2. Discretización de ecuaciones

El proceso de obtención de la solución computacional consiste en 2 pasos que se pueden esquematizar como muestra la figura 2.1

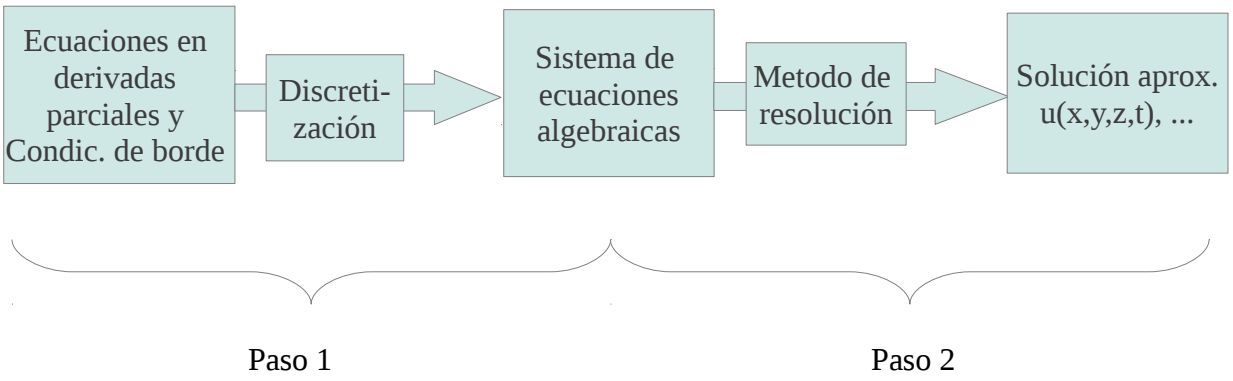

Figura 2.1 – Esquema de proceso de resolución de una EDP.

En el primer paso las ecuaciones que gobiernan el proceso de interes, así como las condiciones de borde, son convertidas a un sistema discreto de ecuaciones algebraicas; este proceso se denomina discretización. Al reeemplazar los términos diferenciales individuales de la EDPs por expresiones algebraicas que conectan valores en nodos de una red finita se introduce un error de truncamiento. En este capítulo veremos como elegir expresiones algebraicas que produzcan los menores errores.

El segundo paso requiere de un método de resolución del sistema de ecuaciones algebraicas. Este paso también puede introducir un error (de solución) pero es generalmente despreciable comparado con el error de truncamiento introducido durante la discretización, a menos que el método sea inestable. En los próximos capítulos discutiremos métodos apropiados para resolver sistemas de ecuaciones algebraicos.

2.1 Discretización

Hay varios métodos para convertir las ecuaciones en derivadas parciales a un sistema de ecuaciones algebraicas. Los mas comunes son el método de diferencias finitas, método de elementos finitos y el método espectral. En la práctica las derivadas temporales son discretizadas casi exclusivamente usando el métodos de diferencias finitas. Las derivadas espaciales son discretizadas usando diferencias finitas, elementos finitos o usando el método espectral.

La discretización se puede dividir en dos categorías: una forma, que da lugar a los métodos de diferencias finitas, es representar la función por su valor en un conjunto discreto de puntos de grilla. La otra forma es el método de los residuos pesados (WRM por su sigla en inglés). Este método es conceptualmente diferente del método de diferencias finitas pues asume que la solución puede ser representada por un cojunto de funciones de prueba. Cuando el conjunto de funciones de prueba forma un conjunto ortogonal esta discretización da lugar al método espectral. Cuando las funciones de prueba son diferentes de cero solamente en una pequeña parte del dominio este método da lugar al método de elementos finitos. Técnicas híbridas usando ambos tipos de discretización también existen, por ejemplo el método pseudo-espectral.

En el curso nos ocuparemos únicamente de la técnica de diferencias finitas. Sin embargo, primero haremos una breve descripción de los métodos a fin de compararlos.

Primero describiremos la formulación general del método de residuos pesados de forma de demostrar la conección entre elementos finitos y el método espectral. El primer paso de un WRM es asumir una solución aproximada de la forma

$$T(x,y,z,t)=T_0(x,y,z,t)+\sum_{j=1}^{J} a_j(t)\Phi_j(x,y,z) \qquad (2.1)$$

donde T_0 se elige para satisfacer las condiciones de borde e iniciales. Las funciones de prueba, o base, $\Phi_j(x,y,z)$ son conocidas. Los coeficientes $a_j(t)$ son desconocidos y deben ser determinados resolviendo el sistema de ecuaciones generado de las EDPs.

Se asume que la ecuación puede ser escrita de la forma

$$L(T^*)=0 \qquad (2.2)$$

donde T* es la solución exacta. Si se sustituye la solución aproximada 2.1 en 2.2 habrá un residuo R, de tal forma que la ecuación queda

$$L(T)=R \qquad (2.3)$$

R es una función contínua de x,y,z y de t en el caso general. Si J es lo suficientemente grande es posible en principio elegir los coeficientes $a_j(t)$ de tal forma que R sea pequeño en el dominio computacional. Los coeficientes $a_j(t)$ se determinan requiriendo que la integral del residuo pesado en el dominio computacional sea cero, o sea

$$\iiint W_m(x,y,z)R\,dx\,dy\,dz=0 \qquad (2.4)$$

m=1...M, lo cual resulta en un sistema de ecuaciones para los $a_j(t)$. Diferentes elecciones de las funciones W_m da lugar a diferentes métodos de residuos pesados. El mas usual es el

Método de Galerkin

En este caso se elige $W_m(x,y,z)=\phi_m(x,y,z)$, o sea que las funciones Wm se eligen de la misma familia que las funciones de prueba. Si las funciones de prueba forman un conjunto completo (2.4) indica que el residuo R es ortogonal a todo miembro del conjunto. Consecuentemente a medida que M tiende a infinito la solución aproximada T convergirá a la solución verdadera T*. Esta elección de funciones de prueba da lugar al método espectral. Si las funciones de prueba no son ortogonales y son diferentes de cero en un subdominio pequeño se tiene el método de elementos finitos.

La diferencia entre los métodos de diferencias finitas, espectral y elementos finitos es ilustrada por la forma en la cual representan la función periódica graficada en la figura 2.2 mediante el uso de 5 unidades de información.

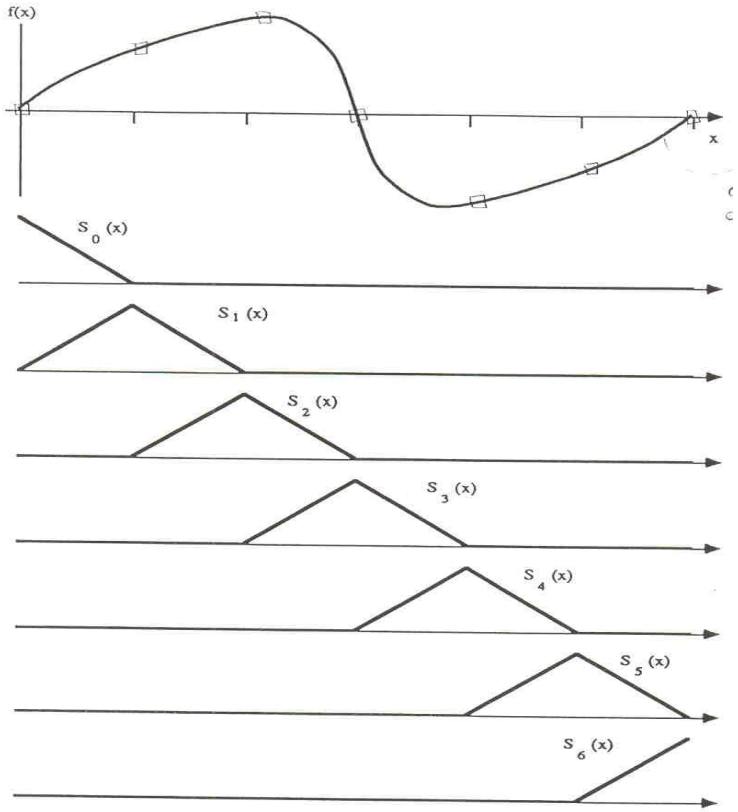

Figura 2.2 – Representación de una función por los differentes métodos.

