Apuntes de Diferencias Finitas

Versión 0.1

Martín Maas

Prólogo

Esta es una primera versión de unos apuntes sobre métodos de Diferencias Finitas para la materia Análisis Numérico del Departamento de Matemática de la FCEyN, UBA. Incluye material de repaso de la materia Cálculo Numérico. El material se encuentra en elaboración, con lo cual se agradece cualquier comentario para corregir, aclarar o mejorar estas notas.

Índice general

1. Discretización de derivadas **5**

2. Problemas de Valores Iniciales para EDO **7**
2.1. Métodos multi-paso 7
2.2. Estabilidad de métodos multi-paso 8
2.3. Cálculos inestables con métodos 0-estables 9
2.4. Estabilidad de polinomios 10

3. Problemas de Valores de Contorno para EDO **12**
3.1. Condiciones de borde de Dirichlet 12
3.2. Condiciones de borde tipo Neumann 15
3.3. Condiciones de borde periódicas 15

4. El método semi-discreto para EDP **18**
4.1. Formulación de los métodos 18
4.2. Error de truncado 19

5. Ecuación del calor **21**
5.1. Método explícito para la ecuación del calor 21
5.2. Método implícito 22
5.3. Método theta 22

6. Análisis de Estabilidad de Von Neumann **23**
6.1. Estabilidad en norma 2 y borde periódico 23
6.2. Ecuación de convección-difusión 23
6.3. Ecuación de reacción-difusión 24

7. Ecuación de transporte **25**
7.1. Método Up-Wind 25
7.2. Condición de CFL 25
7.3. Difusión artificial 25
7.4. Errores de amplitud y de fase 25

8. Teorema de Equivalencia de Lax **26**
8.1. No-suficiencia de la condición espectral 26
8.2. Estabilidad de Lax-Richtmyer 27
8.3. Condiciones suficientes 28
8.4. Cuasi-espectro de Godunov-Riabenki 29

9. Apéndice: Álgebra Lineal **30**

1

Discretización de derivadas

La base de los métodos de diferencias finitas consiste en definir una grilla finita de puntos, y aproximar la derivada de una función por combinaciones lineales de valores de dicha función en nodos vecinos de una grilla.

Definimos las siguientes discretizaciones de la derivada primera, para $h > 0$:

(a) Diferencias Forward: $u'(x) \sim \dfrac{u(x+h) - u(x)}{h}$

(b) Diferencias Backward: $u'(x) \sim \dfrac{u(x) - u(x-h)}{h}$

(c) Diferencias Centradas: $u'(x) \sim \dfrac{u(x+h) - u(x-h)}{2h}$

Queremos estudiar el error local en función de la suavidad de la función u. Para ello, consideramos las expansiones de Taylor

$$u(x \pm h) = u(x) \pm u'(x)h + \frac{u''(\xi)h^2}{2} \qquad \xi \in (x, x \pm h)$$

las cuales son válidas si $u \in C^2$. Reemplazando en las expresiones anteriores podemos obtener expresiones para el error de aproximación, que en los casos de diferencias forward y backward son muy similares:

$$\frac{u(x+h) - u(x)}{h} = u'(x) + \frac{u''(\xi)}{2}h$$

es decir, la aproximación resulta $O(h)$, o de orden 1. En el caso de diferencias centradas necesitamos una expansión de Taylor de orden superior, para lo que pediremos $u \in C^3$, y tenemos

$$u(x \pm h) = u(x) \pm u'(x)h + \frac{u''(x)h^2}{2} \pm \frac{u'''(\xi)h^3}{6} \qquad \xi \in (x, x \pm h)$$

En este caso, observamos que el término de la derivada segunda se cancela en la expresión del error, y que en cambio se obtiene

$$\frac{u(x+h)-u(x-h)}{2h} = u'(x) + O(h^2)$$

es decir, una aproximación de orden 2 para el caso en que $u \in C^3$.

En cuanto a las discretizaciones de la derivada segunda tenemos por supuesto muchas opciones. La discretización más usual viene dada por:

$$\text{Diferencias Centradas:} \quad u''(x) \sim \frac{u(x+h)-2u(x)+u(x-h)}{h^2}$$

Para analizar el error, como estamos esperando tener un orden mayor, hacemos una expansión de Taylor de orden 4

$$u(x \pm h) = u(x) \pm u'(x)h + \frac{u''(x)h^2}{2} \pm \frac{u'''(x)h^3}{6} + \frac{u^{(4)}(\xi)h^4}{12} \qquad \xi \in (x, x \pm h)$$

Observamos que al tener en nuestro operador la suma de $u(x+h)$ y $u(x-h)$, los términos de la expansión de Taylor que aparezcan con signos opuestos se cancelarán, y por lo tanto obtenemos

$$\frac{u(x+h)-2u(x)+u(x-h)}{h^2} = u''(x) + \frac{u^{(4)}(\xi)}{12}h^2 \tag{1.1}$$

es decir, precisión de orden 2.

2

Problemas de Valores Iniciales para EDO

2.1. Métodos multi-paso

Los problemas de valores iniciales que consideraremos son de la forma

$$\begin{cases} U'(t) = & F(t, U(t)), \quad \text{para } t \in (0, t_F) \\ U(0) = & U_0 \end{cases} \tag{2.1}$$

donde tanto U como F son funciones vectoriales.

Comenzamos discretizando el intervalo de tiempo Δt y tomando una cantidad total de pasos temporales N de modo que $N\Delta t = t_F$. Eso nos define una grilla temporal de la forma $\{t_i = i\Delta t\}_{0 \leq i \leq N}$. Sobre esa grilla, definimos nuestra aproximación de diferencias finitas como la sucesión u^i que resulta del método multi-paso de m pasos definido por los parámetros α_j, β_j

$$\sum_{j=0}^{m} \alpha_i u^{n+j} = \Delta t \sum_{j=0}^{m} \beta_i F(t_{n+j}, u^{n+j}). \tag{2.2}$$

En el caso $m = 1$ los métodos son conocidos como métodos de un paso.

Queremos recordar las nociones de consistencia, convergencia y estabilidad. Para ello también necesitaremos definir el error de truncamiento local.

Definición 2.1 (Error de truncamiento local)**.** *Llamaremos τ_i a la diferencia que resulta de evaluar los términos de la ecuación* (2.2) *en la solución exacta del problema* (2.1)*, y normalizando con un factor Δt es decir,*

$$\Delta t \tau_i := \sum_{j=0}^{m} \alpha_i U(t_{n+j}) - \Delta t \sum_{j=0}^{m} \beta_i U'(t_{n+j}).$$

Definición 2.2 (Consistencia)**.** *Diremos que el método dado por la ecuación* (2.2) *es consistente si $\tau_i \to 0$*

Definición 2.3 (Convergencia)**.** *Diremos que el método dado por la ecuación* (2.2) *es convergente a una solución U si, dado $N = \frac{t_F}{\Delta t}$, se tiene que $\|u_N - U(t_F)\| \to 0$ a medida que $\Delta t \to 0$.*

Definición 2.4 (Método convergente)**.** *Diremos que el método multi-paso es convergente si se verifica la convergencia para toda U solución única de un problema de la forma* (2.1) *con F de Lipschitz.*

2.2. Estabilidad de métodos multi-paso

En general, en los métodos multi-paso la consistencia por sí sola no implica la convergencia, sino que necesitamos una condición de estabilidad.

