

UNIVERSIDAD MICHOACANA DE SAN NICOLÁS DE HIDALGO

FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

“Mat. Luis Manuel Rivera Gutiérrez”

“MÉTODOS ESPECTRALES Y APLICACIONES EN FÍSICA”

TESIS

PARA OBTENER EL TÍTULO DE:
LICENCIADO EN CIENCIAS FÍSICO MATEMÁTICAS

PRESENTA:
FRANCISCO ALONSO ESPINOSA CHÁVEZ

ASESOR:
DR. RICARDO BECERRIL BARCENAS

“Instituto de Física y Matemáticas”

U.M.S.N.H

Morelia, Michoacán, Agosto 2010.

Agradecimientos

A mis padres, Francisco Alonso Espinosa Medina y Graciela Chávez Ramírez por todo el apoyo que me han dado en cada decisión que tomo. Al apoyo y paciencia de mis hermanos Arturo y Paty a quienes molestaba cuando no me salía algo.

A mi asesor, Dr. Ricardo Becerril Barcenas, por la paciencia que lo distinguió durante este trabajo.

A mis sinodales, por tomarse el tiempo de leer y darme recomendaciones para mejorar este trabajo y a todo aquel que se interese en leerlo. A todos mis profesores de la Facultad que fueron inspiración.

A mi maestra de matemáticas de segundo año de Secundaria por recomendar-me esta hermosa carrera. A mi maestro de cálculo “Toby” de la preparatoria 3 de la UMSNH por aconsejarme y al poema de Rodrigo Solís “Los sueños de Daniel” por no dejar cosas en el hubiera.

Le agradezco a mis compañeros y amigos de la facultad Heather, Toño, David, Oyuki, Beto, Jared y toda la comunidad FISMAT que de mencionar a todos llenaría una página.

Resumen

Con mucha frecuencia, las ecuaciones diferenciales que aparecen en física no pueden resolverse analíticamente y es necesario hacer uso de los métodos numéricos. Los métodos más conocidos entre estudiantes de licenciatura son las diferencias finitas. Pero en las últimas décadas se ha extendido el uso de los métodos espectrales (ME) en ramas de la física que van, desde dinámica de fluidos hasta relatividad general. Una de las características atractivas de los ME es el alto grado de precisión que pueden proporcionar, además, las soluciones no están dadas sólo en puntos de una red, como con las diferencias finitas, sino en todo el dominio espacial.

A pesar de lo extendido de su uso, en los programas de los cursos de métodos numéricos de las licenciaturas en física e ingeniería, los ME no aparecen y, por otra parte, los libros clásicos del tema no siempre son sencillos de entender en una primera lectura. El propósito de esta tesis es presentar los principios básicos de los ME y aplicarlos en diferentes problemas conocidos que aparecen en física así como mostrar la precisión que pueden alcanzar.

Índice general

Resumen	III
1. Introducción	1
1.1. Aproximación polinomial	1
1.2. Polinomios Ortogonales	2
1.2.1. Problema Sturm-Liouville	2
1.3. Métodos Espectrales	2
2. Base de Fourier y de Chebyshev	5
2.1. Base de Fourier	5
2.1.1. La expansión de Fourier continua	5
2.1.2. La expansión de Fourier discreta	7
2.1.3. Diferenciación	9
2.2. Base de Chebyshev	11
2.2.1. La expansión de Chebyshev continua	13
2.2.2. La expansión de Chebyshev discreta	14
2.2.3. Diferenciación	15
2.3. Error de aliación	19
2.3.1. No linealidades y el error de aliación	21
3. Esquemas Espectrales	27

3.1. Método de Galerkin	27
3.2. Método de Tau	28
3.3. Método de colocación	29
4. Aplicación de los ME en la Física	31
4.1. Ecuación de Helmholtz	31
4.1.1. Chebyshev colocación	32
4.1.2. Chebyshev Tau	33
4.2. Ecuación de Calor	37
4.2.1. Chebyshev Tau	37
4.2.2. Fourier Galerkin	39
4.3. Ecuación de Burgers	41
4.3.1. Chebyshev Tau	42
4.3.2. Fourier colocación	43
5. Conclusiones	47
A. Cuadratura Gauss-Lobatto y puntos de colocación	49
B. Método de descomposición LU	53
C. Matriz tridiagonal con trenza	55
D. Discretización del tiempo	59
D.1. Método de Crank-Nicolson	59
D.2. Método de Adams-Bashforth	59
D.3. Método de Runge-Kutta	60
E. Transformación Hopf-Cole	61

Capítulo 1

Introducción

1.1. Aproximación polinomial

Al hablar de métodos numéricos surge una simple pero importante pregunta, ¿por qué aproximar? la respuesta es obvia, porque es necesario. Supongamos que se desea obtener el valor de $e^{0.5}$, una forma de hacerlo es expresando la función exponencial en una expansión de Taylor, esto es:

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

pero la identidad anterior tiene un inconveniente, no podemos estar sumando infinitamente, por lo que, lo mejor que podemos hacer es tomar un número finito de términos (digamos N), y obtenemos

$$e^x = \sum_{n=0}^N \frac{x^n}{n!}.$$

A la ecuación anterior le llamamos aproximación, y al hacerla estamos introduciendo un error en el resultado generado por los términos con $n > N$ el cual se hará más pequeño entre más grande sea N .

Podemos aproximar una función de varias maneras pero vamos a encontrar que las más simples, aunque no triviales, son aquellas que consisten en polinomios, ya que son más fáciles de evaluar. Una de ellas es truncando expansiones de polinomios ortogonales

$$u_N(x) = \sum_{n=0}^N a_n \Phi_n(x),$$

donde $\Phi_n(x)$ pueden ser los polinomios de Fourier, Chebyshev, Hermite, Legendre, entre varios más. El método de emplear funciones ortogonales para encontrar soluciones numéricas se le llama con mayor frecuencia “métodos espectrales” y estas funciones parten del problema de Sturm-Liouville que se describe a continuación.

1.2. Polinomios Ortogonales

1.2.1. Problema Sturm-Liouville

Todos los polinomios ortogonales ya sean Fourier, Chebyshev, Hermite, Legendre, etc. se obtienen de resolver un solo problema. La razón de esto se debe a que los diferentes polinomios ortogonales son un caso específico de un polinomio ortogonal más general, el cual se obtiene del problema de Sturm-Liouville. No es absolutamente necesario conocer los detalles del teorema de Sturm-Liouville para ser capaces de usar los ME, aunque es útil saber porque ciertas bases son mejores que otras.

El problema de Sturm-Liouville es un problema de segundo orden y valores en la frontera de la forma

$$-\frac{d}{dx} \left(p(x) \frac{du}{dx} \right) + q(x)u = \lambda w(x)u \quad -1 < x < 1$$

+condiciones en la frontera en u (1.2.1)

Las únicas funciones propias que satisfacen dicho problema son

$$-\frac{d}{dx} \left((1-x)^{1+\alpha} - (1+x)^{1+\beta} \frac{du}{dx} \right) = \lambda(1-x)^\alpha(1+x)^\beta u \quad -1 < x < 1 \quad (1.2.2)$$

donde $\alpha, \beta > -1$. Estas funciones propias son los llamados *Polinomios de Jacobi* los cuales se representan por $P_k^{(\alpha, \beta)}(x)$. Donde para el caso de los polinomios de Chebyshev tenemos que $T_k(x) = P_k^{(-1/2, -1/2)}(x)$

1.3. Métodos Espectrales

Los métodos espectrales se utilizan para resolver numéricamente ecuaciones diferenciales de la forma:

$$\frac{\partial u(x, t)}{\partial t} = L(x, t)u(x, t) \quad x \in D, \quad t \geq 0, \quad (1.3.1)$$

$$B(x)u(x, t) = 0, \quad x \in \partial D, \quad t > 0, \quad (1.3.2)$$

$$u(x, 0) = g(x), \quad x \in D, \quad (1.3.3)$$

donde D es el dominio espacial con frontera ∂D , $L(x, t)$ es un operador diferencial, $B(x)$ es el operador de frontera y $g(x)$ es la condición inicial. Lo que se hace para encontrar una solución aproximada con métodos espectrales, es utilizar una serie truncada de funciones ortogonales suaves $\Phi_n(x)$, por lo que la solución se representa como

$$u_N(x, t) = \sum_{n=0}^N \hat{u}_n(t) \Phi_n(x). \quad (1.3.4)$$

Lo siguiente que se debe hacer, es encontrar los coeficientes \hat{u}_n con los cuales tendremos la solución. Existen varias formas de obtener los coeficientes, y cada manera de encontrarlos definen los diferentes esquemas de métodos espectrales. Los más conocidos son: el método de Colocación, el de Galerkin y el de Tau.

Para elegir las funciones ortogonales se toma en cuenta las condiciones de frontera. Para condiciones de frontera periódicas, las funciones de Fourier es la elección natural. Para condiciones de frontera no periódicas, la opción más utilizada la constituyen los polinomios de Chebyshev ya que estos son en realidad una serie de Fourier, por esta razón también se puede aplicar en ellos las transformadas de Fourier rápidas, con las cuales se reduce el número de operaciones, que se ve reflejado en el tiempo que tarda en resolver la ecuación.

Si comparamos los métodos espectrales con los métodos de diferencias finitas y elementos finitos, encontramos que para ecuaciones diferenciales ordinarias o parciales en un dominio simple, y siendo la solución suave, veremos que los métodos espectrales son mucho más precisos. Un ejemplo lo vemos en la figura 1.1 en donde se grafica la solución a la ecuación de Helmholtz como se describe en el capítulo 4.

Es necesario estudiar las bases de funciones ortogonales que elegirán para aplicar los ME. En este trabajo usaremos las bases de Fourier y de Chebyshev, por

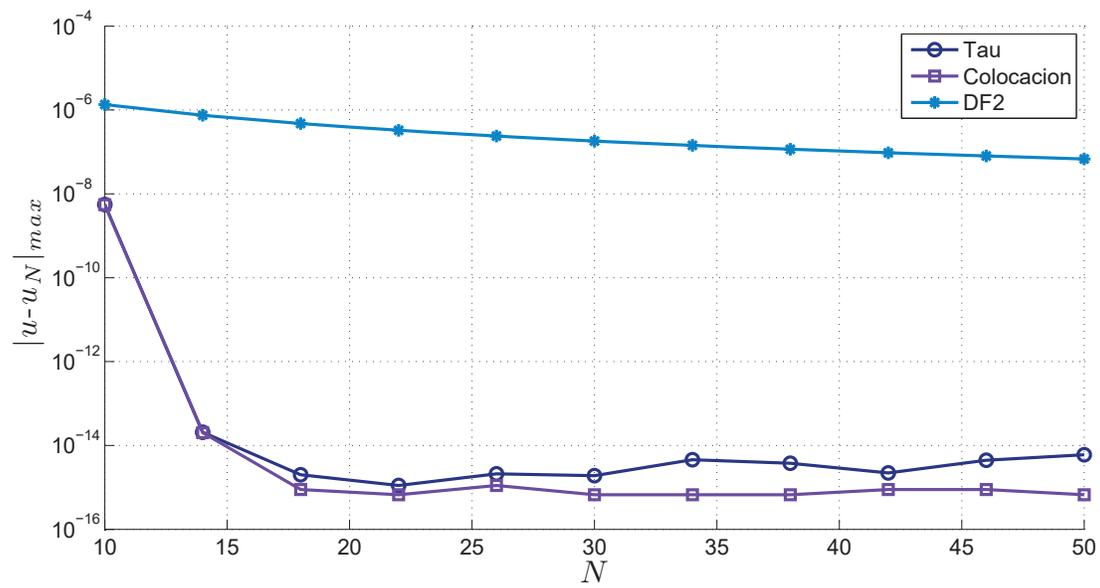


Figura 1.1: Comparación de exactitud de los ME (método de Tau y Colocación) con las diferencias finitas de orden 2.

lo que en el capítulo 2 se hace una exposición breve de las mismas. En el capítulo 3 se describen los esquemas espectrales, los cuales son: el método de Galerkin, método de Tau y método de colocación. En el capítulo 4 se realizan ejemplos para el entendimiento de los ME.

Capítulo 2

Base de Fourier y de Chebyshev

2.1. Base de Fourier

2.1.1. La expansión de Fourier continua

Para resolver problemas con condiciones de frontera periódicas, podemos usar una expansión con la base de Fourier, la cual descompone una función $u(x)$ en una suma de funciones seno y coseno, esta forma trigonométrica es bien conocida. La forma compleja de la serie de Fourier está dada por

$$u(x) = \sum_{n=-\infty}^{\infty} \hat{u}_n e^{inx}. \quad (2.1.1)$$

Como se menciona en el capítulo anterior necesitamos hacer una truncación de esta serie. Si se toma la serie truncada de Fourier que contiene $N + 1$ términos tendremos

$$P_N u(x) = \sum_{n=-N/2}^{N/2} \hat{u}_n e^{inx} \quad (2.1.2)$$

donde asumimos que N es par. Otros autores puede usar diferentes formas de truncar (2.1.1), otra forma sería tomando $n = -N, \dots, N$. Para poder encontrar los coeficientes \hat{u}_n recurrimos a la propiedad de ortogonalidad de la base, por lo que es necesario primero definir ortogonalidad:

Decimos que dos funciones son ortogonales entre ellas cuando su producto punto es igual a cero, esto es

$$(f, g)_\omega = \int_a^b f(x)g^*(x)\omega(x)dx = 0 \quad (2.1.3)$$

donde $[a, b]$ es el dominio de las funciones y $\omega(x)$ es la función peso que normaliza nuestras funciones ortogonales.

Para el caso de Fourier tenemos que la ortogonalidad de dos funciones de esta misma base se expresa de la siguiente manera

$$(e^{inx}, e^{imx}) = \int_0^{2\pi} e^{inx} e^{-imx} dx = \int_0^{2\pi} e^{i(n-m)x} dx = 2\pi\delta_{n,m} \quad (2.1.4)$$

donde la función peso para Fourier es $\omega(x) = 1$.

La expansión en funciones ortogonales, como la series de Fourier, son atractivas porque podemos encontrar los coeficientes fácilmente con una proyección ortogonal. En el caso de Fourier encontramos que

$$(P_N u(x), e^{inx}) = \sum_{m=-N/2}^{N/2} \hat{u}_m (e^{imx}, e^{inx}) = 2\pi\hat{u}_n \quad (2.1.5)$$

por lo que los coeficientes los obtenemos de

$$\hat{u}_n = \frac{1}{2\pi} \int_0^{2\pi} P_N u(x) e^{-inx} dx, \quad n = -N/2, \dots, N/2. \quad (2.1.6)$$

Los coeficientes pueden ser números complejos $\hat{u}_n = a_n - ib_n$ con $a_n, b_n \in \mathbb{R}$. Si nuestra función $u(x)$ se sabe real, encontramos que los coeficientes cumplen con $\hat{u}_{-n} = \hat{u}_n^*$, esto es, $a_{-n} = a_n$ y $-b_{-n} = b_n$. La ecuación (2.1.6) es conocida como la *transformada de Fourier* y la ecuación (2.1.2) la *transformada de Fourier inversa*.

La serie (2.1.2) se dice que converge a la función $u(x)$ si

$$\int_0^{2\pi} |u(x) - P_N u(x)|^2 dx \rightarrow 0 \text{ cuando } N \rightarrow \infty. \quad (2.1.7)$$

Esto es, que la serie de Fourier truncada $P_N u(x)$ se va acercando a $u(x)$ cuando N tiende a infinito. Tomando en cuenta que $(u, u) = \|u\|^2$ definimos el error de truncación $\epsilon_T(N)$ como

$$\epsilon_T(N) = \|u - P_N u\| = \left(\int_0^{2\pi} |u(x) - P_N u(x)|^2 dx \right)^{1/2} \quad (2.1.8)$$

donde no es difícil verificar que

$$\epsilon_T(N) = \left(2\pi \sum_{|n|>N/2} |\hat{u}_n|^2 \right)^{1/2}, \quad (2.1.9)$$

lo cual nos dice que el error de truncación está dominado por los últimos terminos de la serie, de modo que entre más rápido decaigan los coeficientes de la expansión, el error de truncación es más pequeño para una N dada.

Un resultado del análisis de Fourier que nos dice de qué depende que los coeficientes decaigan rápido, es el siguiente:

Si las primeras $(m - 1)$ derivadas de una función $u(x)$ junto con sus extensiones periódicas son continuas y su m -ésima derivada $u^{(m)}(x) \in L^2(0, 2\pi)$ entonces para toda $n \neq 0$ los coeficientes de Fourier decaen como $|\hat{u}_n| \propto \left(\frac{1}{n}\right)^m$.

Donde $L^2(a, b)$ indica que $\|u\|_\omega$ es cuadrado integrable respecto al peso ω en el intervalo $[a, b]$. El resultado anterior nos dice que entre más suave sea la función $u(x)$ más rápidamente converge la aproximación $P_N u(x)$ a ella. En otras palabras, entre más suave sea la función, los coeficientes de Fourier decrecerán más rápido.

