MÉTODOS COMPUTACIONALES en FÍSICA

Patricio Cordero S.

Departamento de Física

Facultad de Ciencias Físicas y Matemáticas

Universidad de Chile

versión 4 de julio de 2013

Índice general

Capítulo 1

Introducción

1.1. Usos del cálculo numérico

En problemas de todas las disciplinas, como ingeniería, economía, ciencias sociales, física, biología, hoy día se utiliza el cálculo numérico en el sentido de lo que se presenta en los próximos capítulos.

Una vez que un problema es planteado en la forma matemática propia a su disciplina éste debe ser reformulado para adecuarlo a lo que es el cálculo numérico. Un caso típico de física básica puede ser la ecuación básica de movimiento unidimensional de una partícula:

$$m\,\frac{dv}{dt} = f(x,v)$$

Se debe comenzar por usar tiempo discretizado para expresar la derivada como cuociente de cantidades finitas. Podría ser, por ejemplo

$$v_k = \frac{x_{k+1} - x_k}{h} \qquad \text{y también} \qquad a_k = \frac{v_{k+1} - v_k}{h}\,, \quad \text{donde} \quad k = 0,1,2\ldots$$

donde $t_k = hk$, $x_k = x(t_k)$, $v_k = v(t_k)$ y la fuerza $f(x,v)$ es una función conocida. De las expresiones anteriores se obtiene instrucciones apropiadas para incluir en un programa computacional:

$$x_{k+1} = x_k + h\,v_k\,, \qquad\qquad v_{k+1} = v_k + \frac{h}{m}\,f_k$$

donde, desde una condición inicial x_0 y v_0 se va, iterativamente, obteniendo los sucesivos valores (x_k, v_k). Este algoritmo se conoce como el algoritmo de Euler. Es fácil generalizar el algoritmo anterior a más dimensiones. Es sencillo, fácil de entender pero, como se verá, puede presentar problemas de estabilidad.

Una vez que el problema ha sido modelado con un conjunto de ecuaciones se debe explorar las implicaciones. De esas implicaciones puede resultar algo esperable pero cuantitativamente no trivial, puede ocurrir que el modelo resulte no ser bueno (dando comportamientos que no son verdaderos) etc. También puede suceder que el modelo dé patrones de comportamientos inesperados pero correctos.

Hoy en día es inconcebible no utilizar cálculo numérico en cada área de la ciencia y la tecnología. A continuación unos pocos ejemplos en física,

- Comportamiento de átomos, núcleos, y el amplio mundo subnuclear de física de partículas
- Dinámica de fluidos: tal como en meteorología, oceonografía, simulación de túneles de viento en el diseño de aviones, en el comportamineto de estrellas etc etc
- Mecánica macroscópica de sólidos: tensiones en estructutras complejas (puentes, barcos..), roturas, grietas, explosiones ..
- Comportamiento de las más variadas moléculas, incluyendo algunas enormes proteinas.
- Astrofísica y cosmología, evolución de galaxias, soluciones de las complicadas ecuaciones de gravitación.

Los esfuerzos computacionales para atacar un problema abarcan desde hacer integrales complicadas, hacer integrales en muchas (a veces demasiadas) dimensiones, estimar funciones partición en sistemas estadísticos, hasta resolver ecuaciones diferenciales ordinarias, ecuaciones diferenciales a derivadas parciales, etc.

Salvo que las ecuaciones del modelo que se estudia sean de naturaleza muy sencilla, lo más probable es que se requiera de una resolución numérica de ellas. Uno de los retos—cuando se resuelve numéricamente un conjunto de ecuaciones—es saber si la solución numérica obtenida es confiable, es decir, si realmente es una solución del problema o si los errores numéricos (o de otro tipo) que produce la metodología numérica usada invalidan total o parcialmente la solución obtenida.

En estas notas se aprenderá algunas técnicas para resolver ecuaciones de diversa naturaleza y en los casos más sencillos veremos también la forma de mantener los errores bajo control. Habrá un permanente trabajo práctico.

1.2. Errores

Al hacer cálculos numéricos introducimos errores que tienen diverso origen. Los más comunes son:

- Errores en la precisión de los datos. Por ejemplo, el valor de π puede ser de baja precisión, $3{,}1416$ en lugar de $3{,}14159265358979323 85\ldots$
- Errores de truncación. Por ejemplo, en lugar del valor e^x se usa $\sum_{k=0}^{N} \frac{x^k}{k!}$ y el N no es lo suficientemente grande. Estos errores aparecen normalmente por la naturaleza iterativa de los métodos y en algún momento es necesario detener la iteración. Por ejemplo, para calcular $\sin x$ se puede usar el siguiente código C,

```
se = 1.0;
x  = 0.3;          /* valor deseado  de x */
x2 = x*x;
u  = x;
for(k = 2; k<N; k+2)
{ u  = -u*x2/k/(k+2);
  se = se + u/k;
}
```

y el resultado analíticos es naturalmente mejor mientras mayor sea N, pero la precisión numérica alcaza su óptimo para N no muy grande.

- Errores de redondeo. Estos se deben al tamaño finito de la información que se guarda en memoria por cada número real. Por ejemplo, cuando calculamos

$$\cos(x) = \frac{d\sin(x)}{dx} \approx \frac{\sin(x+\varepsilon) - \sin(x-\varepsilon)}{2\varepsilon}$$

el resultados es mejor cuanto menor sea ε hasta que se produce un problema al restar dos números demasiado parecidos. Por ejemplo: sin.c

```
            epsilon         dsin(1.0)/dx   cos(1.0)   cos - deriv
 0.04978706836786394446248    0.5400791   0.5403023   0.000223185
 0.00247875217666635849073    0.5403018   0.5403023   0.000000553
 0.00012340980408667956121    0.5403023   0.5403023   0.000000001
 0.00000614421235332820981    0.5403023   0.5403023   0.000000000
 0.00000030590232050182579    0.5403023   0.5403023  -0.000000000
 0.00000001522997974471263    0.5403023   0.5403023  -0.000000000
 0.00000000075825604279119    0.5403023   0.5403023  -0.000000000
 0.00000000003775134544279    0.5403022   0.5403023   0.000000077
 0.00000000000187952881654    0.5402991   0.5403023   0.000003182
 0.00000000000009357622969    0.5401011   0.5403023   0.000201187
 0.00000000000000465888615    0.5442437   0.5403023  -0.003941389
 0.00000000000000023195228    0.4940679   0.5403023   0.046234439
```

Se debe probar con “float” y con “double”.

Estos valores se obtuvieron con el programa que sigue:

```
#include<stdio.h>
#include<math.h>
#include<stdlib.h>
#define     N        14
FILE        *archivo;

main()
{ int         ii;
  double      deriv,epsi,co;
  co = cos(1.0);
  archivo = fopen("sin.dat","wt"); /* w=write  t=text */
  for(ii=1; ii<N-1; ii++)
  { epsi   =  exp(-3.0*ii);
    deriv  =  (sin(1.0+epsi) - sin(1.0-epsi))/2.0/epsi;
    fprintf(archivo,"%26.23f %10.7f %10.7f %12.9f\n",
            epsi,deriv,co,co-deriv);
  }
  fclose(archivo);
}
```

1.3. Tiempo de cálculo

Cuando se programa un cálculo cuyo tiempo de cálculo sabemos que va a ser grande es importante tener alguna idea sobre los elementos que hacen que este cálculo sea lento y conviene

estudiar si hay alguna forma de optimizar. Por ejemplo, si en forma reiterativa en un programa se debe calcular una integral y se va a usar la fórmula

$$\begin{aligned} \int_a^b f(x)\,dx &\approx \left[f(x_0) + 2\sum_{i=1}^{N-1} f(x_i) + f(x_N) \right] h \\ x_i &= a + \frac{b-a}{N}\, i \end{aligned} \tag{1.3.1}$$

debe tenerse presente que este cálculo tarda un tiempo que es $\mathcal{O}(N)$.

El cálculo de la energía de N cargas implica calcular

$$\frac{m}{2}\sum_{k=1}^{N} v_k^2 + \sum_{j=1}^{N}\sum_{k=j+1}^{N} \frac{q_j q_k}{r_{jk}^2}$$

El primer término toma un tiempo $\mathcal{O}(N)$ mientras el segundo tiene un costo $\mathcal{O}(N^2)$. El algoritmo óptimo para invertir una matriz de $N \times N$ es $\mathcal{O}(N^3)$.

1.4. Adimensionalizar

Suele ocurrir que los problemas reales que debemos resolver tienen más parámetros de los necesarios en el sentido de que existe un problema numérico equivalente que se expresa con menos parámetros. Por ejemplo, en el caso de un oscilador armónico

$$m\ddot{x} = -kx, \qquad x(0) = A, \qquad \dot{x}(0) = v_0.$$

si se define $\omega_0^2 = k/m$ el problema aparenta tener tres parámetros de control: ω_0, A y v_0. Sin embargo, si se hace el cambio de variables $x = Az$, $t = t^*/\omega_0$, el problema equivalente es

$$z'' = -z, \qquad z(0) = 1, \qquad z'(0) = \frac{v_0}{A\,\omega_0} \equiv v_0^*.$$

que es un problema con un solo parámetros de control, v_0^*.

Si la adimensionalización se escoge con cuidado se trabaja con cantidades de orden 1 y por tanto se disminuye una fuente de errores.

Capítulo 2

Derivadas e integrales numéricas

2.1. Derivadas

La forma elemental más típica de plantear una derivada es

$$f'(x) \approx \frac{f(x+h)-f(x)}{h} \tag{2.1.1}$$

El desarrollo en serie

$$f(x+h) = f(x) + h\,f'(x) + \frac{h^2}{2}\,f''(x) + \ldots \tag{2.1.2}$$

muestra que en (2.1.1) se desprecia algo que es $\mathcal{O}(h)$. En cambio la siguiente expresión es más precisa,

$$f'(x) = \frac{f(x+h)-f(x-h)}{2h} + \mathcal{O}(h^2) \tag{2.1.3}$$

El error que aquí se indica es un error analítico. Ya se ha comentado que si h es muy pequeño se produce un error de redondeo.

Existe una famila de expresiones para derivadas de cualquier orden. Expresiones simétricas y no simétricas. Usando la notación $f_k \equiv f(x+kh)$, se tiene, por ejemplo, que

$$\begin{aligned} f_{\pm 1} &= f_0 \pm h\,f_0' + \frac{h^2}{2}\,f_0'' \pm \frac{h^3}{3!}\,f_0''' + \mathcal{O}(h^4) \\ f_{\pm 2} &= f_0 \pm 2h\,f_0' + 2h^2\,f_0'' \pm \frac{4h^3}{3}\,f_0''' + \mathcal{O}(h^4) \end{aligned} \tag{2.1.4}$$

de donde

$$f_0'' = \frac{f_2 - 2f_1 + f_0}{h^2} + \mathcal{O}(h) \tag{2.1.5}$$

y también

$$f_0'' = \frac{f_1 - 2f_0 + f_{-1}}{h^2} + \mathcal{O}(h^2) \tag{2.1.6}$$

Hay casos en que no se conoce la función en intervalos regulares. En lugar de intentar un método de interpolación—que se discuten más adelante—se puede usar expresiones como las que siguen,

Es fácil ver que la primera derivada de una función $f(x)$ se puede expresar en términos de $f_{-h_1} = f(x-h_1)$, $f_{h_2} = f(x+h_2)$ y de la propia $f(x)$ como

$$\frac{df}{dx} \approx \frac{h_1^2 f_{h_2} + (h_2^2 - h_1^2)\, f - h_2^2 f_{-h_1}}{h_1 h_2 \,(h_1 + h_2)} + \mathscr{O}(h_1 h_2\, f''') \tag{2.1.7}$$

y con los valores de f en estos mismos tres puntos la segunda derivada se puede escribir

$$\frac{d^2 f}{dx^2} \approx 2\frac{h_1 f_{h_2} - (h_1 + h_2)\, f + h_2 f_{-h_1}}{h_1 h_2 \,(h_1 + h_2)} + \mathscr{O}((h_2 - h_1) f''') \tag{2.1.8}$$

si se compara los errores analíticos en estas expresiones con los asociados a las derivadas simétricas del mismo orden: (2.1.3) y (2.1.6) se observa que los errores son del mimso orden.

♠ *Determine qué derivada es proporcional a*

$$11 f_{-2} - 56 f_{-1} + 114 f_0 - 104 f_1 + 35 f_2$$

e indique el orden del error de la esta expresión para la correspondiente derivada.

♠ *Obtenga el valor de a tal que la siguiente expresión sea la primera derivada f' más un error,*

$$\frac{a f_{-2} - 16 f_{-1} + 36 f_0 - 48 f_1 + 25 f_2}{\text{denom}}$$

Dé expresión para el denominador y para ese error.

2.1.1. Tabla con derivadas a cuatro y cinco puntos

Una derivada de orden n tiene una variedad de expresiones usando $p \geq n+1$ puntos. A continuación algunos ejemplos.

	A cuatro puntos	A cinco puntos
hf'	$\frac{1}{6}(-2f_{\mp1} - 3f_0 + 6f_{\pm1} - f_{\pm2})$	$\frac{1}{12}(f_{-2} - 8f_{-1} + 8f_1 - f_2)$
$h^2 f''$	$f_{-1} - 2f_0 + f_1$	$\frac{1}{12}(-f_{-2} + 16f_{-1} - 30f_0 + 16f_1 - f_2)$
$h^3 f'''$	$\pm(-f_{\mp1} + 3f_0 - 3f_{\pm1} + f_{\pm2})$	$\frac{1}{2}(-f_{-2} + 2f_{-1} - 2f_1 + f_2)$
$h^4 f^{iv}$	no hay	$f_{-2} - 4f_{-1} + 6f_0 - 4f_1 + 2f_2$

Se trata de expresiones tan simétricas como es posible.

2.2. Integración numérica directa

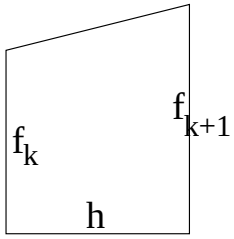
2.2.1. Método trapezoidal

Se desea integrar numéricamente dividiendo el intervalo (a,b) en N intervalos de largo h con los puntos $x_0 = a$, x_1, x_2, ..., x_{N-1}, $x_N = b$. Para obtener este primer algoritmo de integración se comienza por escribir

$$\begin{aligned} f(x) &\approx f_k + (x-x_k)\, f_k' + \frac{1}{2}(x-x_k)^2\, f_k'' + \dots \\ &\approx f_k + (x-x_k)\,\frac{f_{k+1}-f_k}{h} + \mathscr{O}\left((x-x_k)^2\right) \end{aligned} \tag{2.2.1}$$

para integrar en uno solo de los intervalos: desde x_k hasta $x_k + h$,

$$\begin{aligned} \int_{x_k}^{x_k+h} f(x)\,dx &\approx f_k\, h + \frac{h^2}{2}\frac{f_{k+1}-f_k}{h} + \mathscr{O}(h^3) \\ &= \frac{h\,(f_{k+1}+f_k)}{2} + \mathscr{O}(h^3) \end{aligned} \tag{2.2.2}$$



La última expresión es el área del trapecio de la figura. Al sumar k sobre N sitios y tomando en cuenta que $N \sim \frac{1}{h}$ se obtiene, sumando las áreas de los trapecios, que

$$\begin{aligned} \int_a^b f(x)\,dx &= \left(\frac{f_0+f_1}{2} + \frac{f_1+f_2}{2} + \dots \frac{f_{N-1}+f_N}{2}\right) h \\ &= \frac{h}{2}\,(f_0 + 2f_1 + 2f_2 + .. + 2f_{N-1} + f_N) + \mathscr{O}(h^2) \end{aligned} \tag{2.2.3}$$

y el error es de orden $\mathscr{O}(h^2\, f'') = \mathscr{O}(\frac{(b-a)^3\, f''}{N^2})$.

2.2.2. Métodos de Simpson

2.2.2.1. Simpsion $1/3$

Una fórmula algo más precisa es la de Simpson que surge de integrar en forma explícita en x entre $x_{k-1} = x_k - h$ y $x_{k+1} = x_k + h$. La expresión

$$f(x) = f_k + \frac{f_{k+1}-f_{k-1}}{2h}(x-x_k) + \frac{f_{k-1}-2f_k+f_{k+1}}{h^2}\,\frac{(x-x_k)^2}{2} + \mathscr{O}\left((x-x_k)^3\right) + \mathscr{O}\left((x-x_k)^4\right) \tag{2.2.4}$$

```
/*      "simpson.c"
    Programa generico para hacer
    la integral de F(x) desde
    x=a  hasta  x=b  usando
    el metodo de Simpson

    Autor: anonimo
*/

#include<stdio.h>
#include<math.h>
#include<stdlib.h>

#define N    20
#define nu   (N/2)
#define a    0.0//limite inferior
#define b    1.0//limite superior
#define h    ((b-a)/N)

double F(double x) //Aqui se pone
{ return(x*x*x*x); //integrando F
}
```

```
double simpson()
{ int k;
  double sumaPar,sumaImp,xk;
  sumaPar = 0.0;
  sumaImp = 0.0;
  xk      = a;
  for(k=0; k<nu-1; k++)
  {   xk      += h;
      sumaImp += F(xk);
      xk      += h;
      sumaPar += F(xk);
  }
  sumaImp = sumaImp + F(xk+h);
  sumaPar = 2.0*sumaPar +F(a) +F(b);
  return((4.0*sumaImp + sumaPar)*h/3.0);
}

main()
{ double inte;
  inte = simpson();
  printf("integral = %14.11f\n",inte);
}
```

La integración en (x_k-h, x_k+h) de los términos $(x-x_k)^r$ con r impar da cero. De la expresión anterior sobrevive la integración del término con $(x-x_k)^2$, el término cúbico no contribuye al error y el último da un $\mathscr{O}(h^5)$ y se obtiene

$$2h\, f_k + \frac{f_{k-1}-2f_k+f_{k+1}}{h^2}\,\frac{1}{2}\frac{2h^3}{3} = \frac{h}{3}\,(f_{k-1}+4f_k+f_{k+1}) + \mathscr{O}(h^5) \tag{2.2.5}$$

Componiendo esta expresión se obtiene el algoritmo de Simpson "$\frac{1}{3}$" que se aplica con N par,

$$\int_a^b f(x)\,dx \approx \frac{h}{3}\,[f_0 + 4\,(f_1+f_3+..f_{N-1}) + 2\,(f_2+f_4+..+f_{N-2}) + f_N] + \mathscr{O}(h^4) \tag{2.2.6}$$

y el error más precisamente es $\mathscr{O}(h^4\, f^{\mathrm{IV}})$

Viendo la lógica de (2.2.4) y cómo conduce a (2.2.6) resulta obvio obtener expresiones aun más precisas para hacer una integral.

2.2.2.2. Simpson $3/8$

Esta vez el dominio total de integración (a,b) se divide en N intervalos de tamaño h, donde N es múltiplo de 3, $N=3m$. Se comienza buscando una forma aproximada de la integral desde x_i hasta x_{i+3}, donde $x_{i+j} = x_i + jh$.