Ejemplo 2.1. *Se considera el método multipaso dado por*

$$u^{n+2} - 3u^{n+1} + 2u^n = -\Delta t f(u^n)$$

cuyo error de truncado resulta

$$\tau^n = \frac{1}{\Delta t}[u(t_{n+2}) - 3u(t_{n+1}) + 2u(t_n) + \Delta t u'(t_n) = \frac{5}{2}\Delta t u''(t_n) + O(\Delta t)^2$$

es decir que el método es consistente y de orden 1. Sin embargo, el error global no solo no converge sino que explota de manera bastante rápida. Esto se puede ver considerando el problema test

$$u'(t) = 0 \qquad u(0) = 0$$

donde el método toma la forma

$$u^{n+2} - 3u^{n+1} + 2u^n = 0. \tag{2.3}$$

Necesitamos definir dos valores iniciales u^0 y u^1. Si tomamos $u^0 = u^1 = 0$ el método "converge" a la solución exacta que es idénticamente nula, pero en general no podremos disponer del valor exacto de u^1 y normalmente allí cometeremos un error, por ejemplo si obtenemos u^1 a partir de aplicar un método de un paso a u^0. Consideremos entonces el caso donde $u^1 = \Delta t$. La solución calculada podría converger a la solución exacta, puesto que $u^1 = \Delta t \to 0$ con $\Delta t \to 0$, pero sin embargo explota. Para ello veamos que el sistema de diferencias lineal puede resolverse fácilmente de manera explícita dando lugar a la expresión:

$$u^n = 2u^0 - u^1 + 2^n(u^1 - u^0).$$

Observación 2.1. *Para hallar la solución general del sistema lineal de diferencias el procedimiento usual es reemplazar $u^n = lambda^n$ y resolver el polinomio resultante, que se conoce como polinomio característico. En el caso antes mencionado las raíces resultan ser $\lambda_0 = 1$ y $\lambda_1 = 2$.*

Cuadro 2.1: Solución u^N con $u^0 = 0, u^1 = \Delta t$ y varios valores de $\Delta t = \frac{1}{N}$

N	u^N
5	6.2
10	1023
20	$5{,}4 \times 10^4$

Definición 2.5 (0-estabilidad)**.** *Diremos que el método dado por la ecuación* (2.2) *es 0-estable si al aplicarlo al problema* $u' = 0, u(0) = 0$ *se obtiene una sucesión que tiende a* 0*.*

Lema 2.1. *La condición de 0-estabilidad de un método multi-paso es equivalente al "criterio de la raiz".*

Teorema 1 (Equivalencia de Dahlquist)**.** *Un método consistente es convergente si y solo si es 0-estable.*

Demostración. La demostración de que la 0-estabilidad es necesaria puede consultarse en "Notas de Cálculo Numérico, Durán Lasalle Rossi". La suficiencia, en cambio, es algo más complicada de demostrar y está detallada, por ejemplo, en "Hairer, Wanner, Nørsett- Solving Ordinary Differential Equations I: Nonstiff Problems" □

Corolario 2.1. *Los métodos de un paso convergen si y solo si son consistentes.*

Demostración. Alcanza con ver que los métodos de un paso son automáticamente 0-estables. Al aplicar un método dado por la ecuación (2.2) con $m = 1$ al problema test de la definición, resulta

$$\alpha_0 u^n + \alpha_1 u^{n+1} = 0$$

y por ende $u^{n+1} = \frac{\alpha_0}{\alpha_1} u^n$, lo que, considerando que las condiciones iniciales son $u^0 = 0$, implica que la solución numérica es constatemente nula, como buscábamos. □

2.3. Cálculos inestables con métodos 0-estables

Si bien la convergencia está garantizada con la consistencia y la 0-estabilidad, muchas veces en la práctica se necesitan resultados más fuertes que la mera convergencia para Δt suficientemente pequeño. En efecto, el Δt necesario para obtener un error dado, puede ser irrazonablemente pequeño.

Definición 2.6. *Un método de diferencias finitas A-estable si, dado* $a > 0$*, al aplicar el método al problema* $u'(t) = -au(t)$ *se obtiene una solución discreta* u^n *tal que* $u^n \to 0$ *con* $n \to \infty$*.*

Observación 2.2. *La A-estabilidad equivale a pedir que la solución aproximada tenga el mismo comportamiento asintótico que la solución exacta, en el caso de los problemas test* $u'(t) = -au(t)$*.*

Consideremos los dos siguientes métodos:

(a) Euler explícito $\quad \dfrac{u^{n+1}-u^n}{\Delta t} = -au^n$

(b) Euler implícito $\quad \dfrac{u^{n+1}-u^n}{\Delta t} = -au^{n+1}$

y estudiemos su A-estabilidad, es decir, si para un Δt fijo se verifica que $u^n \to 0$ cuando $n \to \infty$.

Proposición 2.1. *Euler explícito es A-estable si* $\Delta t < \frac{2}{a}$.

Demostración. Despejando, se tiene

$$y^{n+1} = (1 - a\Delta t)y^n.$$

Por lo tanto, llamando $\lambda = (1 - a\Delta t)$, tenemos

$$y^n = \lambda^n y_0$$

Para que $\lim_{n\to\infty} y^n = 0$ es necesario que $|\lambda| < 1$, es decir,

$$-1 < 1 - a\Delta t < 1$$

Como $a > 0$ la condición de la derecha se cumple siempre, y por lo tanto la condición de A-estabilidad resulta ser $\Delta t < \frac{2}{a}$. En el caso en que $|\lambda| > 1$ tendremos $y_n \to \infty$ con lo cual el método será inestable. □

Proposición 2.2. *Euler implícito es incondicionalmente A-estable.*

Demostración. Para Euler implícito $\lim_{n\to\infty} y_n = 0$ para todo h. Ejercicio. □

2.4. Estabilidad de polinomios

Diremos que un polinomio es estable si todas sus raíces tienen módulo menor que uno. Naturalmente, nos interesará analizar los polinomios característicos de las ecuaciones en recurrencia que definen los métodos que estamos considerando. Por ejemplo, el siguiente resultado facilita el análisis de estabilidad de los esquemas de 2 pasos.

Lema 2.2. *Dada la ecuación cuadrática* $z^2 + bz + c = 0$ *con* b *y* c *en* $\mathbb{R}$*, las raices están en el círculo unitario* $\Leftrightarrow |c| \leq 1$ *y* $|b| \leq 1 + c$.

Demostración. Dividamos la demostración en dos casos, según el signo de $b^2 - 4c$.

(a) $b^2 \leq 4c$. Las raices en este caso complejas vienen dadas por

$$z_{1,2} = \frac{-b \pm i\sqrt{4c - b^2}}{2}$$

y por lo tanto

$$|z_{1,2}|^2 = \frac{b^2 + 4c - b^2}{4} = c$$

Es decir, en el caso complejo la condición $|c| \leq 1$ es necesaria y suficiente para asegurar que las raíces estarán en el interior del círculo unitario.