2.1.2. La expansión de Fourier discreta

Uno de los puntos clave de las series de Fourier, es la obtención de los coeficientes \hat{u}_n a través de (2.1.6). Sin embargo, con frecuencia es muy difícil encontrar expresiones analíticas de los mismos. Para superar esta dificultad haremos uso de la transformada de Fourier discreta. Su definición se basa en la discretización del dominio espacial $x_j = 2\pi j/N$ con $j = 0, 1, 2, \dots, N - 1$. Los coeficientes de Fourier discretos con respecto a los puntos anteriores son

$$\tilde{u}_n = \frac{1}{N} \sum_{j=0}^{N-1} u(x_j) e^{-inx_j} \quad n = -N/2, \dots, N/2, \quad (2.1.10)$$

esta expresión constituye la definición de la *transformada de Fourier discreta*. Con la ayuda de la relación de ortogonalidad discreta

$$\frac{1}{N} \sum_{j=0}^{N-1} e^{-ipx_j} = \begin{cases} 1 & \text{si } p = Nm, m = 0, \pm 1, \pm 2, \dots \\ 0 & \text{cualquier otro caso} \end{cases} \quad (2.1.11)$$

puede obtenerse la *transformada de Fourier inversa* (inversa a (2.1.10)) que resulta ser

$$u(x_j) = \sum_{n=-N/2}^{N/2} \tilde{u}_n e^{inx_j}, \quad j = 0, \dots, N - 1. \quad (2.1.12)$$

La transformada de Fourier discreta (TFD) se puede ver como un mapeo entre los N números $u(x_j)$ con $j = 0, \dots, N - 1$, y los N números complejos \tilde{u}_n con $n = -N/2, \dots, N/2 - 1$. Esto quiere decir, que si se tienen los $u(x_j)$ se pueden obtener \tilde{u}_n a través de (2.1.10) y si se tienen \tilde{u}_n obtener $u(x_j)$ con (2.1.12). La ecuación anterior y la ecuación (2.1.10) son análogas a las ecuaciones (2.1.6) y (2.1.2) respectivamente. Definimos el polinomio interpolante de $u(x)$ como

$$I_N u(x) = \sum_{n=-N/2}^{N/2} \tilde{u}_n e^{inx} \quad (2.1.13)$$

donde $I_N u(x_j) = u(x_j)$ pero que a diferencia de (2.1.12) podemos tomar cualquier valor de x dentro del intervalo $[0, 2\pi]$.

Para encontrar una relación entre los coeficientes \hat{u}_n y \tilde{u}_n nos valemos de que la ecuación (2.1.1) converge a la función u en todo punto, incluyendo los puntos $\{x_j\}$, sustituimos $u(x_j)$ en (2.1.10) y tendremos

$$\tilde{u}_n = \frac{1}{N} \sum_{j=0}^{N-1} \sum_{k=-\infty}^{\infty} \hat{u}_k e^{i(k-n)x_j} \quad (2.1.14)$$

que al considerar (2.1.11) llegamos a

$$\tilde{u}_n = \hat{u}_n + \sum_{\substack{m=-\infty \\ m \neq 0}}^{\infty} \hat{u}_{n-Nm} \quad n = -N/2, \dots, N/2. \quad (2.1.15)$$

Podemos decir de la ecuación anterior que \tilde{u}_n es una aproximación de \hat{u}_n . ¿Qué tan buena resulta dicha aproximación de \hat{u}_n por \tilde{u}_n ? depende del valor del segundo término del lado derecho de (2.1.15). Otra forma equivalente de escribir (2.1.15) es

$$I_N u = P_N u + R_N u \quad (2.1.16)$$

donde

$$R_N u = \sum_{n=-N/2}^{N/2} \left(\sum_{\substack{m=-\infty \\ m \neq 0}}^{\infty} \hat{u}_{n-Nm} \right) e^{inx} \quad (2.1.17)$$

es el error de discretización también conocido como el *error de aliación*, el cual se estudiará en la sección 3.3.

2.1.3. Diferenciación

Cuando hablamos de ecuaciones diferenciales estamos diciendo que la función solución es diferenciable q veces, donde q es el orden de la ecuación diferencial. En los ME la diferenciación de una función depende de la base que se este usando, por lo que es necesario conocer la diferenciación de ella. Para la base de Fourier tenemos que

$$\frac{d}{dx}(e^{inx}) = ine^{inx} \quad (2.1.18)$$

donde en general

$$\frac{d^q}{dx^q}(e^{inx}) = (in)^q e^{inx}. \quad (2.1.19)$$

Conociendo las derivadas de la base de Fourier podemos aproximar la derivada de nuestra función solución de la siguiente manera

$$(P_N u(x))' = \sum_{n=-N/2}^{N/2} in \hat{u}_n e^{inx} = \sum_{n=-N/2}^{N/2} \hat{u}_n^{(1)} e^{inx} \quad (2.1.20)$$

donde $(P_N u)'$ se le conoce como la *derivada de la proyección de Fourier* y que también podemos escribir como $P_N u'$ ya que, para este caso, la diferenciación y la truncación conmutan.

Notemos que los coeficientes de la expansión de $u'(x)$ dependen directamente de los coeficientes de la expansión de la función $u(x)$, explícitamente $\hat{u}_n^{(1)} = in \hat{u}_n$. En general, si u es q veces diferenciable respecto a x , se tiene que $\hat{u}_n^{(q)} = (in)^q \hat{u}_n$.

La diferenciación aproximada en los puntos de colocación x_j esta dado por la ecuación

$$(D_N^{(1)} u)_j = \sum_{n=-N/2}^{N/2} \tilde{u}_n^{(1)} e^{inx_j}, \quad j = 0, \dots, N-1, \quad (2.1.21)$$

donde

$$\tilde{u}_n^{(1)} = in \tilde{u}_n = \frac{in}{N} \sum_{l=0}^{N-1} u(x_l) e^{-inx_l}, \quad n = -N/2, \dots, N/2. \quad (2.1.22)$$

A $D_N^{(1)} u$ se le conoce como la *derivada de la interpolación de Fourier* de u . Notemos que $D_N^{(1)} u = (I_N u)'$ pero no $D_N^{(1)} u \neq I_N u'$.

Las ecuaciones (2.1.21) y (2.1.22) al combinarse, se pueden escribir de manera matricial de la forma

$$(D_N^{(1)}u)_j = \sum_{l=0}^{N-1} (D_N^{(1)})_{jl}u_l \quad (2.1.23)$$

donde

$$(D_N^{(1)})_{jl} = \frac{1}{N} \sum_{n=-N/2}^{N/2} in e^{in(x_j - x_l)}. \quad (2.1.24)$$

Una propiedad de resolver problemas de esta manera es que ya no nos metemos con los coeficientes, sino directamente con la función evaluada en los puntos de colocación, donde además sabemos que $I_N u(x_j) = u(x_j)$, por lo que no agregamos el error de aliación.

Gottlieb, Hussaini y Orzag (1984) mostraron que el interpolante trigonométrico se puede escribir de la forma

$$I_N u(x) = \sum_{l=0}^{N-1} C_l(x)u(x_l) \quad (2.1.25)$$

donde la función $C_l(x)$ para los puntos $x_l = 2\pi l/N$ es

$$C_l(x) = \frac{1}{N} \sin \left[\frac{N(x - x_l)}{2} \right] \cot \left[\frac{(x - x_l)}{2} \right] \quad (2.1.26)$$

y encontraron que la derivada del interpolante $I_N u(x)$ respecto a x se puede escribir como

$$(I_N u(x))' = \sum_{l=0}^{N-1} C_{l,x}(x)u(x_l) \quad (2.1.27)$$

que al comparar (2.1.23) con (2.1.27) se obtuvo que $(D_N^{(1)})_{jl} = C_{l,x}(x_j)$. Por lo que al derivar (2.1.26) y evaluar en los puntos x_j tenemos

$$(D_N^{(1)})_{jl} = \begin{cases} \frac{1}{2}(-1)^{j+l} \cot \left[\frac{(j-l)\pi}{N} \right], & j \neq l \\ 0, & j = l \end{cases} \quad (2.1.28)$$

y para la segunda derivada de (2.1.26) llegamos a

$$(D_N^{(2)})_{jl} = \begin{cases} \frac{1}{4}(-1)^{j+l} N + \frac{(-1)^{j+l+1}}{2 \sin^2 \left[\frac{(j-l)\pi}{N} \right]} & j \neq l, \\ -\frac{(N-1)(N-2)}{12} & j = l. \end{cases} \quad (2.1.29)$$

Si se desea encontrar la matriz para la q -ésima derivada solamente se ocupa derivar q veces la ecuación (2.1.26) y evaluarse en los puntos de colocación. Para el caso de N impar se ocupan los puntos de colocación

$$x_j = \frac{2j}{N+1}\pi \quad j = 0, \dots, N \quad (2.1.30)$$

donde las matrices de diferenciación para esos puntos se pueden encontrar en el libro de Peyret [10]. Si se desea saber como obtener las funciones $C_l(x)$ vea el capítulo 5 de libro de Boyd [1].

2.2. Base de Chebyshev

Como se menciona en la sección 1.2 la base más utilizada para problemas con condiciones de frontera no periódicas es la base de Chebyshev y en esta sección veremos algunas de las propiedades de los polinomios de Chebyshev y sus transformaciones.

Los polinomios de Chebyshev $T_n(x)$ de grado k se definen en el intervalo $x \in [-1, 1]$ por:

$$T_n(x) = \cos(n \arccos x), \quad n = 0, 1, 2, \dots \quad (2.2.1)$$

por lo que $-1 \leq T_n \leq 1$. De la definición anterior podemos observar que

$$\begin{aligned} T_0(x) &= \cos(0) = 1 \\ T_1(x) &= \cos(\arccos x) = x \\ T_2(x) &= \cos(2 \arccos x) = 2 \cos(\arccos x) - 1 = 2x^2 - 1. \end{aligned}$$

pero los siguientes $T_n(x)$ resultan difícil encontrarlos de esta manera, por lo que para encontrarlos en forma polinomial recurrimos a la identidad trigonométrica

$$\cos(k+1)z + \cos(k-1)z = 2 \cos z \cos kz.$$

Tomando $z = \arccos x$ llegamos a la ecuación de recurrencia siguiente

$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x), \quad n \geq 1 \quad (2.2.2)$$

por lo que, partiendo de $T_0(x)$ y $T_1(x)$ podemos obtener $T_n(x)$ para $n \geq 2$. En la Fig. 2.1 se muestra gráfica de los primeros polinomios.

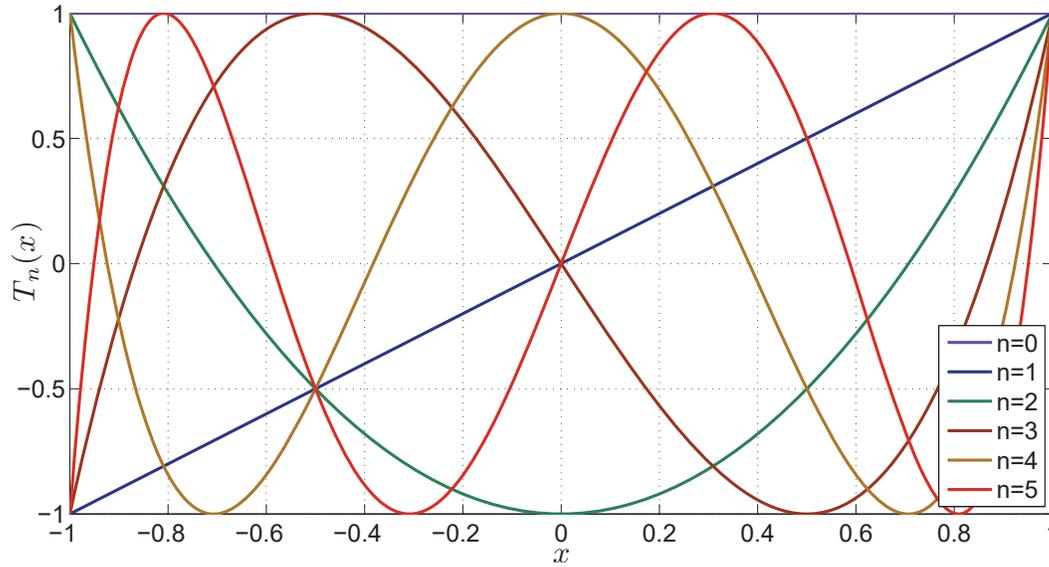


Figura 2.1: Gráfica de los primeros 6 polinomios de Chebyshev.

De la definición (2.2.1) es fácil ver que los polinomios de Chebyshev evaluados en $x = \pm 1$, que son las fronteras de intervalo donde se encuentran definidos, adquieren los valores

$$T_n(\pm 1) = (\pm 1)^n. \quad (2.2.3)$$

Esta identidad nos será útil para implementar algunas condiciones de frontera.

Los polinomios de Chebyshev son ortogonales en el intervalo $[-1, 1]$ con el peso $w(x) = (1 - x^2)^{-1/2}$, esto es

$$(T_n, T_k)_\omega = \int_{-1}^1 \frac{T_n(x)T_k(x)}{\sqrt{1-x^2}} dx = \frac{\pi}{2} c_n \delta_{n,k} \quad (2.2.4)$$

donde $\delta_{n,k}$ es la delta de Kronecker y

$$c_n = \begin{cases} 2, & \text{si } n = 0 \\ 1, & \text{si } n \geq 1 \end{cases}.$$

Para verificar (2.2.4) hacemos el cambio de variable $z = \arccos(x)$ y vemos los siguientes casos

$$n \neq k$$

$$\begin{aligned} \int_0^\pi \cos nz \cos kz dz &= \frac{1}{2} \int_0^\pi [\cos(n+k)z + \cos(n-k)z] dz \\ &= \frac{1}{2} \left[\frac{\sin(n+k)z}{k+n} + \frac{\sin(n-k)z}{n-k} \right]_0^\pi = 0 \end{aligned}$$

$$n = k \neq 0$$

$$\begin{aligned} \int_0^\pi \cos^2 nz dz &= \frac{1}{2} \int_0^\pi (1 + \cos 2nz) dz \\ &= \frac{1}{2} \left[z + \frac{\sin 2nz}{2n} \right]_0^\pi = \frac{\pi}{2} \end{aligned}$$

$$n = k = 0$$

$$\int_0^\pi dz = \pi$$

que se resumen en la ecuación (2.2.4).

2.2.1. La expansión de Chebyshev continua

Al igual que con las series de Fourier, podemos escribir una función $u(x)$ con polinomios de Chebyshev

$$u(x) = \sum_{n=0}^{\infty} \hat{u}_n T_n(x) \quad (2.2.5)$$

La serie infinita la aproximamos con un número $N + 1$ de términos, tal que

$$P_N u(x) = \sum_{n=0}^N \hat{u}_n T_n(x). \quad (2.2.6)$$

Los coeficientes \hat{u}_k los obtenemos a través de la ortogonalidad de los polinomios de Chebyshev de la siguiente manera

$$(P_N u, T_k)_\omega = \sum_{n=0}^N \hat{u}_n (T_n, T_k)$$

obteniendo lo que llamaremos la *transformada de Chebyshev*

$$\hat{u}_n = \frac{2}{\pi c_n} \int_{-1}^1 \frac{u(x) T_n(x)}{\sqrt{1-x^2}} dx \quad (2.2.7)$$

donde la ecuación (2.2.6) es su *transformada de Chebyshev inversa*.

El error por aproximar (2.2.5) con los primeros $N + 1$ términos, al igual que con la serie de Fourier, solo depende de los coeficientes \hat{u}_n con $n > N$, esto es

$$\epsilon_T(N) = \|u - P_N u\|_\omega = \left(\frac{\pi}{2} \sum_{n=N+1}^{\infty} |\hat{u}_n|^2 \right)^{1/2}. \quad (2.2.8)$$

De esto, observamos que entre más rápido decaigan los coeficientes más pequeño será el error. La relación (2.2.8) se obtiene de la misma manera que en (2.1.9) para series de Fourier pero considerando los límites de integración y la función peso correspondiente a los polinomios de Chebyshev.

2.2.2. La expansión de Chebyshev discreta

Por la misma razón que se dio en la sección 2.1.2 es indispensable, en varios casos, discretizar el dominio espacial para evitar dificultades en la obtención de los coeficientes. Los puntos que elegimos para esta discretización son los puntos $x_j = \cos(\pi j/N)$ donde $j = 0, 1, \dots, N-1$, el motivo de la elección de estos puntos en lugar de puntos equidistantes lo podemos ver en el apéndice A. Usando esto, nuestra aproximación en estos puntos es

$$I_N u(x_j) = \sum_{k=0}^N \tilde{u}_k T_k(x_j) = u(x_j). \quad (2.2.9)$$

Para obtener los coeficientes usamos la siguiente relación de ortogonalidad discreta de los polinomios de Chebyshev

$$\sum_{j=0}^N \frac{2}{N \bar{c}_m \bar{c}_j} T_m(x_j) T_n(x_j) = \begin{cases} 1 & n = m \\ 0 & m \neq n \end{cases} \quad (2.2.10)$$

que es válido para $0 \leq m, n \leq N$, donde

$$\bar{c}_k = \begin{cases} 2 & k = 0, N \\ 1 & k = 1, \dots, N-1 \end{cases}$$

y obtenemos

$$\tilde{u}_k = \sum_{j=0}^N \frac{2}{N \bar{c}_k \bar{c}_j} u(x_j) \cos\left(\frac{\pi j k}{N}\right) \quad (2.2.11)$$

que es la *transformada de Chebyshev discreta (TCD)*. La *transformada de Chebyshev inversa* de la ecuación (2.2.11) es la ecuación (2.2.9) y ambas son análogas a las ecuaciones (2.2.7) y (2.2.6) respectivamente.