Se define un polinomio en $y = x_i + \frac{3}{2}h$, de modo que el punto $y=0$ corresponda, como lo muestra la figura 2.1, al punto central del dominio de integración. Para este dominio se define el polinomio

$$p(y) = b_0 + b_1\frac{y}{h} + b_2\left(\frac{y}{h}\right)^2 + b_3\left(\frac{y}{h}\right)^3 \tag{2.2.7}$$

cuyos coeficientes se determinan exigiendo las siguientes cuatro igualdades,

$$p(0) = f_i\,, \qquad p(h) = f_{i+1}\,, \qquad p(2h) = f_{i+2}\,, \qquad p(3h) = f_{i+3}. \tag{2.2.8}$$

Para efectos de saber el valor de la integral, basta con conocer los coeficientes de las potencias pares de y en (2.2.7). En efecto, se obtiene que

$$\int_{-3h/2}^{3h/2} p(y)\,dy = \left(3b_0 + \frac{9}{4}b_2\right) h \tag{2.2.9}$$

De las ecuaciones (2.2.8) se obtiene, en particular, que

$$\begin{aligned} b_0 &= -\frac{1}{16}\left(f_i - 9f_{i+1} - 9f_{i+2} + f_{i+3}\right), \\ b_2 &= \frac{1}{4}\left(f_i - f_{i+1} - f_{i+2} + f_{i+3}\right) \end{aligned} \tag{2.2.10}$$

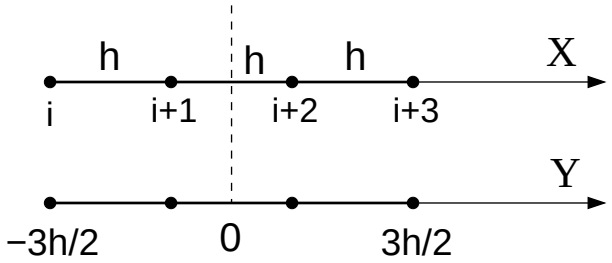


Figura 2.1: *Se integra sobre el dominio $x_i \le x \le x_{i+3}$ de largo $3h$, lo que equivale a integrar usando la variable y en el dominio $-3h/2 \le y \le 3h/2$.*

Lo que lleva a

$$\int_{x_i}^{x_{i+3}} f(x)\,dx \approx \int_{-3h/2}^{3h/2} p(y)\,dy = \frac{3h}{8}\left(f_i + 3f_{i+1} + 3f_{i+2} + f_{i+3}\right) \tag{2.2.11}$$

Si este cálculo se usa en todo el dominio se tiene

$$\begin{aligned} \int_a^b f(x)\,dx &\approx \frac{3h}{8}\left(f_0 + 3f_1 + 3f_2 + f_3 + f_3 + 3f_4 + 3f_5 + f_6 + f_6 + 3f_7 + 3f_8 + f_9 + \ldots\right) \\ &\approx \frac{3h}{8}\left(f_0 + 3f_1 + 3f_2 + 2f_3 + +3f_4 + 3f_5 + 2f_6 + 3f_7 + 3f_8 + 2f_9 + \ldots\right) \\ &\approx \frac{3h}{8}\left[f_0 + 3\sum_{k=0}^{m-1}\left(f_{3k+1} + f_{3k+2}\right) + 2\sum_{k=1}^{m-1} f_{3k} + f_N\right]_{N=3m} \end{aligned} \tag{2.2.12}$$

2.2.3. Discretización no uniforme sencilla

Tanto el método trapezoidal como el de Simpson han sido planteados con discretización uniforme, pero no es necesario proceder de ese modo. En el caso del método trapezoidal se puede tomar cada contribución (2.2.2) con un h propio, y la integral es

$$I = \sum_{k=0}^{N-1} \frac{h_k}{2}\left(f_k + f_{k+1}\right)$$

En el caso del método de Simpson $1/3$ se integró los intervalos de a pares y fue crucial que los dos miembros de cada par fueran iguales, pero distintos pares pueden tener h_k distintos. La integral queda

$$I = \sum_{k=1,3,5\ldots}^{N-1} \frac{h_k}{3}\left(f_{k-1} + 4f_k + f_{k+1}\right)$$

Esta vez la suma va de par en par de intervalos por lo que es necesario sumar solo sobre índices impares. El primero es $k = 1$ y el último es $k = N - 1$.

2.2.4. Limitaciones

Estos métodos no debieran o no pueden ser usados en forma directa si:

- el intevalo de integración es infinito
- la función varía mucho en el intervalo (función con alto contraste)
- hay una divergencia en el intervalo
- ..

2.3. Integración y cambio de variable

2.3.1. Planteamiento y ejemplos

En general para hacer una integral numérica es conveniente hacer algún tipo de cambio de variable. En particular los problemas mencionados antes se superan haciendo un cambio de variable de integración $y = g(x)$, esto es, $dy = g'(x)\,dx$. Genéricamente

$$\begin{aligned} I &= \int_a^b f(x)\,dx \\ &= \int_{g(a)}^{g(b)} \left[\frac{f(x)}{g'(x)}\right]_{x=g^{-1}(y)} dy \end{aligned} \tag{2.3.1}$$

y la segunda forma de la integral se discretiza uniformemente. Nótese que $g(x)$ debe ser monótona para que g' no tenga ceros en el intervalo que interesa. Discretizar uniformemente en la nueva variable y es equivalente a una discretización no uniforme en la variable original x. Otra limitante práctica para $g(x)$ es que debemos conocer la función inversa g^{-1}.

El procedimiento práctico normalmente define x una sola vez—en la rutina un $x = g^{-1}(y)$—el que es usado para calcular $[f(x)/g'(x)]$. Es decir, se genera la secuencia regular de valores y, con cada uno de ellos se calcula x, y se va sumando $f(x)/g'(x)$.

Al hacer un cambio de variable se debe cuidar que los valores de

$$s(y) = \left[\frac{f(x)}{g'(x)}\right]_{x=g^{-1}(y)} \tag{2.3.2}$$

sean finitos en todo el intervalo, en particular en los extremos $g(a)$ y $g(b)$.

Al hacer el cambio de variable $y = g(x)$ se debe cumplir:

a) $g(x)$ es monótona en el intervalo (a,b) original,

b) el nuevo intervalo $(g(a), g(b))$ es finito

c) el nuevo integrando $s(y)$ debe ser regular y de poco contraste.

Como ya se dijo, el cambio de variable equivale a tomar intervalos no regulares en la variable original x. Los puntos regulares y_k en la nueva variable definen puntos $x_k = g^{-1}(y_k)$ en el eje original.

El gran inconveniente de los métodos con cambio de variable presentados hasta aquí es que está limitado a funciones $g(x)$ para las cuales se conoce la función inversa $g^{-1}(y)$. Más adelante, en el capítulo correspondiente a los métodos Monte Carlo se podrá superar este inconveniente.

Ejemplo con intervalo infinito: Consideremos

$$I = \int_0^\infty e^{-x^2+x}\,dx$$

Si se toma $y = e^{-x^2}$ y por tanto $x = \sqrt{-\ln y}$. La integral pasa a ser

$$I = \int_0^1 \left[\frac{e^x}{2x}\right]_{x=\sqrt{-\ln y}} dy$$

que no es aceptable porque se obtiene un integrando divergente.

Pero si se escoge $y = g(x) = e^{-x}$, es decir,

$$s(x) = e^{-x^2+2x}$$

la integral que se debe evaluar es

$$I = \int_0^1 \left[e^{-x^2+2x}\right]_{x=-\ln y} dy$$

El problema ha sido reducido al de una integral en intervalo finito y poco contraste.

Integrandos con mucho contraste: Aun en casos en que no haya divergencias, si la función varía mucho en el intervalo (mucho contraste) se debe hacer el cambio $y = g(x)$ pasando así a una integral sobre la variable y con integrando $[f(x)/g'(x)]_{x=g^{-1}(y)}$ y la función $g(x)$ debe escogerse de tal forma que el nuevo integrando sea lo más plano posible, es decir, con poco contrate. El colmo sería conseguir que fuese una constante, pero en tal caso el problema estaría resuelto antes de comenzar.

Veamos cómo suavizar el integrando con el ejemplo

$$\int_{-1}^1 f(x)\,dx \qquad \text{con} \qquad f(x) = \frac{1}{\tau}e^{-x^2/\tau^2} \quad \text{y} \quad \tau \ll 1.$$

Se trata de buscar un $g(x)$ apropiado. Puesto que g' tiene que tener una forma parecida al integrando $f(x)$ es necesario encontrar una función g con la forma de un escalón redondeado. Escojamos que satisfaga $g(1) = 1$ y $g(-1) = -1$. Por ejemplo, se puede tomar

$$g(x) = \frac{\arctan\left(\frac{x}{a}\right)}{\arctan\left(\frac{1}{a}\right)} \qquad \Leftrightarrow \qquad x = g^{-1}(y) = a\tan\left(y\arctan\frac{1}{a}\right)$$

Se deja como ejercicio ver el a óptimo para cada valor de τ.

Si el integrando tiene muchos picos se subdivide el intervalo de integración para tener integrales con un solo pico en cada segmento y tratar cada caso según lo que convenga.

♠ *Encontrar un cambio de variable apropiado para calcular*

$$\int_1^{40} \frac{dx}{x^2\,(1+x)}$$

2.3.2. Divergencias en el integrando

2.3.2.1. Método 1: regularización del integrando

Si hay divergencias en el intervalo pero aun así la integral es finita, se debe tratar separadamente cada parte. Para ello se redefine intervalos de integración que dejan al punto de divergencia en un extremo para pasar a estudiar la forma de tratar una integral que es divergente en un extremo del intervalo. Tomemos el caso

$$I = \int_0^b f(x)\,dx \qquad \text{con } f(0) = \infty$$

Para que I sea convergente a pesar del valor infinito de f en $x = 0$ es necesario que

$$\lim_{x\to 0} x\,f(x) = 0$$

Para abordar este problema suele ser conveniente hacer el cambio de variable $y = g(x) = x^\alpha$, con $\alpha > 0$ porque $dy = \alpha\,x^{\alpha-1}\,dx$ y

$$I = \int_0^{b^\alpha} \left[\frac{f(x)}{\alpha\,x^{\alpha-1}}\right]_{x=y^{1/\alpha}} dy$$

y se debe escoger α tal que

$$\lim_{y\to 0} \frac{f(x)}{x^{\alpha-1}} \quad \text{sea finito}$$

Pero I es no divergente tan solo si $f(x)$ diverge en el origen más lentamente que $1/x$. Definamos δ, $0 < \delta < 1$, tal que

$$|f(x \approx 0)| < \frac{A}{x^{1-\delta}}$$

Lo que intereza es la contribución a la integral que proviene de una vecindad al origen,

$$\begin{aligned} I_h &= \int_0^{h^\alpha} \left[\frac{f(x)}{\alpha\,x^{\alpha-1}}\right]_{\%} dy \\ &\leq \int_0^{h^\alpha} \left[\frac{A}{\alpha\,x^{1-\delta}\,x^{\alpha-1}}\right]_{\%} dy \\ &\leq \frac{A}{\alpha}\int_0^{h^\alpha} \left[x^{\delta-\alpha}\right]_{\%} dy \\ &\leq \frac{A}{\alpha}\int_0^{h^\alpha} y^{(\delta-\alpha)/\alpha}\,dy \end{aligned} \tag{2.3.3}$$

que es convergente si $\alpha \leq \delta$. Se debe escoger un α positivo menor o igual a δ.

Ejemplo 1: Calcular:

$$\int_0^1 \frac{x^{1/3}}{\sin x}\,dx$$

Cerca del origen el integrando es $f \sim x^{-2/3}$ es decir, $\delta = 1-2/3 = 1/3$ y se puede escoger cualquier α tal que $0 < \alpha \leq 1/3$. Si,por ejemplo, se toma $\alpha = 1/3$ la integral se convierte en

$$3\int_0^1 \frac{y^3}{\sin y^3}\,dy$$

Ejemplo 2: Para calcular

$$\int_0^1 P(x)\,\ln(x)\,dx$$

donde $P(x)$ es una función suave, basta con tomar $y = x^{1/100}$ para tener un integrando $F(y)$ suave.

♠ *Calcular*

$$\int_0^1 \frac{\sin(x)}{\sqrt{1-x^2}}\,dx$$

2.3.2.2. Método 2: tratamiento analítico de la divergencia

Otra forma de tratar integrales que tienen divergencias en el integrando es tratar en forma analítica el trozo que contiene la divergencia. Por ejemplo,

$$\int_0^1 \frac{dx}{(1-x)^{1/3}x^{2/3}}$$

Para tratar la singularidad en $x = 0$ se separa la integral

$$\int_0^h \frac{dx}{(1-x)^{1/3}x^{2/3}} \approx \int_0^h \frac{dx}{x^{2/3}} = 3h^{1/3}$$

se procede en forma similar en el límite superior. El resto de la integral se hace en la forma usual.

2.4. Integral de parte principal

Suele ser necesario calcular la integral

$$I = \int_a^b \frac{f(x)}{x-x_0}\,dx \qquad \text{con} \qquad a \leq x_0 \leq b$$

en que tanto la integral desde a a x_0 como la de x_0 a b son divergentes y $f(x)$ es regular en $x = x_0$. La parte principal de I, $\mathscr{P}(I)$, se define por medio de

$$\int_a^b \frac{f(x)}{x-x_0}\,dx = \mathscr{P}\int_a^b \frac{f(x)}{x-x_0}\,dx + i\pi\, f(x_0)$$

donde la *parte principal* es

$$\mathscr{P}\int_a^b \frac{f(x)}{x-x_0}\,dx = \lim_{h=0}\left[\int_a^{x_0-h}\frac{f(x)}{x-x_0}\,dx + \int_{x_0+h}^b \frac{f(x)}{x-x_0}\,dx\right]$$

Para calcular numéricamente la parte principal se razona a partir de reescribir I en la forma

$$\int_a^b \frac{f(x)}{x-x_0}\,dx = \int_a^b \frac{f(x)-f(x_0)}{x-x_0}\,dx + \int_a^b \frac{f(x_0)}{x-x_0}\,dx \tag{2.4.1}$$

La primera integral, que denotamos I_1, es no singular y se hace en forma estándar, mientras que la segunda integral es

$$\begin{aligned} I_2 = \int_a^b \frac{f(x_0)}{x-x_0}\,dx &= f(x_0)\ln\frac{b-x_0}{a-x_0} \\ &= f(x_0)\left(\ln\frac{b-x_0}{x_0-a} + \ln(-1)\right) \\ &= f(x_0)\left(\ln\frac{b-x_0}{x_0-a} + i\pi\right) \end{aligned} \tag{2.4.2}$$

Se ha aislado el término $i\pi f(x_0)$. La labor de obtener numéricamente la parte principal consiste en evaluar numéricamente la integral I_1 utilizando algún método estándar, para luego sumarle $f(x_0)\ln\frac{b-x_0}{x_0-a}$.

2.5. Problemas

3.1 1. Tal vez sea bueno comenzar por escribir programas de integración trapezoidal y Simpson e intregrar sin cambio de variable

$$\int_0^1 x^n\,dx \quad \text{con} \quad n = 3,4,5\ldots$$

viendo cuánto debe valer N para tener un error de menos del 1 %.

2. Trate de obtener un error menor al 1%.
3. Puede serle útil graficar la función a integrar y la función que resulta después de cada cambio de variable. De esa manera se puede entender la fuente de los posibles errores.
4. Por razones obvias, no se debe calcular la integral por partes, ni hacer algún truco que permita llevarla a una integral analítica.

3.2 La función gamma, $\Gamma(x)$, se define como la siguiente integral

$$\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}\,dt \tag{2.5.1}$$

que converge para todo x positivo, pese a que para $0 < x < 1$ el integrando tiene una divergencia en $t = 0$.

Se pide calcular numéricamente, a partir de la definición anterior, la función Γ para $x = 10$ y $x = 1/2$, valores para los cuales se conocen los resultados analíticos:

$$\Gamma(10) = 9! = 362880 \tag{2.5.2}$$
$$\Gamma(1/2) = \sqrt{\pi} \tag{2.5.3}$$

En cada caso se debe indicar el(los) cambio(s) de variable usado(s), el número de puntos en la discretización, el método de integración (trapezoidal o Simpson), el resultado obtenido y el error cometido respecto al valor analítico.

3.3 Calcule numéricamente las integrales

$$I^{(1)} = \int_0^\infty e^{-x} \ln x \, dx$$

$$I^{(2)} = \int_0^1 \frac{1+x}{1-x^3} \ln \frac{1}{x} \, dx$$

El problema consiste en hacer numéricamente las integrales de arriba con algún cambio de variable para tener un integrando suave en un intervalo finito. Se debe obtener un resultado razonablemente bueno teniedo que evaluar el integrando final el menor número (N) de veces que sea posible. Como criterio de convergencia debe usar alguna cantidad como

$$\text{err} = \frac{I - I_N}{I} < 10^{-q}$$

con $q = 2,3,4,5,6$. Como ya se dijo, una de las metas es conseguir que N sea lo menor posible teniedo un resultado confiable.

En cada caso se debe indicar el (los) cambio(s) de variable usado(s), el número N de puntos en la discretización, el método de integración (trapezoidal o Simpson, nada superior), el resultado obtenido y el error numérico respecto al valor de I. **No** debe usar el conocimiento analítico de la integral sino su propio criterio para estimar ese error. Explique y justifique.

Dibuje el integrando $f(x)$ y separadamente el integrando final $h(y) = [f(x)/g'(x)]_{x=...}$ que haya usado (cada cual es su dominio). Dibujar los valores I_N versus $\frac{1}{N}$ para algunos valores de N.

Por razones obvias, no se permite recurrir a integración por partes, ni hacer algún truco que permita llevarla a una integral analítica.

3.4 Motivación física. Para muchos efectos la fuerza entre átomos puede ser tratada exitosamente con el potencial central, llamado de Lennard-Jones,

$$V = 4V_0 \left[\left(\frac{a}{r}\right)^{12} - \left(\frac{a}{r}\right)^{6} \right] \tag{2.5.4}$$

cuyo valor mínimo es V_0 y se anula cuando r coindice con el radio de Bohr. Una partícula atrapada en este potencial (energía menor que cero), tiene un movimiento en el intervalo $(r_{\min}, r_{\max})$ donde ambos radios son mayores que a. Cuánticamente solo hay un conjunto discreto de energías E_n posibles. Clásicamente $E = \frac{p^2}{2m} + V(r)$ o equivalentemente la magnitud del momentum depende de r en la forma $p(r) = \sqrt{2m(E - V(r))}$. Una forma aproximada

de plantear el problema de encontrar los valores de los niveles cuánticos E_n consiste en exigir la *condición de Bohr-Sommerfeld*

$$\oint \frac{p(r)}{\hbar}\, dr = 2\pi \left(n + \frac{1}{2}\right)$$

con n entero nonegativo. La integral es sobre un ciclo completo de oscilación. El problema se adimensionaliza haciendo las sustituciones

$$E = V_0 \mathscr{E}\,, \qquad r = a\,\rho\,, \qquad V_0 = \frac{\gamma^2 \hbar^2}{2\,a^2\,m}$$

El valor de γ en el caso de la molécula de hidrógeno es $21{,}7$, en el de O_2 es ~ 150.

La condición integral de arriba se convierte en la exigencia que se anule la función

$$F_n(\mathscr{E}_n) = \gamma \int_{\rho_{\min}}^{\rho_{\max}} \sqrt{\mathscr{E}_n - 4\left(\frac{1}{\rho^{12}} - \frac{1}{\rho^6}\right)}\, d\rho - \pi \left(n + \frac{1}{2}\right) \tag{2.5.5}$$

Es decir, el problema consiste en encontrar los ceros de F_n dados $\gamma = 150$ y $n = 0, 1, 2$ con $-1 < \mathscr{E}_n < 0$ sabiendo que

$$\rho_{\min} = \left(\frac{2 - 2\sqrt{\delta_n}}{1 - \delta_n}\right)^{1/6}, \quad \rho_{\max} = \left(\frac{2 + 2\sqrt{\delta_n}}{1 - \delta_n}\right)^{1/6}$$

donde $\delta_n = 1 + \mathscr{E}_n$.

El programa que se diseñe debe ser útil también con otros potenciales $V(r)$.

En la búsqueda de los ceros debe usar el método de la secante (indicando, entre otras cosas, la tolerancia usada y cuántas iteraciones fueron necesarias).

Capítulo 3

Algebra lineal, interpolación, recurrencias y ceros

3.1. Temas de álgebra lineal

Los autovalores de una matriz de $n \times n$ resulta de determinar los ceros del polinomio característico.

3.1.1. Eliminación de Gauss

Se parte con un sistema inicial de ecuaciones

$$\begin{array}{lll} a_{11}x_1 + a_{12}x_2 + \ldots & a_{1n}x_n & = b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots & a_{2n}x_n & = b_2 \\ \ldots & \ldots & \\ a_{n1}x_1 + a_{n2}x_2 + \ldots & a_{nn}x_n & = b_n \end{array} \tag{3.1.1}$$

Este problema puede ser planteado como el de una matriz $\mathbf{A}$ multiplicando a un vector desconocido $\vec{x}$ tal que se obtiene un vector $\vec{b}$:

$$\mathbf{A}\vec{x} = \vec{b} \tag{3.1.2}$$

y se plantea despejar $\vec{x}$.

Hay que tener presente que hay operaciones que cambian la matriz $\mathbf{A}$ que no alteran el conjunto de valores $\{x_j\}$ que constituyen la solución, aunque el orden de ellos puede cambiar. Las operacfiones posibles son

- intercambiar dos filas de $\mathbf{A}$
- multiplicar una fila de $\mathbf{A}$ por algún número λ no nulo
- sumarle a una fila, otyra fila multiplicada por algún número λ

Para resolver (3.1.1) se recurre a las operaciones recién descritas.