Veamos que, la condición $|b| \leq 1+c$ se satisface automáticamente en este caso. Tenemos $b^2 \leq 4c$. Esto en particular implica que $c \geq 0$. Ahora, partiendo de $(1-c)^2 \geq 0$, sumando $4c$ en ambos lados tenemos $(1-c)^2 + 4c \geq 4c$. Distribuyendo el cuadrado y reagrupando, obtenemos $(1+c)^2 \geq 4c$. Como teníamos que $4c \geq b^2$, resulta $(1+c)^2 \geq b^2$, y tomando raiz cuadrada, obtenemos $|1+c| \geq |b|$. Como $c \geq 0$, $|1 + c| = 1 + c$, como queríamos.

(a) $b^2 \geq 4c$. Las raices en este caso son reales, y vienen dadas por

$$z_{1,2} = \frac{-b \pm \sqrt{b^2 - 4c}}{2}$$

y por lo tanto $|z| \leq 1$ equivale a $|-b + \sqrt{b^2 - 4c}| \leq 2$ Tenemos

$$c = z_0 z_1 \qquad b = -(z_0 + z_1)$$

Desarrollando la desigualdad derecha, tenemos

$$|b| \leq 1 + c \quad \Leftrightarrow \quad |z_0 + z_1| \leq 1 + z_0 z_1$$

Ahora bien, si $-1 \leq z_0$ y $-1 \leq z_1$, entonces multiplicando por $1 + z_1$ en la primera hipótesis tenemos

$$-1 - z_1 \leq z_0 + z_0 z_1$$

y sumando z_1 en la desigualdad izquierda tenemos

$$-1 \leq z_1 + z_0 + z_0 z_1$$

Ahora supongamos $z_0 \leq 1$ y $z_1 \leq 1$ y multipliquemos $z_0 \leq 1$ por $(-1 + z_1)$, para obtener

$$-1 \leq -z_0 - z_1 + z_0 z_1$$

Hemos obtenido la desigualdad $|b| \leq 1 + c$ en los dos casos posibles del signo de $b = -(z_0 + z_1)$. Por lo tanto, la desigualdad resulta equivalente a ...

Completar.

□

Ejemplo 2.2. *El método multipaso que resulta de aplicar diferencias centradas en la primera derivada, dado por*

$$\frac{u^{n+1} - u^{n-1}}{2\Delta t} = f(u^n)$$

es 0-estable pero no es A-estable para ningún $a < 0$.

3

Problemas de Valores de Contorno para EDO

Comenzamos esta primer sección con métodos de diferencias finitas para ecuaciones diferenciales ordinarias. Consideraremos por un lado problemas de valores iniciales y problemas de valores de contorno.

3.1. Condiciones de borde de Dirichlet

Consideremos como primer ejemplo la resolución numérica de la ecuación de Poisson en una dimensión, con condiciones de borde de tipo Dirichlet

$$\begin{cases} u_{xx}(x) = & f(x), \quad \text{para } x \in (0,1) \\ u(0) = & \alpha \\ u(1) = & \beta. \end{cases} \tag{3.1}$$

Para ello se considera la malla uniforme $\{x_j = hj,\ j = 0, 1, 2, \ldots m+1\}$ con $h = 1/(m+1)$. Para los puntos de la malla $x_j \in (0,1)$ El esquema de diferencias centradas para la derivada segunda conduce al sistema de ecuaciones:

$$\frac{1}{h^2}\left(U_{j-1} - 2U_j + U_{j+1}\right) = f(x_j) \qquad \text{para } j = 1, 2, 3, \ldots m.$$

Utilizando las condiciones de borde $U_0 = \alpha, U_{m+1} = \beta$ se obtiene el sistema

$$A^h U^h = F^h \tag{3.2}$$

donde $U^h = [U_1, U_2, \ldots, U_m]^T$ es el vector de incógnitas, mientras que la matriz tridiagonal A^h y el vector F^h están dados por:

$$A^h = \frac{1}{h^2}\begin{bmatrix} -2 & 1 & & & & \\ 1 & -2 & 1 & & & \\ & 1 & -2 & 1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & 1 & -2 & 1 \\ & & & & 1 & -2 \end{bmatrix}, \quad F^h = \begin{bmatrix} f(x_1) - \alpha/h^2 \\ f(x_2) \\ f(x_3) \\ \vdots \\ f(x_{m-1}) \\ f(x_m) - \beta/h^2 \end{bmatrix} \tag{3.3}$$

En el camino de demostrar la convergencia del método en cuestión, introducimos algunas definiciones que utilizaremos luego en casos más generales.

Definición 3.1. *Para h fijo, definimos el error puntual de la solución numérica en x_j como*

$$e_j^h = u(x_j) - U_j^h, \qquad e^h = [e_1^h, \cdots, e_m^h]$$

Definición 3.2. *Llamaermos error de truncado local τ_j^h a la magnitud que resulta de evaluar el esquema numérico en la solución exacta u de* (3.1)

$$\tau_j^h = (A^h u)_j - u_{xx}(x_j), \qquad \tau^h = [\tau_1^h, \cdots, \tau_m^h]$$

El error de truncado lo podemos obtener fácilmente, a partir del estudio sobre la precisión de la discretización de las derivadas.

Proposición 3.1. *Si $f \in C^2(0,1)$ existe una constante C independiente de h y de m, tal que $\max_{1\leq j\leq m} \left|\tau_j^h\right| < Ch^2$.*

Demostración. Como $f \in C^2(0,1)$ y $u_{xx} = f$, resulta que $u \in C^4$, y entonces podemos emplear la expresión del error de aproximación de las diferencias centradas (1.1) para obtener la expresión $\tau_j^h = h^2 u^{(4)}(\xi_j)$. Como estamos en un invervalo acotado, la función $u^{(4)}$ alcanza un máximo C lo que concluye la demostración. □

Ahora bien, quisiéramos pasar de una estimación del error de truncado a una estimación del error en la solución. Para ello, observamos que al aplicarle A^h al error, se obtiene el error de truncado:

$$A^h e^h = A^h u - A^h U = A^h u - F = A^h u - u_{xx} = \tau^h,$$

y que suponiendo A^h inversible (lo menos que podemos pedir) resulta

$$e^h = (A^h)^{-1}\tau^h.$$

Por lo tanto, una condición suficiente para tener convergencia es que se verifique la siguiente condición que llamaremos de estabilidad

Definición 3.3. *Una discretización es estable en norma $\|\cdot\|$ si existe C tal que $\|A_h^{-1}\| \leq C \quad \forall h$*

En tal caso, podremos asegurar que el error de la solución (en la norma escogida) está controlado por los errores de truncamiento, ya que resultará:

$$\|e^h\| \leq C\|\tau^h\| \to 0$$

Verifiquemos la condición de estabilidad de la definición 3.3 para el caso particular de la matriz A^h dada por la ecuación (3.3). Vamos a hacerlo en norma 2. Para ello, recordemos cómo se calculaba la norma 2 de una matriz.

Proposición 3.2. *Para toda $A \in \mathbb{R}^{n\times n}$ se tiene $||A||_2 = \sqrt{\rho(A^t A)}$*

Demostración. Demostremos (c). Como la matriz A^tA es simétrica, ello implica que se diagonaliza en una base ortonormal. Ponemos $A^tAv_i = \mu_i v_i$ y tomemos $x = \sum_{i=1}^n \alpha_i v_i$. Luego

$$\|Ax\|_2^2 = \langle Ax, Ax\rangle = \langle x, A^tAx\rangle = \langle \sum_{i=1}^n \alpha_i v_i, \sum_{i=1}^n \mu_i \alpha_i v_i\rangle = \sum_{i=1}^n \mu_i \alpha_i^2$$

y entonces $\|Ax\|_2^2 \leq |\mu_{\max}| \|x\|_2^2$. La demostración concluye notando que la desigualdad se alcanza si tomamos μ_i. □

Corolario 3.1. *Si A es simétrica entonces $||A||_2 = \rho(A)$.*

Entonces tenemos que calcular los autovalores de A^h. Para ello, calculemos los autovalores de una matriz un poco más general. Lo haremos a modo "galerazo", adivinando los autovectores. Para una demostración un poco más deductiva, ver referencia X.

