  puntos nodales. Si se considera otra semidiscretización espacial con  $2M + 2$  puntos nodales, entonces el error local de truncamiento en cada punto nodal  $x_j$  se puede poner como:

$$R_j = \frac{(L_j^1 - L_{2j}^2)}{1 - 1/2^q}, \quad j = 0, \dots, M. \quad (8.67)$$

3. Error espacial. Sea  $U_j^n$  la solución numérica de un problema de valores iniciales y de contorno (8.1) calculada en el nodo  $x_j$  con  $j = 0, \dots, M$  y el instante  $t_n$  con  $n = 0, \dots, N$ . Si se pretende calcular el error espacial de un esquema de orden  $q$ , la malla temporal debe ser lo suficientemente fina como para que el error esté dominado por el error espacial. De esta manera, si  $V_j^n$  es la solución calculada con  $2M + 2$  nodos espaciales en los mismos  $N + 1$  instantes temporales, entonces el error espacial  $E_j^n$  en el instante temporal  $t_n$  y la posición  $x_j$  vale:

$$E_j^n = \frac{(U_j^n - V_{2j}^n)}{1 - 1/2^q}, \quad j = 0, \dots, M, \quad n = 0, \dots, N. \quad (8.68)$$

# Apéndice A

## Nomenclatura

- $P_N$  : Polinomio de grado  $N$
- $f$  : Función continua definida en  $[a, b]$
- $\phi_k$  : Funciones base.
- $\hat{c}_k$  : Coeficientes de la serie truncada.
- $\tilde{c}_k$  : Coeficientes de la serie discreta.
- $\omega$  : Función de peso en el producto interno continuo.
- $E_N$  : Error de interpolación.
- $R_N$  : Error de truncamiento para una aproximación de grado  $N$ .
- $R_L$  : Error de redondeo de Lebesgue para una aproximación de grado  $N$ .
- $I_N$  : Interpolante polinómico de grado  $N$ .
- $\alpha_j$  : Coeficientes de peso en el producto interno discreto.
- $\ell_j$  : Polinomio de Lagrange asociado al punto  $x_j$ .
- $\pi_{N+1}$  : Función de pi de error de truncamiento de un conjunto de  $N + 1$  puntos.

---

$\Lambda_N$	: Constante de Lebesgue de un conjunto de $N + 1$ puntos.
$\lambda_N$	: Función de Lebesgue de un conjunto de $N + 1$ puntos.
$T_k$	: Polinomio de Chebyshev de grado $k$ .
$q$	: Orden del interpolante continuo a trozos.
$\Delta x_j$	: Distancia $x_{j+1} - x_j$ o paso espacial.
$\frac{dU}{dt} = F(U, t)$	: Sistema de EDOS. $U(t) \in \mathbb{R}^N, F : \mathbb{R}^N \times \mathbb{R} \rightarrow \mathbb{R}^N$ .
$U^n$	: Solución numérica en el instante $t_n$ . $U^n \in \mathbb{R}^N$ .
$F^n$	: $F^n = F(U^n, t_n)$ .
$u(x_j, t_n)$	: Solución exacta en el punto $x_j$ y en instante $t_n$ .
$u_j^n$	: Solución numérica en el punto $x_j$ y en instante $t_n$ .
$\Delta t_n$	: Distancia $t_{n+1} - t_n$ o paso temporal.
$R^n$	: Residuo del esquema numérico en el instante $t_n$ .
$T^n$	: Error local temporal de truncamiento en $t_n$ .
$E^n = u_j(t_n) - u_j^n$	: Error global temporal en $t_n$ .
$E_j = u(x_j) - u_j$	: Error global espacial en $x_j$ .
$E_j^n = u(x_j, t_n) - u_j^n$	: Error global espacial y temporal en $(x_j, t_n)$ .
$R_j$	: Error de truncamiento espacial en $x_j$ .
$\alpha(A)$	: Abscisa espectral de la matriz $A$ .
$\rho(A)$	: Radio espectral de la matriz $A$ .
$\mathcal{L}$	: Operador diferencial de un problema de contorno.
$BC$	: Operador diferencial de condiciones de contorno.
$\partial\Omega$	: Puntos frontera de un dominio espacial $\Omega$ .

---

# Bibliografía

- [1] BOYD, J. P. 1999 Chebyshev and Fourier Spectral Methods. Dover Publications, Inc.
- [2] CANUTO, C., HUSSAINI, M.Y., QUARTERONI, A. Y ZANG, T. A. 2006 Spectral Methods. Fundamentals in Single Domains. Springer.
- [3] HAIRER, W., NØRSETT Y WANNER, G. 1993 Solving Ordinary Differential Equations I. Nonstiff Problems. Springer–Verlag.
- [4] HAIRER, W. Y WANNER, G. 1996 Solving Ordinary Differential Equations II. Stiff and Differential –Algebraic Problems. Springer–Verlag.
- [5] ISAACSON, E. AND KELLER, H.B. 1966 Analysis of Numerical Methods.
- [6] KELLER, H.B. 1976 Numerical Solution of Two Point Boundary Value Problems. CBMS–NSF 24. SIAM.
- [7] LAMBERT, J.D. 1991 Numerical Methods for Ordinary Differential Systems. The Initial Value Problem. Wiley.
- [8] PRESS, W. H., TEUKOLSKY, S. A., VETTERLING, W.T., FLANNERY, B.P. 1992 Numerical Recipes in FORTRAN. The Art of Scientific Computing. Cambridge University Press.
- [9] SHAMPINE, L. F. Y GORDON, M. K. 1974 Computer Solution of Ordinary differential Equations. The Initial Value Problem. W. H. Freeman and Company.
- [10] STETTER, H. J. 1973 Analysis of Discretization Methods for Ordinary Differential Equations. Springer–Verlag.
- [11] STOER J. AND BULIRSH R. 1980 Introduction to Numerical Analysis. Springer-Verlag.
- [12] TREFETHEN, L. N. 2013 Approximation Theory and Approximation Practice. SIAM, Philadelphia.