

# 525476 Métodos Espectrales: Sesiones 1 y 2

Leonardo E. Figueroa

CI<sup>2</sup>MA y Departamento de Ingeniería Matemática  
Universidad de Concepción

Viernes 03 y lunes 06 de septiembre de 2021

## Preliminares

## Bibliografía del curso

Referencia	Descripción
[CHQZ06]	Base de la mayor parte del curso. Aspectos teóricos y prácticos, pero no llega a tener ejemplos de código computacional.
[BM97]	Comparable a [CHQZ06] en amplitud y profundidad aunque, a diferencia de éste, discute aproximación del sistema de Navier–Stokes.
[Tre00]	Breve y muy práctico para colocación en una dimensión. Tiene una clara explicación de cómo usar la transformada rápida de Fourier en este contexto.
[Boy01]	Comprehensivo, algo desordenado. Se puede bajar de <a href="http://www-personal.umich.edu/~jpboyd/BOOK_Spectral2000.html">http://www-personal.umich.edu/~jpboyd/BOOK_Spectral2000.html</a> .
[For96]	Breve y de orientación práctica.
[OT13]	Breve artículo que inauguró el método ultraesférico.
[OST20]	Artículo que reseña avances recientes en computación con polinomios ortogonales.
[KS09]	Artículo que desarrolla un método espectral de Fourier con penalización volumétrica para obstáculos sólidos en movimiento

# Detalle de la bibliografía del curso I

- [BM97] Christine Bernardi and Yvon Maday, *Spectral methods*, Handbook of numerical analysis, Vol. V (P. G. Ciarlet and J. L. Lions, eds.), North-Holland, Amsterdam, 1997, pp. 209–485. MR 1470226
- [Boy01] John P. Boyd, *Chebyshev and Fourier spectral methods*, second ed., Dover Publications Inc., Mineola, NY, 2001. MR 1874071 (2002k:65160)
- [CHQZ06] C. Canuto, M. Y. Hussaini, A. Quarteroni, and T. A. Zang, *Spectral methods*, Scientific Computation, Springer-Verlag, Berlin, 2006, Fundamentals in single domains. MR 2223552 (2007c:65001)
- [For96] Bengt Fornberg, *A practical guide to pseudospectral methods*, Cambridge Monographs on Applied and Computational Mathematics, vol. 1, Cambridge University Press, Cambridge, 1996. MR 1386891 (97g:65001)

## Detalle de la bibliografía del curso II

- [KS09] Dmitry Kolomenskiy and Kai Schneider, *A Fourier spectral method for the Navier-Stokes equations with volume penalization for moving solid obstacles*, J. Comput. Phys. **228** (2009), no. 16, 5687–5709. MR 2542911
- [OST20] Sheehan Olver, Richard Mikaël Slevinsky, and Alex Townsend, *Fast algorithms using orthogonal polynomials*, Acta Numer. **29** (2020), 573–699. MR 4189295
- [OT13] Sheehan Olver and Alex Townsend, *A fast and well-conditioned spectral method*, SIAM Rev. **55** (2013), no. 3, 462–489. MR 3089410
- [Tre00] Lloyd N. Trefethen, *Spectral methods in MATLAB*, Software, Environments, and Tools, vol. 10, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2000. MR 1776072 (2001c:65001)

## Presentación: Funciones de forma y de prueba

Los Métodos Espectrales son una clase de discretizaciones espaciales para ecuaciones diferenciales. Involucran espacios de dimensión finita con **funciones de forma** (*trial functions*) y **funciones de prueba** o de peso (*test functions*).

Las funciones de forma se usan para **construir** una aproximación de la solución del problema.

Las funciones de prueba se usan para **asegurar** que la ecuación diferencial y quizás algunas condiciones de contorno sean efectivamente aproximadas. Esto se consigue minimizando, respecto a alguna norma apropiada, el **residuo** resultante de poner una función de forma en el lugar de la solución exacta.

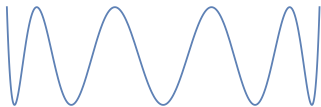
## Presentación: Rasgos distintivos

Lo que distingue a los Métodos Espectrales clásicos de otros métodos de residuos ponderados (que eso es lo que acabamos de describir) es que las funciones de forma y de prueba son funciones de **soporte global** en un **dominio sencillo, infinitamente diferenciables** y casi (o completamente) **ortogonales**; típicamente, autofunciones de problemas de Sturm–Liouville.

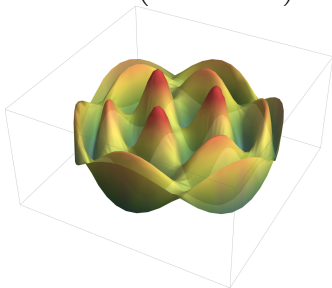
En contraste, en el Método de Elementos Finitos las funciones de forma y de prueba son típicamente de soporte local en un dominio que puede ser muy complejo y tienen regularidad limitada.

# Cuadro ilustrativo de funciones de forma y/o de prueba

## Métodos Espectrales

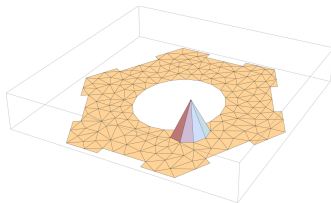
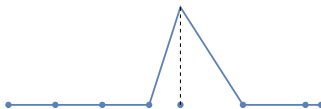


↑ Autofunción de  
 $-\sqrt{1-x^2} \left( \sqrt{1-x^2} u'(x) \right)'$



↑ Autofunción de  
 $-\sqrt{1-x^2-y^2} \operatorname{div} \left( \sqrt{1-x^2-y^2} \nabla u(x,y) \right) - \frac{\partial^2}{\partial \theta^2} u(x,y)$

## Elementos Finitos





## Frontera difusa

A pesar de los contrastes indicados anteriormente, los Métodos Espectrales y de Elementos Finitos tienen mucho en común.

Por ejemplo, buena parte de la teoría de ambos métodos es común.