Proposición 3.3. *Dados $a, b, c \in \mathbb{R}$ definimos la matriz $A \in \mathbb{R}^{n\times n}$ dada por*

$$A = \begin{bmatrix} a & b & & & & \\ c & a & b & & & \\ & c & a & b & & \\ & & \ddots & \ddots & \ddots & \\ & & & c & a & b \\ & & & & c & a \end{bmatrix} \tag{3.4}$$

Llamando $h = 1/(n+1)$ y considerando $w^2 = c/b$, el q-ésimo autovector r^q de A está dado por

$$r_j^q = w^j \sin(q\pi jh), \qquad r^q = [r_1^q, \cdots, r_n^q],$$

y el correspondiente q-ésimo autovalor resulta

$$\lambda^q = a + 2bw\cos(q\pi h).$$

Demostración. Verifiquemos que la expresión dada para los autovectores es correcta y calculemos los autovalores. En efecto, si $1 < j < n$,

$$\begin{aligned}(Ar^q)_j &= cr_{j-1}^q + ar_j^q + br_{j+1}^q \\ &= cw^{j-1}\sin(q\pi(j-1)h) + aw^j\sin(q\pi jh) + bw^{j+1}\sin(q\pi(j+1)h) \\ &= cw^{j-1}\cos(q\pi h)\sin(q\pi jh) + aw^j\sin(q\pi jh) + bw^{j+1}\cos(q\pi h)\sin(q\pi jh) \\ &= r_j^q\left(cw^{-1}\cos(q\pi h) + a + bw\cos(q\pi h)\right) \\ &= r_j^q\lambda^q\end{aligned}$$

para los casos $j = 1, j = n$, observamos que $r_j^q = 0$ y que por lo tanto se satisfacen las mismas expresiones. □

Para concluir la demostración de la estabilidad de A^h nos resta observar que sus autovalores resultan $\lambda^q = \frac{1}{h^2}(-2+2cos(q\pi h))$ y por lo tanto el autovalor de mínimo módulo de A^h es

$$\begin{aligned}\lambda_1 &= \frac{2}{h^2}\left(\cos(q\pi h) - 1\right) \\ &= \frac{2}{h^2}\left(-\frac{1}{2}\pi^2 h^2 + O(h^4)\right) \\ &= \pi^2 + O(h^2)\end{aligned}$$

lo que implica que, como la matriz es simétrica, $\|(A^h)^{-1}\|_2 \leq \frac{1}{\pi^2}$.

3.2. Condiciones de borde tipo Neumann

Por último, consideremos un problema con condiciones de borde tipo Neumann.

$$\begin{cases} u_{xx}(x) = & f(x), \quad \text{para } x \in (0,1) \\ u_x(0) = & XX \\ u_x(1) = & XX. \end{cases} \tag{3.5}$$

Para discretizar el problema utilizaremos diferencias centradas para la derivada segunda, junto con distintas opciones para las derivadas en el borde.

3.3. Condiciones de borde periódicas

Consideremos ahora una variante del problema 3.1 que estabamos analizando, tomando condiciones de borde periódicas en vez de condiciones de Dirichlet. Así, tenemos el problema

$$\begin{cases} u_{xx}(x) = & f(x), \quad \text{para } x \in (0,2\pi) \\ u(0) = & u(2\pi) \end{cases} \tag{3.6}$$

Discretizamos como antes, mediante diferencias centradas para la derivada segunda.

$$\frac{1}{h^2}\left(U_{j-1} - 2U_j + U_{j+1}\right) = f(x_j), \quad 1 \leq j \leq m. \tag{3.7}$$

junto con la condición de borde periódica

$$U_0 = U_m \tag{3.8}$$

Este tipo de problemas se pueden resolver mediante análisis de Fourier, tanto en su versión continua como en su discretización mediante diferencias finitas. Para ello, encontraremos una base de autofunciones de la matriz del problema. La primera observación que nos puede guiar a encontrar estas autofunciones es que, así como

los modos de fourier $e^{i\xi x}$ son autofunciones de la derivada $\frac{\partial}{\partial x}$ con autovalor $i\xi$, es decir, satisfacen

$$\frac{\partial}{\partial x} e^{i\xi x} = i\xi e^{i\xi x},$$

sucede que versiones discretas de $e^{i\xi x}$ se comportan de un modo similar con respecto a los operadores de diferencias finitas. Hagamos los cálculos para los operadores de diferencia centradas para la derivada primera y para la derivada segunda

$$\delta_x(V) = \frac{1}{2h}\left(V_{j+1} - V_{j-1}\right)$$

$$\delta_{xx}(V) = \frac{1}{h^2}\left(V_{j+1} - 2V_j + V_{j-1}\right).$$

Lema 3.1. *considerando una malla equiespaciada $x = \{x_j = jh\}$, las funciones $W_j = e^{i\xi x_j} = e^{i\xi jh}$ son autofunciones de los operadores δ_x, δ_{xx}.*

Demostración. Se tiene que para cualquier h

$$\begin{aligned}
\delta_x(W)_j &= \frac{1}{2h}\left(e^{i(j+1)h\xi} - e^{i(j-1)h\xi}\right) \\
&= \frac{1}{2h}\left(e^{ih\xi} - e^{-ih\xi}\right) e^{ijh\xi} \\
&= \frac{i}{h}\sin(h\xi) e^{ijh\xi} \\
&= \frac{i}{h}\sin(h\xi) W_j.
\end{aligned} \tag{3.9}$$

Para el operador de la derivada segunda, tenemos

$$\begin{aligned}
\delta_{xx}(W)_j &= \frac{1}{h^2}\left(e^{i(j+1)h\xi} - 2e^{ijh\xi} + e^{i(j-1)h\xi}\right) \\
&= \frac{1}{h^2}\left(e^{ih\xi} - 2 + e^{-ih\xi}\right) e^{ijh\xi} \\
&= \frac{2}{h^2}\left(\cos(h\xi) - 1\right) e^{ijh\xi} \\
&= \frac{2}{h^2}\left(\cos(h\xi) - 1\right) W_j \\
&= -\frac{4}{h^2}\sin^2\left(\frac{h\xi}{2}\right) W_j
\end{aligned} \tag{3.10}$$

□

donde la última igualdad es consecuencia de la identidad $\cos(2x) = 1 - 2\sin^2(x)$

Observación 3.1. *En ambos casos, los autovalores discretos son aproximaciones de los autovalores continuos con precisión $O(h^2)$. En detalle, para la derivada primera tememos $g(\xi) = \frac{i}{h}\sin(h\xi) \sim i\xi + O(h^2)$, y para la derivada segunda el autovalor es $g(\xi) = -\frac{4}{h^2}\sin^2\left(\frac{h\xi}{2}\right)$, que aproxima al autovalor exacto con orden $O(h^2)$.*