Para comenzar se divide la primera ecuación por a_{11} y luego cada una de las k-ecuaciones que sigue ($k = 2..n$) se reemplaza a_{kj} por $a_{kj} - a_{k1}a_{1j}/a_{11}$. El resultado es

$$\begin{pmatrix} 1 & a_{12}/a_{11} & a_{13}/a_{11} & ... & a_{1n}/a_{11} \\ 0 & a_{22}-a_{21}a_{12}/a_{11} & a_{23}-a_{21}a_{13}/a_{11} & & a_{2n}-a_{21}a_{1n}/a_{11} \\ .. & .. & & & \\ 0 & a_{n2}-a_{n1}a_{12}/a_{11} & a_{n3}-a_{n1}a_{13}/a_{11} & & a_{nn}-a_{n1}a_{1n}/a_{11} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ .. \\ .. \\ x_n \end{pmatrix} = \begin{pmatrix} b_1/a_{11} \\ b_2 - b_1 a_{21}/a_{11} \\ .. \\ .. \\ b_n - b_1 a_{n1}/a_{11} \end{pmatrix} \tag{3.1.3}$$

A continuación se procede de la misma manera con la submatriz de $(n-1)\times(n-1)$ y así sucesivamente llegándose finalmente a un sistema de la forma

$$\begin{pmatrix} 1 & a'_{12} & a'_{13} & ... & a'_{1n} \\ 0 & 1 & a'_{23} & & a'_{2n} \\ 0 & .. & & & .. \\ .. & .. & & & .. \\ .. & .. & .. & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ .. \\ .. \\ x_n \end{pmatrix} = \begin{pmatrix} b'_1 \\ b'_2 \\ .. \\ .. \\ b'_n \end{pmatrix}$$

Que también puede ser visto como el sistema

$$\begin{array}{lllll} x_1+ & a'_{12}x_2+ & a'_{13}x_3+..+ & a'_{1,n-1}x_{n-1}+ & a'_{1n}x_n & = b'_1 \\ & x_2+ & a'_{23}x_3+..+ & a'_{2,n-1}x_{n-1}+ & a'_{2n}x_n & = b'_2 \\ & & x_3+..+ & a'_{3,n-1}x_{n-1}+ & a'_{3n}x_n & = b'_3 \\ & & & .. & .. & .. \\ & & & .. & .. & .. \\ & & & x_{n-1}+ & a'_{n-1,n}x_n & = b'_{n-1} \\ & & & & x_n & = b'_n \end{array}$$

Formalmente lo que se hizo fue encontrar una matriz no-singular S de modo que $SA = U$ y U es una matriz triangular superior. Esto es, $A = S^{-1}U$ y el problema se ha reducido a $U\vec{x} = S\vec{b}$ que es fácil de resolver. A continuación una forma sencilla e ingenua de hacerlo.

El programa es muy simple, puede ser muy inexacto e incluso inestable. Si se considera los casos

$$\begin{pmatrix} 1 & 3 & 2 \\ 5 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 24 \\ 27 \\ 6 \end{pmatrix} \qquad \begin{pmatrix} 0 & 3 & 2 \\ 5 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 20 \\ 27 \\ 6 \end{pmatrix} \tag{3.1.4}$$

el programa anterior da en forma precisa el resultado $(4,\ 2,\ 7)$ en el primer ejemplo, pero no funciona en el segundo caso. Más adelante veremos un método que resuelve ambos.

Si la matriz $\mathbf{A}$ es no singular la resolución de este sistema da trivialmente el resultado analítico del problema. Si se hace numéricamente hay que hacer algunas consideraciones.

El método presentado en (3.1.3) no es aplicable en forma directa si a_{11} es nulo y si a_{11} en muy chico los errores pueden ser incontrolables. Lo mismo puede decirse si $\tilde{a}_{22} \equiv a_{22} - a_{21}a_{12}/a_{11}$ es muy pequeño o, en general, el primer coeficiente de la primera ecuación, de lo que va quedando, es muy pequeño. También hay que resolver el caso en que ese "primer coeficiente de la primera ecuación de lo que va quedando" es nulo.

Una importante refinación de esto es el método de Gauss con pivoteo que, sin embargo, no se verá aquí. Lo esencial es que se debe permutar filas o columnas de modo de logar que los u_{ii} por lo que se va dividiendo sean lo más grandes posible.

Esto se logra con una matriz de permutación P. Una matriz de permutación de $n \times n$ tiene ceros excepto que tiene un y solo un elemento 1 en cada fila y en cada columna, por ejemplo

$$P = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \tag{3.1.5}$$

Existen $n!$ matrices de permutación de $n \times n$.

Se trabaja con matrices cuadradas, reales simétricas o hermíticas. La notación es

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix}$$

Ella puede multiplicar un vector columna en la forma $A\vec{x}$.

```
// rutina basica que usa el
// metodo de Gauss
void Despejando()
{ for(k=0; k<n-1; k++)
  {  for(i=k+1; i<n; i++)
     { p = a[i][k]/a[k][k];
       for(j=k; j<n+1; j++)
       a[i][j] = a[i][j] - p*a[k][j];
     }
  }
  x[n-1] = a[n-1][n]/a[n-1][n-1];
  for(i=n-2; i>=0; i--)
  { s = 0;
    for(j=i+1; j<n; j++)
    {  s += (a[i][j] * x[j]);
       x[i] = (a[i][n] - s)/a[i][i];
    }
  }
}
```

Figura 3.1: *Versión ingenua del método de Gauss*

El asunto es tener un método para resolver

$$A\vec{x} = \vec{b} \tag{3.1.6}$$

el cual se puede plantear como el sistema de ecuaciones lineales

$$\begin{array}{ccccccc} a_{11}x_1 + & a_{12}x_2 & & a_{1N}x_N & = & b_1 \\ a_{21}x_1 + & a_{22}x_2 & & a_{2N}x_N & = & b_2 \\ .. & .. & .. & .. & = & .. \\ .. & .. & .. & .. & = & .. \\ a_{N1}x_1 + & a_{N2}x_2 & & a_{NN}x_N & = & b_N \end{array} \tag{3.1.7}$$

reemplaza en las $N-1$ ecuaciones restantes. De la primera de estas últimas se depeja x_2 etc. Así se obtiene un sistema triangular de ecuaciones: la primera tiene todas las variables, la segunda tiene desde x_2 en adelante y la última tiene tan solo a x_N. Se llamará b_{ij} a los coeficientes de este sistema triangular. Una vez que se tiene tal sistema se despeja trivialmente x_N de la última ecuación, con la cual ahora se puede despejar x_{N-1} de la penúltima etc.

En general el procedimiento recién descrito no puede ser usado en forma directa porque puede ocurrir que él implique dividir por un número muy pequeño o incluso por cero. Es necesario

usar un procedimiento que no tan solo no acarree tales riesgos sino que además minimice el error.

El procedimiento consiste en intercambiar filas y columnas para minimizar los errores. El siguiente código ilustra esta idea.

3.1.2. Descomposición LU y PLU y uso de la librería gsl_linalg.h

Una forma de llevar a cabo la eliminacion de Gauss consiste en factorizar $\mathbf{A}$ en la forma

$$A = LU$$

donde L es una matriz triangular inferior y U es una matriz triangular superior. Esta descomosición no es única y normalmente se agrega la condición que $L_{kk} = 1$ o bien, $U_{kk} = 1$. La ventaja es que resolver un problema triangular es muy sencillo.

No se verá los algoritmos explícitos para hacer esta descomposición

Para resolver $A\vec{x} = \vec{b}$, esto es, $LU\vec{x} = \vec{b}$, se resuelve primero $L\vec{y} = \vec{b}$ y una vez determinado $\vec{y}$ se resuelve $U\vec{x} = \vec{y}$.

Una matriz cuadrada $\mathbf{A}$ tiene una descomposición LU en dos matrices triandulares, L por *lower* y U por *upper* en la forma

$$PA = LU$$

donde P es una matriz de permutación, L una matriz *unitaria* inferior y U es una matriz triangular superior. La utilidad está en que el sistema $A\vec{x} = \vec{b}$ se convierte en dos problemas triangulares $L\vec{y} = P\vec{b}$ y $U\vec{x} = \vec{y}$ que puede ser resuelto por sustitución inversa.

Ejemplo muy sencillo. Para el problema

$$A\vec{x} = \vec{b} \qquad \begin{pmatrix} 2 & 4 & 1 \\ -10 & -8 & 11 \\ 8 & 22 & 33 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 \\ -20 \\ 2 \end{pmatrix} \tag{3.1.8}$$

se usa la descomposición $A = LU$

$$A = \begin{pmatrix} 1 & 0 & 0 \\ -5 & 1 & 0 \\ 4 & \frac{1}{2} & 1 \end{pmatrix} \begin{pmatrix} 2 & 4 & 1 \\ 0 & 12 & 16 \\ 0 & 0 & 21 \end{pmatrix} \quad \text{esto es} \quad L = \begin{pmatrix} 1 & 0 & 0 \\ -5 & 1 & 0 \\ 4 & \frac{1}{2} & 1 \end{pmatrix} \qquad U = \begin{pmatrix} 2 & 4 & 1 \\ 0 & 12 & 16 \\ 0 & 0 & 21 \end{pmatrix}$$

El problema (3.1.8) se reduce a dos problemas más sencillos, $L\vec{y} = \vec{b}$ (L es triangular) y una vez conocido $\vec{y}$ se resuelve $U\vec{x} = \vec{y}$ donde ahora U es triangular

$$L\vec{y} = \begin{pmatrix} 1 & 0 & 0 \\ -5 & 1 & 0 \\ 4 & \frac{1}{2} & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 1 \\ -20 \\ 2 \end{pmatrix} \qquad \Rightarrow \qquad \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 1 \\ -15 \\ \frac{11}{2} \end{pmatrix}$$

y, conocido $\vec{y}$, se resuelve el segundo problema triangular

$$\begin{pmatrix} 2 & 4 & 1 \\ 0 & 12 & 16 \\ 0 & 0 & 21 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 \\ -15 \\ \frac{11}{2} \end{pmatrix}$$

GaussElim.c

```
#include <stdio.h> // resuelve problema lineal:  Ax=b
#include <math.h>  // A = matrix NxN
#define  N        5
double x[N], b[N]={20,27,6,1,3}; // datos a gusto
double a[N][N]={{0,3,2,1,2},{5,0,1,4,1},{1,1,0,0,4},{4,-2,1,3,0},{8,-9,1,0,2}};
int indc[N];

void reordena()
{ int i, j, k, itmp;
  double c1, pe, pe1, pj;
  double c[N];
  for(i = 0; i < N; ++i) // inicializacion del indice
    indc[i] = i;
  for(i = 0; i < N; ++i) // factores de escala para cada fila
  { c1 = 0;
    for(j = 0; j < N; ++j) // fabs(X) = valor absoluto de X
    { if(fabs(a[i][j]) > c1) c1 = fabs(a[i][j]);}
    c[i] = c1;
  }
  for(j = 0; j < N-1; ++j)//se busca elemento mas grande de cada columna
  { pe1 = 0;
    for(i = j; i < N; ++i)
    { pe = fabs(a[indc[i]][j])/c[indc[i]];
      if(pe > pe1)
      { pe1 = pe;
        k = i;
      }}
    itmp = indc[j]; //Se intercambia filas via indc[]
    indc[j] = indc[k];
    indc[k] = itmp;
    for(i = j+1; i < N; ++i)
    { pj = a[indc[i]][j]/a[indc[j]][j];
      a[indc[i]][j] = pj; //guarda cuocientes de reord bajo la diagonal
      for(k = j+1; k < N; ++k) //Por consistencia se modifica otros elementos
      { a[indc[i]][k] = a[indc[i]][k]-pj*a[indc[j]][k];}
    }} // for j
}

void principal()
{ int i,j;
  reordena();
  for(i = 0; i < N-1; ++i)
  { for(j = i+1; j < N; ++j)
    { b[indc[j]] = b[indc[j]]-a[indc[j]][i]*b[indc[i]];}
  }
  x[N-1] = b[indc[N-1]]/a[indc[N-1]][N-1];
  for(i = N-2; i>=0; i--)
  { x[i]   = b[indc[i]];
    for(j  = i+1; j < N; ++j) { x[i] = x[i]-a[indc[i]][j]*x[j];}
    x[i] = x[i]/a[indc[i]][i];
  }
}

main() // ******************************************************
{ int i;
  principal();
  for(i=0; i<N; i++) printf("%16.8f\n", x[i]);
}
```

Figura 3.2: Código en *C* para aplicar el método de eliminación de Gauss.

que da

$$x_1 = \frac{899}{252} = 3{,}56746 \qquad x_2 = -\frac{403}{252} = -1{,}59921 \qquad x_3 = \frac{11}{42} = 0{,}261905$$

El siguiente código resuelve el problema anterior:

```
#include <stdio.h>
#include <math.h>
#include <gsl/gsl_linalg.h>
main()
{   int s;
    double A[] ={  2.0,  4.0,  1.0,  // matriz a invertir:
                 -10.0, -8.0, 11.0,
                  8.0, 22.0, 33.0 };
    double b[] = { 1.0, -20.0, 2.0};
    gsl_matrix_view m = gsl_matrix_view_array(A, 3, 3);
    gsl_vector_view b = gsl_vector_view_array(b, 3);
    gsl_vector *x = gsl_vector_alloc (3);
    gsl_permutation * p = gsl_permutation_alloc(3);
    gsl_linalg_LU_decomp(&m.matrix, p, &s);
    gsl_linalg_LU_solve(&m.matrix, p, &b.vector, x);
       printf ("x = \n");
    gsl_vector_fprintf(stdout, x, "%g");
    double determinant = gsl_linalg_LU_det(&m.matrix, s);
       printf("Determinante: %lf\n",determinant);
    gsl_permutation_free (p);
}
```

Modificando los datos en este programa se resuelve el segundo ejemplo (3.1.4).

Ejemplo con permutación no trivial:

$$A = \begin{pmatrix} 0 & 1 & 1 & -3 \\ -2 & 3 & 1 & 4 \\ 0 & 0 & 0 & 1 \\ 3 & 1 & 0 & 0 \end{pmatrix} = PLU = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -\frac{3}{2} & \frac{11}{2} & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} -2 & 3 & 1 & 4 \\ 0 & 1 & 1 & -3 \\ 0 & 0 & -4 & \frac{45}{2} \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{3.1.9}$$

3.1.3. Método del gradiente conjugado

Se desea resolver el sistema lineal de ecuaciones

$$\mathbf{A}\vec{x} = \vec{b} \tag{3.1.10}$$

donde **A** es una matriz real, simétrica ($\mathbf{A}^T = \mathbf{A}$), positiva definida, esto es, satisface: $\vec{x} \cdot \mathbf{A}\vec{x} > 0$, para todo $\vec{x}$ real no nulo.

La única solución del problema se denota $\vec{x}_*$.

Por definición dos vectores $\vec{u}$ y $\vec{v}$ son conjugados si

$$\vec{u}\,\mathbf{A}\vec{v} = 0 \tag{3.1.11}$$

Si se define el producto escalar

$$(\vec{r},\ \vec{s}) \equiv \vec{r}\,\mathbf{A}\vec{s} \tag{3.1.12}$$

La relación (3.1.11) expresa que $\vec{u}$ y $\vec{v}$ son ortogonales.

Sea $\{\vec{e}_k\}_{k=1}^N$ un conjunto de N vectores mutuamente ortogonales, es decir, mutuamente conjugados. Ellos constituyen una base en R^n. Con esta base se plantea la expansión

$$\vec{x}_* = \sum_k \alpha_k\, \vec{e}_k \tag{3.1.13}$$

La ecuación (3.1.10) es

$$\vec{b} = \mathbf{A}\vec{x}_* = \sum_k \alpha_k\, \mathbf{A}\vec{e}_k \tag{3.1.14}$$

por lo cual se tiene que

$$\vec{e}_j \cdot \vec{b} = \vec{e}_j \cdot \mathbf{A}\vec{x}_* = \sum_k \alpha_k \vec{e}_j \cdot \mathbf{A}\vec{e}_k = \alpha_j\, \vec{e}_j \cdot \mathbf{A}\vec{e}_j = \alpha_j\, (\vec{e}_j \cdot \vec{e}_j) \tag{3.1.15}$$

Por lo tanto

$$\alpha_j = \frac{\vec{e}_j \cdot \vec{b}}{\vec{e}_j \cdot \mathbf{A}\vec{e}_j} \tag{3.1.16}$$

El método consiste en buscar el vector $\vec{x}$ tal que

$$f(\vec{x}) = \frac{1}{2}\vec{x} \cdot \mathbf{A}\vec{x} - \vec{x} \cdot \vec{b} \tag{3.1.17}$$

sea mínimo, con $\vec{x} \in R^n$.

Para ello se va a generar una secuencia de vectores $\vec{x}_a$ con $a = 0, 1, ...$ para los que $f(\vec{x}_a)$ es cada vez menor. Es obvio que f se anula si se alcanza la solución $\vec{x}_*$ Notemos que

$$\nabla_{\vec{x}} f = \mathbf{A}\vec{x} - \vec{b} \tag{3.1.18}$$

Escogiendo nulo el primer término de la secuencia, $\vec{x}_0 = 0$, el gradiente en ese punto es $\nabla_{\vec{x}} f(\vec{0}) = -\vec{b}$. El primer vector base se toma igual a menos ese gradiente: $\vec{e}_1 = \vec{b}$. El resto de los vectores base deben ser conjugados al gradiente, de ahí el nombre del método.

Definiendo

$$r_a = \vec{b} - \mathbf{A}\vec{x}_a = -\nabla f(\vec{x}_a) \tag{3.1.19}$$

se puede deducir que

$$\vec{e}_{k+1} = \vec{r}_k - \sum_{c \le k} \frac{\vec{e}_c \cdot \mathbf{A}\vec{r}_k}{\vec{e}_c \cdot \mathbf{A}\vec{e}_c}\, \vec{e}_c \tag{3.1.20}$$

El siguiente $\vec{x}$ óptimo es

$$\vec{x}_{k+1} = \vec{x}_k + \alpha_{k+1}\, \vec{e}_{k+1} \tag{3.1.21}$$

3.2. Métodos de interpolación

Se tiene un conjunto de N pares, o datos, $\{x_k, y_k\}_{k=1}^N$ y se desea encontrar una función que describa estos como una función continua $f(x)$. Hay diversos métodos que a continuación se separan en dos grupos. En el primer grupo siempre se cumple que $f(x_k) = y_k$, mientras que en el segundo $f(x_k) \approx y_k$.

3.2.1. Interpolaciones leales

Estos son métodos que definen una función $f(x)$ que satisface

$$f(x_k) = y_k \qquad \text{para todo} \quad k$$

3.2.1.1. Interpolación lineal

En este método se define una recta entre puntos consecutivos

$$f(x_k \le x \le x_{k+1}) = y_k + \frac{y_{k+1}-y_k}{x_{k+1}-x_k}(x-x_k) \tag{3.2.1}$$

Se puede generalizar este método usando tres puntos (x_{k-1}, x_k, x_{k+1}) y definir una curva cuadrática que pase por estos tres puntos. También se puede usar más puntos, pero el método se deteriora porque resultan expresiones que contienen oscilaciones.

3.2.1.2. Interpolación de Lagrange

Este método consiste en construir un polinomio $P(x)$ asociado a un conjunto de $N+1$ pares de valores $\{(x_0,y_0),...(x_N,y_N)\}$. Se define

$$P \approx \prod_{j=1}^{N} y_j \prod_{k=0,k\neq j}^{N} \frac{x-x_k}{x_j-x_k} \tag{3.2.2}$$

esto es, se define los $N+1$ polinomios de orden N

$${}^{(N)}p_j(x) \equiv \prod_{k=0,,k\neq j}^{N} \frac{x-x_k}{x_j-x_k}, \qquad j = 0,\ldots N \tag{3.2.3}$$

con lo cual la interpolación es

$$P(x) = \sum_{j=0}^{N} {}^{(N)}p_j(x)\ y_j \tag{3.2.4}$$

que pasa sobre los N puntos de partida (x_k,y_k), $k = 0,..N$. Este método es particularmente útil cuando se trata de pocos puntos.

En el caso $N = 3$ se debe definir los cuatro polinomios cúbicos

$${}^{(4)}p_0 = \frac{(x-x_1)(x-x_2)(x-x_3)}{(x_0-x_1)(x_0-x_2)(x_0-x_3)}, \qquad {}^{(4)}p_1 = \frac{(x-x_0)(x-x_2)(x-x_3)}{(x_1-x_0)(x_1-x_2)(x_1-x_3)},$$

$${}^{(4)}p_2 = \frac{(x-x_0)(x-x_1)(x-x_3)}{(x_2-x_0)(x_2-x_1)(x_2-x_3)}, \qquad {}^{(4)}p_3 = \frac{(x-x_0)(x-x_1)(x-x_2)}{(x_3-x_0)(x_3-x_1)(x_3-x_2)}.$$

Si, por ejemplo, se asocia a los a los cuatro puntos x_0, x_1, x_2 y x_3 los valores $x_k = k$ e $y_k = \sin x_k$ ($k = 0,1,2,3$), el método anterior para P, usando (3.2.4), da la expresión polinomial para $\sin x$

$$\sin x \approx P = 1{,}207506817\,x - 0{,}355642612\,x^2 - 0{,}0103932197\,x^3$$

Este polinomio no se anula en $x = \pi$ sino en $x = 3{,}112$, pero da el valor correcto de $\sin(x)$ en cada uno de los cuatro puntos x_k usados como datos de entrada.