Más todavía, en ciertos casos ambos métodos coinciden; en efecto:

*Elemento finito con un solo elemento y refinación  $p$  = Método espectral clásico.*

También:

*Métodos espectrales en dominio subdividido imponiendo condiciones de compatibilidad en las interfaces internas = Elemento finito de alto orden.*

## Ejemplos de introducción

## Ejemplos de introducción

Método de Fourier–Galerkin para una ecuación de ondas

## Una ecuación de ondas

$$\frac{\partial u}{\partial t} - \frac{\partial u}{\partial x} = 0 \quad \text{en } (0, T) \times (0, 2\pi), \quad (1a)$$

$$u(t, 0) = u(t, 2\pi) \quad \text{para } t \in (0, T), \quad (1b)$$

$$u(0, x) = u_0(x) \quad \text{para } x \in (0, 2\pi) \quad (1c)$$

Aproximaremos la solución  $u$  de (1) por una función  $u^N$  de la forma

$$u^N(t, x) = \sum_{k=-N/2}^{N/2} a_k(t) \phi_k(x), \quad (2)$$

donde

- ▶ los  $\phi_k$  son unas **funciones de forma** y
- ▶ los  $a_k$  son unos **coeficientes de expansión** que dependen del tiempo

## Imposición débil de ecuación de cuerpo

No puede esperarse en general que  $u^N$  de (2) satisfaga la ecuación de cuerpo (1a); nos conformaremos con exigir que se anule su residuo **testado** contra una colección de **funciones de prueba**

$\psi_k$ :

$$\int_0^{2\pi} \left( \frac{\partial u^N}{\partial t} - \frac{\partial u^N}{\partial x} \right) \overline{\psi_k(x)} dx = 0, \quad (3)$$

para índices  $k$  entre  $-N/2$  y  $N/2$  y todo tiempo  $t \in (0, T)$

Dada la condición de contorno periódica (1b), es natural postular el uso de **polinomios trigonométricos** como funciones de forma y de prueba:

$$\phi_k(x) = \psi_k(x) = e^{ikx}$$

La satisfacción por parte de las funciones de forma y de prueba de las condiciones de contorno esenciales en forma exacta es un rasgo distintivo de los métodos de tipo Galerkin

## Biortogonalidad y sus consecuencias

Debido a la bien conocida relación de biortogonalidad

$$\int_0^{2\pi} \phi_k(x) \overline{\psi_l(x)} dx = \int_0^{2\pi} e^{ikx} \overline{e^{ilx}} dx = 2\pi \delta_{k,l},$$

la ecuación (3) toma la forma

$$\begin{aligned} 0 &= \int_0^{2\pi} \left( \frac{\partial u^N}{\partial t} - \frac{\partial u^N}{\partial x} \right) \overline{e^{ikx}} dx \\ &\stackrel{(2)}{=} \int_0^{2\pi} \left[ \left( \frac{\partial}{\partial t} - \frac{\partial}{\partial x} \right) \sum_{l=-N/2}^{N/2} a_l(t) e^{ilx} \right] \overline{e^{ikx}} dx \\ &= \int_0^{2\pi} \left[ \sum_{l=-N/2}^{N/2} \left( a'_l(t) e^{ilx} - il a_l(t) e^{ilx} \right) \right] \overline{e^{ikx}} dx \\ &= 2\pi (a'_k(t) - ik a_k(t)) \end{aligned}$$

## Sistema de EDO para coeficientes

Así, para  $k \in \{-N/2, \dots, N/2\}$ , el coeficiente  $a_k$  satisface la EDO,

$$a'_k = i k a_k \quad (4)$$

Obtendremos condiciones iniciales para estas EDO imponiendo sobre  $u^N$  débilmente la condición inicial (1c) que satisface  $u$ :

$$\int_0^{2\pi} \underbrace{\left( u^N(0, x) - u_0(x) \right)}_{\text{residuo de condición inicial}} \overline{\psi_k(x)} dx = 0,$$

lo que conduce a

$$a_k(0) = \frac{\int_0^{2\pi} u_0(x) \overline{\psi_k(x)} dx}{\int_0^{2\pi} \phi_k(x) \overline{\psi_k(x)} dx} = \frac{1}{2\pi} \int_0^{2\pi} u_0(x) \overline{\psi_k(x)} dx; \quad (5)$$

equivalentemente, denotando  $S_N := \text{span}(\{\phi_{-N/2}, \dots, \phi_{N/2}\})$ ,

$$u^N(0, x) = \text{proj}_{S_N}(u_0) := \text{proyección ortogonal de } u_0 \text{ sobre } S_N$$

## Aproximación determinada

Resolviendo (4),  $a_k(t) = a_k(0)e^{i k t}$ , donde  $a_k(0)$  se conoce (cf. (5))

Así, la solución aproximada  $u^N$  de (2) queda completamente determinada:

$$u^N(t, x) = \sum_{k=-N/2}^{N/2} a_k(t) \phi_k(x) = \sum_{k=-N/2}^{N/2} a_k(0) e^{i k t} e^{i k x}$$

**Nota:** Salvo en ejemplos de libro, las integrales que son los  $a_k(0)$  deben aproximarse por reglas de cuadratura; suelen funcionar muy bien



## Ejemplo de libro

Consideremos como ejemplo nuestro sistema (1) con

$$u_0(x) = \text{sen}(\pi \cos(x));$$

en este caso,

$$a_k(t) = \text{sen} \left( \frac{k\pi}{2} \right) J_k(\pi) e^{ikt}$$

y

$$u(t, x) = \text{sen}(\pi \cos(x + t)) = \sum_{k \in \mathbb{Z}} a_k(t) e^{ikx}$$

Por lo tanto, para este problema sencillo con su esquema de aproximación *ad hoc* la solución aproximada  $u^N$  de (2) es la **truncación** de la serie de Fourier de la solución exacta o, equivalentemente, **la proyección** de la solución exacta sobre  $S_N$

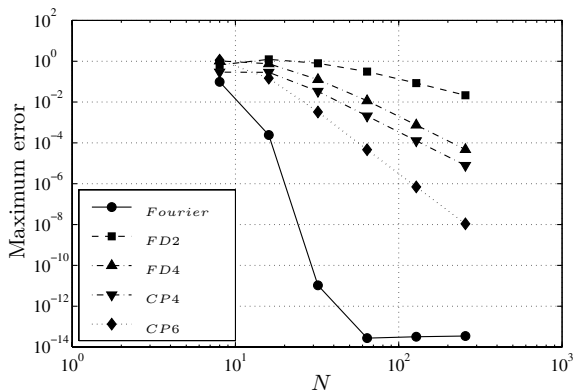
## Convergencia del ejemplo de libro

Por propiedades de las funciones de Bessel  $J_k$ ,

$$(\forall p \in \mathbb{N}_0) \quad k^p a_k(t) \xrightarrow{|k| \rightarrow \infty} 0$$

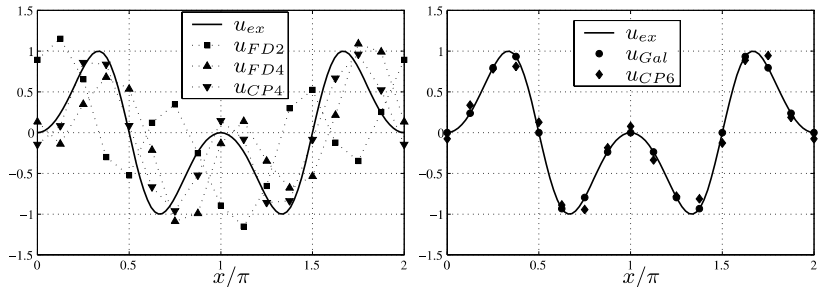
Una consecuencia es que  $u^N \xrightarrow{N \rightarrow \infty} u$  (por ejemplo, en  $L^2$ ) más rápido que cualquier potencia finita de  $1/N$ ; esto se conoce como **convergencia espectral**