Si reemplazamos $u(x_j)$ en su expresión de términos de la serie infinita en (2.2.11) tal que

$$\tilde{u}_k = \sum_{l=0}^N \hat{u}_l \left[\sum_{j=0}^N \frac{2}{N \bar{c}_k \bar{c}_j} T_k(x_j) T_l(x_j) \right] + \sum_{l=N+1}^{\infty} \hat{u}_l \left[\sum_{j=0}^N \frac{2}{N \bar{c}_k \bar{c}_j} T_k(x_j) T_l(x_j) \right]$$

donde el primer término entre corchetes es la ecuación (2.2.10), entonces se puede escribir como

$$\tilde{u}_k = \hat{u}_k + \sum_{l=N+1}^{\infty} \hat{u}_k \left[\sum_{j=0}^N \frac{2}{N\bar{c}_k\bar{c}_j} \cos\left(\frac{\pi jk}{N}\right) \cos\left(\frac{\pi jl}{N}\right) \right]$$

donde si usamos

$$\cos \alpha \cos \beta = \frac{1}{2} [\cos(\alpha + \beta) + \cos(\alpha - \beta)]$$

y consideramos

$$\sum_{j=0}^N \cos\left(\frac{pj\pi}{N}\right) = \begin{cases} N+1 & \text{Si } p = 2mN, m = 0, \pm 1, \pm 2, \dots \\ \frac{1}{2}[1 - (-1)^p] & \text{cualquier otro,} \end{cases} \quad (2.2.12)$$

se puede obtener

$$\tilde{u}_k = \hat{u}_k + \frac{1}{\bar{c}_k} \sum_{m=1}^{\infty} \hat{u}_{2Nm \pm k} \quad k = 0, \dots, N \quad (2.2.13)$$

Así que esta \tilde{u}_n es una aproximación de \hat{u}_n , y es buena en la medida en que el segundo término del lado derecho en (2.2.13) sea despreciable. Al igual que en el caso de Fourier se puede reescribir como en la ecuación (2.1.16) donde $R_N u$ en el caso de Chebyshev es

$$R_N u = \sum_{k=0}^N \frac{1}{\bar{c}_k} \left(\sum_{m=1}^{\infty} \hat{u}_{2Nm \pm k} \right) T_k(x) \quad (2.2.14)$$

que es el error de aliación. La razón de (2.1.16) se debe por la discretización de la aproximación con polinomios de Chebyshev, lo mismo que sucede con la serie de Fourier. En la sección 2.3 se explicará con más detalle sobre el error de aliación.

2.2.3. Diferenciación

Para resolver ecuaciones diferenciales con los polinomios de Chebyshev es necesario la derivada de los polinomios. estos los encontramos como

$$T'_k(x) = \frac{d}{dx}(\cos(k \arccos x)) = \frac{k \sin(k \arccos x)}{\sqrt{1-x^2}}$$

que al hacer el cambio de variable de $x = \cos z$ se simplifica a

$$T'_k(z) = \frac{k \sin kz}{\sin z} \quad 0 \leq z \leq \pi. \quad (2.2.15)$$

Pero la ecuación anterior no está definida en $z = 0$ y $z = \pi$, por lo que buscamos el límite en dichos puntos utilizando la regla de L'Hopital y de ahí encontramos que al regresar a la variable x , obtenemos la evaluación de los puntos de la frontera

$$T'_k(\pm 1) = (\pm 1)^{k+1} k^2. \quad (2.2.16)$$

Este resultado y el de la ecuación (2.2.3) nos serán muy útiles para poner las condiciones de frontera ya sean del tipo Dirichlet, Neumann o combinación de estas.

Es posible encontrar de manera recurrente los polinomios $T_k(x)$ a partir de las $T'_k(x)$ con la identidad trigonométrica

$$\sin(k+1)z - \sin(k-1)z = 2 \cos kz \sin z,$$

combinando esta identidad y la ecuación (2.2.15) obtenemos

$$\frac{T'_{k+1}(x)}{k+1} - \frac{T'_{k-1}(x)}{k-1} = 2T_k(x) \quad k \geq 2. \quad (2.2.17)$$

Ahora bien, en el problema descrito por (1.3.1) a (1.3.3) puede llegar a pedir que la función solución sea continua en sus m derivadas. Si nuestra función u es continua en su primera derivada, se puede escribir una serie truncada de la derivada de la función de la siguiente manera

$$(P_N u(x))' = \sum_{k=0}^N \hat{u}_k T'_k(x) = \sum_{k=0}^{N-1} \hat{u}_k^{(1)} T_k(x) \quad (2.2.18)$$

donde $(P_N u)'$ es la *derivada de la proyección de Chebyshev*. Existe una relación entre los coeficientes \hat{u}_k y $\hat{u}_k^{(1)}$

$$c_k \hat{u}_k^{(1)} = \hat{u}_{k+2}^{(1)} + 2(k+1) \hat{u}_{k+1} \quad 0 \leq k \leq N-1. \quad (2.2.19)$$

La relación anterior nos va a permitir encontrar los coeficientes \hat{u}_k en función de los $\hat{u}_k^{(1)}$. La ecuación (2.2.19) la obtenemos a partir de integrar (2.2.18)

$$P_N u(x) = \int P_N u'(x) dx = \int (\hat{u}_0^{(1)} T_0 + \hat{u}_1^{(1)} T_1 + \cdots + \hat{u}_N^{(1)} T_N) dx$$

considerando

$$\begin{aligned} T_0(x) &= T'_1(x) \\ T_1(x) &= \frac{1}{4} T'_2(x) \end{aligned}$$

y utilizando la ecuación (2.2.17)

$$\begin{aligned}
P_N u(x) &= \hat{u}_0^{(1)} T_1 + \frac{\hat{u}_1^{(1)} T_2}{2} + \frac{\hat{u}_2^{(1)} (T_3 - T_1)}{2} + \frac{\hat{u}_3^{(1)} (T_4 - T_2)}{2} + \cdots + \\
&\quad \frac{\hat{u}_{k-1}^{(1)} (T_k - T_{k-2})}{2} + \frac{\hat{u}_k^{(1)} (T_{k+1} - T_{k-1})}{2} + \frac{\hat{u}_{k+1}^{(1)} (T_{k+2} - T_k)}{2} \\
&\quad + \cdots
\end{aligned}$$

Si agrupamos los coeficientes del polinomio T_1 , se obtiene \hat{u}_1 , ahora si agrupamos los coeficientes del polinomio T_2 , se obtiene \hat{u}_2 , si agrupamos los coeficientes del polinomio T_n , se obtiene \hat{u}_n , es decir

$$\begin{aligned}
T_1 &\longrightarrow \hat{u}_1 = \frac{1}{2}(2\hat{u}_0^{(1)} - \hat{u}_2^{(1)}) &\implies 2\hat{u}_1 &= 2\hat{u}_0^{(1)} - \hat{u}_2^{(1)} \\
T_2 &\longrightarrow \hat{u}_2 = \frac{1}{2}\left(\frac{\hat{u}_1^{(1)}}{2} - \frac{\hat{u}_3^{(1)}}{2}\right) &\implies 4\hat{u}_2 &= 2\hat{u}_1^{(1)} - \hat{u}_3^{(1)} \\
&\vdots \\
T_{N-2} &\longrightarrow \hat{u}_{N-2} = \frac{1}{2}\left(\frac{\hat{u}_{N-3}^{(1)}}{N-2} - \frac{\hat{u}_{N-1}^{(1)}}{N-2}\right) &\implies 2(N-2)\hat{u}_{N-2} &= \hat{u}_{N-3}^{(1)} - \hat{u}_{N-1}^{(1)} \\
T_{N-1} &\longrightarrow \hat{u}_{N-1} = \frac{1}{2}\left(\frac{\hat{u}_{N-2}^{(1)}}{N-1}\right) &\implies 2(N-1)\hat{u}_{N-1} &= \hat{u}_{N-2}^{(1)} \\
T_N &\longrightarrow \hat{u}_N = \frac{1}{2}\left(\frac{\hat{u}_{N-1}^{(1)}}{N}\right) &\implies 2(N)\hat{u}_N &= \hat{u}_{N-1}^{(1)}
\end{aligned}$$

donde $\hat{u}_N^{(1)} = \hat{u}_{N+1}^{(1)} = 0$. La relación \hat{u}_0 no se obtiene de la misma manera, ya que aparece en la constante de integración. Si nuestra función es m veces diferenciable, utilizamos en varias ocasiones la ecuación (2.2.17) y podemos llegar a

$$c_k \hat{u}_k^{(q)} = \hat{u}_{k+2}^{(q)} + 2(k+1)\hat{u}_{k+1}^{(q-1)} \quad 0 \leq k \leq N-q \quad (2.2.20)$$

donde $\hat{u}_{N-q+1}^{(q)} = \hat{u}_{N-q+2}^{(q)} = 0$. Esta relación de recurrencia nos permitirá obtener los coeficientes $\hat{u}_k^{(q)}$ de los coeficientes $\hat{u}_k^{(q-1)}$.

La *derivada del interpolante de Chebyshev* de una función u en los puntos x_j esta definida como la derivada de la serie de Chebyshev discreta de u en los mismos nodos

$$D_N^{(1)} u = (I_N u)'. \quad (2.2.21)$$

esta derivada puede también escribirse en forma matricial como

$$(D_N^{(1)} u)_j = \sum_{l=0}^N (D_N^{(1)})_{jl} u(x_l), \quad j = 0, \dots, N. \quad (2.2.22)$$

Para obtener la matriz $(D_N^{(1)})_{jl}$ se puede hacer de dos maneras. La primera es haciendo

$$(I_N u(x_j))' = \sum_{l=0}^N \tilde{u}_l T_l'(x_j) \quad (2.2.23)$$

y usando la TCD para luego expresar $T_l(x_j)$ y $T_l'(x_j)$ en términos de funciones trigonométricas. Finalmente, aplicar identidades trigonométricas para evaluar las sumas. La segunda manera es recordando la ecuación (2.1.25) que nos dice que el interpolante se puede escribir como

$$I_N u(x) = \sum_{l=0}^N C_l(x) u(x_l) \quad (2.2.24)$$

donde para el caso de Chebyshev la función $C_l(x)$ para los puntos $x_j = \cos \pi j/N$ es

$$C_l(x) = \frac{(-1)^{j+1}(1-x^2)T_N'(x)}{\bar{c}_j N^2(x-x_j)} \quad (2.2.25)$$

y tendremos que $C_l^{(1)}(x_j) = (D_N^{(1)})_{jl}$. La segunda forma es la más sencilla de hacer, ya que para matrices $(D_N^{(q)})_{jl}$ solo hay que derivar q veces la función (2.2.25) que a diferencia de la primera donde se tienen que realizar varios pasos. Para la matriz de la primera derivada del interpolante tenemos

$$(D_N)_{jl} = \begin{cases} \frac{\bar{c}_j}{\bar{c}_l} \frac{(-1)^{j+l}}{x_j - x_l}, & j \neq l, \\ -\frac{x_l}{2(1-x_l^2)}, & 1 \leq j = l \leq N-1, \\ \frac{2N^2+1}{6}, & j = l = 0, \\ -\frac{2N^2+1}{6}, & j = l = N, \end{cases} \quad (2.2.26)$$

y para la segunda derivada

$$(D_N^{(2)})_{jl} = \begin{cases} \frac{(-1)^{j+l}}{\bar{c}_l} \frac{x_j^2 + x_j x_l - 2}{(1-x_j^2)(x_j - x_l)^2}, & \begin{matrix} 1 \leq j \leq N-1 \\ 0 \leq l \leq N, j \neq l \end{matrix} \\ -\frac{(N^2-1)(1-x_j^2)+3}{3(1-x_j^2)^2}, & 1 \leq j = l \leq N-1, \\ \frac{2}{3} \frac{(-1)^l}{\bar{c}_l} \frac{(2N^2+1)(1-x_l)-6}{(1-x_j)^2}, & j = 0, 1 \leq l \leq N, \\ \frac{2}{3} \frac{(-1)^{l+N}}{\bar{c}_l} \frac{(2N^2+1)(1-x_l)-6}{(1-x_j)^2}, & j = N, 1 \leq l \leq N-1, \\ \frac{N^4-1}{15}, & j = l = 0, j = l = N. \end{cases} \quad (2.2.27)$$

para otros puntos de colocación las matrices cambian. Cabe mencionar que existe una forma general de obtener la matriz para cualquier polinomio ortogonal. La explicación de como obtenerla y el algoritmo para implementar se puede encontrar en el libro de Kopriva [8].

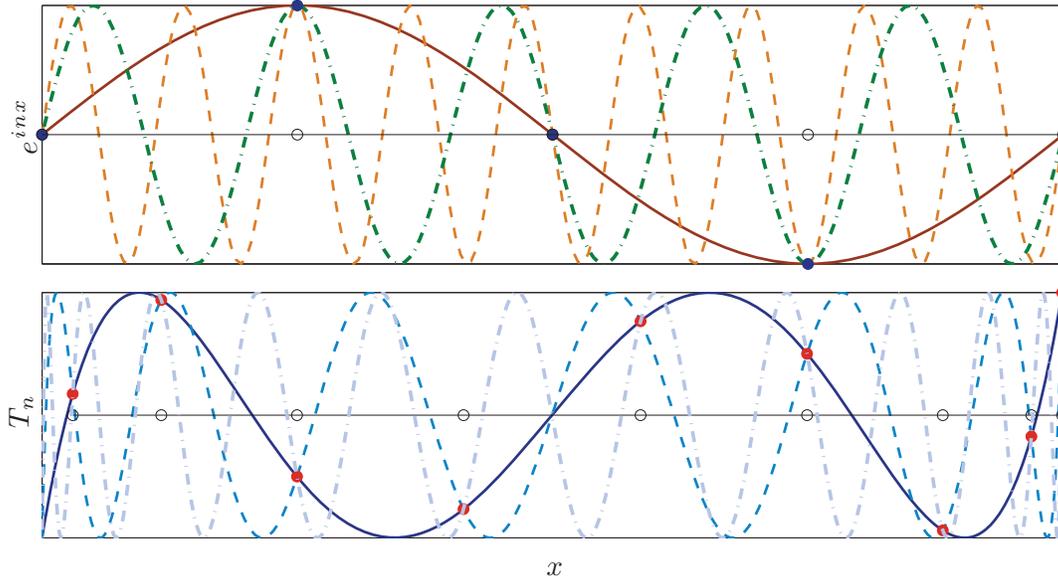


Figura 2.2: Los puntos muestran donde la función original (línea continua) coincide con otras funciones (líneas punteadas), este evento representa el error de aleación. La figura de superior es la representación del error de aleación para la base de Fourier y la figura inferior para la base de Chebyshev.

2.3. Error de aliación

Pudimos ver, tanto con la transformada de Fourier discreta y la transformada de Chebyshev discreta, nos encontramos con un error debido a la discretización del espacio. Estos errores son las ecuaciones (2.1.15) y (2.2.13). Ambas nos muestran que el k -ésimo modo del interpolante de u depende no solo del k -ésimo modo de u , sino también de todos los modos de u que se alían al k -ésimo modo en la red. La Figura 2.2 nos ilustra el efecto de aliación en la base de Fourier y Chebyshev.