3.2.2. Métodos de ajuste suavizado

Son métodos que buscan una función suave que satisfaga

$$f(x_k) \approx y_k$$

3.2.2.1. Mínimos cuadrados

Usando polinomios simples. Versión 1. El objetivo es, dada una lista de valores y_k asociados a puntos dados x_k con $k = 1,\ldots,n$, ajustar una función $f(x,\{\beta\})$ tal que

$$S = \sum_{k=1}^{n} r_k^2 \equiv \sum_{k=1}^{n} \left(y_k - f(x_k,\{\beta\})\right)^2 \tag{3.2.5}$$

tenga un valor mínimo, donde $\{\beta\}$ representa un conjunto de m parámetros que se debe ajustar.

A las diferencias $r_k \equiv y_k - f(x_k,\{\beta\})$ se los llama residuos.

Las condiciones que impone exigir que S tenga el menor valor posible son las m ecuaciones

$$\frac{\partial S}{\partial \beta_s} = 2\sum_k r_k \frac{\partial r_k}{\partial \beta_s} = 0\,, \qquad s = 1,\ldots,m \tag{3.2.6}$$

La forma estándar de abordar este problema consiste en dar una dependencia lineal en los β a las f, del tipo

$$f(x_k,\beta) = \sum_s \beta_s \phi_s(x_k) \tag{3.2.7}$$

Si se define

$$X_{is} = \frac{\partial f(x_i,\beta)}{\partial \beta_s} = \phi_s(x_i) \tag{3.2.8}$$

Acá no se demostrará que los β están dados por

$$\beta = \left(X^T X\right)^{-1} X^T y \tag{3.2.9}$$

Ejemplo. Se toma los mismos mismos cuatro puntos x_k y los mismos $y_k = \sin x_k$ y se define

$$f(x,\{\beta\}) = \beta_1 + \beta_2 x + \beta_3 x^2$$

Nótese que se ha escogido $\phi_1 = 1$, $\phi_2 = x$ y $\phi_3 = x^2$. Evaluado para los cuatro valores de x_k, de modo que cada $r_k = y_k - f(x_k,\{\beta\})$ es lineal en los β y por esto S es cuadrático en los β. Las ecuaciones $\partial S/\partial \beta_s = 0$ son lineales en los β y triviales de resolver. En el ejemplo presente resultan

$$\beta_1 = -0{,}003117966425\,, \qquad \beta_2 = 1{,}256354951\,, \qquad \beta_3 = -0{,}4024121014$$

con lo cual la función seno queda aproximada por

$$\sin x \approx f = -0{,}003117966425 + 1{,}256354951\,x + -0{,}4024121014\,x^2$$

Usando polinomios simples. Versión 2. La idea es ajustar un polinomio $p_m(x)$ de orden m, esto es, se trata de encontrar un $f(x)$ que tenga la forma de un polinomio,

$$p_m(x) = \sum_{k=0}^{m} a_k\, x^k \tag{3.2.10}$$

Para obtener los coeficientes a_k se busca las condiciones para que la desviación cuadrática media sea mínima. Si tuviésemos una fución continua F en lugar de los y_k, se debería requerir que $\chi^2 = \int \left(p_m(x) - F(x)\right)^2 dx$ sea mínimo, pero lo único que se puede hacer es requerir que

$$\chi^2(a) = \sum_{i=0}^{N} \left(p_m(x_i) - y_i\right)^2 \tag{3.2.11}$$

sea mínimo, donde lo que se debe variar son los coeficientes a_i que definen el polinomio. Esto consiste en plantear las $m+1$ ecuaciones lineales

$$\frac{\partial \chi^2}{\partial a_k} = 0 \tag{3.2.12}$$

Métodos para resolver ecuaciones sistemas de lineales ya fueron vistos.

A modo de ejemplo manejable trabajemos el caso $m = 1$, esto es

$$p_1(x) = a_0 + a_1 x$$

de tal modo que

$$\chi^2 = \sum_{i=0}^{N} \left(a_0 + a_1 x - y_i\right)^2$$

y las ecuaciones 3.2.12 son

$$\begin{aligned} (N+1)\, a_0 + c_1\, a_1 - c_3 &= 0 \\ c_1\, a_0 + c_2\, a_1 - c_4 &= 0 \end{aligned}$$

donde

$$c_1 = \sum_{i=0}^{N} x_i\,, \qquad c_2 = \sum_{i=0}^{N} x_i^2\,, \qquad c_3 = \sum_{i=0}^{N} y_i\,, \qquad c_4 = \sum_{i=0}^{N} x_i y_i$$

De todo lo anterior se obtiene que

$$a_0 = \frac{c_1 c_4 - c_2 c_3}{c_1^2 - (N+1)c_2}\,, \qquad a_1 = \frac{c_1 c_3 - (N+1)\, c_4}{c_1^2 - (N+1)c_2}$$

Se puede adivinar que si m es algo mayor, la solución analítica puede llegar a ser muy complicada.

Usando polinomios ortogonales: Esta vez el polinomio p_m se escribe en la forma

$$p_m(x) = \sum_{i=0}^{m} \alpha_i P_i(x) \tag{3.2.13}$$

donde los $P_i(x)$ son polinomios reales que satisfacen relaciones de ortogonalidad de la forma

$$\langle P_k\,|P_\ell\rangle \equiv \int_a^b P_k(x)\,w(x)\,P_\ell(x)\,dx = \mathscr{N}_k\delta_{k\ell}$$

y donde $w(x)$ es el peso que define la ortogonalidad de los polinomios específicos P_i que se esté usando. Más en general, en este contexto se define

$$\langle F\,|G\rangle \equiv \int_a^b F(x)\,w(x)\,G(x)\,dx$$

Sin embargo, para lo que acá interesa no se tiene funciones definidas continuamente, de modo que es necesario cambiar el producto escalar a:

$$\langle F\,|G\rangle \equiv \sum_{i=0}^{N} F(x_i)\,w(x_i)\,G(x_i) \tag{3.2.14}$$

y, para poder usar estas nociones, se necesita polinomios que satisfagan ortogonalidad con esta definición. Se quiere entonces polinomios $U_k(x)$ tales que

$$\langle U_k\,|U_\ell\rangle = \sum_{i=0}^{N} U_k(x_i)\,w(x_i)\,U_\ell(x_i) = \delta_{k\ell}\,\mathscr{N}_k \tag{3.2.15}$$

Los polinomios U_k que se necesita se generan con la siguiente relación de recurencia,

$$U_{k+1}(x) = (x-g_k)\,U_k(x) - h_k\,U_{k-1}(x) \tag{3.2.16}$$

donde los g_k y h_k se obtienen de

$$g_k = \frac{\langle xU_k\,|U_{k-1}\rangle}{\langle U_k\,|U_k\rangle}, \qquad h_k = \frac{\langle xU_k\,|U_{k-1}\rangle}{\langle U_{k-1}\,|U_{k-1}\rangle} \tag{3.2.17}$$

donde $U_0(x) = 1$ y $h_0 = 0$. Se puede demostrar que ellos satisfacen todas las propiedades necesarias sin importar qué se tome como $w(x)$.

En lo que sigue consideraremos el caso $w(x) = 1$ pero es fácil generalizarlo a otros casos.

La aproximación de mínimos cuadrádos se obtiene una vez que se encuentra todos los α_k tales qye χ^2 es mínimo. Se desea que

$$\frac{\partial\chi^2}{\partial\alpha_j} = 0, \qquad \frac{\partial^2\chi^2}{\partial\alpha_j^2} = 0 \qquad \text{para} \quad j = 0,1,...,m \tag{3.2.18}$$

Se puede demostrar que

$$\alpha_j = \frac{\langle U_j\,|f\rangle}{\langle U_j\,|U_j\rangle} \tag{3.2.19}$$

En nuestro caso f representa los valores y_i.

Como miniejemplo considérese los tres polinomios:

$$P_0 = 1, \qquad P_1 = x-2, \qquad P_2 = x^2 - 4x + \frac{10}{3}$$

los puntos $x_0 = 1\,, \quad x_1 = 2\,, \quad x_2 = 3$, y el siguiente producto escalar

$$\langle i|j\rangle = \sum_{k=0}^{2} P_i(x_k)P_j(x_k)$$

Se cumple

$$\langle 0|0\rangle = 3\,, \qquad \langle 0|1\rangle = 0\,, \qquad \langle 0|2\rangle = 0$$
$$\langle 1|1\rangle = 2\,, \qquad \langle 1|2\rangle = 0\,, \qquad \langle 2|2\rangle = \frac{2}{3}$$

3.2.2.2. La aproximación "spline" cúbica

Dados los $N+1$ datos (x_i, y_i), $i = 0, 1 \ldots, N$, se define N funciones polinomiales de orden 3

$$f_k(x) = y_k + c_{k1}(x - x_k) + c_{k2}(x - x_k)^2 + c_{k3}(x - x_k)^3\,, \qquad x_k \leq x \leq x_{k+1}\,, \quad k = 0, \ldots N-1$$

Se debe encontrar las $3N$ incógnitas c_{kj} para tener definido el método. Para ellos se plantea el siguiente sistema de ecuaciones

$$\begin{aligned} f_k(x_{k+1}) &= y_{k+1} && N \text{ ecuaciones} \\ \left.\frac{df_{k-1}}{dx}\right|_{x_k} &= \left.\frac{df_k}{dx}\right|_{x_k} && N-1 \text{ ecuaciones} \\ \left.\frac{d^2 f_{k-1}}{dx^2}\right|_{x_k} &= \left.\frac{d^2 f_k}{dx^2}\right|_{x_k} && N-1 \text{ ecuaciones} \end{aligned} \tag{3.2.20}$$

lo que da un total de $3N-2$ ecuaciones. Si se agrega las condiciones $\left.\frac{d^2 f_0}{dx^2}\right|_{x_0} = 0$ y $\left.\frac{d^2 f_{N-1}}{dx^2}\right|_{x_N} = 0$ para tener las $3N$ ecuaciones necesarias.

Los coeficientes se resuelven como un sistema linea de $3N$ ecuaciones y debe usarse algunos de los métodos ya descritos.

3.2.2.3. Ajuste no paramétrico

Si se tiene datos tipo experimentales $\{x_i, y_i\}_{i=1}^{N}$ y se desea un ajuste sin parámetros se plantea definir una función que represente algún tipo de promedio de los puntos vecinos. Para ello se define una función $K(x)$ que debe cumplir

1. K es finita
2. $\int K$ es finita
3. preferentemente que tenga suporte finito
4. rápida de evaluar
5. continua y de derivada continua

para definir

$$f(x) = \frac{\sum_i y_i K(x-x_i)}{\sum_i K(x-x_i)} \tag{3.2.21}$$

Nótese que la distribución de Gauss no es una buena elección porque tiene soporte infinito. Tampoco resulta una función escalón porque no es diferenciable.

Una posible elección es

$$K(x) = \begin{cases} (a^2-x^2)^2 & \text{si } |x| \le a \\ 0 \end{cases}$$

Con este método se puede evaluar en cualquier punto x dentro del rango cubierto por los datos, pero no es sensato intentar extrapolar.

3.3. Aproximante de Padé

Algunas funciones no pueden ser expresadas en forma polinomial con precisión apropiada, como es el caso de una expansión en serie truncada,

$$f(x) \approx \sum_{n=0}^{S} c_n x^n, \qquad \text{donde} \qquad c_n = \frac{f^{(n)}(0)}{n!} \tag{3.3.1}$$

Lo que puede resultar muy efectivo es aproximar la función como una función racional, esto es, el cuociente de dos polinomios,

$$f(x) \approx R_{M,N}(x) = \frac{\sum_{i=0}^{M} n_i x^i}{1+\sum_{j=1}^{N} d_j x^j} \tag{3.3.2}$$

donde los $M+N+1 = S$ coeficientes n_i y d_j deben ser determinados y para ello se exige que $R_{M,N}(x)$ satisfaga

$$R(0) = f(0); \quad R^{(k)}(0) = f^{(k)}(0) \quad \text{para} \quad k = 1,2..,(M+N) \tag{3.3.3}$$

donde $Y^{(k)}$ se refiere a la derivada de orden k de Y. En (3.3.3) se tiene $M+N+1$ ecuaciones para igual número de incógnitas. En otras palabras, se exige que la expanción en serie de $R_{M,N}(x)$ hasta potencia $M+N$ coincida con la expansión de $f(x)$ la misma potencia.

Por ejemplo

$$f = \frac{1-x^2}{1+x^2}\sin x$$

tiene un desarrollo en serie

$$f \approx x - \frac{13}{6}x^3 + \frac{281}{120}x^5 - \frac{2369}{1008}x^7 + \frac{852913}{362880}x^9 - \frac{93820541}{39916800}x^{11} + \ldots$$

que diverge rápidamente, mientras que la aproximación de Padé

$$f \approx \frac{x - 1{,}009135886x^3}{1+1{,}157530781x^2+0{,}1663166920x^4}$$

representa a la función en un rango más amplio.

En general para obtener los coeficiente de Padé se debe resolver un sistema lineal de ecuaciones, métodos que—como se ha visto—-si no son bien tratados, podrían dar poca precisión. Se sabe que por lo menos se debe usar el método de factorización LU posiblemente seguido por algún método iterativo que mejora la precisión.

3.4. Recurrencias, puntos fijos y ceros

Muchos métodos numéricos hacen uso de métodos iterativos. Es interesante entonces saber decidir cuándo estos métodos son estables y bajo qué condiciones convergen. Una sucesiones puede ser convergente, divergente o tener algún comportamiento más complicado. Un caso clásico es el *mapa logístico*

$$x_{n+1} = A x_n (1 - x_n), \qquad \text{con} \qquad 0 < A \leq 4 \tag{3.4.1}$$

definida para $0 \leq x < 1$. Compruebe, por ejemplo, que para $A = 3{,}1$ los x_n terminan saltando entre dos valores fijos.

3.4.1. Estabilidad

La recurrencia

$$x' = g(x) \tag{3.4.2}$$

tiene punto fijo en el valor $x^\star$ si $x^\star = g(x^\star)$. Interesa saber si ese punto fijo es estable.

Al iterar a partir de un punto muy cercano a $x^\star$ se obtiene

$$\begin{aligned} x^\star + \varepsilon' &= g(x^\star + \varepsilon) \\ &\approx g(x^\star) + \varepsilon g'(x^\star) \\ \varepsilon' &= \varepsilon g'(x^\star) \end{aligned} \tag{3.4.3}$$

de donde se concluye que si $|g'(x^\star)| < 1$ la iteración converge hacia el punto fijo, lo que se conoce como *estabilidad lineal* del punto fijo $x^\star$ de $g(x)$. También se dice que $x^\star$ es un *atractor*, ya que los puntos cercanos son atraidos, via la iteración (3.4.2), hacia $x^\star$.

Por ejemplo, (3.4.1) tiene un punto fijo trivial $x^\star = 0$ y otro, $x^\star = 1 - \frac{1}{A}$. El punto fijo trivial es estable cuando $A < 1$. El segundo punto es estable tan solo si $1 < A < 3$. Se deja como ejercicio hacer un programa que muestre el comportamiento de (3.4.1) en todo el rango permitido de $\mathbf{A}$.

3.4.2. Ceros

3.4.3. Encajonamiento

3.4.3.1. Búsqueda de puntos con distinto signo para f.

La idea es seleccionar dos puntos x_1 y x_2, ($x_1 < x_2$) y si el signo de la función es el mismo en ambos puntos variar la posición de estos puntos hasta que $f_1\, f_2 < 0$.

Si se ha escogido dos semillas cercanas, entonces se procede a alejar las dos semillas, si son semillas muy distantes se busca ir acercándolas. Sólo consideraremos el primer caso.

Se tiene que $f_1\, f_2 > 0$. Se escoge $\alpha > 1$ y se procede como sigue:

```
f1 = f(x1);  f2 = f(x2);
while(f1*f2>0)
{   if (fabs(f1) < fabs(f2))
    { x1 += alpha*(x1-x2);
      f1  = f(x1);}
    else
    { x2 += alpha*(x2-x1);
      f2  = f(x2);}
 }
```

La rutina anterior termina cuando se tiene puntos en los cuales la función tiene distinto signo. A continuación se procede a acercar los dos puntos bajo la condición que la función en todo momento tenga signo distinto. Esto se denomina *encajonar*. Existen diversas estrategias para encajonar.

Encajonamiento sencillo

```
f1 = f(x1);  f2 = f(x2);
do
{  xm = 0.5*(x1+x2);
   fm = f(xm);
   if(f1*fm < 0.0)
   { x2 = xm;  f2 = fm; }
   else
   { x1 = xm;  f1 = fm; }
}while(fabs(x2-x1)>tolerancia);
```

3.4.3.2. Método de Newton y de la secante

Una forma de buscar—en forma que suele ser precisa y rápida—los ceros de una función $f(x)$, cuya ubicación se conoce en forma aproximada, consiste en iterar usando

$$g(x) = x - \frac{f(x)}{f'(x)} \tag{3.4.4}$$

Si x_0 es un cero de $f(x)$, es fácil comprobar que $g'(x_0) = 0$, que implica que x_0 es punto fijo (localmente) estable de $g(x)$. Este es le *método de Newton* para obtener ceros.

Suele ocurrir que se necesite conocer la ubicación precisa de los ceros de una función demasiado complicada para poder tener una forma analítica para $f'(x)$. Esto impide poder hacer uso

de (3.4.4). Existe un método inspirado en el anterior que es fácil de programar y normalmente de convergencia muy rápida. En lugar de (3.4.4) se usa

$$x_{n+1} = x_n - \frac{f_n}{\frac{f_n - f_{n-1}}{x_n - x_{n-1}}} \qquad \text{que se simplifica a} \qquad x_{n+1} = \frac{x_{n-1}\, f_n - x_n\, f_{n-1}}{f_n - f_{n-1}} \tag{3.4.5}$$

y que se conoce como el *método de la secante* y estrictamente corresponde a $x_{n+1} = g(x_n, x_{n-1})$.

Un buen programa no debiera producir error jamás. En el caso de arriba no se ha precavido el caso en que el denominador $f_n - f_{n-1}$ pueda anularse. Tampoco se ha previsto la posibilidad de que jamás se logre convergencia de la secuencia. Conviene poner un contador que no permita que se sobrepase algun número de iteraciones.

Los autores del siguiente algoritmo, publicado en 2007, afirman que es mucho más estable que el método de la secante[1]:

$$x_{n+1} = x_{n-1} - \frac{x_{n-1} f_{n-1}}{f_{n-1} + x_{n-1} \frac{f_n - f_{n-1}}{x_n - x_{n-1}}} \tag{3.4.6}$$

$$= \frac{f_n - f_{n-1}}{x_{n-1} f_n - 2x_{n-1} f_{n-1} + x_n f_{n-1}} x_{n-1}^2 \tag{3.4.7}$$

♠ *Determine todos los ceros reales de* $P_a = x^5 - 3x^4 - 2x^2 + 11x - 1$ *y de* $P_b = 16x^5 - 168x^4 + 657x^3 - 1161x^2 + 891x - 243$ *usando los algoritmos descritos.*

3.4.4. Puntos fijos con más de una variable

Lo anterior se puede generalizar al caso de N variables planteando la relación de recurrencia

$$\vec{x}\,' = \vec{g}(\vec{x}) \tag{3.4.8}$$

Si $\vec{x}_0$ es punto fijo de g entonces

$$\vec{x}_0 + \vec{\varepsilon}\,' = \vec{g}(\vec{x}_0 + \vec{\varepsilon}\,) = \vec{g}(\vec{x}_0\,) + \vec{\varepsilon} \cdot (\nabla \vec{g})_{x_0} \tag{3.4.9}$$

donde el último término en forma más explícita es

$$\varepsilon_j \left(\frac{\partial g_i}{\partial x_j} \right)_{x_0}$$

De (3.4.9) se obtiene entonces que

$$\vec{\varepsilon}\,' = \mathscr{J}_0 \vec{\varepsilon} \tag{3.4.10}$$

El punto $\vec{x}_0$ es un punto fijo estable si los autovalores α_k del Jacobiano $\mathscr{J}$ de $\vec{g}(\vec{x})$, evaluado en $\vec{x}_0$, tienen parte real menor que la unidad, es decir, si se escriben en la forma

$$\alpha_k = e^{-a_k + i b_k}$$

son tales que todos los a_k son positivos.