Lema 3.2. *Los modos de Fourier discretos* $W(k) = \{\frac{e^{kijh}}{\sqrt{2\pi}}\}$ *con* $h = 2\pi/(J+1)$, $k \in \mathbb{Z}$, *son ortogonales con respecto al producto interno* $(v, w) = h\sum_{j=0}^{J} v_j \bar{w}_j$

Demostración. Dado que $e^{k_1 ijh} e^{-k_2 ijh} = e^{(k_1-k_2)ijh}$, alcanza con probar que

$$h\sum_{j=0}^{J} e^{kijh} = 0 \qquad \text{si } k \neq 0 \tag{3.11}$$

ya que, claramente, $(W(k), W(k)) = 1$. Utilizando la suma de la serie geométrica $\sum_{j=0}^{n-1} x^j = (1-x^n)/(1-x)$ obtenemos

$$h\sum_{j=0}^{J} \left(e^{kih}\right)^j = h\frac{1-e^{kih(J+1)}}{1-e^{kih}} = h\frac{1-e^{2\pi ik}}{1-e^{kih}} = 0$$

siempre que $1 - e^{kih} \neq 0$, lo que puede garantizarse cuando $|k| \leq J$. □

Ahora que sabemos que los modos de Fourier discretos forman una base ortonormal de autofunciones del operador, en particular hemos demotrado que

Observación 3.2. *La matriz* B^h *que representa el problema discreto definido en las ecuaciones* (3.7)-(3.8) *se diagonaliza en una base ortonormal.*

Demostración. No es necesario hallar de forma explícita la matriz, sino que alcanza con observar que el cálculo que ya realizamos en la ecuación (3.10) ya lo demuestra. □

Para ver la estabilidad, como sabemos que la matriz del problema se diagonaliza en una base ortonormal, alcanza con acotar inferiormente el mínimo de los autovalores. Llamativamente, el cálculo es el mismo que antes para las condiciones de Dirichlet:

$$\begin{aligned}\lambda_1 &= \frac{2}{h^2}\left(\cos(q\pi h) - 1\right) \\ &= \frac{2}{h^2}\left(-\frac{1}{2}\pi^2 h^2 + O(h^4)\right) \\ &= \pi^2 + O(h^2)\end{aligned}$$

lo que implica que $\|(B^h)^{-1}\|_2 \leq \frac{1}{\pi^2}$.

4

El método semi-discreto para EDP

Para discretizar este problema tenemos que utilizar una grilla temporal y otra espacial. Dados Δt, Δx, un número total de pasos N_t y una cantidad total de puntos N_x, definimos entonces las grillas

$$\{t_i = i\Delta t\}_{0 \leq i \leq N_t} \qquad \{x_j = j\Delta x\}_{0 \leq j \leq N_x}$$

Definamos los errores de discretización y de truncamiento:

Definición 4.1. *El error de discretización en la malla, viene dado por*

$$e_j^n = U(x_j, t_n) - u_j^n \tag{4.1}$$

Definición 4.2. *El error de truncado viene dado por la diferencia que surge al reemplazar la solución exacta en la definición del método discreto, es decir,*

$$T(x_j, t_n) = XXX \tag{4.2}$$

4.1. Formulación de los métodos

Para obtener métodos de resolución numérica de una ecuación en derivadas parciales como

$$U_t = U_{xx} \tag{4.3}$$

un enfoque es un enfoque es considerar primero una malla espacial (con $h = \Delta x$) y aproximar $\frac{\partial^2}{\partial x^2}$ con una matriz A, dejando la derivada temporal sin discretizar. Es decir, aproximamos $U_{xx} \approx A$ y planteamos el problema semi-discreto

$$u' = Au. \tag{4.4}$$

Para resolver este problema, podemos aplicar cualquier método para resolución de ecuaciones diferenciales ordinarias (incluyendo el uso de paquetes de software como las rutinas `ode45,ode23s` de Matlab). Por ejemplo, podemos utilizar un método multipaso de m pasos definido por los parámetros α_i, β_i, y $k = \Delta t$, y obtenemos el esquema de diferencias finitas:

$$\sum_{i=0}^{m} \alpha_i u^{n+i} = k \sum_{i=0}^{m} \beta_i A u^{n+i}.$$

Algunos métodos para $U_t = U_{xx}$, tanto explícitos como implícitos, de orden $O(h^2)$ y diferentes precisiones en Δt (que puede determinarse de la manera descripta) son:

(a) Crank-Nicolson (Adams-Moulton de 1 paso):

$$\frac{u_j^{n+1} - u_j^n}{\Delta t} = \frac{\frac{1}{2}\delta_{xx}(u_j^n) + \frac{1}{2}\delta_{xx}(u_j^n)}{(\Delta x)^2}$$

(b) Adams-Bashforth de 2 pasos:

$$\frac{u_j^{n+1} - u_j^n}{\Delta t} = \frac{3}{2}\delta_{xx}(u_j^n) - \frac{1}{2}\delta_{xx}(u_j^{n-1})$$

(c) Método totalmente implícito de 2 pasos (BDF2):

$$\frac{3}{2}u_j^{n+1} - 2u_j^n + \frac{1}{2}u_j^{n-1} = r\delta_{xx}(u_j^{n+1})$$

4.2. Error de truncado

Si la aproximación espacial es de orden h^p, tenemos que para cualquier función U, vale

$$U_{xx} = AU + Ch^p \tag{4.5}$$

donde $C = C(x,t)$ es independiente de h.

El error de truncado T satisface

$$kT = \sum_{i=0}^{m} \alpha_i U^{n+i} - k \sum_{i=0}^{m} \beta_i A U^{n+i} \tag{4.6}$$

Nos interesa estimar T utilizando que U es la solución exacta de (4.3). Para ello consideremos la expansión de Taylor en el tiempo

$$U(t_{n+i}, \cdot) = U(t_n, \cdot) + U_t(t_n, \cdot)(ik) + U_{tt}(t_n, \cdot)\frac{(ik)^2}{2} + \cdots + O(k^{q+1}).$$

Reemplazando en (4.6) y separando las diferentes potencias de k, se obtiene (llamando $U(t_n, \cdot)$ al vector $U(t_n, x_j)$)

$$
\begin{aligned}
kT = & \sum_{i=0}^{m} \alpha_i U(t_n, \cdot) \\
& + k \sum_{i=0}^{m} i\alpha_i U_t(t_n, \cdot) - \beta_i AU(t_n, \cdot) \\
& \qquad\qquad \vdots \\
& + k^q \sum_{i=0}^{m} \frac{i^q \alpha_i}{q!} \frac{\partial^q}{\partial t^q} U(t_n, \cdot) - \frac{i^{q-1}\beta_i}{(q-1)!} \frac{\partial^{q-1}}{\partial t^{q-1}} AU(t_n, \cdot) \\
& + O(k^{q+1})
\end{aligned}
\tag{4.7}
$$

Si tuvieramos $U_t = AU$ (es decir, la solución exacta del sistema de ecuaciones ordinarias (4.4)), la condición sobre los α_i, β_i para tener orden r en el tiempo, sería que se anulen los coeficientes

$$
d_q = \sum_{i=0}^{m} \frac{i^q \alpha_i}{q!} - \frac{i^{q-1}\beta_i}{(q-1)!}, \qquad \text{para} \qquad 1 \le q \le r,
$$

junto con $d_0 = \sum_{i=0}^{m} \alpha_i$.