# Gráfico de convergencia extraído de [CHQZ06]



**Fig. 1.1.** Maximum errors for the linear hyperbolic problem at  $t = 2\pi$  for Fourier Galerkin and several finite-difference schemes

## Gráficos de resultados extraídos de [CHQZ06]



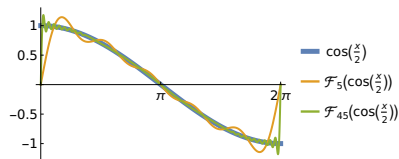
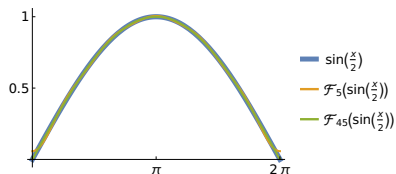
**Fig. 1.2.** Numerical solutions for the linear hyperbolic problem at  $t = 2\pi$  for  $N = 16$  for Fourier Galerkin and several finite-difference schemes

## Ejemplos de introducción

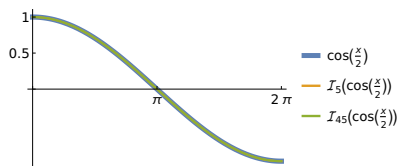
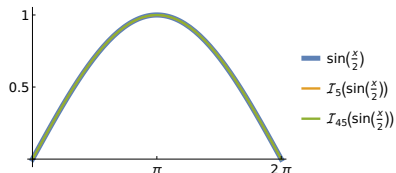
Método de colocación de Chebyshev para la ecuación del calor

# Situaciones noperiódicas

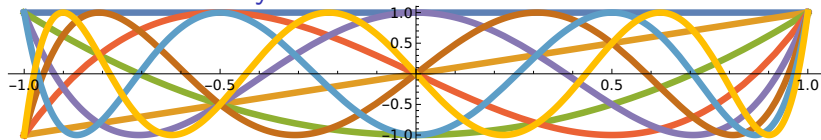
Para las expansiones en **polinomios trigonométricos**, la falta de periodicidad cuenta como discontinuidad, lo que **perjudica** la velocidad de convergencia de proyecciones e interpolaciones.



Para expansiones con **polinomios algebraicos**, la no-periodicidad no importa.



## Polinomios de Chebyshev



Una magnífica base de polinomios es la de los **polinomios de Chebyshev** (del primer tipo):

$$(\forall k \in \mathbb{N}_0) \quad T_k(x) = \cos(k \arccos(x)) \quad (6)$$

Efectivamente son polinomios: Usando la identidad trigonométrica  $\cos(\alpha + \beta) = \cos(\alpha)\cos(\beta) - \sin(\alpha)\sin(\beta)$  obtenemos que

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

Esta es una relación de recurrencia que, inicializada con las fórmulas obvias  $T_0(x) = 1$  y  $T_1(x) = x$ , permite concluir que cada  $T_k$  es un polinomio de grado exactamente  $k$ .

Como consecuencia, para todo  $N \in \mathbb{N}_0$ ,  $\mathbb{P}_N = \text{span}(\{T_0, \dots, T_N\})$ .

## Algunas propiedades fáciles de comprobar

Directo de la definición (6):

$$T_k(1) = 1, \quad T_k(-1) = (-1)^k \\ (\forall x \in [-1, 1]) \quad |T_k(x)| \leq 1$$

Los polinomios de Chebyshev son **polinomios ortogonales**:

$$\int_{-1}^1 T_k(x) T_l(x) \frac{1}{\sqrt{1-x^2}} dx = \int_0^\pi \cos(k\theta) \cos(l\theta) d\theta = \begin{cases} \pi & \text{si } k = l = 0 \\ \pi/2 & \text{si } k = l \neq 0 \\ 0 & \text{si } k \neq l \end{cases}$$

Los polinomios de Chebyshev son **autofunciones de un operador diferencial** de segundo orden (satisfacen un problema de **Sturm–Liouville**):

$$-\sqrt{1-x^2} \left( \sqrt{1-x^2} T'_k(x) \right)' = k^2 T_k(x)$$



## Ecuación del calor y funciones de forma

Consideremos esta ecuación del calor con condiciones de contorno Dirichlet:

$$\frac{\partial u}{\partial t} - \frac{\partial^2 u}{\partial x^2} = 0 \quad \text{en } (0, T) \times (-1, 1) \quad (7a)$$

$$u(t, -1) = 0 \quad \text{para } t \in (0, T) \quad (7b)$$

$$u(t, 1) = 0 \quad \text{para } t \in (0, T) \quad (7c)$$

$$u(x, 0) = u_0(x) \quad \text{para } x \in (-1, 1) \quad (7d)$$

Buscamos una solución aproximada de la forma

$$u^N(t, x) = \sum_{k=0}^N a_k(t) T_k(x) \quad (8)$$

con coeficientes  $a_k(t)$  que debemos determinar; esto es, empleamos a  $\mathbb{P}_N$  como espacio de funciones de forma de esta discretización espacial, con su base de polinomios de Chebyshev  $\{T_k\}_{k=0}^N$ .

## Fijamiento de coeficientes mediante colocación

En cada instante de tiempo la aproximación  $u^N$  tiene  $N + 1$  grados de libertad a determinar.

Para ello impondremos que  $u^N$  satisfaga la ecuación de cuerpo (7a) en forma exacta en  $N - 1$  nodos distintos  $x_j \in (-1, 1)$ :

$$(\forall j \in \{1, \dots, N - 1\}) \quad \frac{\partial u^N}{\partial t}(t, x_j) - \frac{\partial^2 u^N}{\partial x^2}(t, x_j) = 0 \quad (9)$$

Estas condiciones se pueden interpretar como lo que resulta de imponer la ecuación de cuerpo testada contra la base de “funciones” de prueba  $\{\psi_j\}_{j=1}^{N-1}$ , donde  $\psi_j(x) = \delta(x - x_j)$ .

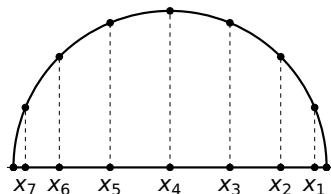
Las 2 condiciones que faltan las obtenemos de imponer las condiciones de contorno (7b) y (7c):

$$u^N(t, -1) = 0, \quad u^N(t, 1) = 0$$

# Nodos

Los  $x_j$  no deben escogerse equiespaciados. Una excelente alternativa es

$$x_j = \cos\left(\frac{j\pi}{N}\right) \quad (10)$$

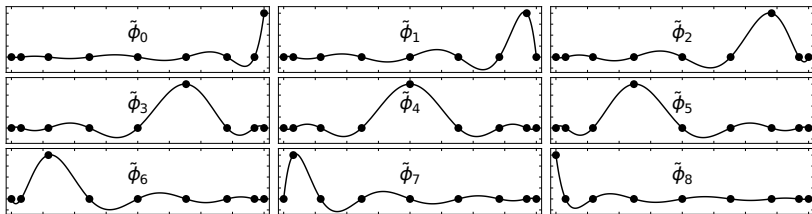


Con esta distribución de nodos, apiñada hacia los extremos del intervalo, se reducen fenómenos tipo Runge. Además, se puede explotar Transformadas de Cosenos Discretas (DCT), parientes de la Transformada Rápida de Fourier (FFT), para acelerar dramáticamente varios cálculos asociados.