La función f(x) está representada por tres discretizaciones diferentes: en el método de diferencias finitas f(x) está representada por su valor exacto en 5 diferentes puntos a lo largo de la dirección

x. En el método espectral la función f(x) se puede aproximar por una serie de Fourier truncada:

$$f(x) \sim a_1 + a_2 \cos x + a_3 \sin x + a_4 \cos 2x + a_5 \sin 2x$$

mientras que en el método de elementos finitos la función puede ser aproximada por una suma finita de funciones S(x):

$$b_0 S_0 + b_1 S_1 + b_2 S_2 + b_3 S_3 + b_4 S_4 + b_5 S_5$$

2.2 Método de diferencias finitas

En esta sección ilustraremos el método de discretización por diferencias finitas considerando la ecuación de difusión en 1D.

$$\frac{\partial T}{\partial t} = \alpha \frac{\partial^2 T}{\partial x^2} \quad 0 \leqslant x \leqslant 1 \tag{2.5}$$

$$\textit{Condiciones de borde}: \; T(0,t) = b, \; T(1,t) = d$$
$$\textit{Condiciones iniciales}: \; T(x,0) = T_0(x), \; 0 \leqslant x \leqslant 1$$

Para discretizar es necesario primero definir una grilla. La figura 2.3 muestra una grilla en el espacio x-t. Los pasos Δt y Δx y el significado del subíndice j y superíndice n están indicados en la figura y T^n_j es el valor de T en el nodo (j,n).

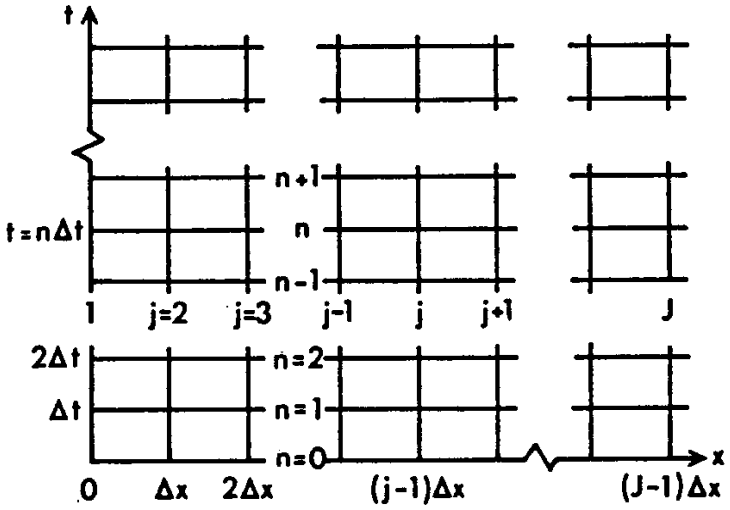

Figura 2.3 – Ejemplo de grilla.

El primer paso en desarrollar un algoritmo para calcular valores de T que aparecen en (2.5) es expresar las derivadas temporales y espaciales de T en el nodo (j,n) en término de los valores de

T en los nodos cercanos. Para ello se usan series de Taylor del tipo

$$T^n_{j+1} = \sum_{m=0}^{\infty} \frac{\Delta x^m}{m!} \left[\frac{\partial^m T}{\partial x^m}\right]^n_j \tag{2.6}$$

$$T^{n+1}_j = \sum_{m=0}^{\infty} \frac{\Delta t^m}{m!} \left[\frac{\partial^m T}{\partial t^m}\right]^n_j \tag{2.7}$$

Estas series pueden ser truncadas en cualquier punto; el error de truncamiento resultante está dominado por el siguiente término en la expansión si Δx<<1 o Δt<<1. Por ejemplo, de (2.6) y (2.7) es posible aproximar una derivada primera espacial o temporal de T usando una **fórmula de diferencias finitas progresiva**

$$T^n_{j+1} = T^n_j + \Delta x \left[\frac{\partial T}{\partial x}\right]^n_j + \frac{\Delta x^2}{2}\left[\frac{\partial^2 T}{\partial x^2}\right]^n_j$$

$$\textit{lo cual resulta en} \quad \left[\frac{\partial T}{\partial x}\right]^n_j = \frac{T^n_{j+1} - T^n_j}{\Delta x} - \frac{\Delta x}{2}\left[\frac{\partial^2 T}{\partial x^2}\right]^n_j \tag{2.8}$$

$$\textit{Analogamente} \quad \left[\frac{\partial T}{\partial t}\right]^n_j = \frac{T^{n+1}_j - T^n_j}{\Delta t} - \frac{\Delta t}{2}\left[\frac{\partial^2 T}{\partial t^2}\right]^n_j$$

o una **fórmula de diferencias finitas regresiva** (para ello se hace una aproximación T(x-Δx)=T(x)-Δx*T'(x)+...)

$$\left[\frac{\partial T}{\partial x}\right]^n_j = \frac{T^n_j - T^n_{j-1}}{\Delta x} + \frac{\Delta x}{2}\left[\frac{\partial^2 T}{\partial x^2}\right]^n_j$$

$$\left[\frac{\partial T}{\partial t}\right]^n_j = \frac{T^n_j - T^{n-1}_j}{\Delta t} + \frac{\Delta t}{2}\left[\frac{\partial^2 T}{\partial t^2}\right]^n_j \tag{2.9}$$

Este tipo de aproximación de primer órden necesita únicamente dos puntos, pero introduce un error del órden de O(Δx) ó O(Δt), asumiendo que Δx<<1 y Δt<<1 y que las derivadas de mayor orden están acotadas.

Las diferencias de mayor orden requieren mas puntos. Ahora veamos una metodología general para construir aproximaciones de mayor orden. Empezamos de la expresión general

$$\left[\frac{\partial T}{\partial x}\right]^n_j = aT^n_{j-1} + bT^n_j + cT^n_{j+1} + O(\Delta x^m) \tag{2.10}$$

donde a, b, c se tienen que determinar y el término $O(\Delta x^m)$ indicará la precisión de la aproximación resultante. Usando (2.6) es posible escribir

$$aT^n_{j-1}+bT^n_j+cT^n_{j+1}=(a+b+c)T^n_j+(-a+c)\Delta x\left[\frac{\partial T}{\partial x}\right]^n_j +(a+c)\frac{\Delta x^2}{2}\left[\frac{\partial^2 T}{\partial x^2}\right]^n_j+(-a+c)\frac{\Delta x^3}{6}\left[\frac{\partial^3 T}{\partial x^3}\right]^n_j \qquad (2.11)$$

Imponiendo a+b+c=0, (-a+c)Δx=1 resulta en

a = c -1/Δx

b = -2c + 1/Δx

para cualquier c. Eligiendo c de tal forma que el tercer término de la derecha en (2.11) sea nulo produce la solución mas precisa posible con tres parámetros a elegir. Esto es,

c = -a = 1/(2Δx)

b = 0

y da lugar a la **fórmula de diferencias finitas centrada**

$$\left[\frac{\partial T}{\partial x}\right]^n_j=\frac{1}{2\Delta x}(-T^n_{j-1}+T^n_{j+1})-\frac{\Delta x^2}{6}\left[\frac{\partial^3 T}{\partial x^3}\right]^n_j+... \qquad (2.12)$$

que tiene un error de truncamiento de $O(\Delta x^2)$. Claramente la aproximación por diferencias finitas centrada produce un error de truncamiento de orden mayor que las aproximaciones progresiva y regresiva. La representación gráfica de las tres aproximaciones se muestra en la figura 2.4: la pendiente AC es la fórmula centrada, AB y BC son las fórmulas regresiva y progresiva, respectivamente.