Pero no se tiene $U_t = AU$, sino $U_t = U_{xx}$. Intercalando AU en (4.3), y utilizando (4.5), se obtiene

$$
U_t = AU + O(h^p). \tag{4.8}
$$

Podemos obtener expresiones de la forma $U_{tt} = AU_t + O(h^p)$ tomando derivadas temporales en la expresión anterior (suponiendo que la constante $C = C(x,t)$ en (4.5) admite derivadas temporales). Reemplazando estas expresiones en (4.7) vemos que los mismos términos que corresponden a los coeficientes $d_q = 0$ se anulan (igual que en el caso de ordinarias), resultando

$$
T \sim O(h^r) + O(h^p).
$$

En definitiva, hemos demostrado el siguiente

Lema 4.1. *Un método semi-discreto preserva el error de truncado espacial y su error de truncado temporal coincide con el del método multipaso aplicado a una ODE.*

5

Ecuación del calor

5.1. Método explícito para la ecuación del calor

La primera ecuación en derivadas parciales que consideraremos es la ecuación del calor. Para aproximar la ecuación $U_t = U_{xx}$ se utiliza el esquema en diferencias finitas que se obtiene al realizar una discretización explícita en la variable temporal y diferencias centradas para la segunda derivada espacial:

$$u_j^{n+1} = ru_{j-1}^n + (1-2r)u_j^n + ru_{j+1}^n \quad \text{donde } r = \frac{\Delta t}{(\Delta x)^2} = k/h^2$$

Proposición 5.1. *El error de discretización satisface la siguiente ecuación de recurrencia:*

$$e_j^{n+1} = re_{j-1}^n + (1-2r)e_j^n + re_{j+1}^n + kT(x_j, t_n)$$

Proposición 5.2. *Suponiendo que U tiene derivadas continuas y acotadas hasta el tercer orden en t, y hasta de orden seis en x, el error de truncado para el problema X puede expresarse como:*

$$T(x_j, t_n) = \frac{h^2}{12}\left(6rU_{tt} - U_{xxxx}\right)_{j,n} + \frac{k^2}{6}U_{ttt}(x_j, t_n + \theta_n k) - \frac{h^4}{360}U_{xxxxxx}(x_j + \theta_j h, t_n)$$

con $-1 < \theta_j < 1$, $0 < \theta_n < 1$.

(a) Pruebe que para $r > 0$ el error de truncado es $O(h^2)$, y que en el caso $r = 1/6$ es $O(h^4)$.

(b) Probar que si $0 < r \leq 1/2$ entonces el error global E_n dado por

$$E_n = \max_{0 \leq j \leq N} \{|e_j^n|\},$$

satisface la siguiente estimación en función del tiempo:

$$E_n \leq tM,$$

donde M es el valor máximo de $|T|$. En particular, observe que si además $t \leq T_f$ y U_{tt} y U_{xxxx} estan acotadas luego

$$E_n \leq CT_f k \tag{5.1}$$

para una constante C independiente de h y k.

5.2. Método implícito

Aplicando diferencias backward en el tiempo se obtiene el método implícito.

5.3. Método theta

Otra opción posible es tomar una combinación de diferencias forward y backward en el tiempo, ponderadas mediante un factor θ.

6

Análisis de Estabilidad de Von Neumann

6.1. Estabilidad en norma 2 y borde periódico

Consideremos problemas de EDP con condiciones de borde periódicas. Cuando estudiamos problemas de valores de contorno, hemos visto que los modos de Fourier formaban una base ortonormal de autovalores de los operadores de diferencias, bajo esas condiciones de borde. Por lo tanto, cabe preguntarse cómo se podrían usar esas herramientas para simplificar el análisis de estabilidad.

6.2. Ecuación de convección-difusión

Consideremos el problema de convección-difusión, dado por

$$u_t + au_x = \mu u_{xx} \tag{6.1}$$

y para resolverlo, el método dado por una discretización explícita en el tiempo y con diferencias centradas de oden 2 en el espacio:

$$\frac{u_j^{n+1} - u_j^n}{\Delta t} + a\frac{u_{j+1}^n - u_{j-1}^n}{2\Delta x} = \mu\frac{u_{j+1}^n - 2u_j^n + u_{j-1}^n}{(\Delta x)^2} \tag{6.2}$$

como antes, consideramos condiciones de borde periódicas (o "ignoramos" las condiciones de borde). En el siguiente lema, establecemos las condiciones para asegurar la estabilidad de (6.2) por el método de Fourier.

Lema 6.1. *Llamando* $r = \frac{\mu\Delta t}{(\Delta x)^2}$, $p = \frac{a}{\mu}\frac{\Delta x}{2}$, *el método dado por* (6.2) *es estable por el método de Fourier, siempre que* $r \leq \frac{1}{2}$ *y* $rp^2 \leq \frac{1}{2}$.

Demo. Escribimos el método en la forma

$$u_j^{n+1} = -a\frac{\Delta t}{2\Delta x}\left(u_{j+1}^n - u_{j-1}^n\right) + \frac{\mu\Delta t}{(\Delta x)^2}\left(u_{j+1}^n - 2u_j^n + u_{j-1}^n\right) + u_j^n$$

Llamando $r = \frac{\mu \Delta t}{(\Delta x)^2}$ y $rp = -a\frac{\Delta t}{2\Delta x}$, es decir, $p = \frac{a}{\mu}\frac{\Delta x}{2}$, utilizando el método de Fourier (y notando $\beta = \xi h/2$), obtenemos

$$\begin{aligned}\lambda &= 1 + r(e^{i\xi h} - 2 + e^{-i\xi h}) - rp(e^{i\xi h} - e^{-i\xi h}) \\ &= 1 - 4r\sin^2(\beta) - 2irp\sin(\xi h)\end{aligned}$$

Tomando el módulo y utilizando la identidad $\sin(2\theta) = 2\sin(\theta)\cos(\theta)$ en la parte imaginaria,

$$\begin{aligned}|\lambda|^2 &= \left[1 - 4r\sin^2(\beta)\right]^2 + 4r^2p^2\sin^2(\xi h) \\ &= 1 + 16r^2\sin^4(\beta) - 8r\sin^2(\beta) + 16r^2p^2\sin^2(\beta)\cos^2(\beta) \\ &= 1 + 8r\sin^2(\beta)\left[2r\sin^2(\beta) - 1 + 2rp^2\cos^2(\beta)\right] \\ &= 1 - 8r\sin^2(\beta)\left[1 - \cos^2(\beta) + \cos^2(\beta) - 2r\sin^2(\beta) - 2rp^2\cos^2(\beta)\right] \\ &= 1 - 8r\sin^2(\beta)\left[(1-2r)\sin^2(\beta) + (1-2rp^2)\cos^2(\beta)\right]\end{aligned}$$

Claramente, si $r \leq \frac{1}{2}$ y $rp^2 \leq \frac{1}{2}$, se tiene $|\lambda| \leq 1$ y por lo tanto estabilidad por el método de Fourier. □

6.3. Ecuación de reacción-difusión

7

Ecuación de transporte

7.1. Método Up-Wind

7.2. Condición de CFL

7.3. Difusión artificial

Varios métodos para el problema del transporte lineal $u_t + au_x = 0$ son en verdad, discretizaciones del problema de convección-difusión (6.1), con un cierto valor de μ añadido artificialmente con la discretización, es decir, $\mu \sim \Delta x$ o bien $\mu \sim \Delta t$, etc. Por ejemplo los métodos up-wind, Lax-Friedrichs y Lax-Wendroff provinen de discretizar esta ecuación con diferencias forward en el tiempo y centradas de orden 2 en el espacio (6.2) y con distintos valores de μ.