# Funciones delta discretas

Asociadas a estos nodos están las **funciones delta discretas**  $\tilde{\phi}_j$ ,  $0 \leq j \leq N$ , que están definidas por las propiedades

$$\begin{aligned} (\forall j \in \{0, \dots, N\}) \quad \tilde{\phi}_j &\in \mathbb{P}_N \\ (\forall i, j \in \{0, \dots, N\}) \quad \tilde{\phi}_j(x_i) &= \delta_{i,j} \end{aligned}$$



## Valores modales y valores nodales

El conjunto  $\{\tilde{\phi}_0, \dots, \tilde{\phi}_N\}$  es otra base de  $\mathbb{P}_N$ , por lo que

$$u^N(t, x) = \sum_{k=0}^N a_k(t) T_k(x) = \sum_{j=0}^N u_j(t) \tilde{\phi}_j(x)$$

para ciertos coeficientes dependientes del tiempo  $u_j(t)$  llamados **valores nodales** ya que

$$u_j(t) = u^N(t, x_j)$$

A los coeficientes  $a_k(t)$  se les denomina **valores modales**.

La expansión modal es preferida para operaciones del cálculo y la nodal para la evaluación y para la aplicación de operadores no lineales.

## Transformación de valores nodales a valores modales

Los valores modales se pueden expresar en términos de los valores nodales de acuerdo a

$$(\forall k \in \{0, \dots, N\}) \quad a_k(t) = \frac{2}{N\bar{c}_k} \sum_{j=0}^N \bar{c}_j^{-1} u_j(t) \cos\left(\frac{\pi j k}{N}\right) \quad (11)$$

donde los pesos  $\bar{c}_k$  están dados por

$$\bar{c}_k = \begin{cases} 2 & \text{si } k \in \{0, N\} \\ 1 & \text{si } 1 \leq k \leq N-1 \end{cases} \quad (12)$$

Empleando una DCT, la aplicación de la fórmula (11) cuesta  $\mathcal{O}(N \log(N))$  operaciones en lugar de las  $\mathcal{O}(N^2)$  que en general cuesta aplicar un producto matriz-vector.

Lo mismo ocurre con la transformación inversa (valores modales a valores nodales). Esto hace atractivo ir saltando de una representación a la otra según la operación que se desee realizar.

## Diferenciación espectral

Como  $u^N(t, x)$  es un polinomio de grado  $\leq N$  respecto a  $x$ ,  $\frac{\partial u^N}{\partial x}(t, x)$  es un polinomio de grado  $\leq N - 1$  respecto a  $x$  y posee su propia expansión análoga a (8)

$$\frac{\partial u^N}{\partial x}(t, x) = \sum_{k=0}^N a_k^{(1)}(t) T_k(x)$$

(obviamente  $a_N^{(1)}(t) = 0$ ; conviene mantener ese término).

Definiendo además, en forma natural,  $a_{N+1}^{(1)}(t) = 0$ , ocurre que por medio de la relación de recurrencia<sup>1</sup>

$$\bar{c}_k a_k^{(1)}(t) = a_{k+2}^{(1)}(t) + 2(k+1)a_{k+1}(t)$$

se pueden calcular todos los  $a_k^{(1)}(t)$  en términos de los  $a_k(t)$  en  $\mathcal{O}(N)$  operaciones.

---

<sup>1</sup>¿Impaciente por probarla? Por identidades trigonométricas, para  $k \geq 1$ ,  $2T_k = \frac{1}{k+1} T'_{k+1} - \frac{1}{k-1} T'_{k-1}$ ; derive (8) y compare términos.

## Segunda diferenciación espectral

Iterando el procedimiento anterior, es posible calcular los coeficientes  $a_k^{(2)}(t)$  de la expansión

$$\frac{\partial^2 u^N}{\partial x^2}(t, x) = \sum_{k=0}^N a_k^{(2)}(t) T_k(x)$$

en términos de los  $a_k(t)$  en  $\mathcal{O}(N) + \mathcal{O}(N) = \mathcal{O}(N)$  operaciones.



# Explotación de colocación

Así, para  $1 \leq j \leq N - 1$ ,

$$\begin{aligned} u_j'(t) &= \sum_{l=0}^N u_l'(t) \tilde{\phi}_l(x_j) \stackrel{(8)}{=} \frac{\partial u^N}{\partial t}(t, x_j) \stackrel{(9)}{=} \frac{\partial^2 u^N}{\partial x^2}(t, x_j) \\ &= \sum_{k=0}^N a_k^{(2)}(t) T_k(x_j) \stackrel{(6),(10)}{=} \sum_{k=0}^N a_k^{(2)}(t) \cos\left(\frac{\pi j k}{N}\right) \end{aligned} \tag{13}$$

Esta transformación lineal que conecta a los  $a_k^{(2)}(t)$  con los  $u_j'(t)$  también puede computarse en  $\mathcal{O}(N \log(N))$  operaciones por medio de una DCT.

## Operador lineal

Sea  $\mathbf{f}_N: \mathbb{R}^{N-1} \rightarrow \mathbb{R}^{N-1}$  el mapa que **recibe** un  $(u_j(t))_{j=1}^{N-1}$  y

1. le prepone y pospone sendos ceros de acuerdo a las condiciones de contorno para formar a  $(u_j(t))_{j=0}^N$  (0 flops),
2. aplica una DCT para obtener  $(a_k(t))_{k=0}^N$  de acuerdo a (11) ( $\mathcal{O}(N \log(N))$  flops),
3. aplica las relaciones de recurrencia de diferenciación espectral para obtener  $(a_k^{(2)})_{k=0}^N$  ( $\mathcal{O}(N)$  flops),
4. aplica otra DCT para obtener  $(u'_j(t))_{j=0}^N$  de acuerdo a (13) ( $\mathcal{O}(N \log(N))$  flops) y
5. quita las (espúreas) primera y última entradas del último resultado **para obtener**  $(u'_j(t))_{j=1}^{N-1}$  (0 flops).

Costo total de una aplicación de  $\mathbf{f}_N$ :  $\mathcal{O}(N \log(N))$  flops.

Como  $\mathbf{f}_N$  es una composición de mapas lineales, es lineal y existe una matriz  $D_N^2 \in \mathbb{R}^{(N-1) \times (N-1)}$  que lo representa. Esta matriz es conocida, pero es llena y aplicarla directamente cuesta  $\mathcal{O}(N^2)$  flops.