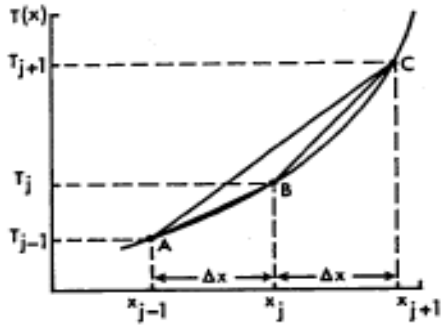

Figura 2.4 - Representación en diferencias finitas de $\partial T/\partial x$.

De la misma forma que hemos obtenido fórmulas en diferencias que aproximan el valor de la derivada primera en un punto, prodríamos obtener expresiones que aproximen el valor de las derivadas de órdenes superiores. Por ejemplo, si sumamos los desarrollos en serie

$$T^n_{j+1} = T^n_j + \Delta x [\frac{\partial T}{\partial x}]^n_j + \frac{\Delta x^2}{2}[\frac{\partial^2 T}{\partial x^2}]^n_j + \frac{\Delta x^3}{6}[\frac{\partial^3 T}{\partial x^3}]^n_j + \frac{\Delta x^4}{24}[\frac{\partial^4 T}{\partial x^4}]^n_j$$
$$T^n_{j-1} = T^n_j - \Delta x [\frac{\partial T}{\partial x}]^n_j + \frac{\Delta x^2}{2}[\frac{\partial^2 T}{\partial x^2}]^n_j - \frac{\Delta x^3}{6}[\frac{\partial^3 T}{\partial x^3}]^n_j + \frac{\Delta x^4}{24}[\frac{\partial^4 T}{\partial x^4}]^n_j \qquad (2.13)$$

se obtiene

$$T^n_{j+1} + T^n_{j-1} = 2T^n_j + \Delta x^2 [\frac{\partial^2 T}{\partial x^2}]^n_j + 2\frac{\Delta x^4}{24}[\frac{\partial^4 T}{\partial x^4}]^n_j \qquad (2.14)$$

de la cual obtenemos la **fórmula de diferencias centradas para la derivada segunda** espacial de T

$$[\frac{\partial^2 T}{\partial x^2}]^n_j = \frac{T^n_{j+1} - 2T^n_j + T^n_{j-1}}{\Delta x^2} + O(\Delta x^2) \qquad (2.15)$$

Apliquemos los resultados obtenidos a la ecuación de difusión en 1D (2.5). Si consideramos diferencias progresivas en el tiempo y diferencias centradas en el espacio la ecuación algebraica a resolver es (Forward Time-Centered Space FTCS)

$$T^{n+1}_j = T^n_j + \frac{\alpha \Delta t}{\Delta x^2}(T^n_{j-1} - 2T^n_j + T^n_{j+1}) \qquad (2.16)$$

Los errores de truncamiento de (2.16) son de O(Δt) y O(Δx^2). Si se considera diferencia centrada en el tiempo se puede construir el siguiente algoritmo explícito

$$T_j^{n+1} = T_j^{n-1} + \frac{2\alpha \Delta t}{\Delta x^2}(T_{j-1}^n - 2T_j^n + T_{j+1}^n) \qquad (2.17)$$

El algoritmo (2.17) es mas preciso que el (2.16) pero mas complicado pues para resolverlo se necesita de tres nivels de datos, n-1, n, n+1, en lugar de dos.

La expresión de la fórmula centrada es simétrica con respecto al nodo donde se quiere calcular la derivada. Es posible también desarrollar una fórmula asimétrica de 3 puntos para la derivada primera. Partiendo de una expresión general de la forma

$$\left[\frac{\partial T}{\partial x}\right]_j^n = aT_j^n + bT_{j+1}^n + cT_{j+2}^n + O(\Delta x^m) \qquad (2.18)$$

donde a,b, c son a determinar. Se expande T_{j+1}^n y T_{j+2}^n en forma de series de Taylor segun j y luego se sustituye en (2.18). Como resultado se obtiene

$$\left[\frac{\partial T}{\partial x}\right]_j^n = (a+b+c)T_j^n + (b\Delta x + c\,2\Delta x)\left[\frac{\partial T}{\partial x}\right]_j^n + \left(\frac{b\Delta x^2}{2} + \frac{c(2\Delta x)^2}{2}\right)\left[\frac{\partial^2 T}{\partial x^2}\right] + \ldots \qquad (2.19)$$

Comparando los lados izquierdo y derecho de (2.19) indica que para obtener el menor error se debe imponer

a+b+c = 0

bΔx+c2Δx = 1

bΔx^2/2 + c(2Δx)^2/2 = 0

Esto da valores

a = -1.5/Δx, b = 2/Δx, c = -0.5/Δx

y

$$\left[\frac{\partial T}{\partial x}\right]_j^n = \frac{-1.5T_j^n + 2T_{j+1}^n - 0.5T_{j+2}^n}{\Delta x} - \frac{\Delta x^2}{3}\left[\frac{\partial^3 T}{\partial x^3}\right]_j^n + \ldots \qquad (2.20)$$

Esta fórmula tiene un error de truncamiento de orden $O(\Delta x^2)$ como la fórmula centrada.

Si incluímos cinco términos en (2.18) (en lugar de 3) es posible derivar otras fórmulas para $\partial T/\partial x$ de mayor orden. No obstante discretizaciones de mayor orden tienen restricciones de estabilidad mayores que aquellas de menor orden (lo veremos mas adeltante). Por ello una estrategia alternativa es elegir los coeficientes en los desarrollos tipo (2.18) de tal forma que reduzcan el error y que al mismo tiempo mejore la estabilidad de la solución.

2.3 Exactitud de la discretización

El proceso de discretización invariablemente lleva a un error, excepto cuando la solución exacta tiene una forma analítica sencilla. Por eso, la fórmula de diferencias centradas es exacta para polinomios de hasta orden cuadrático, mientras que las fórmula progresiva y regresiva son exactas solo para polinomios lineales.

En general la evaluación del término de mayor orden de aquellos no considerados en la aproximación es una buena medida del error cometido si el tamaño de la grilla es chico. La tabla 2.1 resume las fórmulas algebraicas para aproximar $\partial T/\partial x$ en diferentes formas y sus errores de truncamiento asociados al término de mayor orden. Idem tabla 2.2 para $\partial^2 T/\partial x^2$.

Case	Algebraic formula	Truncation error leading term
3PT SYM	$(\bar{T}_{j+1}-\bar{T}_{j-1})/2\Delta x$	$\Delta x^2\bar{T}_{xxx}/6$
FOR DIFF	$(\bar{T}_{j+1}-\bar{T}_j)/\Delta x$	$\Delta x\bar{T}_{xx}/2$
BACK DIFF	$(\bar{T}_j-\bar{T}_{j-1})/\Delta x$	$-\Delta x\bar{T}_{xx}/2$
3PT ASYM	$(-1.5\bar{T}_j+2\bar{T}_{j+1}-0.5\bar{T}_{j+2})/\Delta x$	$-\Delta x^2\bar{T}_{xxx}/3$
5PT SYM	$(\bar{T}_{j-2}-8\bar{T}_{j-1}+8\bar{T}_{j+1}-\bar{T}_{j+2})/12\Delta x$	$-\Delta x^4\bar{T}_{xxxxx}/30$

Tabla 2.1 – Error de truncamiento para diferentes fórmulas de $\partial T/\partial x$

Case	Algebraic formula	Truncation error leading term
3PT SYM	$(\bar{T}_{j-1}-2\bar{T}_j+\bar{T}_{j+1})/\Delta x^2$	$\Delta x^2\bar{T}_{xxxx}/12$
3PT ASYM	$(\bar{T}_j-2\bar{T}_{j+1}+T_{j+2})/\Delta x^2$	$\Delta x\bar{T}_{xxx}$
5PT SYM	$(-\bar{T}_{j-2}+16\bar{T}_{j-1}-30\bar{T}_j$ $+16\bar{T}_{j+1}-\bar{T}_{j+2})/12\Delta x^2$	$-\Delta x^4\bar{T}_{xxxxxx}/90$

Tabla 2.2 – Error de truncamiento para diferentes fórmulas de $\partial^2 T/\partial x^2$

Una forma mas directa de comparar la exactitud de varias fórmulas algebraicas es considerar una función analítica simple y comparar los valores de las derivadas obtenidas analíticamente con aquellas obtenidas de las fórmulas de discretización. Para tener una idea de cuan grande es el error usando diferentes fórmulas las tablas 2.3 y 2.4 comparan el valor calculado de $\partial T/\partial x$ y de $\partial^2 T/\partial x^2$ evaluado en x=1 con el valor de las derivadas de $T=e^x$ calculado analíticamente. El paso espacial es $\Delta x=0.1$. Generalmente, las fórmulas de 3 puntos, simétricas o asimétricas, son considerablemente mas exactas que los esquemas pregresivo y regresivo, pero considerablemente menos exactas que la fórmula de 5 puntos. Comparando las dos últimas columnas es evidente que el término de mayor orden en el desarrollo de Taylor da una muy buena idea del error introducido en la discretización.