Observación 7.1. *Más aún, el método de Lax-Wendroff puede derivarse a partir de calcular el μ que maximiza el orden de truncado del método dado por* (6.2)*, para la ecuación de transporte lineal $u_t + au_x = 0$.*

Observación 7.2. *El análisis de estabilidad para la ecuación de convección-difusión del Lema 6.1 permite obtener condiciones de estabilidad por el método de Fourier para los métodos Up-Wind, Lax-Friedrichs y Lax-Wendroff para la ecuación del transporte.*

7.4. Errores de amplitud y de fase

8

Teorema de Equivalencia de Lax

8.1. No-suficiencia de la condición espectral

La impresión que uno puede tener después de haber estudiado las secciones anteriores, es que si tenemos una discretización de una EDP mediante diferencias finitas, de la forma

$$u^{n+1} = M^\nu u^n$$

alcanza con que

$$\rho(M^\nu) < 1 \qquad \forall \nu$$

para garantizar la convergencia. Esto parece una condición bastante fuerte, ya que sabemos que una matriz cuyo radio espectral es menor que 1 satisfacerá

$$\lim_{n\to\infty} M^n = 0 \tag{8.1}$$

Sin embargo, veamos un contra-ejemplo de que esta condición no ganartiza la convergencia. Para ello, consideremos nuevamente la ecuación de transporte lineal

$$U_t + U_x = 0 \tag{8.2}$$

pero, a diferencia de los análisis anteiores, consideraremos condiciones de borde no-periódicas, específicamente condiciones de Dirichlet homogéneas en el origen

$$U(0,t) = 0 \tag{8.3}$$

junto con condiciones iniciales dadas por una función arbitraria, es decir,

$$U(x,0) = f(x) \tag{8.4}$$

Consideremos la discretización que resulta de tomar diferencias forward en la derivada temporal, y backward en la derivada espacial. Llamando $\nu = \frac{\Delta t}{\Delta x}$, tenemos

$$u_j^{n+1} - u_j^n = \nu(u_{j-1}^n - u_j^n) \qquad u_0^n = 0. \tag{8.5}$$

En términos de una expresión matricial, tenemos

$$u^{n+1} = M^\nu u^n \qquad u_i^0 = f(x_i)$$

donde la matriz M^ν viene dada por

$$M^\nu = \begin{bmatrix} 1-\nu & & & \\ \nu & 1-\nu & & \\ & \ddots & \ddots & \\ & & \nu & 1-\nu \end{bmatrix} \tag{8.6}$$

Como la matriz es triangular superior, sus autovaloes están en la diagonal y por lo tanto el radio espectral es $|1-\nu|$. Claramente,

$$\rho(M) \leq 1 \qquad \text{si y solo si} \qquad 0 \leq \nu \leq 2.$$

Sería un error concluir que hay estabilidad cuando estas condiciónes se satisfacen, porque alugnos de los valores de ν permitidos por este criterio violan la condición de CFL, que era necesaria para la convergencia. Por ejemplo, si tomamos $\nu = 1{,}5$, tendremos $\rho(M) = 0{,}5$ pero sin embargo el método no puede ser convergente.

Observación 8.1. *La matriz M en los casos $\nu > 1$ satisface $\rho(M) < 1$ pero sin embargo $\|M\|_2 > 1$. Por eso el análisis de los autovalores no garantiza la estabilidad en norma 2. Claramente, las matrices no son normales, a diferencia de las matrices que solían aparecer (o estaban implícitas) en los casos de condiciones de borde periódicas. Ver Lema XXX.*

Para obtener una condición suficiente de estabilidad para el esquema propuesto (8.5), podemos recurrir a la norma infinito, observando que los coeficientes son positivos y suman 1 siempre que $\nu \leq 1$, y en ese caso se tiene $\|M\|_\infty \leq 1$. El requerimiento de que $\nu \leq 1$ es por lo tanto necesario y suficiente para la convergencia.

8.2. Estabilidad de Lax-Richtmyer

Por lo visto en la sección anterior, por más que una discretización de diferencias finitas verifique

$$\lim_{n\to\infty} M^n = 0 \qquad \forall M \in \mathcal{M} \tag{8.7}$$

donde $\mathcal{M}$ es la colección de matrices indexada con $h = \Delta x$, los métodos resultantes pueden no ser convergentes. Esto nos motiva a introducir una definición más fuerte de estabilidad, que resultará necesaria y suficiente para la convergencia. El criterio de estabilidad que a partir de ahora tomaremos como definición al analizar EDPs es el siguiente:

Definición 8.1 (Estabilidad de Lax-Richtmyer)**.** *Un método es estable si y solo si existe una constante C independiente de h,Δt, y un cierto τ tal que*

$$\|M^n\| \leq C \qquad \textit{para todo} \qquad 0 \leq n\Delta t \leq T_f$$

y para todo $\Delta t \leq \tau$.

Definición 8.2 (Estabilidad fuerte o práctica)**.** *Una condición que llamamos estabilidad fuerte es tener* $\|M\| \leq 1$*. Esto implica Lax-Richtmyer con* $C = 1$*.*

Teorema 2 (Equivalencia de Lax)**.** *Un método consistente es convergente si y solo si es estable.*

Definición 8.3 (Condición de von Neumann)**.** *Un método satisface la condición de von Neumann si existe una constante* C' *tal que*

$$\rho(M) \leq 1 + C'\Delta t$$

Proposición 8.1. *Si un método es estable (Lax-Richtmyer) en cualquier norma, entonces satisface la condición de von Neumann.*

Demo. Como $\|M^n v\| = |\lambda|^n \|v\|$ para cualquier autovalor λ, se tiene $\|M^n\| \geq |\lambda|$, es decir, $\|M^n\| \geq \rho(M)^n$, cualquiera sea la norma elegida.

Es inmediato ver que si se verifica la condición de estabilidad de Lax-Richtmyer, entonces existe C tal que

$$|\lambda|^n \leq C \qquad \text{para todo} \qquad 0 \leq n\Delta t \leq T_f,$$

y por lo tanto,

$$|\lambda| \leq K^{\Delta t/T_f}.$$

Como la función $f(x) = K^x$ es convexa, usando

$$f(tx_1 + (1-t)x_2) \leq tf(x_1) + (1-t)f(x_2)$$

con $x_1 = 1$, $x_2 = 0$ y $t = \Delta t/T_f$, obtenemos

$$|\lambda| \leq K^{\Delta t T_f} \leq \Delta t/T_f C + (1 - \Delta t/T_f) = 1 + (C-1)/T_f \Delta t$$

que es lo que queríamos demostrar, con la constante $C' = (C-1)/T_f$. □

Observación 8.2. *Para negar la estabilidad (en cualquier norma), alcanza con que* $|\lambda| > \alpha > 1$*, con cierto* α *fijo a lo largo del camino de refinamiento* $h(\Delta t)$*.*

Observación 8.3. *Otra condiciones necesaria para tener estabilidad es la condición de CFL para problemas hiperbólicos. Un método consistente que no la verifica, tiene que ser inestable, como consecuencia del Teorema de Lax.*

8.3. Condiciones suficientes

Veamos ahora algunas condiciones suficientes para tener estabilidad, es decir, que garantizan la estabilidad en alguna norma.