# Sistema de Ecuaciones Diferenciales Ordinarias (EDO)

Las condiciones iniciales sobre los coeficientes  $u_j(t)$ ,  $1 \leq j \leq N - 1$ , se fijan recogiendo la condición inicial (7d) del problema original por colocación:

$$u^N(0, x_j) = u_j(0) = u_0(x_j)$$

Recolectando  $\mathbf{u}(t) = (u_j(t))_{j=1}^{N-1}$  y  $\mathbf{u}_0 := (u_0(x_j))_{j=1}^{N-1}$ , nuestro problema reducido a un Problema de Valores Iniciales (PVI) vectorial que podemos remitir a nuestro paquete para EDO numéricas favorito:

$$\begin{cases} \mathbf{u}'(t) = \mathbf{f}_N(\mathbf{u}(t)) \\ \mathbf{u}(0) = \mathbf{u}_0 \end{cases}$$

## Ejemplo con solución exacta conocida

Si en el problema (7) fijamos la condición inicial como

$$u_0(x) = \text{sen}(\pi x)$$

entonces la solución exacta es

$$u(t, x) = e^{-\pi^2 t} \text{sen}(\pi x)$$

y admite la expansión en serie de Fourier–Chebyshev

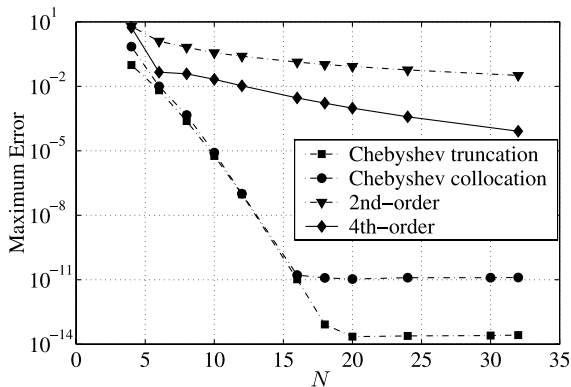
$$u(t, x) = \sum_{k=0}^{\infty} b_k(t) T_k(x) \quad \text{con} \quad b_k(t) = \frac{2}{c_k} \text{sen}\left(\frac{\pi}{2}\right) J_k(\pi) e^{-\pi^2 t}$$

donde a su vez

$$c_k = \begin{cases} 2 & \text{si } k = 0 \\ 1 & \text{si } k \geq 1 \end{cases}$$

Notar que en general los coeficientes  $b_k(t)$  de la serie de Fourier–Chebyshev de la solución exacta  $u$  **no** coinciden con los coeficientes  $a_k(t)$  de la solución aproximada  $u^N$ .

## Gráfico de convergencia extraído de [CHQZ06]



**Fig. 1.3.** Maximum errors for the heat equation problem at  $t = 1$  for Chebyshev collocation and several finite-difference schemes. The Chebyshev truncation result is shown for comparison

## Ejemplos de introducción

Método de Galerkin con integración numérica de Legendre para  
una ecuación de advección-difusión-reacción

## Una ecuación de advección-difusión-reacción

Consideremos el problema estacionario de advección-difusión-reacción

$$\mathcal{F}(u)' + \gamma u = f \quad \text{en } (-1, 1) \quad (14a)$$

Aquí, el *flujo* de advección-difusión está definido por

$$\mathcal{F}(u) = -\nu \frac{\partial u}{\partial x} + \beta u \quad (14b)$$

y la ecuación de cuerpo (14a) se complementa con una condición de contorno prescribiendo a la función incógnita en  $x = -1$

$$u(-1) = 0 \quad (14c)$$

y otra condición de contorno prescribiendo al flujo en  $x = 1$

$$\mathcal{F}(u)(1) + g = 0 \quad (14d)$$

Aquí,  $g$  es una constante y los coeficientes  $\nu$  (difusión),  $\beta$  (advección/transporte) y  $\gamma$  (reacción), junto al término fuente  $f$ , pueden variar en el dominio  $(-1, 1)$ . Además, suponemos que  $\nu \geq \bar{\nu}$  para alguna constante  $\bar{\nu} > 0$ .

## Formulación variacional

Mediante integración por partes podemos deducir la siguiente formulación variacional para el problema (14): Hallar  $u \in V := \{v \in H^1(-1, 1) \mid v(-1) = 0\}$  que satisfaga

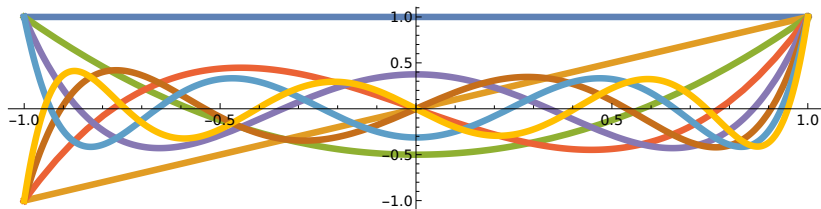
$$\begin{aligned} (\forall v \in V) \quad \int_{-1}^1 \nu u' v' - \int_{-1}^1 \beta u v' + \int_{-1}^1 \gamma u v \\ = \int_{-1}^1 f v + g v(1) \end{aligned} \tag{15}$$

Notemos que la condición de contorno (14c) se ha impuesto al espacio  $V$  (condición de contorno *esencial*), mientras que la condición de contorno (14d) se ha incorporado en forma débil (condición de contorno *natural*).

Queremos discretizar esta formulación variacional mediante un **método de Galerkin**, consistente en sustituir el espacio  $V$  por un subespacio de dimensión finita apropiado. Para construirlo emplearemos los...



# Polinomios de Legendre

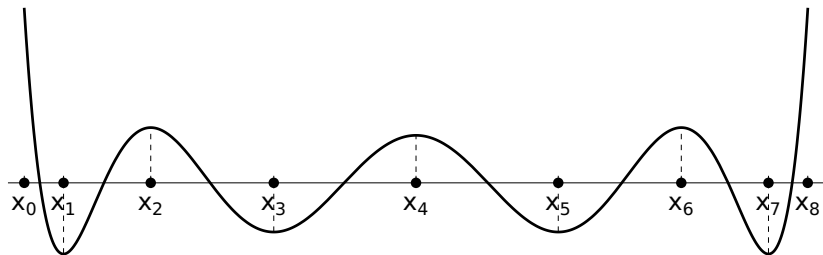


Si sometemos la sucesión de los monomios  $(x^k)_{k \in \mathbb{N}_0}$  al proceso de ortogonalización de Gram–Schmidt respecto al producto interior de  $L^2(-1, 1)$  (esto es,  $\langle f, g \rangle = \int_{-1}^1 f g$ ) pero normalizando por la evaluación en  $x = 1$  en lugar de por norma, obtenemos la sucesión de los **polinomios de Legendre**  $(L_k)_{k \in \mathbb{N}_0}$ .

Probaremos más adelante que, para cada  $k \in \mathbb{N}_0$ , el polinomio de Legendre  $L_k$  así definido es de grado exactamente  $k$  y que posee  $k - 1$  puntos extremales (máximos o mínimos locales) en  $(-1, 1)$ .