[1] C. Hu, *Computing in Science and Engineering* v.9, #5, p.78 (2007)

3.4.5. Método de la secante en varias variables

Se desea encontrar los ceros de

$$\vec{F}(\vec{r}) = \begin{pmatrix} F_1(\vec{r}) \\ .. \\ F_n(\vec{r}) \end{pmatrix} \tag{3.4.11}$$

donde $\vec{r} = (x_1, x_2, \ldots x_n)$. Es decir, se busca puntos $\vec{r}_0$ en los que todas las funciones $F_n(\vec{r})$ se anulan simultáneamente.

El método de Newton generalizado a este problema se plantea a partir de definir

$$\vec{r}\,' = \vec{g}(\vec{r}) \equiv \vec{r} \; - \; \mathbf{J}^{-1}(\vec{r})\,\vec{F}(\vec{r}) \tag{3.4.12}$$

donde **J** es el Jacobiano $\frac{\partial F_i}{\partial x_j}$ de $\vec{F}$. En efecto, el lado izquierdo se escribe en torno a un cero de $\vec{F}$: $\vec{g}(\vec{r}_0 + \vec{\varepsilon}\,) \approx \vec{g}(\vec{r}_0) + \vec{\varepsilon}\,'$. El correspondiente lado derecho es $\vec{r}_0 + \vec{\varepsilon}$ más el producto de las expansiones de $\mathbf{J}^{-1}$ y de $\vec{F}$. Pero la expansión de este último es $\vec{F}(\vec{r}_0\,) = 0$ más $\mathbf{J}_0\,\vec{\varepsilon}$, donde el índice cero indica que el Jacobiano es evaluado en $\vec{r}_0$. Siendo $\mathbf{J}_0\,\vec{\varepsilon}$ una cantidad de primer orden, no es necesario expandir el factor $\mathbf{J}^{-1}(\vec{r}\,)$ sino que basta con tomar sencillamente $\left(\mathbf{J}_0^{-1}\right)$. Con todo lo anterior se ve que a primer orden, el lado derecho de (3.4.12) se anula: el Jacobiano $\mathscr{J}$ de la función $\vec{g}(\vec{r}\,)$ evaluado en los ceros de $\vec{F}$ (no confundirlo con **J** que es el Jaconbiano de $\vec{F}$) tiene autovalores nulos. Esto garantiza que los ceros de $\vec{F}$ se comportan como puntos fijos estables de la recurrencia que define (3.4.12).

En el caso de dos funciones F_1 y F_2 en las variables x e y se tiene

$$\mathbf{J} = \begin{pmatrix} \frac{\partial F_1}{\partial x} & \frac{\partial F_1}{\partial y} \\ \frac{\partial F_2}{\partial x} & \frac{\partial F_2}{\partial y} \end{pmatrix}, \qquad \det(\mathbf{J}) = \Delta \qquad \& \qquad \mathbf{J}^{-1} = \frac{1}{\Delta}\begin{pmatrix} \frac{\partial F_2}{\partial y} & -\frac{\partial F_1}{\partial y} \\ -\frac{\partial F_2}{\partial x} & \frac{\partial F_1}{\partial x} \end{pmatrix}$$

El método de la secante se obtiene a partir de lo anterior usando, en lugar de **J**, la matriz de cantidades tipo

$$\frac{\partial F_i}{\partial x_j} \approx \frac{F_i(x_j^{\nu}) - F_i(x_j^{\nu-1})}{x_j^{\nu} - x_j^{\nu-1}} \tag{3.4.13}$$

donde el índice ν se refiere a la iteración ν-ésima. Para no hacer muy pesada la notación, se ha usado como argumento de F_i solamente la variable x_j que se está cambiando.

3.5. Problemas

2.1 Haga un programa utilizando el método de gradiente conjugado que resuelva el sistema con seis incógnitas x_k: $\mathbf{A}\vec{x} = \vec{b}$ donde

$$\mathbf{A} = \begin{pmatrix} 4 & 1 & 2 & 3 & 2 & -2 \\ 1 & 3 & 2 & 1 & 0 & 1 \\ 2 & 2 & 1 & 3 & 1 & 3 \\ 3 & 1 & 3 & 5 & 2 & -1 \\ 2 & 0 & 1 & 2 & 1 & 2 \\ -2 & 1 & 3 & -1 & 2 & 5 \end{pmatrix} \quad \text{y} \quad \vec{b} = \begin{pmatrix} 26 \\ 21 \\ 38 \\ 40 \\ 26 \\ 35 \end{pmatrix}$$

En C el resultado de $9/10$ puede dar cero. Se debe escribir 9.0/10.0.

2.2 (a) Se trata de probar el método de ajuste no paramétrico que se definió en clases a partir de una función $K_a(x)$ con soporte en $-a \leq x \leq a$ que tenga forma de campana anulándose en los puntos extremos. Considere el archivo "DatosT2.dat" de N pares (x_a, y_a) que puede bajar de ucursos. Defina dos funciones $K_a(x)$ diferentes y por cada una de ellas utilice 3 valores de a. Esto da seis funciones $F_\beta(x)$ con $\beta = 1,2,.,6$.

(b) Además compare los seis ajustes anteriores con el interpolación de Lagrange F_L. ¿Cuán suave es esta interpolación? Si fuese interesante incluya un detalle (un *zoom*) que muestre una pequeña zona de la interpolación y los puntos dados.

(c) Para cada uno de estas F_β y de F_L obtenga la suma $S = \frac{1}{N}\sum_k (y_k - F(x_k))^2$. Un buen ajuste debe ser suave y tener un S_β tan pequeño como sea posible. Comente.

2.3 (a) Haga un programa que busque los ceros del polinomio: $P = x^7 - 3x^6 - 8x^5 + 20x^4 + 15x^3 - 13x^2 + 24x - 36$ que incluya tres rutinas diferentes para buscar ceros: método de Newton, de la secante, algún método de encajonamiento. Aleatoriammente el programa debe generar al menos 30 semillas diferentes, que se ubiquen entre -3 y 3, con las que se debe probar todas las rutinas "busca cero" y por cada una de ellas debe anotar qué cero de P obtuvo y cuántas iteraciones fueron necesarias. Es muy importante que cada una de las rutinas que buscan ceros tengan un límite al número de iteraciones que hace (por ejemplo, no hacer más de 50 iteraciones). Su informe debe tener una tabla con columnas que den la semilla, el algoritmo, la raíz obtenida y el número de iteraciones; puede decir "ninguna" ya que un algoritmo puede fallar con algunas semillas. En C el generador de semillas puede ser: `semilla = -3.0 + 6.0*drand48();`

(b) Usando el algoritmo visto en clases encuentre los ceros simultáneos de $F_1(x,y) = x^4 + y^4 - 10$ y $F_2(x,y) = x^3 y - x y^3 - 0{,}5y - 2{,}0$. Esto es, debe encontrar pares $p_j = (x_j, y_j)$ en los cuales ambas funciones son nulas. Indique cuántas iteraciones fueron necesarias en cada caso.

Capítulo 4

Ecuaciones diferenciales ordinarias

4.1. Reducción a ecuaciones de primer orden

El problema de resolver

$$\frac{d^2g}{d\xi^2} = F(\xi, g, g') \tag{4.1.1}$$

puede ser replanteado en la forma

$$\frac{d\vec{y}}{d\xi} = \vec{f}(\xi, \vec{y}) \quad + \quad \text{alguna condición inicial} \tag{4.1.2}$$

donde

$$\vec{y} = \begin{pmatrix} g \\ y_2 \end{pmatrix}, \qquad \vec{f} = \begin{pmatrix} y_2 \\ F \end{pmatrix} \tag{4.1.3}$$

es decir

$$\begin{aligned} \frac{dg}{d\xi} &= y_2 \\ \frac{dy_2}{d\xi} &= F(\xi; (g, y_2)) \end{aligned} \tag{4.1.4}$$

Se puede adivinar de lo anterior que siempre se puede reducir un problema de ecuaciones diferenciales ordinarias a un sistema de ecuaciones de primer orden.

Por ejemplo $\ddot{\vec{x}} = \frac{1}{m}\vec{F}$ puede plantearse definiendo

$$\vec{y} = \begin{pmatrix} \vec{x} \\ \vec{v} \end{pmatrix} \qquad \vec{f} = \begin{pmatrix} \vec{v} \\ \frac{1}{m}\vec{F} \end{pmatrix} \tag{4.1.5}$$

con lo que el problema consiste en resolver

$$\frac{d\vec{x}}{dt} = \vec{v}, \qquad \frac{d\vec{v}}{dt} = \frac{1}{m}\vec{F} \tag{4.1.6}$$

A continuación se verá una serie de métodos para abordar (4.1.2).

4.1.1. Método directo simple (de Euler)

Este método consiste en plantear (4.1.2) en la forma

$$y'_{n+1} = \frac{y_{n+1} - y_n}{h} - \mathscr{O}\left(\frac{h}{2}y''\right) = f_n$$

y de aquí definir la recurrencia

$$y_{n+1} = y_n + h\, f_n + \mathscr{O}(h^2) \tag{4.1.7}$$

Una forma para tratar de mejorar la precisión podría consistir en aproximar $y' \to (y_{n+1} - y_{n-1})/2h$ con lo que la ecuación a iterar es

$$y_{n+1} = y_{n-1} + 2h\, f_n + \mathscr{O}(h^3) \tag{4.1.8}$$

pero ambas recurrencias sufren del mismo problema de estabilidad.

Estabilidad de (4.1.8): Sea $\bar{y}$ la solución exacta del problema discreto. Se define ε_n tal que $y_n = \bar{y_n} + \varepsilon_n$ con lo cual (4.1.8) pasa a ser

$$\begin{aligned} \bar{y}_{n+1} + \varepsilon_{n+1} &= \bar{y}_{n-1} + \varepsilon_{n-1} + 2h\, f(\xi_n, \bar{y}_n + \varepsilon_n) \\ &= \bar{y}_{n-1} + \varepsilon_{n-1} + 2h\left(f(\xi_n, \bar{y}_n) + \frac{\partial f(\xi_n, \bar{y}_n)}{\partial y}\,\varepsilon_n\right) \end{aligned} \tag{4.1.9}$$

de donde

$$\varepsilon_{n+1} = \varepsilon_{n-1} + 2h\,\frac{\partial f(\xi_n, \bar{y}_n)}{\partial y}\,\varepsilon_n \tag{4.1.10}$$

que es la ecuación que da la forma como se propaga el error. Si la función f no es muy sensible a y se puede razonar suponiendo que es constante y en tal caso se trabaja con $\varepsilon_{n+1} = \varepsilon_{n-1} + 2\gamma\varepsilon_n$. Esta última ecuación se puede resolver suponiendo que

$$\varepsilon_n = \varepsilon_0\, \lambda^n$$

porque (4.1.10) se recuce a $\lambda^2 - 2\gamma\lambda - 1 = 0$ que da las raices $\lambda_\pm = \gamma \pm \sqrt{1+\gamma^2}$ y entonces

$$\varepsilon_n = A\lambda_+^n + B\lambda_-^n \tag{4.1.11}$$

• Si $\gamma > 0$ entonces $\lambda_+ > 1$ y λ_+^n crece con n y el error crece.

• Si $\gamma < 0$ se obtiene $\lambda_\pm = -|\gamma| \pm \sqrt{1+\gamma^2}$ y $\lambda_- < -1$ y λ_-^n crece con n cambiando de signo. La solución numérica que se obtiene va oscilando en torno a la solución $\bar{y}$ con amplitud creciente.

Es decir, el método es incondicionalmente inestable y no sirve.

4.1.2. Método implícito

La ecuación (4.1.2) se plantea

$$\left(\frac{dy}{d\xi}\right)_{n+\frac{1}{2}} = f\left(\xi_{n+\frac{1}{2}}, y_{n+\frac{1}{2}}\right) \tag{4.1.12}$$

El lado izquierdo se reescribe como

$$\frac{y_{n+\frac{1}{2}+\frac{1}{2}} - y_{n+\frac{1}{2}-\frac{1}{2}}}{2\frac{h}{2}} + \mathcal{O}(h^2) = \frac{y_{n+1}-y_n}{h} \tag{4.1.13}$$

y el lado derecho se reemplaza por el promedio de los valores en n y en $n+1$, $\frac{1}{2}[f_n + f_{n+1}]$ con lo cual resulta

$$y_{n+1} = y_n + \frac{h}{2}(f_n + f_{n+1}) + \mathcal{O}(h^3) \tag{4.1.14}$$

La incógnita, y_{n+1} aparece en ambos lados, por tanto, en cada paso de iteración se debe buscar el cero de la función

$$G(z) = z - y_n - \frac{h}{2}[f(\xi_n, y_n) + f(\xi_n + h, z)]$$

Estabilidad del método implícito: En forma análoga a como se procedió en el caso anterior se plantea

$$\bar{y}_{n+1} + \varepsilon_{n+1} = \bar{y}_n + \varepsilon_n + \frac{h}{2}(f(\xi_n, \bar{y}_n + \varepsilon_n) + f(\xi_{n+1}, \bar{y}_{n+1} + \varepsilon_{n+1}) \tag{4.1.15}$$

Expandiendo y trabajando un par de pasos se obtiene que

$$\varepsilon_{n+1} = \varepsilon_n + \frac{h}{2}(\partial_y f_n \varepsilon_n + \partial_y f_{n+1} \varepsilon_{n+1}) \tag{4.1.16}$$

El último ε_{n+1} puede ser reemplazado por ε_n ya que la diferencia es de más alto orden. Entonces

$$\varepsilon_{n+1} \approx \varepsilon_n \left[1 + \frac{h}{2}(\partial_y f_n + \partial_y f_{n+1})\right] \tag{4.1.17}$$

Lo crucial es el signo del paréntesis redondo en la expresión anterior. Si es negativo el error decrece.

Como se ve, lo que importa es el signo de $dy/d\xi$. Si la función es creciente, el error crece (aunque puede que porcentualmente crezca menos que la función) y si la función decrece el error decrece. El método puede funcionar y puede no funcionar: es condicionalmente estable.

4.1.3. Algoritmos Runge-Kutta

Esta vez (4.1.2) se plantea en la forma

$$\begin{aligned} \frac{d\vec{y}}{d\xi} &= \vec{f}(\xi, \vec{y}) \\ \vec{y}(0) &= \vec{y}_0 \end{aligned}$$

Y se usa dos expansiones de Taylor,

$$\vec{y}_{n+1} = \vec{y}_n + h\vec{y}_n{}' + \frac{h^2}{2}\vec{y}_n{}'' + \mathcal{O}(h^3) \tag{4.1.18}$$

$$\vec{y}_{n+\frac{1}{2}}{}' = \vec{y}_n{}' + \frac{h}{2}\vec{y}_n{}'' + \mathcal{O}(h^2) \tag{4.1.19}$$

De la última, multiplicada por h, se obtiene

$$\frac{h^2}{2}\vec{y}_n{}'' = \left(\vec{y}_{n+\frac{1}{2}}{}' - \vec{y}_n{}'\right) h + \mathcal{O}(h^3) \tag{4.1.20}$$

Que se reescribe utilizando la ecuación original

$$\frac{h^2}{2}\vec{y}_n{}'' = \left(\vec{f}_{n+\frac{1}{2}} - \vec{y}_n{}'\right) h + \mathcal{O}(h^3) \tag{4.1.21}$$

Al reemplazar esta expresión en (4.1.18) se cancelan las primeras derivadas y se obtiene

$$\vec{y}_{n+1} = \vec{y}_n + h\,\vec{f}(\xi_n + \frac{h}{2}, \vec{y}_n + \frac{h}{2}\,\vec{f}_n) \tag{4.1.22}$$

Este resultado final conocido como **RK2**, tradicionalmente se reescribe en la forma

$$\begin{aligned} \vec{k}_1 &= h\,\vec{f}(\xi_n, \vec{y}_n) \\ \vec{k}_2 &= h\,\vec{f}(\xi_n + \frac{h}{2}, \vec{y}_n + \frac{1}{2}\vec{k}_1) \qquad \textbf{RK2} \\ \vec{y}_{n+1} &= \vec{y}_n + \vec{k}_2 + \mathcal{O}(h^3) \end{aligned} \tag{4.1.23}$$

Este método es explícito, el error es orden h^3 y puede hacerse estable. La desventaja es que la función f debe ser llamada dos veces en cada iteración. Otra ventaja es que, puesto que avanza paso a paso y no requiere de información anterior, se puede ir ajustando el paso h a medida que se avanza en la integración.

Siguiendo un camino semejante se obtiene algoritmos de más alto orden.

RK3:

$$\begin{aligned} \vec{k}_1 &= h\,\vec{f}(\xi_n, \vec{y}_n) \\ \vec{k}_2 &= h\,\vec{f}(\xi_n + \frac{h}{2}, \vec{y}_n + \frac{1}{2}\vec{k}_1) \\ \vec{k}_3 &= h\,\vec{f}(\xi_n + h, \vec{y}_n - \vec{k}_1 + 2\vec{k}_2) \\ \vec{y}_{n+1} &= \vec{y}_n + \frac{1}{6}\left(\vec{k}_1 + 4\vec{k}_2 + \vec{k}_3\right) + \mathcal{O}(h^4) \end{aligned} \tag{4.1.24}$$

RK4:

$$\begin{aligned} \vec{k}_1 &= h\,\vec{f}(\xi_n, \vec{y}_n) \\ \vec{k}_2 &= h\,\vec{f}(\xi_n + \frac{h}{2}, \vec{y}_n + \frac{1}{2}\vec{k}_1) \\ \vec{k}_3 &= h\,\vec{f}(\xi_n + \frac{h}{2}, \vec{y}_n + \frac{1}{2}\vec{k}_2) \\ \vec{k}_4 &= h\,\vec{f}(\xi_n + h, \vec{y}_n + \vec{k}_3) \\ \vec{y}_{n+1} &= \vec{y}_n + \frac{1}{6}\left(\vec{k}_1 + 2\vec{k}_2 + 2\vec{k}_3 + \vec{k}_4\right) + \mathcal{O}(h^5) \end{aligned} \tag{4.1.25}$$

Ventajas: es h^5, es estable, permite adaptar el paso. Tiene amplia aplicabilidad.

Desventajas: se debe calcular f cuatro veces en cada iteración.

Sobre el paso variable. Para que el método sea preciso la magnitud de los $\vec{k}_a$ deben ser mucho menores que $\delta \equiv \|\vec{y}_{n+1} - \vec{y}_n\|$. Si se va detectando que tales magnitudes se acercan a δ se debe escoger h más chico. Por el contrario, si $\|\vec{k}_a\| \sim \delta$ entonces se debe agrandar h.

4.1.4. Estabilidad de RK4 en el caso $y' = \lambda y$

Un análisis completo de estabilidad depende de la ecuación particular que se quiera tratar, pero la tendencia general de cada esquema de integración puede sondearse estudiando lo que ocurre en el caso de la sencilla ecuación $y' = \lambda y$. En lo que sigue se estudiará para esta ecuación la estabilidad de RK4. Calcularemos los k_a y finalmente usaremos la expresión para y_{n+1} dada en (4.1.25). En las expresiones de los k_a se reemplaza f por λ multiplicando al segundo argumento de f que, genéricamente es y, obtieniéndose

$$\begin{aligned} k_1 &= h\lambda\, y_n \\ k_2 &= h\lambda \left(y_n + \frac{1}{2}k_1\right) = h\lambda \left(1 + \frac{h}{2}\lambda\right) y_n \\ k_3 &= h\lambda \left(y_n + \frac{h}{2}\lambda \left(1 + \frac{h}{2}\lambda\right) y_n\right) \\ &= h\lambda \left(1 + \frac{h}{2}\lambda \left(1 + \frac{h}{2}\lambda\right)\right) y_n \\ k_4 &= h\lambda \left(1 + h\lambda \left(1 + \frac{h}{2}\lambda \left(1 + \frac{h}{2}\lambda\right)\right)\right) y_n \end{aligned} \tag{4.1.26}$$

Al reemplazar estos valores en la expresión para y_{n+1} y escribiendo $y_{n+1} = \bar{y}_{n+1} + \varepsilon_{n+1}$ y similarmente $y_n = \bar{y}_n + \varepsilon_n$ (donde los $\bar{y}$ son la solución exacta de la ecuación discreta, se obtiene que

$$\frac{\varepsilon_{n+1}}{\varepsilon_n} = 1 + h\lambda + \frac{(h\lambda)^2}{2} + \frac{(h\lambda)^3}{6} + \frac{(h\lambda)^4}{24} \tag{4.1.27}$$

Para que haya estabilidad este cuociente tiene que tener un valor absoluto menor que 1 y puede comprobarse que esto requiere que

$$-2{,}7853 < h\lambda < 0$$

Por ejemplo, si $\lambda = -1$ entonces la estabilidad está garantizada con $0 < h < 2{,}7853$, que da un amplio margen para tener una ecuación absolutamente estable aun cuando, si h no es pequeño, la solución va a ser posiblemente poco confiable.