Con ayuda de la ecuación (2.1.16) podemos ver que la norma del error de interpolación se puede escribir como

$$\|u - I_N u\|^2 = \|u - P_N u\|^2 + \|R_N u\|^2 \quad (2.3.1)$$

donde observamos que el error debido a la interpolación es mayor que el de truncación. Esta declaración es cierta pero falta mencionar que la influencia de la aliación en la exactitud del problema es asintóticamente del mismo orden que la del error de truncación, esto quiere decir, que el error de aliación va haciéndose

más pequeño cuando N crece (esto se explica en la sección 5.1 del libro de Canuto, et. al. [4]). Consideremos un ejemplo para observar que el comportamiento del error de aliación en cierto N . Tomemos la función $u(x) = \frac{3}{5-4\cos x}$ donde se conoce que para la base de Fourier $\hat{u}_k = 2^{-|k|}$ y comparamos con los coeficientes de la transformada de Fourier discreta \tilde{u}_k en la Cuadro 2.1 y 2.2. En dichas tablas observamos que la diferencia de los coeficientes va haciéndose menor mientras n se acerque al cero. Al comparar ambas tablas vemos que el error provocado por la aliación en los coeficientes con mismo índice es menor cuando $N = 52$ que cuando $N = 32$, lo cual concuerda con que el error de aliación decrece cuando crece N . Si tomamos distintos valores de N mayores a 52 se observa que la diferencia en los n cercanos a cero dejan de hacerse menor y oscilan en valores de orden 10^{-16} . El hecho de que no disminuya el error se debe a los errores de redondeo de la computadora. Si se toman $N > 100$ se tiene que el error en la mayoría de los \tilde{u}_n es del orden 10^{-16} , los cuales son errores muy pequeño que podemos despreciar.

n	\hat{u}_n	\tilde{u}_n	$ \tilde{u}_n - \hat{u}_n $
-16	1.5259e-05	3.0518e-05 - 2.8064e-17i	1.5259e-05
-15	3.0518e-05	3.8147e-05 + 8.1966e-16i	7.6294e-06
-14	6.1035e-05	6.4850e-05 + 6.1930e-16i	3.8147e-06
⋮	⋮	⋮	⋮
-1	0.5	5.0000e-01 + 1.3010e-17i	5.8208e-10
0	1	1.0000	4.6566e-10
1	0.5	5.0000e-01 - 1.3010e-17i	5.8208e-10
⋮	⋮	⋮	⋮
14	6.1035e-05	6.4850e-05 - 6.1930e-16i	3.8147e-06
15	3.0518e-05	3.8147e-05 - 8.1966e-16i	7.6294e-06
16	1.5259e-05	3.0518e-05 + 2.8064e-17i	1.5259e-05

Cuadro 2.1: Comparación entre los coeficientes de la expansión de Fourier continua \hat{u}_n y los coeficientes de la expansión de Fourier discreta \tilde{u}_n . Donde se toma $N = 32$.

n	\hat{u}_n	\tilde{u}_n	$ \tilde{u}_n - \hat{u}_n $
-26	1.4901e-08	2.9802e-08 - 1.7124e-15i	1.4901e-08
-25	2.9802e-08	3.7253e-08 + 6.2930e-16i	7.4506e-09
-24	5.9605e-08	6.3330e-08 + 4.4836e-16i	3.7253e-09
⋮	⋮	⋮	⋮
-1	0.5	5.0000e-01-5.6045e-18i	5.5514e-16
0	1	1.0000	2.2204e-16
1	0.5	5.0000e-01+5.6045e-18i	5.5514e-16
⋮	⋮	⋮	⋮
24	5.9605e-08	6.3330e-08 - 4.4836e-16i	3.7253e-09
25	2.9802e-08	3.7253e-08 - 6.2930e-16i	7.4506e-09
26	1.4901e-08	2.9802e-08 + 1.7124e-15i	1.4901e-08

Cuadro 2.2: Comparación entre los coeficientes de la expansión de Fourier continua \hat{u}_n y los coeficientes de la expansión de Fourier discreta \tilde{u}_n . Donde se toma $N = 52$.

2.3.1. No linealidades y el error de aliación

Cuando tratamos una ecuación diferencial donde se tiene un término no lineal sucede que el error de aliación en dicho término no decrece como es común. Deja de ser correcto que el error de discretización en el término no lineal se vaya haciendo pequeño cuando N crezca. Esto lo podemos ver mejor al graficar el valor absoluto de los coeficientes en forma logarítmica donde después de cierto k dejan de acercarse al cero, esto lo ilustra la Figura 2.3.

Este incremento en los coeficientes nos puede provocar dos problemas: la exactitud de la función aproximada con la función solución no se hará menor cuando se haga más grande N y para problemas evolutivos no tendremos estabilidad numérica. Esto último quiere decir que, en cada paso evolutivo nuestra solución numérica perderá exactitud, por lo que después de cierto tiempo nuestro modelo numérico ya no funciona.

Existen varias formas de tratar el error de aliación en el término no lineal y evitar las inestabilidades numéricas, una de ellas es mediante la técnica de la *regla de los 3/2*. Para explicar esta regla tomaremos el único caso de no linealidad que

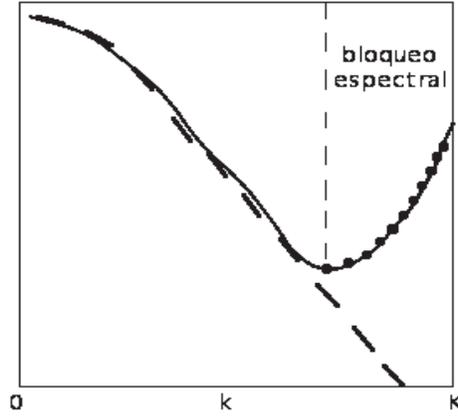


Figura 2.3: La línea punteada son el valor absoluto de los coeficientes en $t = 0$ y la línea continua a un tiempo después. La línea vertical es la frontera de los números de onda (Fourier) o el grado del polinomio (Chebyshev) entre la parte del espectro que decrece y la parte donde crece debido al ruido numérico y el de aliación. Los coeficientes corruptos están marcados con puntos.

se maneja en esta tesis que es de tipo

$$s(x) = u(x)v(x). \quad (2.3.2)$$

Analícemos de manera detallada como se comporta la no linealidad cuando se trate en un problema periódico.

Los coeficientes de la expansión en series de Fourier de la función $s(x)$ se dan por la ecuación (2.1.6), que al usar la ecuación (2.3.2) y expandir también las funciones $u(x)$ y $v(x)$, podemos llegar a

$$\hat{s}_k = \frac{1}{2\pi} \int_0^{2\pi} \left(\sum_{m=-N/2}^{N/2} \hat{u}_m e^{imx} \right) \left(\sum_{n=-N/2}^{N/2} \hat{v}_n e^{inx} \right) e^{-ikx} dx \quad (2.3.3)$$

$$= \sum_{m=-N/2}^{N/2} \sum_{n=-N/2}^{N/2} \left(\frac{1}{2\pi} \int_0^{2\pi} e^{i(m+n-k)x} dx \right) \quad (2.3.4)$$

$$= \sum_{m=-N/2}^{N/2} \sum_{n=-N/2}^{N/2} \hat{u}_m \hat{v}_n \delta_{m+n,k} \quad (2.3.5)$$

que finalmente nos da

$$\hat{s}_k = \sum_{n=-N/2}^{N/2} \hat{u}_{k-n} \hat{v}_n. \quad (2.3.6)$$

Consideramos que los coeficientes discretos de Fourier sean

$$\tilde{s}_k = \frac{1}{M} \sum_{j=0}^{M-1} s(x_j) e^{-inx_j} \quad (2.3.7)$$

donde $M > N$ y los puntos $x_j = 2\pi j/M$. Realizamos los mismos pasos hechos para obtener la ecuación (2.1.15) y llegamos a

$$\tilde{s}_k = \hat{s}_k + \sum_{\substack{m=-\infty \\ m \neq 0}}^{\infty} \hat{s}_{k+Mm} \quad (2.3.8)$$

donde al sustituir (2.3.6) en la parte del error de aliación tenemos que

$$\tilde{s}_k = \hat{s}_k + \sum_{\substack{m=-\infty \\ m \neq 0}}^{\infty} \sum_{n=-N/2}^{N/2} \hat{u}_{k+mM-n} \hat{v}_n \quad (2.3.9)$$

donde escogemos M tal que $|k + mM - n| > N/2$ para $|k|, |p| \leq N/2$ tomando en cuenta esto el valor mínimo que puede tener M para satisfacer lo anterior en toda k sería cuando $k = -N/2$ y $n = N/2$ y $m = 1$, esto es

$$-\frac{N}{2} - \frac{N}{2} + M > \frac{N}{2} \implies M > \frac{3N}{2} \quad (2.3.10)$$

De igual manera analicemos el comportamiento del término no lineal cuando se usa la base de Chebyshev. Tomemos la ecuación (2.3.2) podemos expandirla con polinomios de Chebyshev tal que sus coeficientes sean

$$\hat{s}_k = \frac{2}{\pi c_k} \int_{-1}^1 \left(\sum_{m=0}^N \hat{u}_m T_m \right) \left(\sum_{n=0}^N \hat{v}_n T_n \right) \frac{T_k(x)}{\sqrt{1-x^2}} dx \quad (2.3.11)$$

donde usamos la propiedad trigonométrica

$$\cos \alpha \cos \beta = \frac{1}{2} [\cos(\alpha + \beta) + \cos(\alpha - \beta)]$$

y la relación de ortogonalidad (2.2.4) para obtener

$$\hat{s}_k = \frac{1}{2} \left(\sum_{n+m=k} \hat{u}_m \hat{v}_n + \sum_{|m-n|=k} \hat{u}_m \hat{v}_n \right) \quad n, m \leq N. \quad (2.3.12)$$

que podemos reescribir como

$$\hat{s}_k = \frac{1}{2} \left(\sum_{n=0}^N \hat{u}_{k-n} \hat{v}_n + \sum_{n=0}^N \hat{u}_{n \pm k} \hat{v}_n \right) \quad 0 \leq k \leq N \quad (2.3.13)$$

donde en el segundo término de la derecha se pide que cuando sea el caso $n - k$ esta diferencia sea mayor o igual a cero. Para encontrar el error de aliación en este término no lineal utilizamos la transformada de Chebyshev discreta (2.2.11) y encontramos que

$$\tilde{s}_k = \frac{2}{M\bar{c}_k} \sum_{j=0}^M \frac{1}{\bar{c}_j} s(y_j) T_k(y_j) \quad (2.3.14)$$

donde utilizamos $M \geq N$ por esto los nuevos puntos $y_j = \cos(\pi j/M)$. El error de aliación que se obtiene por la discretización es

$$\tilde{s}_k = \hat{s}_k + \frac{1}{\bar{c}_k} \sum_{m=1}^{\infty} \hat{s}_{2Mm \pm k} \quad k = 0, \dots, N \quad (2.3.15)$$

y al sustituir (2.3.13) en (2.3.15) tenemos que

$$\tilde{s}_k = \hat{s}_k + \frac{1}{\bar{c}_k} \sum_{m=1}^{\infty} \frac{1}{2} \left(\sum_{n=0}^N \hat{u}_{2Mm \pm k - n} \hat{v}_n + \sum_{n=0}^N \hat{u}_{\pm 2Mm \pm k + n} \hat{v}_n \right) \quad (2.3.16)$$

donde $k, n \leq N$ y para el segundo término entre paréntesis solo se toman en cuenta los coeficientes con $\pm 2Mm \pm k + n \geq 0$. Ahora vamos a pedir que

$$2M \pm k - n > N \quad \pm 2M \pm k + n > N$$

y obtenemos que el peor caso es

$$2M - N - N > N \implies M > \frac{3N}{2}.$$

Tanto como con la base de Fourier y la base de Chebyshev podemos tomar cualquier $M \geq 3N/2$ tal que si filtramos los coeficientes que pueden contribuir al error de aliación de la forma

$$\hat{u}_m = \begin{cases} \hat{u}_m & |m| \leq N/2 \\ 0 & \text{otro caso} \end{cases} \quad (2.3.17)$$

para la base de Fourier y de la forma

$$\hat{u}_m = \begin{cases} \hat{u}_m & m \leq N \\ 0 & \text{otro caso} \end{cases} \quad (2.3.18)$$

para la base de Chebyshev. Con este tipo de filtrado tenemos que el término que define el error de aliación estará compuesto de un coeficiente igual a cero por lo que se elimina el error de aliación, quedándonos así

$$\tilde{s}_k = \hat{s}_k. \quad (2.3.19)$$

Esta claro que se están realizando mucho más operaciones pero lo bueno es que ya no tendremos un error que va creciendo con el tiempo. Cabe mencionar que la regla de los $3/2$ es para una dimensión. Para dos o tres dimensiones la regla cambia a $9/4$ y $27/8$ respectivamente, lo que resulta un costo muy alto para eliminar el error de aliación, por lo que se recomienda buscar otro método para estas dimensiones. Esta regla también cambia para no linealidades diferentes a la cuadrática y hay que ver si el costo no es muy grande, de ser así se recomienda buscar otra.

En la tesis de maestría de Susana Hernández [7] se resuelve con ME la ecuación de Swift-Van Hook, la cual contiene términos no lineales más complicados y elimina el error de aliación.

Capítulo 3

Esquemas Espectrales

El método espectral para resolver ecuaciones diferenciales parciales consiste en buscar soluciones de la forma

$$P_N u(x, t) = \sum_{n=0}^N \hat{u}_n(t) \Phi_n(x)$$

donde Φ_n es una base ortogonal de funciones suaves. La manera de encontrar los coeficientes determina el método espectral. Los más usados son el de Galerkin, el Tau y el de Colocación. Estos métodos son bosquejados brevemente en este capítulo.

3.1. Método de Galerkin

El método de Galerkin es un método donde las funciones de la base se eligen de tal forma que de entrada la solución propuesta satisfaga las condiciones de frontera.

Siguiendo la filosofía de los ME, se busca una solución en forma de una serie truncada de esa base, y como $P_N u$ es una aproximación de la ecuación diferencial a resolver, al sustituirla en ésta, el resultado no será igual a cero. A esta diferencia la llamamos *residuo* que representamos con $R(x, t)$ y es nuevamente una función que depende del espacio y del tiempo. En general, de la ecuación (1.3.1) llegamos a

$$\frac{\partial P_N u}{\partial t} - L(x, t) P_N u(x, t) - f(x, t) = R(x, t) \quad (3.1.1)$$

donde $P_N u$ es la truncación de la expansión de u en la base ortogonal elegida Φ_n ,

esto es $P_N u = \sum_{n=0}^N \hat{u}_n \Phi_n$.

La exactitud de la aproximación dependerá de que tan chico sea el valor del residuo $R(x, t)$. Uno esperaría que el residuo tendiera a cero cuando N tienda a infinito. Como toda función, el residuo puede expandirse en la base elegida y sus coeficientes r_n calcularse con la proyección usual $r_n = (R, \Phi_n)$. Cualquier función queda representada fundamentalmente con los primeros términos de su expansión y el error, como hemos dicho, está en los términos con $|n| > N/2$ para la serie de Fourier y para los polinomios de Chebyshev $n > N$. Por tanto, exigiendo que los primeros términos de la expansión del residuo se anulen, el residuo se minimiza, esto es, se exige que

$$(R, \Phi_n)_\omega = \int R(x, t) \Phi_n(x) \omega(x) dx = 0 \quad (3.1.2)$$

donde $\Phi_n(x)$ es la base ortogonal. Con la ecuación (3.1.2) tendremos las N ecuaciones diferenciales ordinarias para los coeficientes \hat{u}_n y para problemas evolutivos ocuparemos las condiciones iniciales para los coeficientes $\hat{u}_n(t = 0)$, éstos se obtienen de tomar

$$\hat{u}_n(0) = \int_a^b u(x, 0) \Phi_n(x) dx \quad (3.1.3)$$

donde $[a, b]$ depende del intervalo para el cual la base esta definida. Para evaluar los $\hat{u}_n(0)$ podemos utilizar ya sea el método de Simpson o el método de cuadratura para la base correspondiente, para más detalles vea el apéndice A. Y para evolucionar los coeficientes se puede implementar el método de Runge-Kutta que se explica en el apéndice D.

3.2. Método de Tau

El método de Tau es una versión modificada del método de Galerkin. La diferencia radica en la forma en manejar las condiciones de frontera. La expansión se propone de la siguiente forma

$$P_N u(x, t) = \sum_{n=0}^{N+K} \hat{u}_n \Phi_n(x) \quad (3.2.1)$$

donde K es el número de condiciones de frontera $B(x)u(x, t) = 0$. Es en esos últimos K términos donde se incluyen las condiciones de frontera. Específicamente,

con Tau realizamos los mismos pasos hechos con Galerkin solo que la relación entre los coeficientes (3.1.2) será válida para $0 \leq n \leq N$. Las condiciones de frontera nos conducen a ecuaciones algebraicas para los K coeficientes restantes.

Para ejemplificar, tomemos los dos tipos de frontera más conocidos, tipo Dirichlet y tipo Neumann para un problema de una dimensión usando la base de Chebyshev y recordando las identidades (2.2.3) y (2.2.16) tendremos que

$$u(\pm 1, t) = \sum_{n=0}^{N+K} (\pm 1)^n \hat{u}_n = b_{\pm} \quad (\text{Dirichlet}) \quad (3.2.2)$$

$$u'(\pm 1, t) = \sum_{n=0}^{N+K} n^2 (\pm 1)^{n+1} \hat{u}_n = b_{\pm} \quad (\text{Neumann}) \quad (3.2.3)$$

donde b_{\pm} es en general una función del tiempo $b_{\pm}(t)$. Con las condiciones de frontera, ya sea Dirichlet, Neumann o una combinación de estas, obtenemos las K ecuaciones faltantes para obtener los coeficientes y resolver la ecuación diferencial. El número de condiciones de frontera depende de las dimensiones espaciales del problema, es decir para una dimensión serán $K = 2$ para dos dimensiones serán $K = 4$ y en general el número de condiciones serán $K = 2\alpha$ donde α es la dimensión del problema.

3.3. Método de colocación

Partimos de la propuesta de solución como serie truncada

$$P_N u(x) = \sum_{n=0}^N \hat{u}_n \Phi_n(x)$$

y definimos una serie de puntos x_0, x_1, \dots, x_N interiores del dominio que llamamos *puntos de colocación*. Luego se pide *colocar* la función en esos puntos, es decir, se requiere que la ecuación diferencial se satisfaga exactamente en los puntos de colocación, esto es, $R(x_j) = 0$ donde R es el residuo.