## Nodos de Gauss–Lobatto de Legendre

Fijemos un  $N \in \mathbb{N}$ ,  $N \geq 2$ . Denotamos a los **puntos extremales** de  $L_N$ , ordenados de menor a mayor, por  $x_j$ ,  $1 \leq j \leq N - 1$ . Se puede probar que estos puntos extremales efectivamente son  $N - 1$  en número y que todos yacen en  $(-1, 1)$ . Además, definimos directamente a  $x_0 = -1$  y  $x_N = 1$ . A los miembros de la colección  $(x_j)_{j=0}^N$  se les denomina *nodos de Gauss–Lobatto de Legendre*.



**Figura:** En este ejemplo con  $N = 8$ , los nodos de Gauss–Lobatto de Legendre consisten en los 7 puntos extremales de  $L_8$  en  $(-1, 1)$  en unión con  $x_0 = -1$  y  $x_8 = 1$

## Cuadratura de Gauss–Lobatto–Legendre

Se puede probar que existen pesos positivos  $w_j$ ,  $0 \leq j \leq N$ , tales que la *fórmula de cuadratura de Gauss–Lobatto–Legendre*

$$\int_{-1}^1 p(x) \, dx \sim \sum_{j=0}^N p(x_j) w_j$$

es exacta para polinomios de grado menor o igual que  $2N - 1$ .

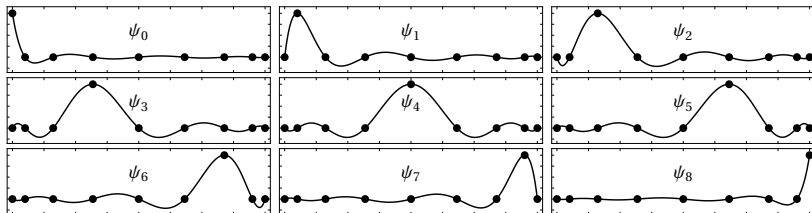
### Cuadro comparativo

Cuadratura	Gauss–Lobatto–Legendre de $N + 1$ nodos	Gauss–Legendre de $N + 1$ nodos
Nodos	Puntos extremales de $L_N$ en unión con $\{-1, 1\}$	Raíces de $L_{N+1}$ ; todas yacen en $(-1, 1)$
Exacta en	$\mathbb{P}_{2N-1}$	$\mathbb{P}_{2N+1}$

# Funciones delta discretas

Definimos **funciones delta discretas**  $\psi_j$ ,  $0 \leq j \leq N$ , de la misma forma que lo hicimos en la diapositiva 28 pero ahora asociadas a los nodos de Gauss–Lobatto–Legendre:

$$\begin{aligned} (\forall j \in \{0, \dots, N\}) \quad \psi_j &\in \mathbb{P}_N \\ (\forall i, j \in \{0, \dots, N\}) \quad \psi_j(x_i) &= \delta_{ij} \end{aligned}$$



## Galerkin puro

La envoltura lineal  $V_N := \text{span}(\{\psi_1, \dots, \psi_N\})$  ( $\psi_0$  no participa) es un subespacio de  $V$ , así que sirve para plantear un esquema de Galerkin para la formulación variacional (15): Hallar los coeficientes  $u_l$ ,  $1 \leq l \leq N$ , tales que  $u^N = \sum_{l=1}^N u_l \psi_l \in V_N$  satisfice

$$\begin{aligned} (\forall v \in V_N) \quad \int_{-1}^1 \nu u^{N'} v' - \int_{-1}^1 \beta u^N v' + \int_{-1}^1 \gamma u^N v \\ = \int_{-1}^1 f v + g v(1) \end{aligned}$$

Equivalentemente,

$$\begin{aligned} (\forall j \in \{1, \dots, N\}) \quad \int_{-1}^1 \nu u^{N'} \psi_j' - \int_{-1}^1 \beta u^N \psi_j' \\ + \int_{-1}^1 \gamma u^N \psi_j = \int_{-1}^1 f \psi_j + g \delta_{j,N} \end{aligned} \tag{16}$$

## Galerkin con integración numérica

En el **método de Galerkin con integración numérica** (G-NI) aproximaremos las integrales involucradas en el esquema de Galerkin puro (16) mediante cuadraturas de Gauss–Lobatto–Legendre, quedando con la formulación: Hallar los coeficientes  $u_l$ ,  $1 \leq l \leq N$ , tales que  $u^N = \sum_{l=1}^N u_l \psi_l \in V_N$  satisface

$$(\forall j \in \{1, \dots, N\})$$

$$\begin{aligned} \sum_{k=0}^N \left( \nu u^{N'} \psi_j' \right) (x_k) w_k - \sum_{k=0}^N \left( \beta u^N \psi_j' \right) (x_k) w_k \\ + \sum_{k=0}^N \left( \gamma u^N \psi_j \right) (x_k) w_k = \sum_{k=0}^N (f \psi_j) (x_k) w_k + g \delta_{j,N} \end{aligned} \quad (17)$$

## Sistema lineal de ecuaciones

El esquema G-NI (17) puede escribirse como el sistema lineal de ecuaciones

$$(\forall j \in \{1, \dots, N\}) \quad \sum_{l=1}^N K_{j,l} u_l = b_j,$$

donde, explotando la propiedad  $\psi_j(x_k) = \delta_{j,k}$  de las funciones delta discretas,

$$K_{j,l} = \sum_{k=0}^N (\nu \psi'_l \psi'_j)(x_k) w_k - (\beta \psi'_j)(x_l) w_l + \gamma(x_j) w_j \delta_{l,j}$$
$$b_j = f(x_j) w_j + g \delta_{j,N}$$

## Flujo numérico

Sea  $I_N$  el operador de interpolación a  $\mathbb{P}_N$  en los nodos de Gauss–Lobatto de Legendre; esto es, dada una función continua  $\varphi: [-1, 1] \rightarrow \mathbb{R}$ ,  $I_N(\varphi)$  se define como aquel único miembro de  $\mathbb{P}_N$  que satisface

$$(\forall j \in \{0, \dots, N\}) \quad I_N(\varphi)(x_j) = \varphi(x_j)$$

Así podemos definir al *flujo numérico*

$$\mathcal{F}^N(u^N) := I_N(\mathcal{F}(u^N))$$

Mediante este flujo numérico las dos primeras sumas en el esquema G-NI (17) se pueden reescribir como

$$\begin{aligned} & \sum_{k=0}^N \left( \nu u^{N'} \psi'_j \right)(x_k) w_k - \sum_{k=0}^N \left( \beta u^N \psi'_j \right)(x_k) w_k \\ & \stackrel{(14b)}{=} - \sum_{k=0}^N \left( \mathcal{F}(u^N) \psi'_j \right)(x_k) w_k = - \sum_{k=0}^N \left( \mathcal{F}^N(u^N) \psi'_j \right)(x_k) w_k \end{aligned} \tag{18}$$