Si $\lambda > 0$ se puede superar la limitación de estabilidad integrando en la dirección opuesta, esto es, usando $h < 0$. Se comienza desde una "condición final" y, por cierto no se llega a la condición inicial que se da como dato. Por lo que debe volver a integrarse con otra "condición final" escogiéndola usando una rutina de Newton apropiada.

4.2. Integradores multipaso

4.2.1. Presentación

Nuevamente considérese la ecuación genérica

$$y' = f(\xi, y) \tag{4.2.1}$$

Al integrarla entre ξ_n y ξ_{n+1} se obtiene

$$y_{n+1} - y_n = \int_{\xi_n}^{\xi_{n+1}} \underbrace{f(\xi, y(\xi))}_{\mathscr{F}(\xi)} \, d\xi \tag{4.2.2}$$

El integrando será denotado $\mathscr{F}(\xi)$ y debe tenerse presente que es un valor de $y' = f$.

En este tipo de algoritmo debe tenerse una condición inicial: $y_0 = y(\xi_0)$. Puesto que $f(\xi, y)$ es una función conocida, entonces de (4.2.1) se tiene también $y'(0)$. El valor $y_1 = y(\xi_0 + h)$ se puede obtener, al menos en una aproximación de bajo orden, por medio de $y_1 \approx y_0 + h y'(0)$. Si se necesitara más puntos iniciales (por ejemplo y_0, y_1, y_2), estos algoritmos deben obtener esos primeros valores con alguna estrategia diferente como puede ser RK4 con paso suficientemente fino.

Al aproximar $\mathscr{F}(\xi) \approx \mathscr{F}_n$ la integral en (4.2.2) resulta valer $h\mathscr{F}_n + \mathscr{O}(h^2)$ y se obtiene

$$y_{n+1} = y_n + h\mathscr{F}_n + \mathscr{O}(h^2)$$

que es el algoritmo de Euler, que ya se sabe que es inestable.

4.2.2. Algoritmo predictor de Adams-Bashforth

Los métodos AB que se discuten a continuación se basan en la predicción del valor de F en el intervalo que se requiere en (4.2.2) usando como información valores anteriores: $\mathscr{F}_n$, $\mathscr{F}_{n-1}$ ...

Si se toma como aproximación que $\mathscr{F}(\xi) = a\xi + b$ y que exige que tal expresión sea válida en $\xi = \xi_{n-1}$ y en $\xi = \xi_n$ se obtiene

$$\mathscr{F}(\xi) = \frac{-\xi + \xi_n}{h} \mathscr{F}_{n-1} + \frac{\xi - \xi_{n-1}}{h} \mathscr{F}_n + \mathscr{O}(h^2) \tag{4.2.3}$$

Si se extiende la validez de la expresión anterior al intervalo siguiente, en él se puede hacer la integral

$$\int_{\xi_n}^{\xi_{n+1}} \mathscr{F} \, d\xi = h\left(\frac{3}{2}\mathscr{F}_n - \frac{1}{2}\mathscr{F}_{n-1}\right) + \mathscr{O}(h^2) \tag{4.2.4}$$

Es decir, se usa el conocimiento de $\mathscr{F}$ en (ξ_{n-1}, ξ_n) para extrapolar al intervalo (ξ_n, ξ_{n+1}) y hacer la integral recién descrita. Esta extrapolación conduce al integrador **AB3**,

$$y_{n+1} = y_n + \frac{h}{2}(3\mathscr{F}_n - \mathscr{F}_{n-1}) + \mathscr{O}(h^3) \qquad \textbf{AB3} \tag{4.2.5}$$

Claramente acá se ha usado la extrapolación como una forma de predecir el comportamiento de la función.

A continuación, versiones más precisas. En la que sigue se aproxima $\mathscr{F} = a\xi^2 + b\xi + c$ y los coeficientes (a, b, c) se determinan exigiendo que den los valores $\mathscr{F}_{n-2}$, $\mathscr{F}_{n-1}$ y $\mathscr{F}_n$. Una vez que se tiene tales coeficientes se tiene una forma cuadrática para F que se integra en el intervalo (ξ_n, ξ_{n+1}). Finalmente se obtiene:

$$y_{n+1} = y_n + \frac{h}{12}(23\mathscr{F}_n - 16\mathscr{F}_{n-1} + 5\mathscr{F}_{n-2}) + \mathscr{O}(h^4) \qquad \textbf{AB4} \tag{4.2.6}$$

En forma semejante pero ahora tomando en cuenta $\mathscr{F}_{n-3}$, $\mathscr{F}_{n-2}$, $\mathscr{F}_{n-1}$ y $\mathscr{F}_n$ se obtiene:

$$y_{n+1} = y_n + \frac{h}{24}\left(55\mathscr{F}_n - 59\mathscr{F}_{n-1} + 37\mathscr{F}_{n-2} - 9\mathscr{F}_{n-3}\right) + \mathscr{O}(h^5) \qquad \textbf{AB5} \qquad (4.2.7)$$

Estos son métodos explícitos de alto orden (error pequeño) y rápidos porque $\mathscr{F}$ se evalua una sola vez en cada paso. Pero se extrapola en lugar de interpolar como lo hace Runge-Kutta y por tanto falla si $\mathscr{F}$ es muy variable. Por su propia naturaleza el paso h debe permanecer fijo.

4.2.3. Estimador de Adams-Moulton

En este caso se estima f_{n+1} y en el caso AM3 se usa tan solo ξ_n y ξ_{n+1}.

AM3: Se toma

$$\mathscr{F}(\xi) = \frac{\xi_{n+1}-\xi}{h}\mathscr{F}_n + \frac{\xi-\xi_n}{h}\mathscr{F}_{n+1} + \mathscr{O}(h^2)$$

que permite obtener que la integral de la derecha sea

$$\int_{\xi_n}^{\xi_{n+1}} \mathscr{F}(\xi)\,d\xi = \frac{\mathscr{F}_n + \mathscr{F}_{n+1}}{2}h + \mathscr{O}(h^3)$$

por lo cual

$$y_{n+1} = y_n + \frac{h}{2}\left(\mathscr{F}_n + \mathscr{F}_{n+1}\right) + \mathscr{O}(h^3) \qquad \textbf{AM3} \qquad (4.2.8)$$

Este es un método implícito y es equivalente a alguno de lo métodos que ya se había visto. Para determinar y_{n+1} se debe buscar el cero de

$$G(z) = y_n + \frac{h}{2}\left(\mathscr{F}_n + f(\xi_{n+1}, z)\right) - z$$

Por definición $G(y_{n+1}) = 0$.

AM4: Esta vez $f(\xi)$ en el intervalo $(n, n+1)$ se aproxima con una parábola que pasa por los valores $\mathscr{F}_{n-1}$, $\mathscr{F}_n$ y $\mathscr{F}_{n+1}$ y se obtiene

$$I = \frac{h}{12}\left(5\mathscr{F}_{n+1} + 8\mathscr{F}_n - \mathscr{F}_{n-1}\right) + \mathscr{O}(h^4)$$

y entonces

$$y_{n+1} = y_n + \frac{h}{12}\left(5\mathscr{F}_{n+1} + 8\mathscr{F}_n - \mathscr{F}_{n-1}\right) + \mathscr{O}(h^4) \qquad \textbf{AM4} \qquad (4.2.9)$$

AM5: En forma análoga que en el caso anterior pero usando un polinomio cúbico con la información de $(n-2, n-1, n, n+1)$ y se obtiene

$$y_{n+1} = y_n + \frac{h}{24}\left(9\mathscr{F}_{n+1} + 19\mathscr{F}_n - 5\mathscr{F}_{n-1} + \mathscr{F}_{n-2}\right) + \mathscr{O}(h^5) \qquad \textbf{AM5} \qquad (4.2.10)$$

Puesto que estos métodos son implícitos se debe determinar un cero con métodos como el de la secante y eso conlleva un riesgo. Pero en §4.2.4 se muestra una variante muy exitosa.

4.2.4. Método predictor-corrector

Este método consiste en mezclar métodos explícito y otro implícito del mismo orden.

La versión más sencilla de predictor-corrector es la *regla trapezoidal de Nystrom* que se puede enunciar como sigue:

p: Primero se predice $y_{n+1}^{(P)} = y_{n-1} + 2h\mathscr{F}_n$.

e: Con el valor recién obtenido de y_{n+1} se evalua $\mathscr{F}_{n+1}$

c: El valor de y_{n+1} se corrije calculando $y_{n+1} = y_n + \frac{h}{2}(\mathscr{F}_n + \mathscr{F}_{n+1})$

En el método general de esta serie y que hace uso de los algoritmos AB y AM se procede como sigue:

P: Usando AB se obtiene $y_{n+1}^{(P)}$ a partir de la información de puntos anteriores;

E: el último valor obtenido de y_{n+1} se usa para calcular $\mathscr{F}_{n+1}^{(E)} = \mathscr{F}(\xi_{n+1}, y_{n+1})$;

C: el último valor, $\mathscr{F}_{n+1}^{(E)}$, se usa en AM para obtener $y_{n+1}^{(C)}$.

En el paso C se usa Adams-Moulton como si fuese un método explícito. El paso P se hace una sola vez, pero los pasos (EC) se pueden hacer sucesivamente para lograr algún tipo de convergencia, lo que se denota **P(EC)**^{n}. El paso E es el único que hace uso de la función f de la ecuación que se está resolviendo.

4.3. Predictor-corrector de Gear

Aquí se da una versión sencilla de esta estrategia predictor-corrector para la ecuación de Newton $\ddot{r} = a$, donde $a(r,v)$ es una función conocida. En general todas estas cantidades tienen varias componentes.

En la etapa predictiva se calcula

$$\begin{array}{rcccccccc} r_{n+1} & = & r_n & + & hv_n & + & \frac{h^2}{2}a_n & + & \frac{h^3}{6}b_n \\ hv_{n+1} & = & & & hv_n & + & 2\frac{h^2}{2}a_n & + & 3\frac{h^3}{6}b_n \\ \frac{h^2}{2}a_{n+1} & = & & & & & \frac{h^2}{2}a_n & + & 3\frac{h^3}{6}b_n \\ \frac{h^3}{6}b_{n+1} & = & & & & & & & \frac{h^6}{6}b_n \end{array} \tag{4.3.1}$$

Con los actuales valores en r_{n+1} y v_{n+1} se puede calcular un valor corregido a_{n+1}^c, lo que define un *error*

$$\Delta_{n+1} = \frac{h^2}{2}\left(a_{n+1}^c - a_{n+1}\right)$$

El único lugar donde interviene el lado derecho de la ecuación de movimiento es en el cálculo de

a^c_{n+1}. Una vez determinado este error se calcula nuevas cantidades

$$\begin{array}{lll} r_{n+1} & = r_{n+1} & + \; c_0\,\Delta_{n+1} \\ hv_{n+1} & = nv_{n+1} & + \; c_1\,\Delta_{n+1} \\ \frac{h^2}{2}a_{n+1} & = \frac{h^2}{2}a_{n+1} & + \; c_2\,\Delta_{n+1} \\ \frac{h^3}{6}b_{n+1} & = \frac{h^3}{6}b_{n+1} & + \; c_3\,\Delta_{n+1} \end{array} \tag{4.3.2}$$

C.W. Gear (en trabajos de 1966 y 1971) determinó los valores óptimos para los coeficientes c_j en diversos casos. Dependen, por ejemplo, de si la ecuación que se resuelve es de primer orden (como las ecuaciones de Hamilton) o de segundo orden (Newton) y también depende del orden de la expansión que se haga (en (4.3.1) se expandió hasta $b = r'''$). En el caso que aquí se ha presentado los coeficientes de Gear son

$$c = \begin{pmatrix} 1/6 \\ 5/6 \\ 1 \\ 1/3 \end{pmatrix} \tag{4.3.3}$$

Al revés que los métodos predictor multipaso vistos antes, éste determina directamente los valores en $n+1$ conociendo los valores en n.

4.4. Métodos de Verlet y variaciones

Se trata de resolver ecuaciones de Newton

$$\ddot{x} = \mathscr{A}(x,t) \qquad \text{donde } \mathscr{A} \text{ es la fuerza dividida por la masa} \tag{4.4.1}$$

sin pasar a ecuaciones de primer orden. Estos métodos se definen cuando $\mathscr{A}(x,t)$ no depende de las velocidades. Sin embargo es generalizable al caso en que existe una fuerza viscosa lineal en la velocidad $\ddot{x} = a_0(x,t) - cv$. Se usa $v = \frac{x_{n+1}-x_{n-1}}{2h}$.

L. Verlet[1] presentó su algoritmo por primera vez en su trabajo *Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules*, Phys. Rev. **159**, 98-103 (1967).

4.4.1. Propiamente Verlet

La ecuación (4.4.1) se discretiza,

$$\frac{x_{n+1} - 2x_n + x_{n-1}}{h^2} + \mathscr{O}(h^2) = \mathscr{A}_n$$

con lo cual,

$$x_{n+1} = 2x_n - x_{n-1} + h^2\,\mathscr{A}_n + \mathscr{O}(h^4) \tag{4.4.2}$$

En cada iteración se evalua $\mathscr{A}$ una sola vez y el error es orden h^4. La velocidad no aparece.

[1] Físico francés nacido en 1931

Las iteraciones son $(x_{n-1}, x_n) \to x_{n+1}$, pero si las condiciones iniciales son x_0 y v_0 se puede integrar con RK4 desde x_0 hasta x_1 y luego se procede con (4.4.2) o bien usando $x(h) = x(0) + h\,v(0) + \frac{h^2}{2}\mathscr{A}(0) + \frac{h^3}{6}\mathscr{A}'(0)$.

Nótese que igualmente se puede despejar x_{n-1} y la ecuación es la misma, es decir, el algoritmo es reversible en el tiempo.

Para evaluar la velocidad se puede hacer

$$v_n = \frac{x_{n+1} - x_{n-1}}{2h} + \mathscr{O}(h^2) \tag{4.4.3}$$

que es un error muy grande frente a h^4.

Si bien (4.4.2) aparece con un error de truncamiento pequeño, suele tener error de redondeo grande porque a una diferencia de dos números grandes (orden 1), $2x_n - x_{n-1}$, se le suma una cantidad de segundo orden, $h^2\,a_n$. Para evitar esta fuente de error existe el método *leapfrog*.

4.4.2. Estabilidad del método de Verlet

Sea $\bar{x}$ la solución exacta de (4.4.2) con h fijo y definamos

$$x_n = \bar{x_n} + \varepsilon_n$$

donde los ε representan desviaciones que se han introducido. La sustitución de esta definición en la ecuación de Verlet da

$$\varepsilon_{n+1} - \left(2 + h^2\,a'_n\right)\,\varepsilon_n + \varepsilon_{n-1} = 0 \tag{4.4.4}$$

donde a_n se escribió como $a(\bar{x_n} + \varepsilon_n) \approx a(\bar{x_n}) + \varepsilon_n\,a'_n$.

Caso de fuerza armónica: A continuación se analiza el caso de una fuerza armónica,

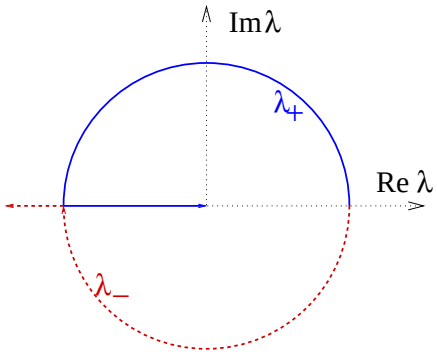
$$a(x) = -\omega^2 x \qquad\qquad a'(x) = -\omega^2$$

entonces

$$\varepsilon_{n+1} - 2\,(1-R)\,\varepsilon_n + \varepsilon_{n-1} = 0 \tag{4.4.5}$$

donde $2R = h^2\omega^2$. Se reemplaza $\varepsilon_n = \varepsilon_0\,\lambda^n$ lo que inmediatamente conduce a

$$\lambda^2 - 2(1-R)\,\lambda + 1 = 0 \tag{4.4.6}$$



cuyas raices son

$$\lambda_\pm = 1 - R \pm \sqrt{R^2 - 2R} \tag{4.4.7}$$

por lo que λ puede ser complejo.

Si $R = 0$ entonces $\lambda_\pm = 1$.

Si $R = 1$ entonces $\lambda_\pm = \pm i$.

Si $R = 2$ entonces $\lambda_\pm = -1$.

Si $R \to \infty$ entonces $\lambda_\pm = 1 - R \pm R\sqrt{1 - \frac{2}{R}}$ concluyéndose que $\lambda_+ = 0$ y que $\lambda_- = -\infty$.

En resumen, cuando R crece desde cero hasta infinito, λ_+ recorre primero la semicircunferencia $\mathrm{Im}\,\lambda > 0$ unitaria del plano complejo y luego avanza desde $\lambda_+ = -1$ hasta $\lambda_+ = 0$ mientras que λ_- recorre primero la semicircunferiencia $\mathrm{Im}\,\lambda < 0$ unitaria del plano complejo y luego avanza desde $\lambda_- = -1$ hasta $\lambda_- = -\infty$.

Si $R \leq 2$ se tiene $|\lambda_\pm| = 1$ que garantiza estabilidad, mientras que $R > 2$ implica $|\lambda_-| > 1$ que garantiza inestabilidad.

Puesto que $R = \frac{h^2\omega^2}{2}$, la condición $R \leq 2$ corresponde a

$$h \leq \frac{2}{\omega}$$

y si se reemplaza $\omega = \frac{2\pi}{T}$ entonces la condición es

$$h \leq \frac{T}{\pi} \tag{4.4.8}$$

Otros casos: Si se trata de otras fuerzas que provienen de un potencial $U(x)$, la cota máxima para h está dada por el período mínimo, es decir, por la frecuencia máxima, que es proporcional a $\sqrt{U''(x_{\min})}$.

Si bien de esta manera se obtiene estabilidad, eso no quiere decir que se tenga confiabilidad, es decir, precisión aceptable. Normalmente se debe usar un h bastante menor que el que garantiza estabilidad.

4.4.3. Leapfrog

Puesto que

$$\frac{v_{n+\frac{1}{2}} - v_{n-\frac{1}{2}}}{h} + \mathscr{O}(h^2) = F_n$$

se obtiene que

$$v_{n+\frac{1}{2}} = v_{n-\frac{1}{2}} + hF_n + \mathscr{O}(h^3)$$

Además

$$\frac{x_{n+\frac{1}{2}+\frac{1}{2}} - x_{n+\frac{1}{2}-\frac{1}{2}}}{h} + \mathscr{O}(h^2) = v_{n+\frac{1}{2}}$$

que conduce a

$$x_{n+1} = x_{n-1} + hv_{n+\frac{1}{2}} + \mathscr{O}(h^3)$$

En resumen las ecuaciones que siguen deben usarse en el orden indicado

$$\begin{aligned} v_{n+\frac{1}{2}} &= v_{n-\frac{1}{2}} + h\,a_n + \mathscr{O}(h^3) \\ x_{n+1} &= x_n + h\,v_{n+\frac{1}{2}} + \mathscr{O}(h^3) \end{aligned} \tag{4.4.9}$$

Las posiciones se determinan con n entero y las velocidades con índice semientero.

4.5. Algoritmos simplécticos

4.5.1. Operadores de traslación

Se define la notación

$$e^{a\frac{d}{d\xi}} \equiv 1 + a\frac{d}{d\xi} + \frac{a^2}{2!}\frac{d^2}{d\xi^2} + \ldots + \frac{a^n}{n!}\frac{d^n}{d\xi^n} + \ldots + \ldots \tag{4.5.1}$$

de donde se obtiene que

$$e^{a\frac{d}{d\xi}} f(\xi) = \sum_{n=0}^{\infty} \frac{a^n}{n!}\frac{d^n}{d\xi^n} f(\xi) = f(\xi + a) \tag{4.5.2}$$

La acción del operador $e^{a\frac{d}{d\xi}}$ sobre una función $f(\xi)$ es producir la función trasladada $f(\xi + a)$.