Case	Algebraic formula	$\left[\frac{d\bar{T}}{dx}\right]_j$	Error	Leading term in T.E.
Exact	—	2.7183	—	—
3PT SYM	$(\bar{T}_{j+1}-\bar{T}_{j-1})/2\Delta x$	2.7228	0.4533×10^{-2}	0.4531×10^{-2}
FOR DIFF	$(\bar{T}_{j+1}-\bar{T}_j)/\Delta x$	2.8588	0.1406×10^{-0}	0.1359×10^{-0}
BACK DIFF	$(\bar{T}_j-\bar{T}_{j-1})/\Delta x$	2.5868	-0.1315×10^{-0}	-0.1359×10^{-0}
3PT ASYM	$(-1.5\bar{T}_j+2\bar{T}_{j+1}-0.5\bar{T}_{j+2})/\Delta x$	2.7085	-0.9773×10^{-2}	-0.9061×10^{-2}
5PT SYM	$(\bar{T}_{j-2}-8\bar{T}_{j-1}+8\bar{T}_{j+1}-\bar{T}_{j+2})/12\Delta x$	2.7183	-0.9072×10^{-5}	-0.9061×10^{-5}

Tabla 2.3 – Comparación de fórmulas para evaluar dT/dx en x=1.0.

Case	Algebraic formula	$\left[\frac{d\bar{T}^2}{dx^2}\right]_j$	Error	Leading term in T.E.
Exact	—	2.7183	—	—
3PT SYM	$(\bar{T}_{j-1}-2\bar{T}_j+\bar{T}_{j+1})/\Delta x^2$	2.7205	0.2266×10^{-2}	0.2265×10^{-2}
3PT ASYM	$(\bar{T}_j-2\bar{T}_{j+1}+\bar{T}_{j+2})/\Delta x^2$	3.0067	0.2884×10^{-0}	0.2718×10^{-0}
5PT SYM	$(-\bar{T}_{j-2}+16\bar{T}_{j-1}-30\bar{T}_j+16\bar{T}_{j+1}-\bar{T}_{j+2})/12\Delta x^2$	2.7183	-0.3023×10^{-5}	-0.3020×10^{-5}

Tabla 2.4 - Comparación de fórmulas para evaluar d^2T/dx^2 en x=1.0.

Notar que en el caso de la derivada segunda, la fórmula asimétrica de 3 puntos tiene un error bastante mayor que la fórmula simétrica de 3 puntos (tabla 2.4). Esto es consistente con el hecho de que el término de mayor orden del desarrollo de Taylor que queda sin usar es de $O(\Delta x)$ (tabla 2.2).

Como vimos en las tablas anteriores existe una gran correlación entre error calculado y el error

del truncamiento. Por lo tanto es esperable que el error calculado se reduzca con Δx como se muestra en las tablas 2.1 y 2.2. El error calculado E será entonces de la forma

$E = A(\Delta x)^k$

donde k es el exponente del tamaño de grilla en el término de mayor orden del truncamiento (columna de la derecha en tablas 2.1 y 2.2).

Graficando en escala logarítmica, la razón de convergencia de la solución para varios casos sigue la razón de convergencia implicada por el error de truncamiento de las tablas 2.1 y 2.2 (figura 2.5). Como resultado, la razón de convergencia puede ser estimada del error de truncamiento aún cuando la solución exacta es desconocida.

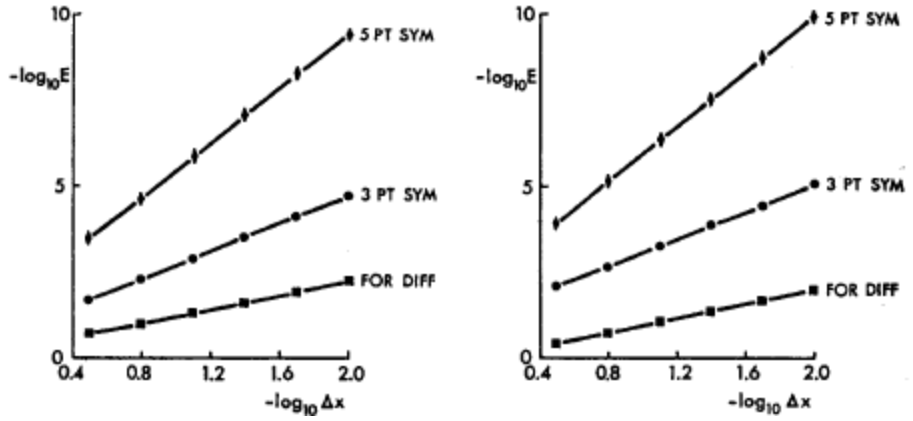

Figura 2.5 – Convergencia de los resultados de la evaluación de dT/dx (izquierda) y de d^2T/dx^2 (derecha). $E = |[dT/dx]_{DF} / [dT/dx]_{exact} - 1|$ para dT/dx y análogo para d^2T/dx^2. Se verifica que el error disminuye con Δx y la pendiente es k.

De los resultados anteriores parecería que siempre se debería usar una fórmula de mayor órden en una grilla bien fina (alta resolución) lo cual es, sin embargo, no tan obvio. Primero que nada, la evaluación de una fórmula de mayor orden toma en cuenta mas puntos y por lo tanto es menos económico que la evaluación de una fórmula de menor orden. Desde una perspectiva práctica la precisión a la que puede llegarse en un tiempo de ejecución, o sea la eficiencia computacional, es más importante que la precisión por si sola; ésta siempre puede aumentarse afinando la grilla. Segundo, las fórmulas de mayor orden muestran un incremento significativo en la precisión frente a fórmulas de menor órden cuando la grilla es fina. No obstante, a veces los problemas pueden resolverse con grillas relativamente gruesas, o tenemos limitaciones por memoria computacional o de tiempo de cálculo.

La superioridad de las fórmulas de mayor orden también depende de cuan suave es la solución exacta. Si la solución es discontínua en el rango en el cual la fórmula algebraica es evaluada las fórmulas de mayor orden no son significativamente mas exactas. Esto puede ser ilustrado con la función

$$y = tgh[k(x-1)] \qquad (2.21)$$

La figura 2.6 muestra la función (2.21) para valores de k=1, 5, 20. Es evidente que el gradiente en x=1 crece con k. La figura 2.7 muestra la primera y segunda derivada de y con respecto a x evaluada en x=0.96 usando las fórmulas simétricas de 3 y 5 puntos para valores decrecientes de Δx y k=5, 20. Se observa claramente que la fórmula de 5 puntos sólo produce precisión superior si la grilla es suficientemente fina. Para valores intermedios de Δx la fórmula de 5 puntos produce un error mayor en el cálculo de la derivada primera.

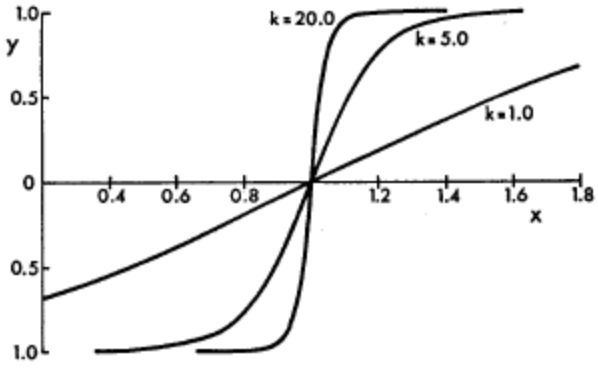

Figura 2.6 – Función analítica y=tgh[k(1-x)].