## Cuadratura e integración por partes

Como tanto el flujo numérico  $\mathcal{F}^N(u^N)$  como cada función delta discreta  $\psi_j$  pertenecen a  $\mathbb{P}_N$ , podemos usar la exactitud de la cuadratura de Gauss–Lobatto–Legendre de  $N + 1$  nodos en  $\mathbb{P}_{2N-1}$  e integración por partes para obtener

$$\begin{aligned} - \sum_{k=0}^N \left( \mathcal{F}^N(u^N) \psi_j' \right) (x_k) w_k &= - \int_{-1}^1 \mathcal{F}^N(u^N) \psi_j' \\ &= \int_{-1}^1 \mathcal{F}^N(u^N)' \psi_j - \left[ \mathcal{F}^N(u^N) \psi_j \right] \Big|_{x=-1}^{x=1} \\ &= \sum_{k=0}^N \left( \mathcal{F}^N(u^N)' \psi_j \right) (x_k) w_k - \mathcal{F}(u^N)(1) \psi_j(1) \end{aligned} \tag{19}$$

## Forma alternativa de las ecuaciones del esquema G-NI

Combinando (18) y (19), sustituyendo el resultado en el esquema G-NI (17) y explotando la propiedad  $\psi_j(x_k) = \delta_{j,k}$  de las funciones delta discretas, obtenemos que sus ecuaciones equivalen a:

$$\begin{aligned} (\forall j \in \{1, \dots, N\}) \\ \left( \mathcal{F}^N(u^N)' + \gamma u^N \right)(x_j) w_j - \mathcal{F}(u^N)(1) \delta_{j,N} \\ = f(x_j) w_j + g \delta_{j,N} \end{aligned} \quad (20)$$

## G-NI reinterpretado como colocación

Para  $j \in \{1, \dots, N-1\}$ , (20) se reduce a

$$\left(\mathcal{F}^N(u^N)' + \gamma u^N - f\right)(x_j) = 0$$

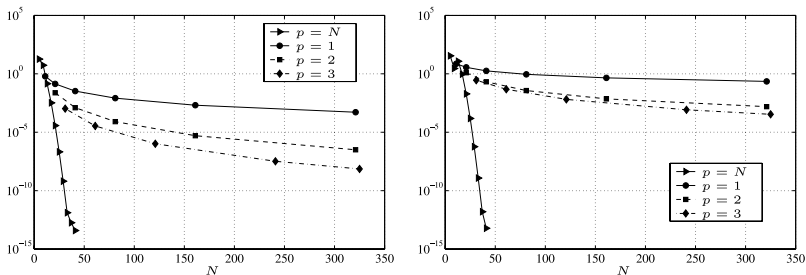
Esto es, en los nodos de cuadratura interiores ocurre que estamos imponiendo la ecuación de cuerpo (14a) **por colocación** pero con el flujo numérico  $\mathcal{F}^N(u^N)$  en lugar del flujo  $\mathcal{F}(u^N)$ .

Para  $j = N$ , (20) se reduce a

$$\left(\mathcal{F}^N(u^N)' + \gamma u^N - f\right)(1) - \frac{1}{w_N} \left(\mathcal{F}(u^N)(1) + g\right) = 0$$

Esto es, en  $x = 1$  ocurre que estamos imponiendo por colocación una combinación lineal particular de la ecuación de cuerpo y la condición de contorno (14d). Puede probarse que  $1/w_N \sim N^2$  como  $N \rightarrow \infty$ , lo que, junto al hecho de que el residuo de la ecuación de cuerpo se impone como cero en nodos interiores cada vez más cercanos a  $x = 1$ , manifiesta que estamos imponiendo la condición de contorno (14d) de forma cada vez más estricta

# Gráficos de convergencia extraídos de [CHQZ06]



**Fig. 1.4.** Comparison between the accuracy of the G-NI solution (corresponding to the curve  $p = N$ ) and the finite-element solutions of order  $p = 1, 2$  and  $3$  versus  $N$  which represents the total number of nodal values. The maximum error between the numerical solution and the exact one  $u(x) = \cos(3\pi(1+x)) \cdot \sin(\pi/5 \cdot (x+0.5)) + \sin(\pi/10)$  (*left*) and the absolute value of the error on the flux at  $x = 1$  (*right*)

## Ejemplos de introducción

Método tau de Legendre para la ecuación de Poisson

## Ecuación de Poisson en un cuadrado

Consideremos la ecuación de Poisson

$$\mathcal{M}(u) := -\Delta u = f \quad \text{en } (-1, 1)^2 \quad (21a)$$

complementada con las condiciones de contorno abajo, arriba, a la izquierda y a la derecha, respectivamente:

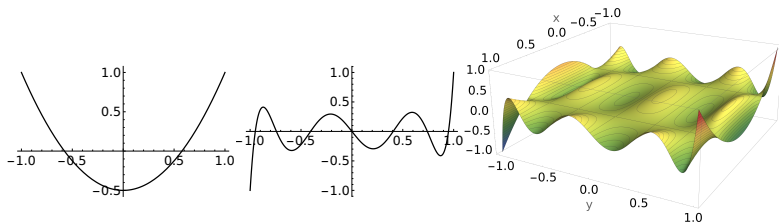
$$\begin{aligned}\mathcal{B}^{(1)}(u) &:= u(\cdot, -1) = 0 \\ \mathcal{B}^{(2)}(u) &:= u(\cdot, 1) = 0 \\ \mathcal{B}^{(3)}(u) &:= u(-1, \cdot) = 0 \\ \mathcal{B}^{(4)}(u) &:= u(1, \cdot) = 0\end{aligned} \quad (21b)$$

## Funciones de forma

Buscamos una aproximación  $u^N$  para la solución  $u$  de la forma

$$u^N(x, y) = \sum_{k,l=0}^N a_{k,l} L_k(x) L_l(y) \quad (22)$$

Esto es, el espacio de forma es la envoltura lineal de la colección de productos tensoriales de polinomios de Legendre  $L_k \otimes L_l$ ,  $0 \leq k, l \leq N$ .



**Figura:** De izquierda a derecha: Gráficos de  $L_2$ , de  $L_7$  y de  $L_2 \otimes L_7$ .

## Funciones de prueba/test: Primer tipo

Para fijar los  $(N + 1)^2$  grados de libertad  $a_{k,l}$  de  $u^N$  nos ayudaremos con dos tipos de funciones test.

El **primer tipo** consiste en las funciones

$$\psi_{k,l} = \frac{2k+1}{2} L_k \otimes \frac{2l+1}{2} L_l, \quad 0 \leq k, l \leq N-2$$

Son las mismas funciones base del espacio de forma, pero menos en número (grado parcial máximo  $N - 2$  en vez de  $N$ ) y reescaladas.