En forma similar se puede definir el operador $\exp\left[\vec{b}\cdot\nabla_{\vec{z}}\right]$,

$$\begin{aligned} f(\vec{z}+\vec{b}) &= f(\vec{z}) + (\vec{b}\cdot\nabla) f(\vec{z}) + \frac{1}{2!}\vec{b}\cdot\nabla\left(\vec{b}\cdot\nabla\right) f(\vec{z}) + \ldots & (4.5.3)\\ &= f(\vec{z}) + \sum_j b_j \frac{\partial}{\partial z_j} f(\vec{z}) + \frac{1}{2!}\sum_{j,k} b_j b_k \frac{\partial^2}{\partial z_j \partial z_k} f(\vec{z}) + \ldots & (4.5.4)\\ &= \exp\left[\vec{b}\cdot\nabla_{\vec{z}}\right] f(\vec{z}) & (4.5.5) \end{aligned}$$

4.5.2. Ecuaciones de movimiento

En mecánica las ecuaciones de movimiento de un sistema pueden escribirse en la forma

$$\frac{d}{dt}\begin{pmatrix} \vec{r} \\ \vec{v} \end{pmatrix} = \begin{pmatrix} \vec{v} \\ \vec{F} \end{pmatrix} \tag{4.5.6}$$

donde $\vec{F}$ es la fuerza total $\vec{F}$ sobre la partícula dividida por la masa de ella m. Estas mismas ecuaciones pueden escribirse también usando el operador Liouvileano,

$$\mathscr{L} \equiv \vec{v}\cdot\nabla_r + \vec{F}\cdot\nabla_v \tag{4.5.7}$$

donde los ingredientes con que se construye $\mathscr{L}$ están en el vector $\begin{pmatrix} \vec{r} \\ \vec{v} \end{pmatrix}$ sobre el cual va a actuar.

La ecuación (4.5.6) toma la forma

$$\frac{d}{dt}\begin{pmatrix} \vec{r} \\ \vec{v} \end{pmatrix} = \mathscr{L}\begin{pmatrix} \vec{r} \\ \vec{v} \end{pmatrix} \tag{4.5.8}$$

La solución formal del movimiento entonces es

$$\begin{pmatrix} \vec{r}(\varepsilon) \\ \vec{v}(\varepsilon) \end{pmatrix} = e^{\varepsilon\mathscr{L}}\begin{pmatrix} \vec{r} \\ \vec{v} \end{pmatrix}\Bigg|_{t=0} \tag{4.5.9}$$

El problema es que esta operación no puede hacerse en forma sencilla. Sin embargo, si las fuerzas dependen solo de la posición conviene escribir

$$e^{\varepsilon \mathscr{L}} = e^{\varepsilon A + \varepsilon B} \tag{4.5.10}$$

donde

$$A = \vec{v} \cdot \nabla_r, \qquad B = \vec{F} \cdot \nabla_v \tag{4.5.11}$$

El operador $e^{\varepsilon A}$ actua sólo sobre los vectores posición, es una simple traslación de la posición, mientras que el operador $e^{\varepsilon B}$ actua sólo sobre los vectores velocidad y es una traslación de velocidades, es decir,

$$e^{\varepsilon A}\begin{pmatrix} \vec{r} \\ \vec{v} \end{pmatrix} = \begin{pmatrix} \vec{r} + \varepsilon \vec{v} \\ \vec{v} \end{pmatrix}, \qquad e^{\varepsilon B}\begin{pmatrix} \vec{r} \\ \vec{v} \end{pmatrix} = \begin{pmatrix} \vec{r} \\ \vec{v} + \varepsilon \vec{F} \end{pmatrix} \tag{4.5.12}$$

Nótese que en $\vec{r} + \varepsilon \vec{v}$ el vector velocidad es el que aparece en la segunda compente y en $\vec{v} + \varepsilon \vec{F}$ el argumento de $\vec{F}$ es el vector $\vec{r}$ que está en la primera componente. Así entonces, si se define la notación

$$\vec{r}_n \equiv \vec{r}(\varepsilon n) \qquad \text{y} \qquad \vec{v}_n \equiv \vec{v}(\varepsilon n) \tag{4.5.13}$$

resulta

$$e^{\varepsilon A}\begin{pmatrix} \vec{r}_n \\ \vec{v}_m \end{pmatrix} = \begin{pmatrix} \vec{r}_{n+1} \\ \vec{v}_m \end{pmatrix} = \begin{pmatrix} \vec{r}_n + \varepsilon \vec{v}_m \\ \vec{v}_m \end{pmatrix}, \qquad e^{\varepsilon B}\begin{pmatrix} \vec{r}_n \\ \vec{v}_m \end{pmatrix} = \begin{pmatrix} \vec{r}_n \\ \vec{v}_{m+1} \end{pmatrix} = \begin{pmatrix} \vec{r}_n \\ \vec{v}_m + \varepsilon F_n \end{pmatrix} \tag{4.5.14}$$

Integrar el movimiento es no trivial porque

$$e^{\varepsilon (A+B)} \neq e^{\varepsilon A} e^{\varepsilon B}$$

Una forma conocida de encontrar una solución aproximada es

$$e^{\varepsilon (A+B) + \mathscr{O}(\varepsilon^3)} = e^{\frac{\varepsilon}{2} B} e^{\varepsilon A} e^{\frac{\varepsilon}{2} B} \tag{4.5.15}$$

4.5.3. Construcción del algoritmo $\mathscr{O}(\varepsilon^3)$

A continuación se muestra en detalle como actúa el operador compuesto (4.5.15). El operador a la extrema derecha actua sobre (r_n, v_n):

$$\begin{pmatrix} \vec{r}_n \\ \vec{v}_{n+\frac{1}{2}} \end{pmatrix} = e^{\frac{\varepsilon}{2} B}\begin{pmatrix} \vec{r}_n \\ \vec{v}_n \end{pmatrix} = \begin{pmatrix} \vec{r}_n \\ \vec{v}_n + \frac{\varepsilon}{2}\vec{F}(\vec{r}_n) \end{pmatrix} = \begin{pmatrix} \vec{r}_n \\ \vec{v}_n + \frac{\varepsilon}{2}\vec{F}_n \end{pmatrix}$$

El siguiente operador actua sobre este resultado

$$\begin{aligned} \begin{pmatrix} \vec{r}_{n+1} \\ \vec{v}_{n+\frac{1}{2}} \end{pmatrix} &= e^{\varepsilon A}\begin{pmatrix} \vec{r}_n \\ \vec{v}_{n+\frac{1}{2}} \end{pmatrix} = \begin{pmatrix} \vec{r}_n + \varepsilon \vec{v}_{n+\frac{1}{2}} \\ \vec{v}_{n+\frac{1}{2}} \end{pmatrix} = \begin{pmatrix} \vec{r}_n + \left(\vec{v}_n + \frac{\varepsilon}{2}\vec{F}_n\right)\varepsilon \\ \vec{v}_n + \frac{\varepsilon}{2}\vec{F}_n \end{pmatrix} \\ &= \begin{pmatrix} \vec{r}_n + \varepsilon \vec{v}_n + \frac{\varepsilon^2}{2}\vec{F}_n \\ \vec{v}_n + \frac{\varepsilon^2}{2}\vec{F}_n \end{pmatrix} \end{aligned}$$

Y finalmente

$$\begin{pmatrix} \vec{r}_{n+1} \\ \vec{v}_{n+1} \end{pmatrix} = e^{\frac{\varepsilon}{2}B} \begin{pmatrix} \vec{r}_{n+1} \\ \vec{v}_{n+\frac{1}{2}} \end{pmatrix} = \begin{pmatrix} \vec{r}_n + \vec{v}_n \varepsilon + \frac{\varepsilon^2}{2}\vec{F}_n \\ \vec{v}_n + \frac{\varepsilon}{2}\left(\vec{F}_n + \vec{F}_{n+1}\right) \end{pmatrix} \tag{4.5.16}$$

En esta expresión final se ve que es crucial que $\vec{F}_n$ dependa tan solo de las posiciones, porque la parte superior de la ecuación permite calcular $\vec{r}_{n+1}$, con el cual se obtiene $\vec{F}_{n+1}$ y así se tiene los ingrediente para calcular explícitamente $\vec{v}_{n+1}$.

4.5.4. El Jacobiano asociado

Por simplicidad se calcula el Jacobiano asociado al algoritmo unidimencional

$$\begin{aligned} x_{n+1} &= r_n + v_n \varepsilon + \frac{\varepsilon^2}{2} F_n \\ v_{n+1} &= v_n + \frac{\varepsilon}{2}\left(F_n + F_{n+1}\right) \end{aligned} \tag{4.5.17}$$

Se comprueba que

$$\begin{aligned} \frac{\partial x_{n+1}}{\partial x_n} &= 1 + \frac{\varepsilon^2}{2}\frac{\partial F_n}{\partial x_n} \\ \frac{\partial x_{n+1}}{\partial v_n} &= \varepsilon \\ \frac{\partial v_{n+1}}{\partial x_n} &= \frac{\varepsilon}{2}\left(\frac{\partial F_n}{\partial x_n} + \frac{\partial F_{n+1}}{\partial x_{n+1}}\frac{\partial x_{n+1}}{\partial x_n}\right) \\ \frac{\partial v_{n+1}}{\partial v_n} &= 1 + \frac{\varepsilon}{2}\frac{\partial F_{n+1}}{\partial x_{n+1}}\frac{\partial x_{n+1}}{\partial v_n} \end{aligned}$$

que conduce a que

$$J = \frac{\partial x_{n+1}}{\partial x_n}\frac{\partial v_{n+1}}{\partial v_n} - \frac{\partial x_{n+1}}{\partial v_n}\frac{\partial v_{n+1}}{\partial x_n} = 1$$

El Jacobiano vale uno a todo orden y en cualquier dimensión. Esta es una característica de los algoritmos simplécticos.

- Revizar validez en dimensión mayor.

4.5.5. Nuevamente el algoritmo de Verlet

El algoritmo simpléctico (4.5.16) escrito por componentes es

$$\begin{aligned} \vec{r}_{n+1} &= \vec{r}_n + \vec{v}_n \varepsilon + \frac{\varepsilon^2}{2}\vec{F}_n \\ \vec{v}_{n+1} &= \vec{v}_n + \frac{\varepsilon}{2}\left(\vec{F}_n + \vec{F}_{n+1}\right) \end{aligned}$$

Si a la primera de ellas se le resta una réplica con $n \to n-1$ se obtiene

$$\vec{r}_{n+1} - \vec{r}_n = \vec{r}_n - \vec{r}_{n-1} + \varepsilon\left(\vec{v}_n - \vec{v}_{n-1}\right) + \frac{\varepsilon^2}{2}\left(\vec{F}_n - \vec{F}_{n-1}\right)$$

pero, de acuerdo a la segunda de las ecuaciones, se puede reemplazar $\vec{v}_n - \vec{v}_{n-1}$ por $\frac{\varepsilon}{2}\left(\vec{F}_{n-1} + \vec{F}_n\right)$ por lo que se obtiene

$$\vec{r}_{n+1} - 2\vec{r}_n + \vec{r}_{n-1} = \varepsilon^2 \vec{F}_n + \mathscr{O}(\varepsilon^4)$$

que es el algoritmo de Verlet obtenido como consecuencia del algoritmo simpléctico (4.5.16).

4.5.6. Algoritmos simplécticos de más alto orden

Un teorema establece cómo construir algoritmos simplécticos de más alto orden. Una familia de ellos toman la forma

$$e^{(A+B)\,\varepsilon + \mathscr{O}(\varepsilon^{n+1})} = \prod_{j=1}^{N} e^{a_j A \varepsilon}\, e^{b_j B \varepsilon} \qquad \text{aquí se define } n \tag{4.5.18}$$

Esta forma general es muy importante porque es trivial demostrar que una traslación de posición sin trasladar las velocidades o vice versa es una transformación con Jacobiano unitario: $\det \frac{\partial(\vec{r}',\vec{v}')}{\partial(\vec{r},\vec{v})} = 1$. Esta propiedad es necesaria y suficiente para que se conserve el volumen del espacio de fase, que es una propiedad básica de las evoluciones hamiltoneanas.

Los coeficientes $\{a_j, b_j\}$ se determinan exigiendo dos condiciones: que n sea maximal (es decir, minimizando el error por truncamiento) y que haya invariancia a la inversión temporal. La primera exigencia se traduce en muchas condiciones dependiendo del valor de N. De todas esas condiciones las que siempre se deben cumplir son

$$\sum a_j = 1\,, \qquad \sum b_j = 1$$

Para ver cómo exigir invariancia temporal se debe recordar primero que

$$\left(e^{tA_1}\, e^{tA_2} \ldots e^{tA_K}\right)^{-1} = e^{-tA_K} \ldots e^{-tA_2}\, e^{-tA_1} \tag{4.5.19}$$

Cuando se expande (4.5.18) toma una forma como (4.5.19). Si se resume (4.5.19) como $U^{-1}(t) = V(t)$, simetría a la inversión temporal quiere decir que $V(t) = U(-t)$, lo que hace necesario que $A_1 = A_K$, $A_2 = A_{K-1}$ etc.

El algoritmo (4.5.15) que se ha analizado corresponde a $(N = 2, n = 2, a_1 = 0, a_2 = 1, b_1 = b_2 = \frac{1}{2})$. Un caso superior [encontrado por Forest y Ruth, Physica D, 43, 105 (1990)]: $(N = 4, n = 4, a_1 = 0, a_2 = a_4 = \theta, a_3 = 1 - 2\theta, b_1 = b_4 = \frac{\theta}{2}, b_2 = b_3 = \frac{1-\theta}{2})$ y $\theta = (2 - 2^{1/3})^{-1}$.

Muchísimo más sobre esto puede verse en el artículo de Omeyan, Mryglod y Folk en Phys. Rev. E **66**, 026701 (2002) y en referencias que ahí se citan.

4.6. Recomendación final

- Si f se puede evaluar rápidamente conviene usar RK4.
- Si se necesita ir adaptando el paso también debe usarse RK4.
- Si es una ecuación de Newton conservativa se debe usar alguno de los algoritmos simplécticos.
- Si f es muy lento de evaluar conviene usar un método predictor corrector.

4.7. Problemas

1. Considere la ecuación de un oscilador armónico forzado

$$\ddot{x} + \omega^2 x = A \sin(kx - \Omega t) \tag{4.7.1}$$

Ella proviene de la ecuación de una partícula cargada moviéndose en un campo magnético uniforme a lo largo del eje Z al que se superpone una electrostática plana que se propaga a lo largo del eje X. Ver Physcis Today de noviembre 1988, p27. Esencialmente la misma ecuación es mencionada también en Physics Today, marzo 1987, p9.

2. Integrar la evolución de la ecuación de van der Pol

$$\ddot{x} = (1 - x^2)\dot{x} - x \tag{4.7.2}$$

Representa a un oscilador armónico más un término que podría representar un freno viscoso, pero para $|x| < 1$ ese término acelera en lugar de frenar.

3. Integrar la evolución del oscilador de Duffing

$$\ddot{x} = x - x^3 - a\dot{x} + b\cos\omega t \tag{4.7.3}$$

4. Se trata de calcular, usando RK4, el comportamiento de un péndulo amortiguado por un roce viscoso γ y forzado. Este sistema consiste en una vara ideal rígida de largo L, de masa despreciable, en cuyo extremo hay una masa m. El punto de apoyo $\mathscr{O}$ no está fijo sino que oscila verticalmente con amplitud A y frecuencia ν. Se puede demostrar que la ecuación para el ángulo φ es

$$\ddot{\varphi} + \gamma\dot{\varphi} + \omega_0^2 \sin\varphi = -\frac{(2\pi\nu)^2 A}{L}\cos(2\pi\nu t)$$

donde $\omega_0^2 = \frac{g}{L}$, $g = 9{,}81$ y $L = 1$m.

   a) Tome $A = 0{,}07$m, $\gamma = 0{,}1\text{seg}^{-1}$, $\varphi(0) = 10^o$ y $\dot{\varphi}(0) = 0$ y estudie la amplitud angular asintótica ($t \to \infty$) como función de la frecuencia en el rango: $0{,}9 \le \nu \le 1{,}1$ con $\delta\nu = 0{,}001$. (*Amplitud angular* es el valor máximo que toma $\varphi(t)$).

   b) Tome $A = 0{,}3$m, $\gamma = 0{,}4\text{seg}^{-1}$, $\dot{\varphi}(0) = 0$ y $\nu = 3\text{seg}^{-1}$. Obtenga la evolución de φ tanto cuando $\varphi(0)$ está alrededor de 10^o como cuando está alrededor de 170^o.

5. Modifique el algoritmo de Verlet original para el caso en que la ecuación de movimiento bidimensional pueda tener una fuerza de roce viscoso lineal, $\ddot{\vec{r}} = \vec{a}(\vec{r}) - c\vec{v}$. Con ese nuevo algoritmo integre numéricamente la órbita en el caso

$$\vec{a} = -k_1\vec{r} - k_2 r^2\vec{r} \tag{4.7.4}$$

y tome $k_1 = 1$, $k_2 = 0{,}05$ y $c = 0{,}1$. Use condiciones iniciales $(x = 4{,}0, y = 0{,}0)$, $(v_x = 0{,}0, v_y = 1{,}0)$ y dibuje la órbita durante tiempo suficiente para que corte al eje X ocho veces.

6. Estudie el movimiento de un péndulo plano, que en lugar de hilo tiene un resorte de constante elástica k y largo natural R. La masa del punto material es $m = 1$. Use coordenadas cartesianas para integrar usando el algoritmo de Verlet. Las ecuaciones son

$$\begin{aligned} \ddot{x} &= -k\frac{(r-R)\,x}{r} \\ \ddot{y} &= -k\frac{(r-R)\,y}{r} - g \end{aligned}$$

Use: $k = 5$, $R = 4$, $g = 1$, tome condiciones iniciales: $(x = 0, y = -R - g/k, v_x = 0{,}1, v_y = 0)$ e integre hasta $t = 20$. Calcule la evolución del punto material con RK4 y con Verlet usando el mismo $dt = \varepsilon$. Escoja este incremento de modo que ambos algoritmos den soluciones parecidas pero distinguibles (al menos en la parte final). Para hacer esta comparación dibuje $x(t)$, $y(t)$ y $y(x)$. Compruebe que usando un dt varias veces más chico ambos algoritmos dan esencialmente la misma solución. ¿Cual de las dos soluciones originales estaba más cerca de la solución más precisa?

7. Integre la ecuación

$$\ddot{\vec{r}}(t) = -\frac{\vec{r}}{r^3}$$

directamente en coordenadas cartesianas $(x(t), y(t))$ usando como condiciones iniciales:

$$\vec{r}(0) = 2\,\hat{\imath} \qquad \vec{v}(0) = 0{,}1\,\hat{\jmath}$$

En todos los casos integre hasta poco más allá de completar una vuelta y dibujando la órbita en el plano (x, y) usando $N \leq 5000$ aun cuando el resultado sea insatisfactorio.

Convierta las ecuaciones a un sistema de primer orden e integre usando RK4.

8. Para obtener el algoritmo para integrar la evolución de una cadena unidimensional de N osciladores amortiguados usando Verlet se comienza con las ecuaciones de movimiento del sistema conservativo. El lagrangeano del sistema (naturalmente sin amortiguar) es

$$L = \sum_{a=0}^{N} \frac{m\,\dot{q}_a^2}{2} - \sum_{a=1}^{N} \frac{k}{2}\,(q_{a+1} - q_a)^2 \tag{4.7.5}$$

Aquí los q_a son las desviaciones del punto de equilibrio de cada oscilador. La ecuación de movimiento genérica es $m\,\ddot{q}_a = k\,(q_{a+1} + q_{a-1} - 2q_a)$ de ahí que si se agrega amortiguación la ecuación queda $m\,\ddot{q}_a = k\,(q_{a+1} + q_{a-1} - 2q_a) - c\,\dot{q}_a$ o equivalentemente

$$\ddot{q}_a = \omega^2\,(q_{a+1} + q_{a-1} - 2q_a) - \eta\,\dot{q}_a \tag{4.7.6}$$

Al discretizar usando el algoritmo de Verlet se obtiene

$$\frac{q_a^{n+1} - 2q_a^n + q_a^{n-1}}{\varepsilon^2} = \omega^2\left(q_{a+1}^n + q_{a-1}^n - 2q_a^n\right) - \eta\,\frac{q_a^{n+1} - q_a^{n-1}}{2\varepsilon}$$

El problema consiste en tratar *dos medios*: la interacción entre las partículas de la 0 a la A está caracterizada por un ω_1 y de la $A+1$ hasta la última por un ω_2 (pero se usará el mismo

η). La partícula A es la frontera entre ambos medios y es la que satisface una ecuación diferente.