Proposición 8.2. *Si* M *es una matriz normal, la condición de von Neumann también es suficiente.*

Demo. Para matrices normales, la norma 2 coincide con el radio espectral, y por lo tanto

$$\|M^n\|_2 = \rho(M^n) \leq (1 + C'\Delta t)^n \leq e^{C'T_f} \leq C$$

donde C es independiente de $\Delta t, \Delta x$. □

Corolario 8.1. *Si la matriz M tiene como autovectores a los modos de Fourier discretos $W(\xi) = e^{i\xi jh}$, entonces la condición de von Neumann sobre los autovalores $\lambda(\xi)$ es suficiente.*

Demo. Es inmediato, ya que como los modos de Fourier discretos forman una base ortonormal, M es una matriz normal. □

Observación 8.4. *Esto permite justificar el análisis de estabilidad por el método de Fourier, donde reemplazamos $u_j^n = \lambda^n e^{i\xi jh}$ y buscamos una cota sobre λ de la forma*

$$|\lambda(\xi)| \leq 1 + C'\Delta t$$

Vale la pena recordar que el método de Fourier ignora las condiciones de borde del problema, o, equivalentemente, las supone periódicas.

8.4. Cuasi-espectro de Godunov-Riabenki

Una condición necesaria para la estabilidad, que en particular nos permitirá justificar que el criterio de Von Neumann es *necesario* para la estabilidad de problemas con condiciones de borde generales, fue provista por Godunov-Riabenki en 1963. El punto de partida es una debilitación de la noción de autovectores y autovalores de una matriz, para considerar un conjunto más grande llamado cuasi-espectro – consecuentemente, la restricción sobre el nuevo cuasi-espectro resultará más fuerte que una cota sobre el radio espectral.

Definición 8.4. *Un punto λ está en el cuasi-espectro de una una familia de matrices $\{Q_h\}$ si para cada $\varepsilon > 0$ existe h_0 tal que, para todo $h < h_0$ existe un cuasi-autovector u tal que*

$$\|Q_h u - \lambda u\| \leq \varepsilon \|u\|$$

Teorema 3 (Condición de Godunov-Riabenki)**.** *Para la estabilidad de un problema de la forma*

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

es necesario que el cuasi-espectro de $\{Q_h\}$ esté contenido en el disco unitario.

Corolario 8.2. *Para un problema de valores iniciales con condiciones de borde no-periódicas, la condición de Von Neumann es necesaria para la estabilidad.*

Demostración. Consideremos un ejemplo suficientemente ilustrativo. □

9

Apéndice: Álgebra Lineal

Teorema 4 (de Gerschgorin)**.** *Sea $A \in \mathbb{C}^{n\times n}$ y sea $R_i = \sum_{j\neq i} |a_{ij}|$. Entonces, todo autovalor λ de A satisface $|\lambda - a_{i,i}| \leq R_i$ para algún i.*

Definición 9.1. *Una matriz $A \in \mathbb{R}^{n\times n}$ se dice estrictamente diagonal-dominante (por columnas) si para todo $1 \leq k \leq n$ se tiene que $|a_{kk}| > \sum_{j\neq k} |a_{kj}|$.*

Teorema 5. *Una matriz estrictamente diagonal dominante es inversible.*

Dem. Supongamos que no lo es, y que por lo tanto existe un vector $x \neq 0$ tal que $Ax = 0$. Sea k tal que $x_k = \|x\|_\infty$. Entonces

$$0 = \sum_{j=1}^{n} a_{kj}x_j \;\Rightarrow\; a_{kk}x_k = \sum_{j\neq k} a_{kj}x_j \;\Rightarrow\; a_{kk} = \sum_{j\neq k} a_{kj}\frac{x_j}{x_k}$$

Tomando valor absoluto y usando la desigualdad triangular,

$$|a_{kk}| \leq \sum_{j\neq k} |a_{kj}| \left|\frac{x_j}{x_k}\right| \leq \sum_{j\neq k} |a_{kj}|$$

lo que es una contradicción, porque A era estrictamente diagonal dominante. □

Lema 9.1. *Para una matriz A inversible, una norma vectorial $\|\cdot\|$ y su norma matricial inducida $\|\cdot\|$, tenemos la expresión*

$$\|A^{-1}\|^{-1} = \inf_{x\neq 0} \frac{\|Ax\|}{\|x\|}. \tag{9.1}$$

Dem. Tenemos la definición

$$\|A^{-1}\| = \sup_{x\neq 0} \frac{\|A^{-1}x\|}{\|x\|}.$$

Para empezar, intentemos sacarnos de encima la inversa en el lado derecho. Para eso, como A es inversible, llamemos $y = A^{-1}x$. Nos queda entonces

$$\|A^{-1}\| = \sup_{Ay\neq 0} \frac{\|y\|}{\|Ay\|}.$$

Ahora, es inmediato ver que para cualquier conjunto $C \subset \mathbb{R}$, vale $\sup(C)^{-1} = \text{ínf}(C^{-1})$. En efecto, esto es porque $a \geq c \; \forall c \in C$ si y solo si $a^{-1} \leq c^{-1} \; \forall c \in C$.

Tenemos entonces que

$$\|A^{-1}\|^{-1} = \inf_{Ay \neq 0} \frac{\|Ay\|}{\|y\|}.$$

Ahora bien, como x era un vector arbitrario y A es inversible, resulta que y también lo es. Llegamos entonces a la expresión (9.1) que queríamos demostrar. □

Teorema 6 (de Varah)**.** *Sea A estrictamente diagonal dominante por columnas, y llamemos $\alpha = \text{mín}_k \left\{ |a_{kk}| - \sum_{i \neq k} |a_{k,j}| \right\}$. Entonces $\|A^{-1}\|_\infty \leq \alpha$.*

Dem. Por el Lema 9.1, tenemos que

$$\|A^{-1}\|_\infty^{-1} = \inf_{x \neq 0} \frac{\|Ax\|_\infty}{\|x\|_\infty}.$$

Por ende, alcanza con demostrar que

$$\alpha \|x\|_\infty \leq \|Ax\|_\infty \quad \forall x \in \mathbb{R}^n.$$

Para ello, tomemos un vector x cualquiera y tomemos la coordenada que realiza la norma infinito del vector, es decir, $x_k = \|x\|_\infty$. Entonces

$$0 < \alpha < |a_{kk}| - \sum_{j \neq k} |a_{k,j}|$$

$$0 < \alpha |x_k| < |a_{kk}||x_k| - \sum_{j \neq k} |a_{k,j}||x_k|$$

$$\leq |a_{kk}||x_k| - \left| \sum_{j \neq k} a_{kj} x_k \right|$$

$$\leq \left| \sum_{j=1}^{n} a_{k,j} x_k \right| \leq \max_k \left| \sum_{j=1}^{n} a_{k,j} x_k \right| = \|Ax\|_\infty$$

lo que concluye la demostración. □