Podemos escribir el método de colocación espectral siguiendo la filosofía del método de Galerkin y de Tau de minimizar el residuo a través de (3.1.2), pero en el método de colocación se proyecta el residuo sobre $\delta(x - x_j)$ siendo x_j los puntos de colocación, esto es, se fuerza a que la ecuación se satisfaga en los puntos

de colocación al exigir que

$$(R, \delta(x - x_j)) = \int_a^b R(x)\delta(x - x_j)dx = Res(x_j) = 0 \quad (3.3.1)$$

para $j = 0, 1, 2, \dots, N$. Los índices $j = 0, N$ corresponderían a las condiciones de frontera.

Capítulo 4

Aplicación de los ME en la Física

En este capítulo se realizan ejemplos esenciales para el entendimiento de los ME y sus esquemas explicados en el capítulo anterior. Las ecuaciones que analizaremos son: la ecuación de Helmholtz, la ecuación de calor y la ecuación de Burgers. Cada una nos ayudará a entender tres aspectos de la aplicación de los ME las cuales son: problemas con independencia temporal (Helmholtz), con dependencia temporal (Calor y Burgers) y el manejo de términos no lineales (Burgers). Con estos ejemplos se pretende que el lector tenga las bases suficiente para aplicar los ME para cualquier ecuación diferencial en una dimensión.

4.1. Ecuación de Helmholtz

La ecuación de Helmholtz es una ecuación diferencial parcial que aparece en fenómenos ondulatorios. Esta ecuación se escribe:

$$(-\nabla^2 + \lambda)u(\vec{r}) = f(\vec{r}) \quad (4.1.1)$$

donde λ es una constante y u solo depende del espacio. Para entender su presencia en fenómenos ondulatorios tomemos la ecuación de onda

$$\left(\nabla^2 - \frac{1}{c^2} \frac{\partial^2}{\partial t^2}\right)u(\vec{r}, t) = 0 \quad (4.1.2)$$

y realizamos una separación de variables

$$u(\vec{r}, t) = A(\vec{r})T(t) \quad (4.1.3)$$

llegamos que

$$(\nabla^2 + k^2)A(\vec{r}) = 0, \quad (4.1.4)$$

$$\left(\frac{\partial^2}{\partial t} + k^2 c^2\right)T(t) = 0 \quad (4.1.5)$$

donde (4.1.4) es la forma homogénea de (4.1.1). La ecuación de Helmholtz se ve envuelta en problemas de diferentes áreas tales como la radiación electromagnética, sismología, acústica, entre otras.

4.1.1. Chebyshev colocación

Consideramos la ecuación (4.1.1) para una dimensión con la condición de frontera tipo Dirichlet

$$-\frac{\partial^2 u}{\partial x^2} + \lambda u = f(x) \quad (4.1.6)$$

$$u(\pm 1) = 0, \quad (4.1.7)$$

sustituyendo en (4.1.6) la expansión

$$u_N(x) = \sum_{n=0}^N \hat{u}_n T_n(x) \quad (4.1.8)$$

tenemos

$$-\sum_{n=0}^{N-2} \hat{u}_n \frac{\partial^2}{\partial x^2} T_n(x_j) + \lambda \sum_{n=0}^{N-2} \hat{u}_n T_n(x_j) = f(x_j) \quad (4.1.9)$$

donde $x_j = \cos(\pi j/N)$ con $j = 1, \dots, N-1$. No se consideran los puntos x_0 y x_N ya que son los puntos en la frontera. Las dos ecuaciones para esos puntos las obtendremos de las condiciones de frontera, por lo que de (4.1.7) tendremos

$$u(\pm 1) = \sum_{n=0}^N \hat{u}_n T_n(\pm 1) = \sum_{k=0}^N (\pm 1)^k \hat{u}_n \quad (4.1.10)$$

donde hicimos uso de la propiedad (2.2.3). Con (4.1.9) y (4.1.10) completamos nuestro sistema de ecuaciones para los coeficientes \hat{u}_n y este lo podemos escribir de manera matricial como $Ax = b$, donde

$$A = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ -T_0''(x_1) + \lambda T_0(x_1) & -T_1''(x_1) + \lambda T_1(x_1) & \cdots & -T_N''(x_1) + \lambda T_N(x_1) \\ -T_0''(x_2) + \lambda T_0(x_2) & -T_1''(x_2) + \lambda T_1(x_2) & \cdots & -T_N''(x_2) + \lambda T_N(x_2) \\ \vdots & \vdots & \vdots & \vdots \\ -T_0''(x_{N-1}) + \lambda T_0(x_{N-1}) & -T_1''(x_{N-1}) + \lambda T_1(x_{N-1}) & \cdots & -T_N''(x_{N-1}) + \lambda T_N(x_{N-1}) \\ 1 & -1 & \cdots & (-1)^N \end{pmatrix},$$

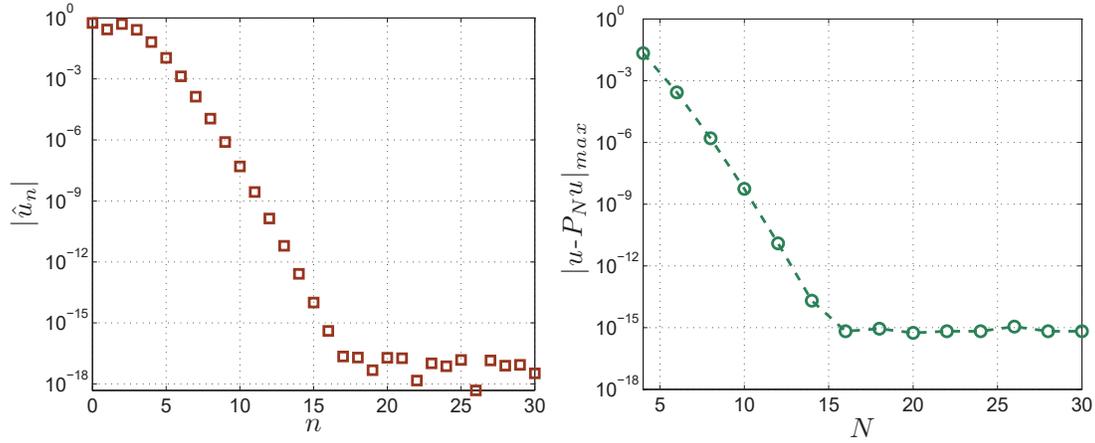


Figura 4.1: Para el método de colocación observamos en la figura izquierda el decaimiento de los coeficientes en función de n para $N = 30$. La figura derecha nos muestra como cae el error de la aproximación con la función exacta con respecto a N .

$$x = \begin{pmatrix} \hat{u}_0 \\ \hat{u}_1 \\ \hat{u}_2 \\ \vdots \\ \hat{u}_{N-1} \\ \hat{u}_N \end{pmatrix}, \quad b = \begin{pmatrix} 0 \\ f(x_1) \\ f(x_2) \\ \vdots \\ f(x_{N-1}) \\ 0 \end{pmatrix}.$$

Para resolver el sistema $Ax = b$ utilizamos el método de descomposición LU que se describe en el apéndice B y obtenemos los coeficientes de nuestra expansión.

En la figura 4.1 se consideró $\lambda = 10^5$ y se tomó que para la función $f(x) = [(\lambda - 1)x^2 + 4x - (\lambda + 1)]e^{-x}$ tenemos que la solución exacta es $u(x) = (x^2 - 1)e^{-x}$. Vemos que los coeficientes dejan de decaer debido a los errores de redondeo de la computadora por lo que el error de la solución también se ve afectada. El error $|u - P_N u|_{max}$ se obtiene por medio de la ecuación (2.2.8). Para la figura 4.1 la ecuación (2.2.8) se integró con el método de cuadratura para la base de Chebyshev.

4.1.2. Chebyshev Tau

Ahora resolveremos las mismas ecuaciones (4.1.6) y (4.1.7) pero en esta ocasión con el esquema de Tau. Buscamos entonces su aproximación con los polinomios

de Chebyshev tomando en cuenta que

$$u_N(x) = \sum_{n=0}^N \hat{u}_n T_n(x), \quad (4.1.11)$$

$$u''_{N-2}(x) = \sum_{n=0}^{N-2} \hat{u}_n^{(2)} T_n(x), \quad (4.1.12)$$

$$f(x) = \sum_{n=0}^N \hat{f}_n T_n(x) \quad (4.1.13)$$

por lo que, al sustituir las ecuaciones (4.1.11)-(4.1.13) en (4.1.6) y utilizando la propiedad de ortogonalidad para minimizar el residuo podemos llegar a:

$$-\hat{u}_n^{(2)} + \lambda \hat{u}_n = \hat{f}_n \quad 0 \leq n \leq N-2 \quad (4.1.14)$$

donde el máximo valor de $n_{max} = N-2$ ya que $u''(x)$ es de orden $N-2$. Las ecuaciones restantes para resolver el sistema las obtenemos de las condiciones de frontera

$$\sum_{n=0}^N \hat{u}_n = 0, \quad \sum_{n=0}^N (-1)^n \hat{u}_n = 0$$

si se considera N como un número par las ecuaciones anteriores se pueden reescribir como

$$\sum_{\substack{n=0 \\ \text{n par}}}^N \hat{u}_n = 0, \quad \sum_{\substack{n=1 \\ \text{n impar}}}^{N-1} \hat{u}_n = 0. \quad (4.1.15)$$

Ahora el propósito es encontrar los coeficientes a partir de las ecuaciones (4.1.14) y (4.1.15) una forma de hacerlo es comenzar a resolver (4.1.14) utilizando la ecuación de recurrencia (2.2.20) para $q=1$ y $q=2$, esto es

$$c_n \hat{u}_n^{(1)} = \hat{u}_{n+2}^{(1)} + 2(n+1) \hat{u}_{n+1} \quad 0 \leq n \leq N-1 \quad (4.1.16)$$

$$c_n \hat{u}_n^{(2)} = \hat{u}_{n+2}^{(2)} + 2(n+1) \hat{u}_{n+1}^{(1)} \quad 0 \leq n \leq N-2 \quad (4.1.17)$$

recordando que $c_0 = 2$ y $c_m = 1$ donde $m \geq 1$. Tomamos la ecuación (4.1.14) y cambiamos los índices de n a $n-1$ y por otro lado de n a $n+1$ quedando

$$\begin{aligned} \hat{u}_{n-1}^{(2)} &= \lambda \hat{u}_{n-1} - \hat{f}_{n-1} & 1 \leq n \leq N-1 \\ \hat{u}_{n+1}^{(2)} &= \lambda \hat{u}_{n+1} - \hat{f}_{n+1} & -1 \leq n \leq N-3 \end{aligned}$$

estas dos ecuaciones las sustituimos en (4.1.17) y llegamos a

$$2n\hat{u}_n^{(1)} = c_{n-1}(\lambda\hat{u}_{n-1} - \hat{f}_{n-1}) - (\lambda\hat{u}_{n+1} - \hat{f}_{n+1}) \quad 1 \leq n \leq N-3 \quad (4.1.18)$$

donde quitamos los términos $\hat{u}_n^{(1)}$ con $n = -1, 0$ porque no están definidos y pedimos los términos $\hat{u}_n^{(1)} = 0$ con $n \geq N-2$. Entonces llegamos a:

$$2n\hat{u}_n^{(1)} = c_{n-1}(\lambda\hat{u}_{n+1} - \hat{f}_{n+1}) - \beta_{n+1}(\lambda\hat{u}_{n+1} - \hat{f}_{n+1}) \quad 1 \leq n \leq N-1 \quad (4.1.19)$$

donde introducimos

$$\beta_n = \begin{cases} 1 & 0 \leq n \leq N-2 \\ 0 & n > N-2 \end{cases}$$

para poder tener coeficientes con $n > N-2$. Ahora, usamos la ecuación (4.1.16) para poner $\hat{u}_n^{(1)}$ en términos de \hat{u}_n . Con la ayuda de (4.1.19), pero acomodando los índices como:

$$\begin{aligned} 2(n-1)\hat{u}_{n-1}^{(1)} &= c_{n-2}(\lambda\hat{u}_{n-2} - \hat{f}_{n-2}) - \beta_n(\lambda\hat{u}_n - \hat{f}_n) & 2 \leq n \leq N \\ 2(n+1)\hat{u}_{n+1}^{(1)} &= c_n(\lambda\hat{u}_n - \hat{f}_n) - \beta_{n+2}(\lambda\hat{u}_{n+2} - \hat{f}_{n+2}) & 0 \leq n \leq N-2 \end{aligned}$$

y sustituyéndolas en (4.1.16) con el índice modificada de n a $n-1$

$$\begin{aligned} 2n\hat{u}_n &= \frac{c_{n-1}}{2(n-1)} \left\{ c_{n-2}(\lambda\hat{u}_{n-2} - \hat{f}_{n-2}) - \beta_n(\lambda\hat{u}_n - \hat{f}_n) \right\} \\ &\quad - \frac{\beta_n}{2(n+1)} \left\{ c_n(\lambda\hat{u}_n - \hat{f}_n) - \beta_{n+2}(\lambda\hat{u}_{n+2} - \hat{f}_{n+2}) \right\} \quad 2 \leq n \leq N \end{aligned}$$

donde se volvió a agregar β_n para poder ajustar los índices.

Observamos que en esta última ecuación c_{n-1} , c_n siempre son igual a 1 pues $n = 2, \dots, N$. También veamos que $\beta_n\beta_{n+2} = \beta_{n+2}$, por lo que obtenemos:

$$\begin{aligned} 2n\hat{u}_n &= \frac{1}{2(n-1)} \left[c_{n-2}(\lambda\hat{u}_{n-2} - \hat{f}_{n-2}) - \beta_n(\lambda\hat{u}_n - \hat{f}_n) \right] \\ &\quad - \frac{\beta_n}{2(n+1)}(\lambda\hat{u}_n - \hat{f}_n) + \frac{\beta_{n+2}}{2(n+1)}(\lambda\hat{u}_{n+2} - \hat{f}_{n+2}), \quad (4.1.20) \end{aligned}$$

donde $2 \leq n \leq N$. Despejando los coeficientes y simplificando

$$\begin{aligned} &-\frac{\lambda c_{n-2}}{4n(n-1)}\hat{u}_{n-2} + \left[1 + \frac{\lambda\beta_n}{2(n^2-1)} \right] \hat{u}_n - \frac{\lambda\beta_{n+2}}{4n(n+1)}\hat{u}_{n+2} \\ &= -\frac{c_{n-2}}{4n(n-1)}\hat{f}_{n-2} + \frac{\beta_n}{2(n^2-1)}\hat{f}_n - \frac{\beta_{n+2}}{4n(n+1)}\hat{f}_{n+2}, \quad (4.1.21) \end{aligned}$$

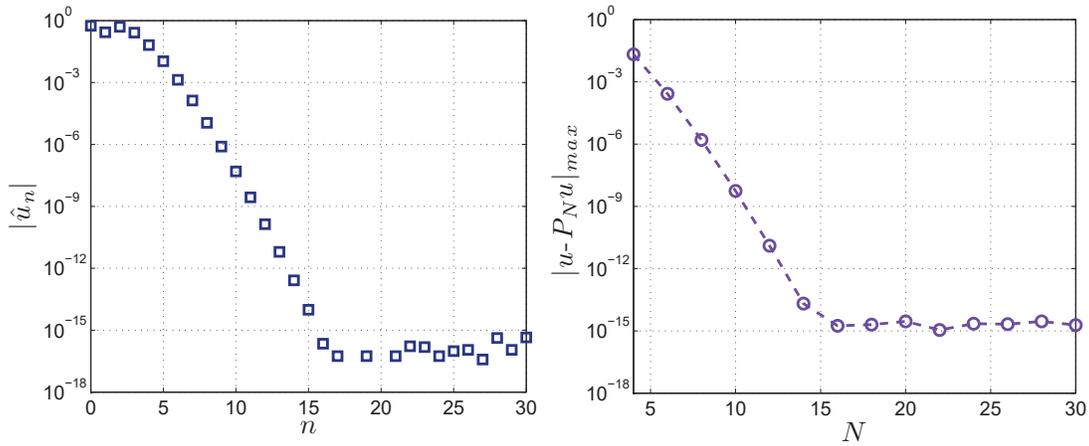


Figura 4.2: Para el método de Tau observamos en la figura izquierda el decaimiento de los coeficientes en función de n para $N = 30$. La figura derecha nos muestra como cae el error que existe de la aproximación con la función exacta con respecto a N .

con $2 \leq n \leq N$. En el libro de Canuto [4] esta relación tiene errores tipográficos y en la expresión (4.1.21) fueron corregidos.

Si ponemos el sistema de ecuaciones (4.1.21) de manera matricial, donde los términos pares forman una matriz y los impares otra, tenemos un arreglo de la forma:

$$\begin{pmatrix} 1 & 1 & 1 & 1 & \cdots & 1 & 1 \\ p_1 & q_1 & r_1 & & & & \\ & p_2 & q_2 & r_2 & & & 0 \\ & & \ddots & \ddots & \ddots & & \\ 0 & & & p_{m-1} & q_{m-1} & r_{m-1} & \\ & & & & p_m & q_m & \end{pmatrix} \begin{pmatrix} \omega_0 \\ \omega_1 \\ \omega_2 \\ \vdots \\ \omega_{m-1} \\ \omega_m \end{pmatrix} = \begin{pmatrix} F_0 \\ F_1 \\ F_2 \\ \vdots \\ F_{m-1} \\ F_m \end{pmatrix} \quad (4.1.22)$$

donde ω_m representa ya sea \hat{u}_{2n} o \hat{u}_{2n+1} . El sistema (4.1.22) se puede resolver de la manera descrita en el apéndice C y con esto obtenemos la solución de la ecuación de Helmholtz.