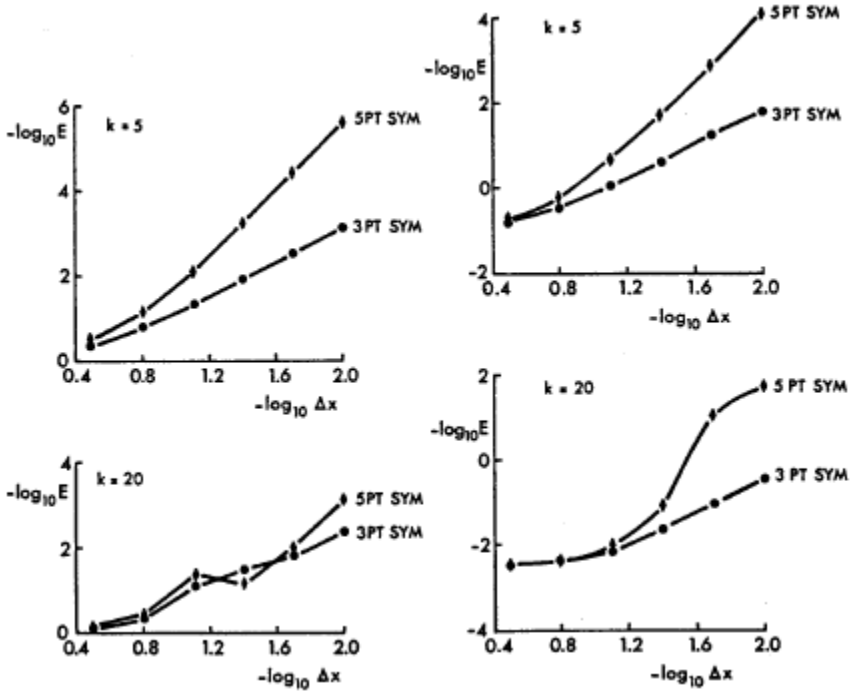

Figura 2.7 – Convergencia de $[dy/dx]_{DF}$ (izquierda) y $[d^2y/dx^2]_{DF}$ (derecha): influencia de la suavidad de la solucion. $E=|[dT/dx]_{DF}/[dT/dx]_{exact}-1|$ para la columna de la izquierda y analoga para para columna de la derecha.

Para terminar es instructivo mencionar la precisión de las diferentes aproximaciones por diferencias finitas en el caso de movimientos ondulatorios ya que están presentes en muchos de los problemas de dinámica de fluidos geofísicos. Para ello consideraremos una onda progresiva de la forma

$$T(x,t)=\Re\exp[im(x-qt)]=\cos(m(x-qt)) \tag{2.22}$$

donde *R* denota parte real, m=2π/λ es el número de onda y q es la velocidad de propagación de la onda. En el nodo (j,n) el valor exacto de la primera y segunda derivada es

$$\begin{aligned}\frac{\partial T}{\partial x}&=-m\sin[m(x_j-qt_n)]\\ \frac{\partial^2 T}{\partial x^2}&=-m^2\cos[m(x_j-qt_n)]\end{aligned} \tag{2.23}$$

Sustituyendo (2.22) en la fórmula de tres puntos

$$\left[\frac{\partial T}{\partial x}\right]_j^n \sim \frac{T_{j+1}^n - T_{j-1}^n}{2\Delta x} = \frac{\cos[m(x_j - qt_n) + m\Delta x] - \cos[m(x_j - qt_n) - m\Delta x]}{2\Delta x}$$
$$\left[\frac{\partial T}{\partial x}\right]_j^n \sim \frac{-m\sin[m(x_j - qt_n)]\sin(m\Delta x)}{m\Delta x} \qquad (2.24)$$

Por lo tanto el cociente entre la derivada calculada y la exacta es

$$AR(1)_{3PT} = \frac{[\partial T/\partial x]_j^n}{[\partial T/\partial x]} = \frac{\sin(m\Delta x)}{m\Delta x} \qquad (2.25)$$

Para encontrar la derivada segunda usamos la aproximacion de derivadas centrada y sustituimos la ecuacion (2.22), lo cual resulta en

$$\frac{T_{j-1}^n - 2T_j^n + T_j^n}{\Delta x^2} = -m^2\left(\frac{[\sin(m\Delta x/2)]}{m\Delta x/2}\right)^2 \cos[m(x_j - qt_n)] \qquad (2.26)$$

El cociente entre la derivada segunda calculada y la exacta queda

$$AR(2)_{3PT} = \left(\frac{[\sin(m\Delta x/2)]}{m\Delta x/2}\right)^2 \qquad (2.27)$$

De acuerdo a (2.25) el uso de la aproximación en diferencias finitias cambia la amplitud de la derivada. Para longitudes de onda largas, $\lambda > 20\Delta x$, la amplitud de la derivada primera se reduce en un factor entre 0.984 a 1.0. No obstante, cuando hay menos de 4 nodos en una longitud de onda (ondas cortas) la amplitud de la derivada es menor que 0.64 de su valor correcto. Para la derivada segunda en el caso $\lambda = 20\Delta x$ diferencias finitas reduce la amplitud en 0.992, mientras que para el caso $\lambda = 2\Delta x$ la amplitud se reduce por un factor 0.405. Esto muestra que las longitudes de onda largas son mejor representadas que las longitudes de onda corta, lo cual sigue siendo válido para formulas de diferencias finitas de mayor órden. La tabla 2.5 compara los cocientes entre derivadas calculadas y exactas para fórmulas de 3 y 5 puntos. Por lo tanto para resolver adecuadamente movimientos ondulatorios es necesario al menos 4 nodos en una longitud de onda. De otra forma, el movimiento será gravemente distorsionado.

Derivative	Scheme	Amplitude ratio Long wavelength ($\lambda = 20\,\Delta x$)	Amplitude ratio Short wavelength ($\lambda = 4\Delta x$)
$\frac{d\bar{T}}{dx}$	3PT SYM	0.9836	0.6366
	5PT SYM	0.9996	0.8488
Derivative	Scheme	Long wavelength ($\lambda = 20\Delta x$)	Short wavelength ($\lambda = 2\Delta x$)
$\frac{d^2\bar{T}}{dx^2}$	3PT SYM	0.9918	0.4053
	5PT SYM	0.9999	0.5404

Tabla 2.5 – Cocientes de amplitud para una onda progresiva.

2.4 Ecuación de Difusión

Como dijimos al comienzo la base del método de diferencias finitas es la construcción de una grilla, reemplazo de las EDPs por ecuaciones algebraicas e identificación de un algoritmo para resolver el sistema de ecuaciones. El diagrama conceptual se muestra en la figura 2.8.

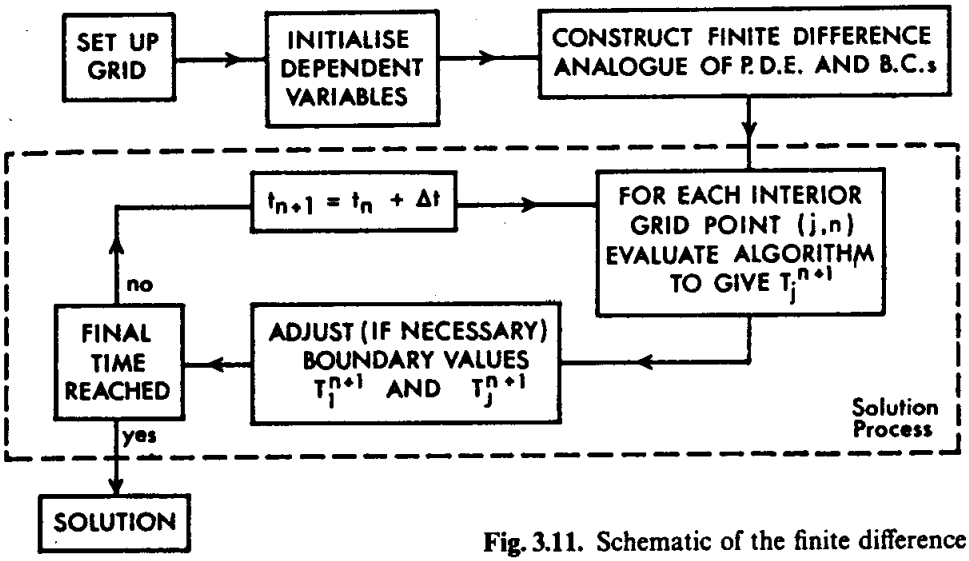

Fig. 3.11. Schematic of the finite difference solution process

Figura 2.8 – Esquema de solucion de EDPs por diferencias finitas.