Las empleamos para imponer la ecuación de cuerpo (21a) según

$$(\forall k, l \in \{0, \dots, N-2\}) \quad \int_{(-1,1)^2} \mathcal{M}(u^N) \psi_{k,l} = \int_{(-1,1)^2} f \psi_{k,l} \quad (23)$$

Notemos que no empleamos integración por partes. Además, las condiciones (23) son  $(N - 1)^2$  en número, insuficientes para fijar los  $(N + 1)^2$  grados de libertad  $a_{k,l}$ , pues estamos reservando ecuaciones para la imposición de las condiciones de contorno.



## Funciones de prueba/test: Segundo tipo

El **segundo tipo** de función test consiste en las funciones

$$\chi_k^{(1)}(x) = \frac{2k+1}{2} L_k(x), \quad 0 \leq k \leq N$$

$$\chi_k^{(2)}(x) = \frac{2k+1}{2} L_k(x), \quad 0 \leq k \leq N$$

$$\chi_l^{(3)}(y) = \frac{2l+1}{2} L_l(y), \quad 0 \leq l \leq N$$

$$\chi_l^{(4)}(y) = \frac{2l+1}{2} L_l(y), \quad 0 \leq l \leq N$$

Las usamos para imponer las condiciones de contorno (21b) según

$$\begin{aligned} (\forall (i, k) \in \{1, 2\} \times \{0, \dots, N\}) \quad & \int_{-1}^1 \mathcal{B}^{(i)}(u^N)(x) \chi_k^{(i)}(x) dx = 0 \\ (\forall (i, l) \in \{3, 4\} \times \{0, \dots, N\}) \quad & \int_{-1}^1 \mathcal{B}^{(i)}(u^N)(y) \chi_l^{(i)}(y) dy = 0 \end{aligned} \tag{24}$$

## Redundancia al rescate

¿Ya tenemos suficientes condiciones?

$$\underbrace{(N-1)^2}_{(23)} + \underbrace{4(N+1)}_{(24)} = (N+1)^2 + 4 > \underbrace{(N+1)^2}_{\text{grados de libertad}}$$

¡Ahora son demasiadas!

**Pero** las  $4(N+1)$  condiciones (24) vienen con redundancias: Equivalen a  $4N$  ecuaciones linealmente independientes<sup>2</sup>. Así:

$$\underbrace{(N-1)^2}_{(23)} + \underbrace{4N}_{(24) \text{ destilado}} = (N+1)^2 = \underbrace{(N+1)^2}_{\text{grados de libertad}}$$

---

<sup>2</sup>En efecto, para  $i \in \{1, 2, 3, 4\}$ ,  $\mathcal{B}^{(i)}(u^N) \in \mathbb{P}_N$ . Las condiciones (24) indican que  $\mathcal{B}^{(i)}(u^N)$  es ortogonal a una base de  $\mathbb{P}_N$ ; luego,  $\mathcal{B}^{(i)}(u^N) \equiv 0$ . Esto equivale a requerir que  $\mathcal{B}^{(i)}(u^N)$  se anule en  $N+1$  puntos distintos de  $[-1, 1]$  y nada impide fijar a los extremos como dos de estos puntos. De esta forma, las condiciones (24) equivalen a requerir que  $u^N$  se anule en  $N-1$  puntos interiores distintos y en los dos extremos de cada lado. Como los extremos de los lados son esquinas del cuadrado, en cada una de las cuatro esquinas estamos exigiendo dos veces que  $u^N$  se anule (una vez por cada lado que incide en la esquina), lo que deja de manifiesto que sobran exactamente cuatro condiciones.

## Fórmulas de diferenciación espectral

Se puede probar que

$$\frac{\partial^2 u^N}{\partial x^2} = \sum_{k,l=0}^N a_{k,l}^{(2,0)} L_k \otimes L_l \quad y \quad \frac{\partial^2 u^N}{\partial y^2} = \sum_{k,l=0}^N a_{k,l}^{(0,2)} L_k \otimes L_l$$

donde los coeficientes  $a_{k,l}^{(2,0)}$  y  $a_{k,l}^{(0,2)}$  de las expansiones de estas derivadas y los coeficientes  $a_{k,l}$  de la expansión (22) de  $u^N$  están conectados por

$$\begin{aligned} a_{k,l}^{(2,0)} &= \left(k + \frac{1}{2}\right) \sum_{\substack{p=k+2 \\ p+k \text{ es par}}}^N [p(p+1) - k(k+1)] a_{p,l} \\ a_{k,l}^{(0,2)} &= \left(l + \frac{1}{2}\right) \sum_{\substack{q=l+2 \\ q+l \text{ es par}}}^N [q(q+1) - l(l+1)] a_{k,q} \end{aligned} \tag{25}$$

## Sistema de...

Definiendo

$$(\forall k, l \in \{0, \dots, N-2\}) \quad f_{k,l} = \int_{(-1,1)^2} f \psi_{k,l}$$

y explotando la propiedad (que probaremos más adelante)

$$(\forall j \in \mathbb{N}_0) \quad \int_{-1}^1 L_j^2(x) dx = \frac{2}{2j+1}$$

concluimos que los coeficientes  $a_{k,l}$  de (22) satisfacen el gran sistema de ecuaciones lineales...

... ecuaciones

$$(\forall k, l \in \{0, \dots, N-2\}) \quad - \left( a_{k,l}^{(2,0)} + a_{k,l}^{(0,2)} \right) = f_{k,l} \quad (26a)$$

$$(\forall k \in \{0, \dots, N\}) \quad \sum_{l=0}^N (-1)^l a_{k,l} = 0 \quad (26b)$$

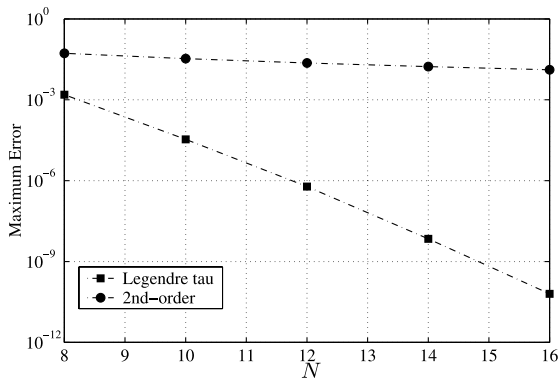
$$(\forall k \in \{0, \dots, N\}) \quad \sum_{l=0}^N a_{k,l} = 0 \quad (26c)$$

$$(\forall l \in \{0, \dots, N\}) \quad \sum_{k=0}^N (-1)^k a_{k,l} = 0 \quad (26d)$$

$$(\forall l \in \{0, \dots, N\}) \quad \sum_{k=0}^N a_{k,l} = 0 \quad (26e)$$

Por supuesto, las ecuaciones que involucran a los  $a_{k,l}^{(2,0)}$  y  $a_{k,l}^{(0,2)}$  se entienden en términos de los  $a_{k,l}$  por medio de la fórmula de diferenciación espectral (25).

## Gráfico de convergencia extraído de [CHQZ06]



**Fig. 1.5.** Maximum errors for the Poisson problem for Legendre tau and second-order finite-difference schemes