Para encontrar la solución con el método de Tau, al igual que con el método de colocación se tomó la función $f(x) = [(\lambda - 1)x^2 + 4x - (\lambda + 1)]e^{-x}$ donde se obtiene que la función solución es $u(x) = (x^2 - 1)e^{-x}$. La exactitud que obtenemos con el método de Tau la podemos observar en la Fig. 4.2 donde a partir de $N = 16$ el error es cercano al orden de 10^{-15} . La ventaja de haber resuelto la ecuación

de Helmholtz con el método de Tau que tiene respecto al método de colocación es que el número de operaciones es menor, por lo tanto, el tiempo en resolverse también lo fue.

4.2. Ecuación de Calor

Una ecuación diferencial parcial importante es la ecuación de calor que fue propuesta por Fourier en 1807 en su memoria de propagación del calor en los cuerpos sólidos. Esta ecuación se derivada de la ley de Fourier y la conservación de la energía, donde la ley de Fourier dice que el flujo de la energía térmica a través de una superficie es proporcional al gradiente negativo de temperatura a través de la misma superficie.

La ecuación de calor describe la distribución de calor (o variación de temperatura) en una región a través del tiempo. Esta se escribe de manera general

$$\frac{\partial u}{\partial t} = \nu \nabla^2 u \quad (4.2.1)$$

donde ν es el coeficiente de difusión térmica.

Además, la ecuación de calor está relacionada con el estudio del movimiento Browniano a través de la teoría de probabilidades, entre otras cosas como la difusión química, etc.

4.2.1. Chebyshev Tau

Tomemos la ecuación de calor descrita por (4.2.1) en una dimensión, esto es

$$\frac{\partial u}{\partial t} = \nu \frac{\partial^2 u}{\partial x^2} \quad (4.2.2)$$

que al aproximarla con polinomios de Chebyshev obtenemos un residuo, tal que

$$R(x, t) = \sum_{n=0}^{N-2} (\dot{\hat{u}}_n(t) - \nu \hat{u}_n^{(2)}(t)) T_n(x).$$

Al minimizar este residuo con (3.1.2) llegamos a

$$\dot{\hat{u}}_n(t) = \nu \hat{u}_n^{(2)}(t) \quad 0 \leq n \leq N - 2 \quad (4.2.3)$$

donde para obtener $\hat{u}_n^{(2)}$ en términos de \hat{u}_k hacemos uso de la relación de recurrencia

$$c_n \hat{u}_n^{(1)} = \hat{u}_{n+2}^{(1)} + 2(n+1)\hat{u}_{n+1} \quad 0 \leq n \leq N-1 \quad (4.2.4)$$

$$c_n \hat{u}_n^{(2)} = \hat{u}_{n+2}^{(2)} + 2(n+1)\hat{u}_{n+1}^{(1)} \quad 0 \leq n \leq N-2. \quad (4.2.5)$$

Faltan dos ecuaciones para resolver el sistema de ecuaciones y estas las obtenemos de las condiciones de frontera. Las condiciones de frontera que proponemos son

$$\frac{\partial u}{\partial x}(-1, t) - u(-1, t) = 0 \quad \frac{\partial u}{\partial x}(1, t) + u(1, t) = 0 \quad (4.2.6)$$

donde al hacer la expansión con los polinomios de Chebyshev y usar las identidades (2.2.3) y (2.2.16) llegamos a

$$\sum_{n=0}^N (-1)^n (2n^2 + 1) \hat{u}_n = 0, \quad \sum_{n=0}^N (2n^2 + 1) \hat{u}_n = 0$$

que al considerar N como un número par se pueden reescribir como

$$\hat{u}_N = -\frac{1}{2N^2 + 1} \sum_{\substack{n=0 \\ n \text{ par}}}^{N-2} (2n^2 + 1) \hat{u}_n, \quad (4.2.7)$$

$$\hat{u}_{N-1} = -\frac{1}{2(N-1)^2 + 1} \sum_{\substack{n=1 \\ n \text{ impar}}}^{N-3} (2n^2 + 1) \hat{u}_n. \quad (4.2.8)$$

Con las ecuaciones (4.2.3), (4.2.7) y (4.2.8) tenemos los coeficientes para cualquier tiempo y para evolucionar la ecuación en el tiempo, necesitamos los coeficientes al tiempo inicial por lo que usamos la condición inicial

$$P_N u(x, 0) = \sum_{n=0}^N \hat{u}_n(0) T_n(x) = g(x) \quad (4.2.9)$$

donde al recordar (2.2.10) obtenemos los coeficientes como

$$\hat{u}_k(0) = \frac{2}{\pi c_k} \int_{-1}^1 \frac{g(x) T_k(x)}{\sqrt{1-x^2}} dx. \quad (4.2.10)$$

La condición inicial que se elige para este problema es $g(x) = 1$. Para esta condición inicial si se toma $\hat{u}_0 = 1$ y los demás coeficientes iguales a cero se cumple la condición inicial sin la necesidad de resolver (4.2.10).

Existen varios métodos con los cuales podemos evolucionar los coeficientes en el tiempo pero el que usaremos en esta tesis es el método de Runge-Kutta de orden 4. En la Fig. 4.4 se ponen las gráficas de la solución a varios tiempos y el error de la aproximación respecto a N en diferentes tiempos, con esta última gráfica podemos observar que la exactitud de nuestra función solución no crece a cada paso evolutivo, por lo que nos da la seguridad de que nuestra aproximación tendrá estabilidad temporal, esto es, el error no crece con el tiempo.

La solución exacta para la ecuación de calor, con condición inicial $g(x) = 1$ y condiciones de frontera (4.2.7) y (4.2.8) que se uso para comparar la solución aproximada es

$$u(x, t) = 4 \sum_{n=1}^{\infty} \frac{\sec \alpha_n}{3 + 4\alpha_n^2} e^{-4\nu\alpha_n^2 t} \cos \alpha_n x \quad (4.2.11)$$

donde α_n son las raíces positivas de

$$\alpha \tan \alpha = \frac{1}{2}.$$

El método usado para encontrar las raíces fue el de Newton-Raphson el cual se puede encontrar en cualquier libro de Métodos Numéricos.

4.2.2. Fourier Galerkin

Al igual que en el método de Tau tomamos

$$\frac{\partial u}{\partial t} = \nu \frac{\partial^2 u}{\partial x^2} \quad (4.2.12)$$

y sustituimos la expansión (2.1.2) de u consiguiendo la siguiente relación

$$R(x, t) = \sum_{n=-N/2}^{N/2} (\dot{\hat{u}}_n - \hat{u}_n^{(2)}) e^{inx} \quad (4.2.13)$$

que al minimizar con (3.1.2) llegamos a

$$\dot{\hat{u}}_n(t) = \nu \hat{u}_n^{(2)}(t) \quad 0 \leq n \leq N - 2. \quad (4.2.14)$$

Recordemos la relación (2.1.19) y al tomar en cuenta que las condiciones de frontera son periódicas tenemos

$$\dot{\hat{u}}_n(t) = -n^2 \nu \hat{u}_n(t) \quad 0 \leq n \leq N \quad (4.2.15)$$

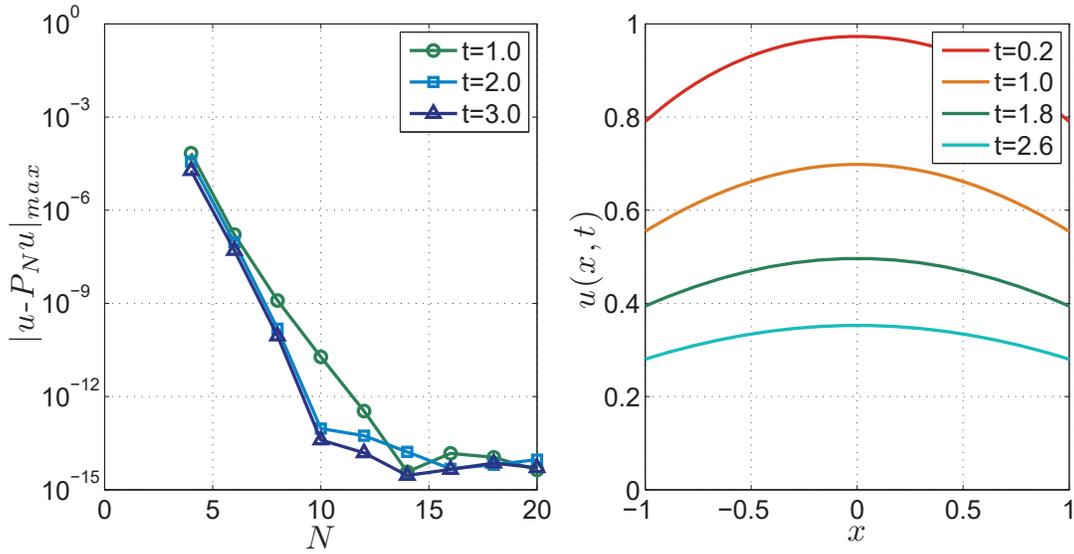


Figura 4.3: La gráfica derecha representa la función $u(x, t)$ que nos indica la temperatura en una barra metálica delgada a lo largo de su longitud. Para tal gráfica se usó $N = 30$. La gráfica izquierda muestra el error máximo de la función respecto a N a los tiempos $t = 1,0$, $t = 2,0$ y $t = 3,0$.

Con esta relación tenemos nuestro sistema de ecuaciones casi completo, solo nos falta obtener los coeficientes al tiempo $t = 0$ que nos lo da la condición inicial

$$u(x, 0) = u_0(x). \quad (4.2.16)$$

De ahí podemos aplicar el método de Runge-Kutta o algún método que nos permita obtener los coeficientes a cualquier tiempo.

La solución analítica que usamos para comparar la exactitud del método Fourier Galerkin es

$$u_b(x, t) = \frac{1}{\sqrt{4\pi\nu t}} \sum_{n=-\infty}^{\infty} e^{-(x-2\pi n)^2/4\pi t} \quad (4.2.17)$$

donde también es solución

$$u(x, t) = c + u_b(x - ct, t + t_0) \quad (4.2.18)$$

de la cual para el tiempo $t = 0$ con $t_0 = 0,1$ y $c = 0$ sacamos la condición inicial de nuestra aproximación.

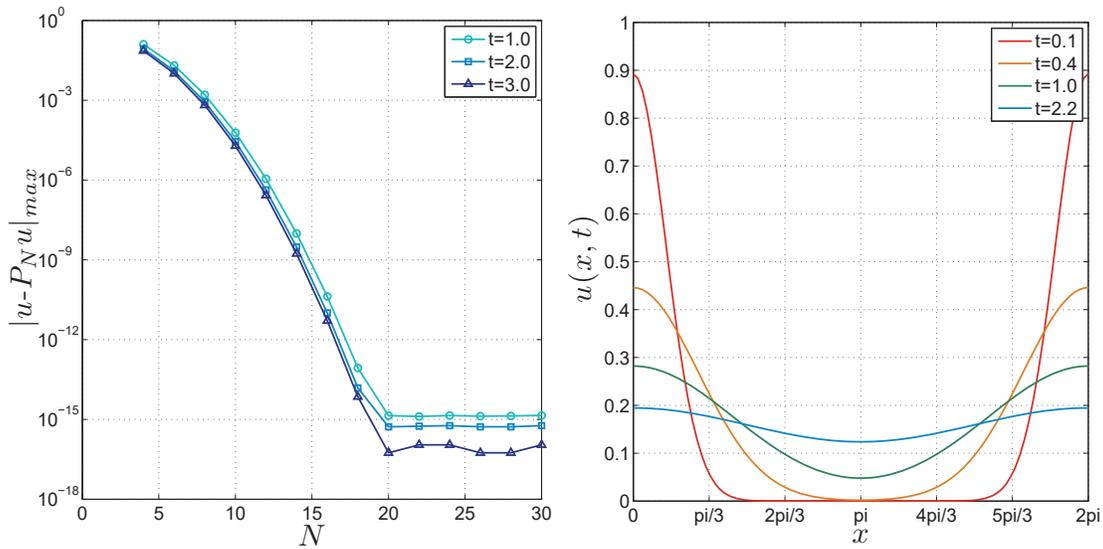


Figura 4.4: La gráfica derecha representa la función $u(x, t)$ que nos indica la temperatura de una barra metálica delgada a lo largo de su longitud, para esta gráfica se uso $N = 30$. La gráfica izquierda muestra el error máximo de la función respecto a N a los tiempos $t = 1$, $t = 2$ y $t = 3$.

En la Figura 4.4 vemos la gráfica del error de la aproximación para diferentes N (izquierda) y la gráfica de la función (derecha). En la gráfica del error respecto a N esta para tres tiempos y observamos que no aumenta, indicándonos que la evolución temporal no afecta la exactitud de la aproximación.

4.3. Ecuación de Burgers

La ecuación de Burgers fue propuesta por el holandés Johannes M. Burgers como una simplificación del modelo de turbulencia de Navier Stokes. Esta ecuación se escribe como

$$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} - \nu \frac{\partial^2 u}{\partial x^2} = 0 \quad x \in \Omega \quad (4.3.1)$$

donde Ω es el dominio espacial, ν es una constante positiva que podemos considerarla como la viscosidad del fluido y u como la velocidad de este. Además, la ecuación (4.3.1) se complementa con una condición inicial

$$u(x, 0) = u_0(x) \quad (4.3.2)$$

y con las condiciones de frontera.

Existen varias maneras de tratar el término no lineal, para el esquema de Tau o Galerkin podemos hacer uso del teorema de convolución o también podemos utilizar las transformadas discretas y sus inversas. La segunda forma consiste en pasar los coeficientes tanto de u como de u_x al espacio físico mediante la transformada discreta inversa y realizamos el producto uu_x en el espacio físico. Después con la transformada discreta pasamos del espacio físico al espacio espectral. Al haber aplicado la transformada discreta añadimos el error de aliación, por lo que, si nos interesa tener mayor exactitud aplicamos alguna técnica como la regla de los 3/2 explicada en la sección 2.3. Para el esquema de Colocación trabajaremos en el espacio físico por lo que se hace el producto directamente sin añadir el error de aliación.

4.3.1. Chebyshev Tau

Al aplicar la aproximación con los polinomios de Chebyshev a la ecuación (4.3.1) tendremos que hay un residuo

$$R(x, t) = \frac{\partial u_N}{\partial t} + u_N \frac{\partial u_N}{\partial x} - \nu \frac{\partial^2 u_N}{\partial x^2} \quad (4.3.3)$$

$$= \sum_{n=0}^{N-2} (\dot{\hat{u}}_n - \hat{P}_n - \nu \hat{u}_n^{(2)}) T_n(x) \quad (4.3.4)$$

Donde \hat{P}_n son los coeficientes del término no lineal $(-u_N \partial_x u_N)$. Este residuo, deseamos minimizarlo por lo que pedimos que $R(x, t)$ sea ortogonal al espacio de los polinomios de Chebyshev y usando la propiedad de ortogonalidad tenemos

$$\int_{-1}^1 \frac{R(x, t) T_k(x)}{\sqrt{1-x^2}} dx = \sum_{n=0}^{N-2} (\dot{\hat{u}}_n + \hat{P}_n - \nu \hat{u}_n^{(2)}) \int_{-1}^1 \frac{T_n(x) T_k(x)}{\sqrt{1-x^2}} dx \quad (4.3.5)$$

entonces, resulta que al minimizar el residuo tendremos que

$$\dot{\hat{u}}_k = \hat{P}_k + \nu \hat{u}_k^{(2)}, \quad 0 \leq k \leq N-2 \quad (4.3.6)$$

Las otras dos ecuaciones faltantes se obtienen a partir de las condiciones de frontera. Digamos que nuestras condiciones de frontera sean $u(\pm 1, t) = 0$, entonces tendremos lo mismo que en la ecuación (4.2.6), esto es:

$$\hat{u}_N = - \sum_{\substack{n=0 \\ \text{n par}}}^{N-2} \hat{u}_n, \quad \hat{u}_{N-1} = - \sum_{\substack{n=1 \\ \text{n impar}}}^{N-3} \hat{u}_n \quad (4.3.7)$$

Para resolver el sistema (4.3.6) proponemos para el término no lineal la fórmula de Adams-Bashforth y para el término lineal el esquema Crank-Nicolson por lo que tenemos:

$$\begin{aligned} \frac{\hat{u}_n(t + \Delta t) - \hat{u}_n(t)}{\Delta t} &= \frac{1}{2}(\hat{P}_n(t - \Delta t) - 3\hat{P}_n(t)) \\ &+ \frac{\nu}{2}(\hat{u}_n^{(2)}(t) + \hat{u}_n^{(2)}(t + \Delta t)) \end{aligned} \quad (4.3.8)$$

despejando los términos con tiempo $(t + \Delta t)$ llegamos a

$$-\hat{u}_n^{(2)}(t + \Delta t) + \frac{2}{\nu\Delta t}\hat{u}_n(t + \Delta t) = f_n(\hat{u}_n, \hat{P}_n, t, t - \Delta t) \quad (4.3.9)$$

donde $0 \leq n \leq N - 2$ y la función f_n esta descrita por

$$f_n = \frac{1}{\nu} \left[\hat{P}_n(t - \Delta t) - 3\hat{P}_n(t) \right] + \hat{u}_n^{(2)}(t) + \frac{2}{\nu\Delta t}\hat{u}_n(t) \quad (4.3.10)$$

Si en (4.3.10) ponemos que $\frac{2}{\nu\Delta t} = \lambda$ habremos llegado a la ecuación (4.1.14), por lo que podemos resolver nuestro sistema en cada paso a partir de la solución de la ecuación de Helmholtz. En la Figura 4.5 observamos lo dicho en la Sección 3.3 acerca del aumento en los coeficientes cuando usamos la TCD y también vemos el cambio que genera en el error. La función solución exacta con la que comparamos nuestra aproximación es

$$u(x, t) = -\frac{\int_{-\infty}^{\infty} \sin[\pi(x - \eta)]f(x - \eta)e^{-\eta^2/4\nu t}d\eta}{\int_{-\infty}^{\infty} f(x - \eta)e^{-\eta^2/4\nu t}d\eta} \quad (4.3.11)$$

donde

$$f(x - \eta) = e^{-\cos(\pi(x-\eta))/2\pi\nu}.$$

También se puede resolver la ecuación de Burgers con el esquema de Tau sin tener que usar el método de Crank-Nicolson y el de Adam-Bashforth para recurrir a la solución de la ecuación de Helmholtz. Esto se hace partiendo de (4.3.6) en la cual podemos utilizar el método de Runge-Kutta y tratar los coeficientes del término no lineal como se explico anteriormente.