El procedimiento puede ejemplificarse con la ecuación de difusión en 1D. Asumamos un caño

aislado térmicamente excepto en los bordes que tiene una temperatura inicial T(x,0)=0 C. A t=0 el caño se pone en contacto con fuentes de calor a T=100 C en sus extremos (figura 2.9). El problema es encontrar numéricamente la evolución de la temperatura T(x,t) para todos los puntos del caño.

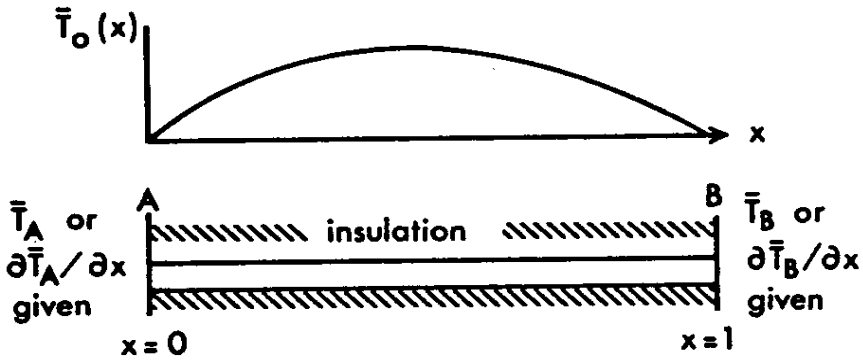

Figura 2.9 – Ilustración del problema a resolver usando la ecuación de difusión.

Si usamos es discretización FTCS, el algoritmo que resuelve la ecuación de difusión es

$$T_j^{n+1} = T_j^n + \frac{\alpha \Delta t}{\Delta x^2}(T_{j-1}^n - 2T_j^n + T_{j+1}^n)$$

Esta ecuación se aplica a todos los puntos interiores, j=2..., J-1. Los valores de T en los puntos en las fronteras están dados por las condiciones de borde. La solución se encuentra aplicando esta ecuación iterativamente para n=1,2...., hasta el tiempo máximo de integración.

2.5 Consistencia, estabilidad y convergencia

El problema de la estabilidad es fundamental en la resolución numérica de EDPs. En ausencia de experiencia computacional sería muy difícil adivinar que la inestabilidad sería un problema. Por ejemplo, L .F. Richardson, quien fue el primero en usar el método de diferencias finitas para resolver EDPs no descubrió el problema de la inestabilidad (ver *Weather Prediction by Numerical Processes*). Sin embargo, el problema de la estabilidad domina el diseño de cualquier algortimo computacional.

La relación entre estabilidad y convergencia empezó a aparecer en los trabajos de Courant, Friedrichs y Lewy en la década de 1920, fue identificada mas claramente por von Neumann en 1940 y llevada a una forma organizada por Lax y Richtmyer recień en la década del '50 en el teorema de Equivalencia de Lax.

Una cuestión fundamental en la resolución de ecuaciones por métodos computaciones es qué garantía puede ser dada de que la solución computacional esté cerca de la solución exacta de la EDP original y, en qué circustancias la solución computacional coincidirá con la solución exacta. La segunda parte de esta pregunta puede contestarse requiriendo que la solución aproximada **converja** a la a la solución exacta a medida que la grilla (Δt, Δx) tienda a cero. Por otro lado, es muy difícil de establecer la convergencia de la solución directamente por lo que en general se sigue la ruta indirecta indicada en la figura 2.10. La ruta indirecta requiere que el sistema de ecuaciones algebraicas formado en el proceso discretización sea **consistente** con las EDPs originales. Consistencia implica que el proceso de discretización pueda ser revertido, a través de un desarrollo de Taylor, para recuperar las ecuaciones originales. Asimismo, el algoritmo usado para resolver las ecuaciones algebraicas debe ser **estable**. Luego se invoca la seudo-ecuación

CONSISTENCIA + ESTABILIDAD = CONVERGENCIA (2.28)

para implicar la convergencia. Las condiciones bajo las cuales (2.28) es válida son dadas por el teorema de equivalencia de Lax.

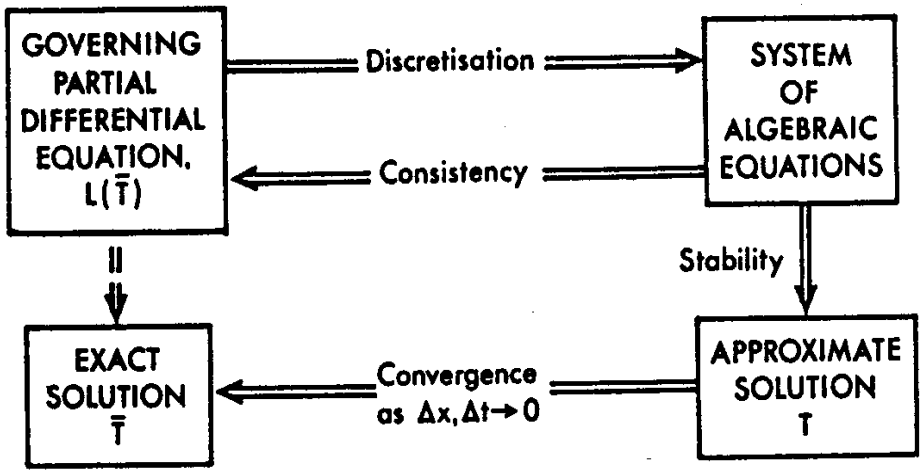

Figura 2.10 – Relación conceptual entre consistencia, estabilidad y convergencia.

2.5.1 Convergencia

La solución del sistema de ecuaciones algebraica que aproxima EDPs se dice convergente si la solución aproximada se acerca a la solución exacta a medida que el tamaño del espaciado de la grilla tiende a cero. O sea, se requiere que $T_j^n \rightarrow \bar{T}(x_j, t_n)$ cuando Δt, Δx → 0.

La diferencia entre la solución exacta de la EDP y la solución exacta del sistema de ecuaciones algebraicas se denomina error de la solución, y se calcula como $e_j^n = \bar{T}(x_j, t_n) - T_j^n$.

La solución exacta del sistema de ecuaciones algebraicas es la solución aproximada de la EDP y se obtiene cuando los cálculos numéricos no incluyen ningún tipo de error, ni siquiera de redondeo. La magnitud del error e^n_j depende del tamaño de la grilla (Δt, Δx) y de los valores de las derivadas de mayor orden en ese nodo no consideradas.

Teorema de Equivalencia de Lax

"Dado un problema de valor inicial lineal bien planteado y una aproximación por diferencias finitas consistente, la estabilidad es una condición necesaria y suficiente para la convergencia de la solución numérica."

Notas:

- A pesar de que el teorema está expresado en términos de una aproximación en diferencias finitas es aplicable a cualquier procedimiento de discretización que da lugar a valores desconocidos en nodos, por ejemplo el método de elementos finitos.
- La mayoría de los flujos reales son no-lineales y son problemas de frontera o mixtos por lo que el teorema de equivalencia de Lax no puede ser aplicado rigorosamente. En esos casos el teorema debe ser interpretado como aquel que provee las condiciones necesarias, pero no siempre suficientes, para la convergencia.

Convergencia numérica

Para las ecuaciones que gobiernan el movimiento de los fluídos es usualmente imposible demostrar teóricamente su convergencia. No obstante, para problemas que tienen soluciones exactas, como la ecuación de difusión, es posible determinar cómo cambia la solución numérica a medida que se refina la grilla. Convergencia implica que el error de la solución tiende a cero con el espaciado de la grilla.

La figura 2.11 muestra la dependencia del error cuadrático medio con el tamaño de Δx calculado con el programa dffsn1.f. Notar que a medida que Δx disminuye es necesario disminuir también el Δt en un factor de 4 para mantener s constante. De la figura se observa que la razón de convergencia va como Δx^2 para todo s, excepto para s=1/6 que va como Δx^4. Como se verá mas adelante la razón de convergencia para s=1/6 es menor pues el error de truncamiento se reduce.

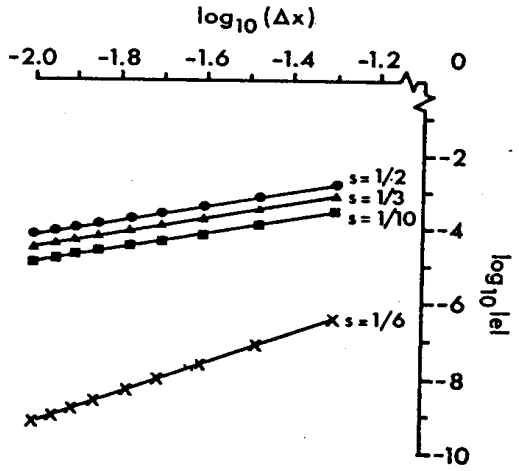

Figura 2.11 – Convergencia numérica para al esquema FTCS de la ecuación de difusión.

2.5.2 Consistencia

De acuerdo al teorema de Lax la consistencia es necesaria para que la solución aproximada converja a la solución de la EDP. El sistema de ecuaciones algebraicas generado por el proceso de discretización se dice consistente con la EDP original si, en el límite del espaciado de la grilla tendiendo a cero, el sistema algebraico es equivalente a la EDP en cada punto.

El procedimiento para verificar la consistencia requiere sustituir la solución exacta en las ecuaciones algebraicas que provienen de la discretización y la expansión de los valores en cada nodo como series de Taylor con respecto a un punto. Para que exista consistencia la expresión resultante debe consistir de la EDP original mas un residuo. El residuo debe ser tal que tienda a cero a medida que el espaciado de la grilla se reduce.

Para ilustrar el procedimiento consideramos la ecuación de difusión y la discretización FTCS. En ese caso la ecuación discretizada es

$$T_j^{n+1} = T_j^n + \frac{\alpha \Delta t}{\Delta x^2}(T_{j-1}^n - 2T_j^n + T_{j+1}^n) = sT_{j-1}^n + (1-2s)T_j^n + sT_{j+1}^n \quad . \qquad (2.29)$$

donde $s = \frac{\alpha \Delta t}{\Delta x^2}$. Sustituyendo la solución exacta de la ecuación de difusión en el nodo (j,n) que llamaremos $\hat{T}$, resulta

$$\hat{T}_j^{n+1} = s\hat{T}_{j-1}^n + (1-2s)\hat{T}_j^n + s\hat{T}_{j+1}^n \quad . \qquad (2.30)$$

Ahora determinamos cuan cercana es (2.30) de la ecuación de difusión (EDP) original en el nodo (j,n). Sustituyendo un desarrollo de Taylor en cada término de (2.30) produce

$$[\frac{\partial \hat{T}}{\partial t}]_j^n - \alpha[\frac{\partial^2 \hat{T}}{\partial x^2}]_j^n + E_j^n = 0 \qquad (2.31)$$

donde

$$E_j^n = 0.5\Delta t[\frac{\partial^2 \hat{T}}{\partial t^2}]_j^n - \alpha(\frac{\Delta x^2}{12})[\frac{\partial^4 \hat{T}}{\partial x^4}]_j^n + O(\Delta t^2, \Delta x^4) \quad . \qquad (2.32)$$

Se puede observar que (2.31) difiere de la ecuación de difusión en un término E^n_j que se denomina error de truncamiento. Este error tiende a cero a medida que el espaciado de la grilla se hace mas chico en cada punto (x_j,t_n). Entonces, en el límite $\Delta t, \Delta x \rightarrow 0$ el algoritmo (2.29) es equivalente a la ecuación de difusión y por lo tanto la ecuación algebraica es consistente con la EDP original.

Ejercicio: Demostrar que el error de truncamiento E^n_j tiende a cero mas rápidamente para s=1/6 que para cualquier otro valor de s a medida que el espaciado de la grilla disminuye (consistente con la figura 2.11).

2.5.3 Estabilidad

La estabilidad se puede pensar como la tendencia a decaer con el tiempo de cualquier perturbación espontánea, como errores de redondeo, en la solución de las ecuaciones algebraicas.

Es fácil demostrar que aún cuando las ecuaciones algebraicas de diferencias finitas sean equivalentes a las EDPs a medida que el espaciado de la grilla disminuye eso no significa que la solución de la ecuación en diferencias finitas converja a la solucion de la EDP. Para mostrar un ejemplo consideremos el programa diffsn1.f que resuelve la ecuación de difusión usando el algoritmo FTCS.

La figura 2.12 muestra la evolución de la solución producida por dffsn1.f para un valor de s=0.5 y otro valor de s=0.6 (Δx=0.1 en ambos casos). Claramente las figuras muestran un comportamiento muy diferente. Para s=0.6 aparece una oscilación que se origina en la línea de simetría y que se propaga hacia las fronteras que no tiene significado físico. La amplitud de la oscilación crece con el tiempo por lo que la solución numérica para s=0.6 no converge a la solución de la ecuación de difusión pues los errores numéricos crecen con el tiempo. El crecimiento de errores en cualquier momento del cálculo está asociado a la inestabilidad numérica.

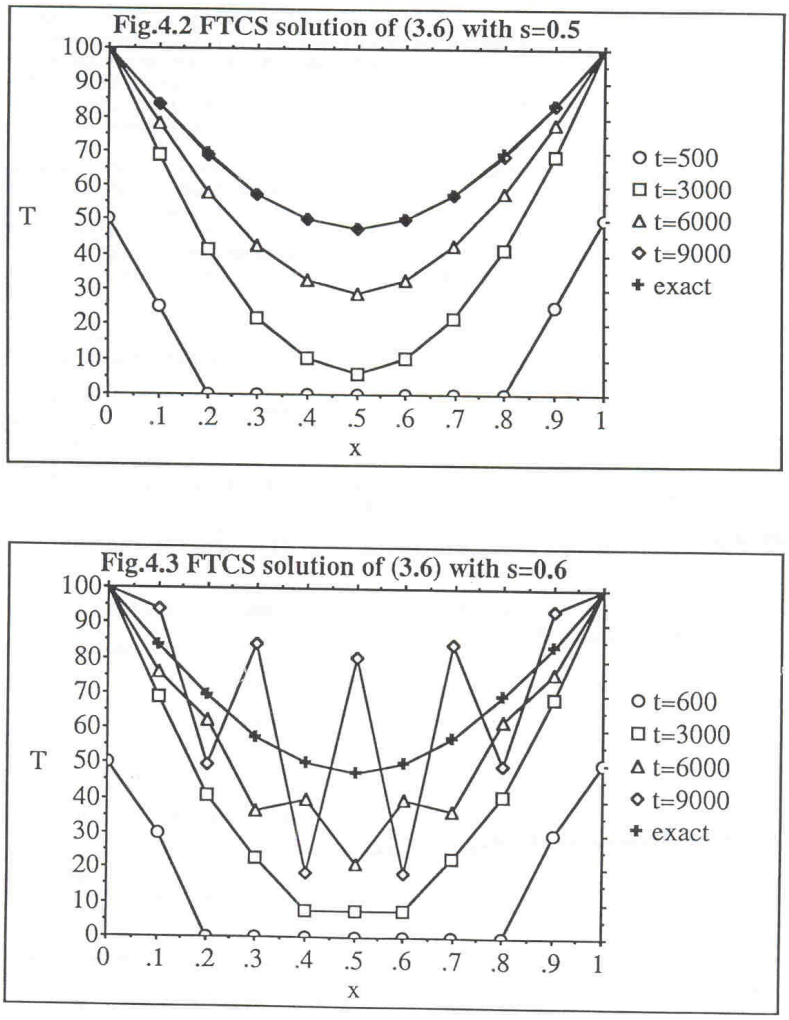

Figura 2.12 – Solución de la ecuación de difusión por FTCS usando s=0.5 (arriba) y s=0.6 abajo.