4.3.2. Fourier colocación

Para resolver con el método de colocación con fronteras periódicas, la ecuación de Burgers se escribe como

$$\frac{\partial \vec{u}}{\partial t} = -\vec{u} \boxtimes D_N^{(1)}\vec{u} + \nu D_N^{(2)}\vec{u} \quad (4.3.12)$$

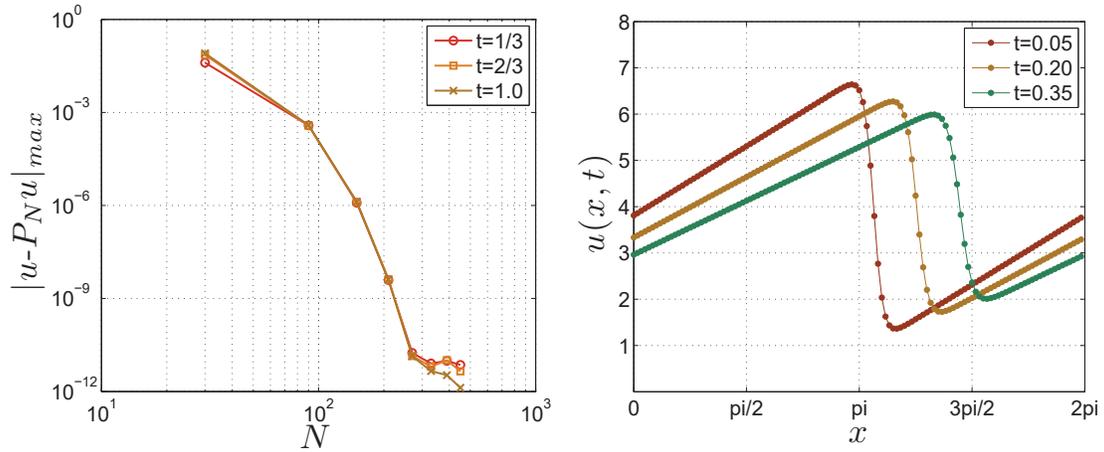


Figura 4.5: En la gráfica de la izquierda tenemos a los coeficientes obtenidos usando la TCD y a los coeficientes usando la TCD y la regla de los 3/2 para $t=1$ y $N=128$. En la gráfica de la derecha se encuentra la comparación del error $|u - P_N u|_{max}$ respecto a N .

donde $D_N^{(1)}$ y $D_N^{(2)}$ son las matrices de diferenciación (2.1.28) y (2.1.29) respectivamente, $\vec{u} = (u(x_0, t), u(x_1, t), \dots, u(x_N, t))^T$ y donde el producto

$$\vec{u} \boxtimes \vec{v} = \begin{pmatrix} u_0 v_0 \\ u_1 v_1 \\ \vdots \\ u_N v_N \end{pmatrix}. \quad (4.3.13)$$

La ecuación de Burgers (4.3.12) es de la forma $\dot{y} = f(y, t)$ por lo que podemos aplicar el método de Runge-Kutta para evolucionar la ecuación diferencial.

Para poder comparar la aproximación es necesario tener una solución analítica. Hopf y Cole mostraron que la transformación

$$u = -2\nu \frac{1}{\phi} \frac{\partial \phi}{\partial x} \quad (4.3.14)$$

reduce la ecuación de Burgers a la ecuación de calor

$$\frac{\partial \phi}{\partial t} - \nu \frac{\partial^2 \phi}{\partial x^2} = 0 \quad (4.3.15)$$

y debido a esta relación podemos obtener soluciones exactas de forma más sencilla, por ejemplo, de la solución analítica usada en la sección 4.2.2 tendremos

$$\phi_b(x, t) = \frac{1}{\sqrt{4\pi\nu t}} \sum_{n=-\infty}^{\infty} e^{-(x-2\pi n)^2/4\pi t} \quad (4.3.16)$$

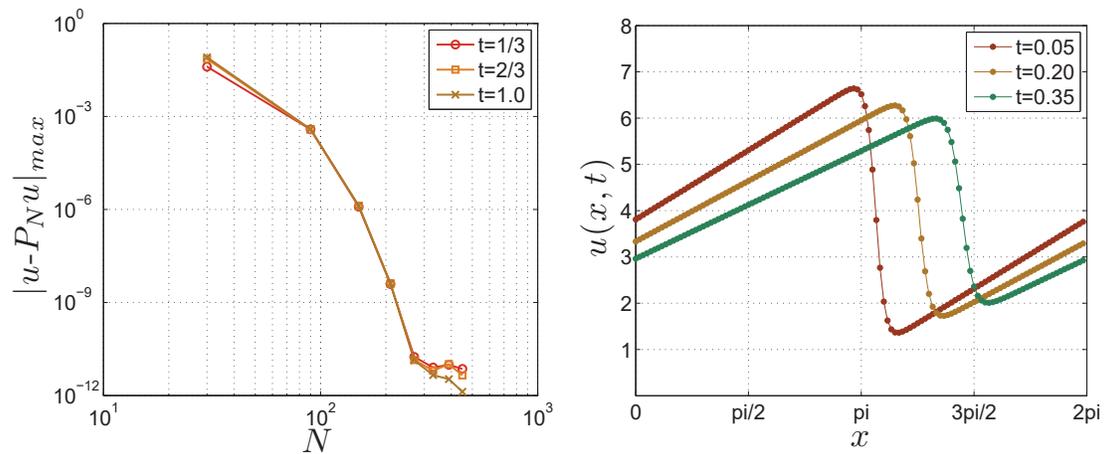


Figura 4.6: La gráfica de la izquierda nos muestra el error máximo entre la solución analítica y la solución aproximada con el método de Fourier Colocación. La gráfica de la derecha nos muestra como es el comportamiento de la ecuación de Burgers en diferentes tiempos donde se tomó $N = 30$ y $\nu = 0,01$.

donde nuestra función solución puede escribirse como

$$u(x, t) = c + u_b(x - ct, t + t_0). \quad (4.3.17)$$

La Figura 4.6 nos muestra la exactitud del método de Colocación donde utilizamos para la solución analítica con $-50 \leq n \leq 50$, $t_0 = 1,0$ $c = 4,0$. Los paso de tiempo utilizados en el fueron de $dt = 10^{-5}$.

Capítulo 5

Conclusiones

1. El propósito de esta tesis es hacer un bosquejo de los métodos espectrales, tal que si existe algún interesado en aprender estos métodos puedan recurrir a esta tesis y tener una idea más clara de su aplicación dado que la bibliografía de los métodos espectrales son difíciles de entender en una primera lectura.
2. Se estudiaron la base de Fourier y la base de Chebyshev para hacer las expansiones de nuestra función solución para problemas de condiciones de frontera periódicas y no periódicas, respectivamente.
3. También se realizó un análisis de el error de aliación para términos no lineales donde nos encontramos que el error aumenta a cada paso de tiempo y utilizamos la regla de los $3/2$ para evitar dicho aumento.
4. Se aplicaron los ME a las ecuaciones de Helmholtz, Calor y Burgers en las cuales comprobamos que la exactitud es muy grande en dichas ecuaciones.
5. Se hizo una comparación del error tomando el error de aliación y cuando se quita mediante la regla de los $3/2$, y se encontró que para el tiempo $t=0.1$ la diferencia es casi imperceptible.
6. En particular, el aprender métodos espectrales será para mi una herramienta muy útil para el desarrollo de mis estudios posteriores dado que mi interés es el control de sistemas dinámicos, en especial en el área de fluidos.

Algunos temas como continuación de esta tesis serían:

1. Utilizar condiciones de frontera semi-infinitas con funciones de Chebyshev racionales $TL_n(x)$ o las funciones de Laguerre, también considerar las con-

diciones de frontera infinitas con las funciones de Chebyshev racionales $TB_n(x)$, las funciones de Hermite o funciones Sinc.

2. Mostrar como implementar los ME en problemas de dos o más dimensiones para los diferentes esquemas.
3. Mostrar como implementar los ME en diferentes geometrías a las cartesianas, por ejemplo, geometrías cilíndricas, esféricas, etc.

Apéndice A

Cuadratura Gauss-Lobatto y puntos de colocación

Los operadores de truncación de Fourier y Chebyshev requieren que evaluemos integrales para resolver proyecciones ortogonales. Pero sabemos que resolver integrales analíticamente no es una tarea fácil. Por esta razón las integrales son comúnmente aproximadas con cuadratura. Esta cuadratura debe de conservar las propiedades de ortogonalidad y de exactitud espectral.

La regla de cuadratura esta dada por

$$\int_a^b f(x)\omega(x)dx \approx Q[f] + E = \sum_{j=0}^N f(x_j)\omega_j + E \quad (\text{A.1})$$

donde los $\{x_j\}$ son los puntos de colocación y los ω_j los pesos correspondientes a cada punto x_j . Para que la propiedad de ortogonalidad se cumpla es necesario que

$$\int_a^b \Phi_n(x)\Phi_m(x)^*dx = \sum_{j=0}^N \Phi_n(x_j)\Phi_m(x_j)^*\omega_j = C\delta_{n,m} \quad (\text{A.2})$$

donde $\Phi(x)$ puede ser tanto la base de Fourier como de Chebyshev y C es una constante.

Para la base de Fourier tendremos que

$$\int_0^{2\pi} e^{inx}e^{-imx}dx = \sum_{j=0}^N e^{in(x_j)}e^{-im(x_j)}\omega_j = 2\pi\delta_{n,m} \quad (\text{A.3})$$

donde si consideramos $x_j = 2\pi j/N$ como los puntos de la red dividida en N partes iguales junto con la relación

$$\sum_{j=0}^{N-1} e^{2il\pi/N} = N\delta_{l,\pm pN} \quad p = 0, \pm 1, \pm 2, \dots \quad (\text{A.4})$$

llegamos a

$$\sum_{j=0}^{N-1} e^{i(n-m)x_j} \omega_j = 2\pi \delta_{n-m,\pm pN} \quad (\text{A.5})$$

donde el valor de la base evaluada en x_0 es igual cuando se hace en x_N . Usamos (A.4) en (A.5) para obtener que $\omega_j = 2\pi/N$. Con esto tenemos la formula de la cuadratura de Fourier dada por

$$Q_F = \frac{2\pi}{N} \sum_{j=0}^{N-1} f(x_j). \quad (\text{A.6})$$

De la misma forma podemos obtener la formula de cuadratura para la base de Chebyshev

$$\int_{-1}^1 T_k(x)T_l(x)\omega(x)dx = \sum_{j=0}^N T_k(x_j)T_l(x_j)\omega_j = \frac{\pi}{2}c_k\delta_{k,l} \quad (\text{A.7})$$

donde $c_0 = 2$ y $c_k = 1$ para $1 \leq k \leq N$. Si consideramos los puntos de colocación $x_j = \cos \pi j/N$ (puntos Gauss-Lobatto) y tomamos la relación

$$\sum_{j=0}^N \frac{1}{\bar{c}_j} T_k(x_j)T_l(x_j) = \frac{N}{2}\bar{c}_k\delta_{k,l} \quad (\text{A.8})$$

donde $\bar{c}_0 = \bar{c}_N = 2$ y $\bar{c}_k = 1$ para $1 \leq k \leq N-1$. Por lo que al combinar las ecuaciones (A.7) y (A.8) llegamos a que los pesos $\omega_j = \pi/\bar{c}_j N$. Esto nos da que la formula de cuadratura de Chebyshev con los puntos Gauss-Lobatto es

$$Q_{CGL} = \frac{\pi}{N} \sum_{j=0}^N \frac{1}{\bar{c}_j} f(x_j). \quad (\text{A.9})$$

Existen otros conjuntos de puntos para los polinomios de Chebyshev, estos son

$$x_j = \cos \frac{(2j-1)\pi}{2N+2} \quad \omega_j = \frac{\pi}{N+1} \quad (\text{A.10})$$

$$x_j = \cos \frac{2j\pi}{2N+1} \quad \omega_j = \begin{cases} \frac{\pi}{2N+1} & j = 0 \\ \frac{\pi}{N+1} & j = 1, \dots, N \end{cases} \quad (\text{A.11})$$

donde (A.10) son los puntos de Gauss y (A.11) son los puntos de Gauss-Radau con sus respectivos pesos ω_j .

En esta tesis se utilizaron los puntos de Gauss-Lobatto los cuales son los puntos extremos de los polinomios de Chebyshev, esto es, cuando $x = \pm 1$ y los puntos de Gauss son los puntos ceros de los polinomios de Chebyshev. Se pueden escoger que los puntos sean equidistantes como se hizo con Fourier pero ocurre el fenómeno de Runge [11] que genera oscilaciones cerca de las fronteras y el error se hace mayor cuando se aumenta el grado del polinomio. La forma de minimizar estas oscilaciones es utilizando cualquiera de los tres conjuntos de puntos ya antes mencionados.

Apéndice B

Método de descomposición LU

La descomposición LU es una descomposición de una matriz $A_{N \times N}$ (no se confunda con la N de truncación) en un producto de dos matrices, una matriz triangular inferior L y una matriz triangular superior U . Esta descomposición se utiliza para resolver sistemas de ecuaciones lineales $Ax = b$ con un costo menor de número de operaciones que con la eliminación gaussiana, por lo que, nos ahorra tiempo. También se utiliza para obtener más rápido la matriz inversa y su determinante.

Para explicarlo tomamos la descomposición de una matriz $A_{3 \times 3}$

$$\begin{pmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{pmatrix} \begin{pmatrix} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

donde tendremos que

$$\begin{pmatrix} l_{11}u_{11} & l_{11}u_{12} & l_{11}u_{13} \\ l_{21}u_{11} & l_{21}u_{12} + l_{22}u_{22} & l_{21}u_{13} + l_{22}u_{23} \\ l_{31}u_{11} & l_{31}u_{12} + l_{32}u_{22} & l_{31}u_{13} + l_{32}u_{23} + l_{33}u_{33} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

la cual nos da 12 variables desconocidas para un sistema de 9 ecuaciones. Debido a esto es necesario poner algunas restricciones en L y U . Por ejemplo podemos pedir que los $l_{ii} = 1$ y con esta forma el sistema de ecuaciones ya tiene solución. En general tendremos tres tipos de ecuaciones que describen la matriz LU

$$l_{i1}u_{1j} + l_{i2}u_{2j} + \cdots + l_{ii}u_{ij} = a_{ij} \quad i < j \quad (\text{B.1})$$

$$l_{i1}u_{1j} + l_{i2}u_{2j} + \cdots + l_{ii}u_{jj} = a_{ij} \quad i = j \quad (\text{B.2})$$

$$l_{i1}u_{1j} + l_{i2}u_{2j} + \cdots + l_{ij}u_{jj} = a_{ij} \quad i > j \quad (\text{B.3})$$

donde para obtener los l_{ij} y u_{ij} hacemos los siguientes pasos

1. Primero, haces $l_{ii} = 1$ para $i = 1, \dots, N$ dado que en general tendremos $N^2 + N$ variables desconocidas y al restringir los l_{ii} tendremos N^2 variables desconocidas para N^2 ecuaciones.
2. Para cada $j = 1, 2, \dots, N$ haces los siguientes dos pasos:

- a) Para $i = 1, 2, \dots, j$ obtienes los u_{ij} a partir de (B.1), (B.2) y los l_{ii} , esto es

$$u_{ij} = a_{ij} - \sum_{k=1}^{i-1} l_{ik} u_{kj}. \quad (\text{B.4})$$

- b) Segundo, para $i = j + 1, j + 2, \dots, N$ se usa (B.3) para obtener l_{ij} de la forma

$$l_{ij} = \frac{1}{u_{jj}} \left(a_{ij} - \sum_{k=1}^{j-1} l_{ik} u_{kj} \right). \quad (\text{B.5})$$

Podemos observar que cada a_{ij} solo es usado una vez, por lo que, tanto l_{ij} como u_{ij} puede ser colocado en el elemento de A que ya ha sido utilizado.

Al tener ya las matrices L y U , sólo nos queda resolver

$$Ax = (LU)x = L(Ux) = Ly = b \quad (\text{B.6})$$

donde primero resolvemos $Ly = b$ para obtener y que se obtiene a partir de

$$y_1 = \frac{b_1}{l_{11}} \quad (\text{B.7})$$

$$y_i = \frac{1}{l_{ii}} \left(b_i - \sum_{j=1}^{i-1} l_{ij} y_j \right) \quad (\text{B.8})$$

para $i = 2, \dots, N$. Luego resolvemos $Ux = y$ para obtener x de la forma

$$x_N = \frac{y_N}{u_{NN}} \quad (\text{B.9})$$

$$x_i = \frac{1}{u_{ii}} \left(y_i - \sum_{j=i+1}^N u_{ij} x_j \right) \quad (\text{B.10})$$

para $i = N - 1, \dots, 1$, esto es, primero obtenemos x_N , luego x_{N-1} así sucesivamente hasta x_1 .

Apéndice C

Matriz tridiagonal con trenza

Mostramos un procedimiento para resolver un sistema lineal $MW = F$ donde la matriz M es de tamaño $(n + 1) \times (n + 1)$, tridiagonal y con una trenza (una fila con entradas diferentes a cero, c_n 's), W y F son vectores de dimensión $n + 1$, por lo que, el sistema tiene la forma

$$MW = F \rightarrow \begin{pmatrix} c_0 & c_1 & c_2 & c_3 & \cdots & c_{n-1} & c_n \\ p_1 & q_1 & r_1 & & & & \\ & p_2 & q_2 & r_2 & & & 0 \\ & & \ddots & \ddots & \ddots & & \\ 0 & & & p_{n-1} & q_{n-1} & r_{n-1} & \\ & & & & p_n & q_n & \end{pmatrix} \begin{pmatrix} \omega_0 \\ \omega_1 \\ \omega_2 \\ \vdots \\ \omega_{n-1} \\ \omega_n \end{pmatrix} = \begin{pmatrix} F_0 \\ F_1 \\ F_2 \\ \vdots \\ F_{n-1} \\ F_n \end{pmatrix}$$

que se puede escribir mejor de la siguiente manera

$$c_0\omega_0 + c_1\omega_1 + \cdots + c_n\omega_n = F_0, \quad (\text{C.1})$$

$$p_i\omega_{i-1} + q_i\omega_i + r_i\omega_{i+1} = F_i, \quad i = 1, \dots, n \quad (\text{C.2})$$

$$p_n\omega_{n-1} + q_n\omega_n = F_n. \quad (\text{C.3})$$

El primer paso que debemos hacer para resolver el sistema es introducir la relación de recurrencia

$$\omega_{i+1} = x_i\omega_i + y_i, \quad i = 0, \dots, n - 1. \quad (\text{C.4})$$

Si obtenemos las cantidades x_i , y_i y ω_0 , con ayuda de (C.4) podemos encontrar las demás ω 's. Para obtener x_i y y_i , sustituimos (C.4) en (C.2) y encontramos que

$$p_i\omega_{i-1} + q_i\omega_i + r_i(x_i\omega_i + y_i)F_i$$

que es igual a

$$\omega_i + \frac{p_i}{q_i + r_i x_i} \omega_{i-1} = \frac{F_i - r_i y_i}{q_i + r_i x_i} \quad (\text{C.5})$$

ahora, de (C.4) llegamos a

$$\omega_i = x_{i-1} \omega_{i-1} + y_{i-1} \quad (\text{C.6})$$

y si ahora comparamos la ecuación (C.5) con (C.6) nos dice que

$$x_{i-1} = \frac{-p_i}{q_i + r_i x_i}, \quad y_{i-1} = \frac{F_i - r_i y_i}{q_i + r_i x_i} \quad i = 1, \dots, n-1. \quad (\text{C.7})$$

Si conociéramos los valores de x_{n-1} y y_{n-1} con la ayuda de (C.7) seríamos capaces de calcular el resto de los x_i y y_i disminuyendo el valor del índice i . Por lo que, necesitamos x_{n-1} y y_{n-1} que obtenemos a partir de (C.3) y hacemos que $i = n-1$ en (C.4), de donde se tienen

$$p_n \omega_{n-1} + q_n \omega_n = F_n, \quad \omega_n = x_{n-1} \omega_{n-1} + y_{n-1} \quad (\text{C.8})$$

despejando ω_n de la primera relación en (C.8)

$$\omega_n = -\frac{p_n}{q_n} \omega_{n-1} + \frac{F_n}{q_n}. \quad (\text{C.9})$$

Ahora de (C.9) y la segunda expresión en (C.8) uno concluye que

$$x_{n-1} = -\frac{p_n}{q_n} \quad \text{y} \quad y_{n-1} = \frac{F_n}{q_n}. \quad (\text{C.10})$$

Ahora que ya tenemos x_{n-1} y a y_{n-1} , (C.7) nos permite calcular las demás x_i y y_i . Sólo nos falta encontrar ω_0 para conocer el resto de las ω 's via (C.9). Para encontrar ω_0 tomamos la ecuación (C.1) y escribimos las $\omega_1, \dots, \omega_n$ en términos de ω_0 de la siguiente manera

$$\omega_i = \xi_i \omega_0 + \eta_i \quad i = 0, 1, \dots, n \quad (\text{C.11})$$

donde las ξ 's y η 's se desconocen en principio. Pero observamos que para $i = 0$ tenemos que

$$\xi_0 = 1, \quad \eta_0 = 0. \quad (\text{C.12})$$

Si sustituimos (C.11) en (C.4) y cambiamos $i \rightarrow i-1$, se llega a

$$\omega_i = x_{i-1} \xi_{i-1} \omega_0 + x_{i-1} \eta_{i-1} + y_{i-1} \quad (\text{C.13})$$

que si comparamos con (C.11) podemos ver que

$$\xi_i = x_{i-1}\xi_{i-1} \quad (\text{C.14})$$

$$\eta_i = x_{i-1}\eta_{i-1} + y_{i-1} \quad (\text{C.15})$$

donde $i = 1, \dots, n$. Volvemos a sustituir a (C.11) pero ahora en (C.1) y resulta

$$c_0\omega_0 + c_1(\xi_1\omega_0 + \eta_1) + \dots + c_n(\xi_n\omega_0 + \eta_n) = F_0$$

o escrita en forma compacta

$$\Xi\omega_0 + H = F_0 \quad (\text{C.16})$$

donde

$$\Xi = \sum_{i=0}^n c_i \xi_i, \quad H = \sum_{i=0}^n c_i \eta_i \quad (\text{C.17})$$

y finalmente encontramos

$$\omega_0 = \frac{F_0 - H}{\Xi}. \quad (\text{C.18})$$

Con esto, resolvemos el sistema lineal con una matriz tridiagonal con trenza y se puede resumir en el siguiente algoritmo:

1. Calcule x_i, y_i para $i = 0, \dots, n - 1$ de las ecuaciones (C.7) y (C.10).
2. Calcule ω_0 de acuerdo a:
 - a) Toma $\xi_0 = 1$ y $\eta_0 = 0$.
 - b) Consigue ξ_i, η_i para $i = 1, \dots, n$ a partir de (C.14) y (C.15).
 - c) Calcula Ξ y H de (C.17).
 - d) Obtienes ω_0 de (C.18).
3. Finalmente, calcula ω_i para $i = 1, \dots, n$ a partir de la ecuación (C.11).

Apéndice D

Discretización del tiempo

En este apéndice mencionamos los métodos de discretización del tiempo que se emplearon en el transcurso de esta tesis, los cuales fueron: el método de Crank-Nicolson, Adams-Bashforth y Runge-Kutta. Estos métodos resuelven problemas de tipo

$$\dot{y} = f(y, t) \tag{D.1}$$

Los métodos antes mencionados para la discretización del tiempo son muy usados en junto con los métodos Espectrales.

D.1. Método de Crank-Nicolson

El método de Crank-Nicolson es un método de diferencias finitas centrada. Si tenemos una ecuación diferencial parcial de una dimensión de la forma de la ecuación (D.1) y definimos $y(i\Delta x, n\Delta t) = y_i^n$, vemos que con la ecuación de Crank-Nicolson, que es el promedio de el método de Euler hacia delante en n y el método de Euler hacia atrás en $n + 1$, tendremos

$$y_i^{n+1} = y_i^n + \frac{\Delta t}{2}(f_i^{n+1} + f_i^n). \tag{D.1}$$

Dado de que es un método de dos pasos nos brinda una precisión de $O(\Delta t^3)$.

D.2. Método de Adams-Bashforth

Los métodos de Adams Bashforth son métodos explícitos de pasos multiples. El más sencillo de esta familia de métodos es el de dos pasos (AB2), para un

problema tipo (D.1) donde definimos $y(i\Delta x, n\Delta t) = y_i^n$ tenemos

$$y_i^{n+1} = y_i^n + \frac{\Delta t}{2}(2f_i^n - f_i^{n-1}) \quad (\text{D.1})$$

donde tenemos que el error es de $O(\Delta t^3)$. Para una mayor exactitud se consideran más pasos, los más usados de estos métodos son AB3 y AB4.

D.3. Método de Runge-Kutta

Los métodos de Runge-Kutta son una familia de métodos iterativos implícitos y explícitos para la aproximación a soluciones de ecuaciones diferenciales ordinarias con el fin de resolver problemas del valor inicial, esto es

$$\dot{y} = f(y, t) \text{ con } y(0) = y_0 \quad (\text{D.1})$$

para encontrar $y_1 = y(\Delta t)$ con $\Delta t > 0$.

El miembro de esta familia que se usa en esta tesis es el método Runge-Kutta de orden cuarto que esta dado por

$$y_{n+1} = y_n + \frac{\Delta t}{6}(k_1 + 2k_2 + 2k_3 + k_4) \quad (\text{D.2})$$

donde

$$\begin{aligned} k_1 &= f(t_n, y_n) && \text{(pendiente al inicio del intervalo)} \\ k_2 &= f\left(t_n + \frac{\Delta t}{2}, y_n + \frac{\Delta t}{2}k_1\right) && \text{(pendiente del punto medio del intervalo usando } k_1) \\ k_3 &= f\left(t_n + \frac{\Delta t}{2}, y_n + \frac{\Delta t}{2}k_2\right) && \text{(pendiente del punto medio del intervalo usando } k_2) \\ k_4 &= f(t_n + \Delta t, y_n + \Delta tk_3) && \text{(pendiente al final del intervalo).} \end{aligned}$$

Como mencionamos anteriormente este método es de cuarto orden, por lo que, el error por paso es del orden $O(\Delta t^5)$.

Apéndice E

Transformación Hopf-Cole

En este apéndice probaremos que la ecuación de Burgers podemos reducirla a la ecuación de calor mediante la transformación de Hopf-Cole

$$u = -2\nu \frac{1}{\phi} \frac{\partial \phi}{\partial x}. \quad (\text{E.1})$$

Tomamos la ecuación de Burgers y sustituimos (E.1) en ella

$$\begin{aligned} \frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} &= \nu \frac{\partial^2 u}{\partial x^2} \\ -2\nu \frac{\partial}{\partial t} \left(\frac{1}{\phi} \frac{\partial \phi}{\partial x} \right) + 4\nu^2 \frac{1}{\phi} \frac{\partial \phi}{\partial x} \frac{\partial}{\partial x} \left(\frac{1}{\phi} \frac{\partial \phi}{\partial x} \right) &= -2\nu^2 \frac{\partial^2}{\partial x^2} \left(\frac{1}{\phi} \frac{\partial \phi}{\partial x} \right) \end{aligned}$$

dividimos entre 2ν

$$\frac{\partial}{\partial t} \left(\frac{1}{\phi} \frac{\partial \phi}{\partial x} \right) - 2\nu \frac{1}{\phi} \frac{\partial \phi}{\partial x} \frac{\partial}{\partial x} \left(\frac{1}{\phi} \frac{\partial \phi}{\partial x} \right) = \nu \frac{\partial^2}{\partial x^2} \left(\frac{1}{\phi} \frac{\partial \phi}{\partial x} \right)$$

y realizamos las derivadas correspondientes en cada término de la ecuación, obteniendo

$$\begin{aligned} &\left(-\frac{1}{\phi^2} \frac{\partial \phi}{\partial t} \frac{\partial \phi}{\partial x} + \frac{1}{\phi} \frac{\partial}{\partial t} \frac{\partial \phi}{\partial x} \right) - 2\nu \left[\frac{1}{\phi^2} \frac{\partial \phi}{\partial x} \frac{\partial^2 \phi}{\partial x^2} - \frac{1}{\phi^3} \left(\frac{\partial \phi}{\partial x} \right)^3 \right] \\ &= \nu \left[\frac{1}{\phi} \frac{\partial}{\partial x} \left(\frac{\partial^2 \phi}{\partial x^2} \right) - \frac{1}{\phi^2} \frac{\partial \phi}{\partial x} \frac{\partial^2 \phi}{\partial x^2} + \frac{2}{\phi^3} \left(\frac{\partial \phi}{\partial x} \right)^3 - \frac{2}{\phi^2} \frac{\partial^2 \phi}{\partial x^2} \frac{\partial \phi}{\partial x} \right] \end{aligned}$$

Ahora, sumando los términos iguales

$$\begin{aligned} -\frac{1}{\phi^2} \frac{\partial \phi}{\partial t} \frac{\partial \phi}{\partial x} + \frac{1}{\phi} \frac{\partial}{\partial t} \frac{\partial \phi}{\partial x} &= \nu \frac{1}{\phi} \frac{\partial}{\partial x} \left(\frac{\partial^2 \phi}{\partial x^2} \right) - \nu \frac{1}{\phi^2} \frac{\partial \phi}{\partial x} \frac{\partial^2 \phi}{\partial x^2} \\ -\frac{1}{\phi^2} \frac{\partial \phi}{\partial x} \left(\frac{\partial \phi}{\partial t} - \nu \frac{\partial^2 \phi}{\partial x^2} \right) &= -\frac{1}{\phi} \frac{\partial}{\partial x} \left(\frac{\partial \phi}{\partial t} - \nu \frac{\partial^2 \phi}{\partial x^2} \right) \end{aligned}$$

donde tomamos en cuenta que $\frac{\partial}{\partial t} \frac{\partial \phi}{\partial x} = \frac{\partial}{\partial x} \frac{\partial \phi}{\partial t}$, por lo llegamos a

$$\left[\frac{\partial}{\partial x} \left(\frac{1}{\phi} \right) + \frac{1}{\phi} \frac{\partial}{\partial x} \right] \left(\frac{\partial \phi}{\partial t} - \nu \frac{\partial^2 \phi}{\partial x^2} \right) = 0 \quad (\text{E.2})$$

que nos muestra que (E.2) es igual a cero, si y solo si

$$\frac{\partial \phi}{\partial t} = \nu \frac{\partial^2 \phi}{\partial x^2} \quad (\text{E.3})$$

y (E.3) es la ecuación de calor, por lo tanto, queda demostrado que la ecuación de Burgers se puede escribir como la ecuación de calor si se hace la transformación de Hopf-Cole.

Bibliografía

- [1] Boyd J. P., “Chebyshev and Fourier Spectral Methods”. Dover New York (2001).
- [2] Briggs W. L. , Henson V. E., “The DFT: An Owner’s Manual for the Discrete Fourier Transform” SIAM (1995).
- [3] Brigham E. O., “The Fast Fourier Transform”. Prentice-Hall (1974).
- [4] Canuto C. , Hussaini M. Y., Quarteroni A., Zang T. A., “Spectral Methods: Fundamentals in Single Domains”. Springer (2006).
- [5] Chu E. , “Discrete and Continuous Fourier Transforms: Analysis, Applications and Fast Algorithms”. CRC Press (2008).
- [6] Gottlieb D., “Numerical Analysis of Spectral Methods: Theory and Applications”. Society for Industrial and Applied Mathematics (1993).
- [7] Hernández S., “Sobre la inestabilidad de ruptura en convección inducida por tensión superficial”, Tesis de maestría del Instituto de Física y Matemáticas de UMSNH, (2010).
- [8] Kopriva D. A., “Implementing Spectral Methods for Partial Differential Equations”. Springer (2009).
- [9] Mason J.C., “Chebyshev polynomials”. Chapman & Hall/CRC (2003).
- [10] Peyret R., “Spectral Methods for Incompressible Viscous Flow”. Springer (2002).
- [11] Runge C., “ber empirische Funktionen und die Interpolation zwischen quidistanten Ordinaten”. Zeitschrift fr Mathematik und Physik 46: 224243.
- [12] Trefethen L., “Spectral Methods in MATLAB”. SIAM (2000).