Cuando se corrió el programa diffsn1 tanto Δx como la difusividad (α=0.00001) estaban fijos por lo que al aumentar s aumentó el paso de integración Δt. El resultado sugiere que la inestabilidad numérica ocurrirá cuando el Δt supere un valor crítico. La pregunta es cómo podemos encontrar ese valor. Existen varios métodos para eso, pero nosotros consideraremos dos métodos muy comunes: el método de la matriz y el método de von Neumann. Ambos métodos están basados en predecir si existirá un crecimiento del error entre la solución verdadera del algoritmo numérico y la solución calculada, incluyendo errores de redondeo.

1) Método de la matriz: Esquema FTCS

Sea $\xi_j^n = T_j^n - \tilde{T}_j^n$ el error entre la solución verdadera del algoritmo numérico y la solución realmente calculada. Entonces se puede mostrar que ξ^n_j satisfacerá las mismas ecuaciones algebraicas que T en el interior del dominio.

Para la ecuación de difusión ξ^n_j satisfacerá (2.29) en el interior

$$\xi_j^{n+1} = s\,\xi_{j-1}^n + (1-2s)\,\xi_j^n + s\,\xi_{j+1}^n \qquad (2.33)$$

La condición de frontera para ξ^n_j es que $\xi^n_1 = \xi^n_J = 0$ para todo n pues los errores en las fronteras son nulos. Por lo tanto sustituyendo ξ^n_j en la ecuación (2.33) se obtiene para j=2, 3,... J-1

$$\begin{aligned}
\xi_2^{n+1} &= (1-2s)\xi_2^n + s\xi_3^n \ , \\
\xi_3^{n+1} &= s\xi_2^n + (1-2s)\xi_3^n + s\xi_4^n \ , \\
\xi_j^{n+1} &= s\xi_{j-1}^n + (1-2s)\xi_j^n + s\xi_{j+1}^n \ , \\
\xi_{J-1}^{n+1} &= s\xi_{J-2}^n + (1-2s)\xi_{J-1}^n \ .
\end{aligned}$$

En forma matricial la ecuaciones se pueden escribir como

$$\xi_j^n = A\xi^n \qquad n = 0,1,... \qquad (2.34)$$

donde A es una matrix cuadrada de dimensión (J-2) y ξ^n un vector de dimensión (J-2)

$$\underline{A} = \begin{bmatrix}
(1-2s) & s & & & & \\
s & (1-2s) & s & & & \\
 & s & . & s & & \\
 & & . & . & . & \\
 & & & . & . & . \\
 & & & & s & (1-2s)
\end{bmatrix}, \quad \xi = \begin{bmatrix}
\xi_2^n \\ . \\ \xi_j^n \\ . \\ . \\ \xi_{J-1}^n
\end{bmatrix}$$

Se puede mostrar que los errores ξ^n están acotados cuando n aumenta si los autovalores λ_m de la matriz A son todos distintos y

$$|\lambda_m| \leqslant 1 \tag{2.35}$$

Los autovalores de la matriz (tri-diagonal) A son

$$\lambda_m = 1 - 4s \sin^2\left(\frac{m\pi}{2(J-1)}\right) \qquad m = 1,2,\ldots,J-2 \tag{2.36}$$

La condición de estabilidad (2.35) entonces implica que s deber ser tal que cumpla

$$-1 \leqslant 1 - 4s \sin^2\left(\frac{m\pi}{2(J-1)}\right) \leqslant 1 \tag{2.37}$$

La desigualdad a la derecha de la ecuación se cumple automáticamente, mientras que la desigualdad de la izquierda requiere

$$0.5 \geqslant s \sin^2\left(\frac{m\pi}{2(J-1)}\right) \tag{2.38}$$

lo cual es válido para todo m si s≤0.5. Por lo tanto el algoritmo FTCS es estable para s≤0.5.

2) Método de von Neumann: Esquema FTCS

El análisis de von Neumann es el mas comunmente usado para determinar el criterio de estabilidad pues es el más fácil de usar y el que tiende a dar mejores resultados. El método es estrictamente válido solamente para establecer condiciones necesarias y suficientes de estabilidad para problemas de valor inicial lineales de coeficientes constantes.

En el método de von Neumann el error ξ^n_j es expandido en series de Fourier. Luego, la estabilidad o inestabilidad del algoritmo computacional se determina considerando si cada uno de los componentes de Fourier del error decae o se amplifica cuando se incrementa un paso de tiempo. Por lo tanto el vector error inicial ξ^0 es expresado como una serie de Fourier compleja finita de tal forma que en x_j el error es

$$\xi_j^0 = \sum_{m=1}^{J-2} a_m e^{i\theta_m j} \qquad j = 2,3,\ldots J-1 \tag{2.39}$$

donde $\theta_m = m\pi\,\Delta x$.

Dada esa forma del error inicial la solución ξ^n_j apropiada al esquema FTCS de la ecuación de difusuión (2.33) será de la forma

$$\xi_j^n = (G)^n e^{i\theta j} \tag{2.40}$$

donde el subíndice m ya no se usa pues es suficiente con estudiar la propagación del error debido a un solo término en (2.39) ya que el algoritmo computacional es lineal. La dependencia temporal de la componente de Fourier del error está contenida en el coeficiente complejo G (el supraíndice n indica elevación de G a la potencia n).

Sustituyendo (2.40) en (2.33) se obtiene

$$(G)^{n+1} e^{i\theta j} = s(G)^n e^{i\theta(j-1)} + (1-2s)(G)^n e^{i\theta j} + s(G)^n e^{i\theta(j+1)} \tag{2.41}$$

Eliminando $(G)^n e^{i\theta j}$

$$G = (1-2s) + s\left(e^{i\theta} + e^{-i\theta}\right) = 1 - 2s(1-\cos\theta) = 1 - 4s\sin^2(\theta/2) \tag{2.42}$$

El término G puede interpretarse como el factor amplificador del modo m de Fourier del error en cada paso temporal pues de (2.40) vale

$$\frac{\xi_j^{n+1}}{\xi_j^n} = G \quad . \tag{2.43}$$

Notemos que G depende de s y de θ, o sea que G(s,θ) depende del tamaño de la grilla y del modo de Fourier considerado pues $s=\frac{\alpha \Delta t}{\Delta x^2}$ y $\theta_m = m\pi \Delta x$. Los errores permanecerán acotados si el valor absoluto de G, llamado ganancia, es siempre menor que la unidad para todos los modos de Fourier. Por lo tanto el requerimiento de estabilidad es

$$|G| \leqslant 1 \quad \text{para todo } \theta \tag{2.44}$$

Entonces, de (2.42) el requerimiento de estabilidad para el esquema FTCS es

$$-1 \leqslant 1 - 4s \sin^2(\theta/2) \leqslant 1 \tag{2.45}$$

lo cual se cumple si s≤0.5. Notar que el resultado es el mismo que para el método de la matriz.

Las ecuaciones que gobiernan la atmósfera son nolineales, tienen coeficientes variables y las condiciones de borde son complicadas por lo que en ese caso el método de von Neumann puede aplicarse solo localmente y con las nolinealidades temporalmente fijas. En este caso mas general el método provee condiciones necesarias, pero no siempre suficientes, para la estabilidad.

Bibliografía principal

- Computational Techniques for Fluid Dynamics 1, Fletcher.