Las partículas de la 1 a la $A-1$ interactuan igual que en el caso anterior. Lo mismo con las partículas de la $A+1$ en adelante.

$$\begin{array}{ll} \ddot{q}_{a<A} & = \omega_1^2 \left(q_{a+1} + q_{a-1} - 2q_a\right) - \eta\, \dot{q}_a \\ \ddot{q}_{a>A} & = \omega_2^2 \left(q_{a+1} + q_{a-1} - 2q_a\right) - \eta\, \dot{q}_a \end{array} \tag{4.7.7}$$

y debe ser discretizada tomando esto encuenta. La partícula A que hace de frontera, en cambio, satisface

$$\ddot{q}_A = \omega_2^2\, q_{A+1} + \omega_1^2\, q_{A-1} - (\omega_1^2 + \omega_2^2) q_A - \eta\, \dot{q}_A \tag{4.7.8}$$

y usted debe discretizarla siguiendo un patrón semejamte al ya usado.

Considere $a = 0,1,.,500$, es decir, 501 puntos, donde $A = 250$, $\omega_1^2 = 2$, $\omega_2^2 = 1$ y $\eta = 0{,}008$. El punto $a = 0$ satisface $q_0(t) = \sin \omega t$ tan solo si $t \leq \frac{2\pi}{\omega}$, después de eso permanece cero para siempre. Tome $\omega = 0{,}1$. El punto $a = 500$ está siempre quieto. De modo que lo que se debe evolucionar son las partículas $a = 1,2,\ldots,499$. La condición inicial es $q_a = 0$, $\dot{q}_a = 0$.

Haga seis gráficos $q_a(t)$ versus a para un t fijo (cada uno de los cuales representa instantáneas del sistema) separadas por $\Delta t = 120$seg. Es decir, el sistema en $t = 120$, $t = 240$ .. hasta $t = 720$. Para que quede más claro lo que representa cada figura, es mejor que dibuje lo dicho con línea sólida y con línea más débil o de puntos el estado del sistema unos 4 segundos antes, de esa manera se verá el sentido de la evolución.

Capítulo 5

Problemas de condiciones de borde y problemas de autovalores

5.1. Introducción

En este capítulo se aprenderá a resolver ecuaciones lineales de la forma

$$\frac{d^2y}{dx^2} - R(x)\,y = S(x) \tag{5.1.1}$$

donde $S(x)$ es un término inhomogéneo y R es una función real. Cuando R es negativo la solución de la ecuación homogénea (esto es, con $S = 0$) son oscilantes con número local de onda $\sqrt{-R}$, mientras que cuando R es positivo la solución está dominada por exponenciales reales tipo $e^{\pm x\sqrt{R}}$.

Si se tiene una ecuación de la forma

$$f'' + A(x)\,f' + B(x)\,f = C(x)$$

se hace el reemplazo $f(x) = y(x)\,\exp[-\frac{1}{2}\int^x A(u)\,du]$ y se obtiene una ecuación de la forma (5.1.1).

Ejemlos de ecuaciones tipo (5.1.1)

E1 Si se busca el potencial electrostático V generado por una distribución de carga $\rho(\vec{r})$ se debe plantear la ecuación de Poisson $\nabla^2 V = -\rho/\varepsilon_0$ que, si hay simetría esférica, puede simplificarse a

$$\frac{1}{r^2}\frac{d}{dr}\left[r^2\frac{dV}{dr}\right] = -\frac{\rho}{\varepsilon_0} \tag{5.1.2}$$

y si se hace la sustitución $V(r) = \phi(r)/(\varepsilon_0 r)$ se llega a

$$\frac{d^2\phi}{dr^2} = -r\rho \tag{5.1.3}$$

que es de la forma (5.1.1) con $R = 0$ y $S = -r\rho$.

.

E2 Similarmente, el problema radial asociado a la ecuación de Schrödinger es

$$\frac{d^2\Psi_R}{dr^2} + k^2(r)\,\Psi_R = 0\,, \qquad k^2(r) = \frac{2m}{\hbar^2}\left[E - \frac{\ell(\ell+1)\hbar^2}{2mr^2} - V(r)\right]$$

que tiene la forma (5.1.1) con $R(r) = -k^2(r)$ y $S = 0$.

Ambos ejemplos podrían ser resueltos con métodos ya vistos excepto que hay situaciones que complican tal metodología.

El método que se describe en lo que sigue es típico que las condiciones de borde deban imponerse en *puntos diferentes* y, en el caso de una ecuación de onda o tipo Schrödinger, hay un problema de autovalores que resolver.

5.2. El algoritmo de Numerov

Expandiendo $y(x \pm h)$ hasta $\mathcal{O}(h^6)$ se obtiene

$$y_{n+1} - 2y_n + y_{n-1} = \left(y_n'' + \frac{h^2}{12} y_n^{IV}\right) h^2 + \mathcal{O}(h^6) \tag{5.2.1}$$

Por otro lado, tomando la segunda derivada de (5.1.1) resulta

$$\begin{aligned} y_n^{IV} &= \frac{d^2}{dx^2}\left[Ry + S\right]_n \\ &= \frac{(Ry)_{n+1} - 2(Ry)_n + (Ry)_{n-1}}{h^2} + \frac{S_{n+1} - 2S_n + S_{n-1}}{h^2} + \mathcal{O}(h^2) \end{aligned} \tag{5.2.2}$$

Esta expresión para y_n^{IV} y la expresión para la segunda derivada que da la ecuación original, (5.1.1) se sustituyen en (5.2.1) y se reordena, obteniéndose la expresión básica para el algoritmo de Numerov:

$$\left(1 - \frac{h^2}{12}R_{n+1}\right) y_{n+1} - 2\left(1 + \frac{5h^2}{12}R_n\right) y_n + \left(1 - \frac{h^2}{12}R_{n-1}\right) y_{n-1} = \frac{h^2}{12}\left(S_{n+1} + 10S_n + S_{n-1}\right) + \mathcal{O}(h^6) \tag{5.2.3}$$

A partir de esta expresión se puede despejar ya sea y_{n+1} o y_{n-1} para tener una relación de recurrencia que resuelve hacia adelante o hacia atrás la ecuación con un error $\mathcal{O}(h^6)$. En cada paso de iteración las funciones R y S son llamadas una sola vez.

Si en la ecuación de partida $\frac{d^2y}{dx^2} - Ry = S$ la función S es función de x y también de $y(x)$ el método anterior sigue siendo válido. En tal caso la ecuación (5.1.1) no es lineal.

5.3. Problemas asociados a las condiciones de borde

5.3.1. Integración directa de un problema con condiciones de borde

Primer ejemplo: Se verá cómo resolver el caso especial de (5.1.3) en que la densidad de carga tiene una forma exponencial, $\rho = -\frac{1}{8\pi}e^{-r}$. Una densidad de carga esféricamente simétrica implica

una carga total $Q = 4\pi \int_0^\infty r^2 \rho\, dr$. En el caso actual arroja $Q = -1$. La función ϕ satisface

$$\frac{d^2\phi}{dr^2} = \frac{r\exp[-r]}{8\pi} \tag{5.3.1}$$

Puesto que ρ es una función suave, el potencial V es finito en el origen, lo que implica—que $\phi \propto rV$—que $\phi(0) = 0$. Por otro lado la forma asintótica de V en infinito tiene que ser $V(r \sim \infty) = \frac{Q}{4\pi\varepsilon_0 r}$ que, en el caso actual, implica que $\phi(\infty) = -\frac{1}{4\pi}$.

Este problema se puede intentar resolver en la forma

$$\phi_{n+1} = 2\phi_n - \phi_{n-1} + \frac{h^2}{12}\left(S_{n+1} + 10S_n + S_{n-1}\right) \tag{5.3.2}$$

donde $S = \frac{r}{8\pi} e^{-r}$. Para proseguir se escoge un $r_{\max}$ suficientemente grande para considerar que ya la solución en ese valor tiene un valor asintótico muy cercano a $-\frac{1}{4\pi}$.

Para integrar se da inicialmente un valor arbitrario a $\phi_1 = \phi(h)$. El valor en $r_{\max}$ va a depender del valor arbitrario dado a ϕ_1. Si se integra desde infinito hacia la izquierda partiendo del valor $\phi_\infty = -\frac{1}{4\pi}$ no se obtendría cero en $r = 0$.

El rango $0 \le r \le r_{\max}$ es dividido en N intervalos de largo h. La rutina que usa (5.3.2) comienza con valores conocidos para los dos primeros puntos: ϕ_0 y ϕ_1 para calcular ϕ_2 y continuar hacia la derecha. La rutina toma como datos iniciales conocidos los valores $S_{\text{izq}} = S(r = 0)$ y $S_{\text{cen}} = S(r = h)$ y luego entra en un ciclo

```
for(k=2;k<=N;k++)
{ r    = k*h;
  Sder = S(r);
  F[k] = 2*F[k-1] - F[k-2] + c0*(Sder + 10*Scen + Sizq);
  Sizq = Scen;
  Scen = Sder;
}
```

Al integrar tomando en cuenta una sola condición de borde se hace inevitable incorporar características de la solución general de la ecuación. En el caso particular que se está viendo la ecuación homogénea tiene como solución general $\phi_h = a + br$ lo que implica que la solución general de la ecuación completa—de la forma $\phi_{\text{gen}} = \phi_p + \phi_h$—crece linealmente con r. El algoritmo (5.3.2) arroja, entonces, una solución particular ϕ_p que asintóticamente crece linealmente con r, lo que no es trivial compatibilizar con la condición $\phi(\infty) = -\frac{1}{4\pi}$.

Una forma algo brutal de obtener la solución buscada consiste en tomar un cierto valor $r_{\max}$ como si ya fuese infinito y definir

$$\phi(r) = \phi_p(r) + \frac{\phi_{\text{m}} - \phi_p(r_{\max})}{r_{\max}}\, r \tag{5.3.3}$$

donde ϕ_{m} es el valor que se quiere imponer para ϕ en $r = r_{\max}$, ya que así es automático que $\phi(0) = 0$ y que $\phi(r_{\max}) = \phi_{\text{m}}$. En resumen, se integra imponiendo tan solo $\phi_p(0) = 0$ y una vez que se tiene esta solución numérica, que llamamos ϕ_p, se calcula ϕ usando (5.3.3). Los valores que toma ϕ_p dependen del valor arbitrario inicialmente dado a $\phi(h)$, pero ese efecto es borrado al usar

(5.3.3). La forma lineal de *reparar* la solución tiene sentido tan solo porque la solución general de la ecuación homogénea en este ejemplo es lineal.

Más en general, para integrar cualquier ecuación de la forma (5.1.1), con $S \neq 0$, dadas las condiciones $y(a) = y_a$ y $y(b) = y_b$, se puede comenzar desde $x = a$ hacia la derecha, obteniéndose una función $\tilde{y}$ que satisface la ecuación y la condición de borde en $x = a$. Pero esta solución no satisface la ecuación de borde en $x = b$. Por otro lado se usa una solución de la ecuación homogénea, $y_h(x)$ que en $x = a$ satisfaga $y_h(a) = 0$. Con $\tilde{y}$ e y_h se construye la función $y(x)$ buscada utilizando la expresión

$$y(x) = \tilde{y}(x) + \frac{y_b - \tilde{y}(b)}{y_h(b)} y_h(x) \tag{5.3.4}$$

Es fácil comprobar que en efecto esta función $y(x)$ satisface la ecuación diferencial y además satisface ambas condiciones de borde. En efecto, puesto que fue construida como combinación lineal de una solución particular $\tilde{y}$ de la ecuación inhomogénea y una solución de la ecuación homogénea, está garantizado que y es solución de la ecuación lineal original.

En los casos en que $b = \infty$ se debe escoger un x_{max}. En tales casos es delicado escoger el valor que debe colocarse en lugar de y_b en la expresión anterior. Lo más sano es estudiar el comportamiento asintótico analítico de la solución, descrito por una función y_{asint} e imponer que $y(x_{\mathsf{max}}) = y_{\mathsf{asint}}(x_{\mathsf{max}})$. Por ejemplo, resolviendo (5.3.1), en lugar de tomar $\phi_m = -\frac{1}{4\pi}$ en (5.3.3) se puede primero ver que la forma asintótica debe ser de la forma $-\frac{1}{4\pi} + \beta e^{-x}$ y, para tener consistencia con $x \sim \infty$, es necesario que $\beta = \frac{1}{4\pi}$, viéndose que es conveniente usar $\phi_m = -\frac{1}{4\pi} + \frac{1}{4\pi} e^{-x_{\mathsf{max}}}$.

Al integrar hacia la derecha desde $x = a$ se debe dar un valor arbitrario a $y(a+h)$. Debe ponerse cuidado con el valor que se escoja para evitar que $\tilde{y}$ pueda tomar valores muy grandes y en general debiera ser $y(a) + \mathcal{O}(h)$. De otro modo, los dos términos en (5.3.4) serían muy grandes de signo opuesto y la solución $y(x)$ resultaría poco confiable.

- En algunos problemas suele ser más conveniente integrar de derecha a izquierda.

- Si las soluciones de la ecuación homogénea son muy dispares, este método puede no ser muy preciso.

En lugar de usar un método de reparación, también puede definirse una función $F(z) = \phi(r_{\mathsf{max}}) - y_b$ donde $z = \phi_1$ y aplicar una rutina busca cero para alcanzar la solución. Esto no es eficiente y puede ser muy inestable.

5.3.2. Uso de una función de Green

5.3.2.1. El problema

Existe una aplicación menos conocida del algoritmo de Numerov y que puede ser muy útil en casos que los otros métodos no arrojan precisón suficiente, lo que sucede si las soluciones de la ecuación homogénea tienen comportamientos muy diferentes. Considérese nuevamente la ecuación general

$$\left(\frac{d^2}{dx^2} - R(x)\right) y = S(x) \qquad y(a) = y_a\,, \quad y(b) = y_b \tag{5.3.5}$$

El método que se verá a contuación debiera ser aplicable aun si el intervalo (a,b) es infinito. Existen dos maneras de abordar este infinito: una es hacer un cambio de variable, por ejemplo, si se hace el cambio $x = \tan(\zeta)$ se puede cubrir todo el eje real x cubriendo un rango perfectamente finito de ζ: $-\frac{\pi}{2} \leq \zeta \leq \frac{\pi}{2}$.

Otra posibilidad se presenta si existe forma de obtener analíticamente la forma asintótica de la solución buscada y, por ejemplo, en lugar de la condición $y(\infty) = 0$—imposible de imponer en un método numérico—se pueda decir $y(x \sim \infty) = x^{-4}$ o lo que corresponda.

5.3.2.2. Papel de la función de Green

Para resolver este problema con el concepto de función de Green se plantea primero resolver otro problema:

$$\left(\frac{d^2}{dx^2} - R(x)\right) f(x) = \tilde{S}(x) \qquad f(a) = 0\,, \quad f(b) = 0 \tag{5.3.6}$$

donde la función $\tilde{S}(x)$ se relaciona con $S(x)$ en la forma que se especifica más adelante.

Este último problema se plantea como el problema de encontrar una función de Green $G(x,x')$ que satisfaga

$$\left(\frac{d^2}{dx^2} - R(x)\right) G(x,x') = \delta(x-x') \qquad \text{tal que} \qquad G(a,x') = 0\,, \quad G(b,x') = 0 \tag{5.3.7}$$

De la propiedad básica de $\delta(x-x')$, la ecuación anterior implica que

$$\int_{x=x'-\varepsilon}^{x=x'+\varepsilon} \left(\frac{d^2}{dx^2} - R(x)\right) G(x,x')\,dx = 1 \tag{5.3.8}$$

Una vez que se tenga $G(x,x')$ se puede ver que $f(x)$ se puede escribir en la forma

$$f(x) = \int G(x,x')\,\tilde{S}(x')\,dx' \tag{5.3.9}$$

ya que trivialmente esta función $f(x)$ satisface (5.3.6).

5.3.2.3. Hacia la solución del problema original

Una vez que se haya resuelto todo lo anterior se plantea que

$$\begin{aligned} y(x) &= f(x) + \alpha x + \beta \\ \tilde{S}(x) &= S(x) + (\alpha x + \beta)\,R(x) \end{aligned} \tag{5.3.10}$$

debido a lo siguiente. La acción del operador $\frac{d^2}{dx^2} - R(x)$ sobre $y(x)$ es la acción sobre $f(x)$ más la acción sobre $\alpha x + \beta$. De la primera acción, resulta $\tilde{S}(x)$ y de la segunda se obtiene $-(\alpha x + \beta)\,R(x)$, de tal modo que la acción de $\frac{d^2}{dx^2} - R(x)$ sobre $y(x)$ da $S(x)$ que es lo que se buscaba.

En principio el problema está resuelto siempre y cuando se escoja α y β de tal modo que $y(x)$ satisfaga las condiciones de borde indicadas en (5.3.5).

Ya que $f(x)$ se anula en ambos bordes, $y(a) = \alpha a + \beta$ y $y(b) = \alpha b + \beta$ lo que determina que

$$\alpha = \frac{y_b - y_a}{b-a} \qquad \beta = \frac{b y_a - a y_b}{b-a}$$

quedando así totalmente definida la solución $y(x)$ obtenida por medio de la función de Green $G(x,x')$.

5.3.2.4. Construcción numérica de la función de Green

El asunto crucial, entonces, es obtener $G(x,x')$ y esto se aborda comenzando por construir dos soluciones especiales de

$$g'' - R(x)\, g = 0 \tag{5.3.11}$$

Se integra numéricamente (5.3.11) usando el algoritmo de Numerov desde a hacia la derecha a partir del valor $g(a) = 0$ lo que numéricamente define la función $g_-(x)$. Similarmente se obtiene una función $g_+(x)$ integrando (5.3.11) hacia la izquierda desde b a partir de $g_+(b) = 0$.

Definiendo el Wronskiano

$$W(x) = {g_+}'(x)\, g_-(x) - {g_-}'(x)\, g_+(x)$$

se puede demostrar en forma muy sencilla que W es una constante, esto es, $W' = 0$. Se escoge normalizar las soluciones $g_\pm$ de modo que $W = 1$.

El hecho que $W \neq 0$ garantiza que se tiene soluciones $g_\pm$ linealmente independientes.

Se define

$$G(x,x') = \begin{cases} g_-(x) g_+(x') & x < x' \\ \\ g_+(x) g_-(x') & x > x' \end{cases} \tag{5.3.12}$$

lo que implica que $G(a,x') = 0$ y que $G(b,x') = 0$.

Veamos, por otro lado, que el límite $\varepsilon \to 0$ de la integral

$$\Delta(\varepsilon) \equiv \int_{x'-\varepsilon}^{x'+\varepsilon} \left(\frac{d^2}{dx^2} - R \right) G(x,x')\; dx$$

no es trivial. Para ver esto se supondrá que $R(x)G(x,x')$ es suave en el rango $x' - \varepsilon \leq x \leq x' + \varepsilon$ por lo que la contribución del término con R en la integral anterior se anula en el límite, mientras que el otro término es la integral de una derivada: $\frac{d^2G}{dx^2}$ por lo cual la integral anterior vale

$$\Delta(\varepsilon) = \left. \frac{dG}{dx} \right|_{x=x'+\varepsilon} - \left. \frac{dG}{dx} \right|_{x=x'-\varepsilon} = W = 1 \tag{5.3.13}$$

por lo cual (5.3.7) y (5.3.8) se cumplen y se comprueba que la función G descrita es la función de Green del problema.

5.3.2.5. La solución

La solución del problema original (5.3.5) es

$$y(x) = \int G(x,x')\tilde{S}(x')\,dx' + \frac{y_b - y_a}{b-a}x + \frac{by_a - ay_b}{b-a} \qquad \text{donde} \tag{5.3.14}$$

$$\tilde{S}(x) = S(x) + \left(\frac{y_b - y_a}{b-a}x + \frac{by_a - ay_b}{b-a}\right) R(x) \tag{5.3.15}$$

Fundiendo ambas expresiones en una se tiene que

$$y(x) = \int G(x,x')\left\{S(x') + \left(\frac{y_b - y_a}{b-a}x' + \frac{by_a - ay_b}{b-a}\right) R(x')\right\} dx' + \frac{y_b - y_a}{b-a}x + \frac{by_a - ay_b}{b-a} \tag{5.3.16}$$

5.4. Problemas de autovalores

Estos son problemas de ecuaciones lineales homogéneas de segundo orden con condiciones de borde en puntos diferentes. El problema genérico que se analizará plantea determinar λ tal que

$$-\frac{d^2y}{dx^2} + q(x)y = \lambda y \qquad a \le x \le b \tag{5.4.1}$$

con $y(a) = 0$ y $y(b) = 0$. No se excluye que $a = -\infty$ ni que $b = \infty$. Puede ocurrir que esta ecuación no tenga solución para cualquier valor de λ. Los λ permitidos se los denomina los *autovalores* del operador

$$-\frac{d^2}{dx^2} + q(x)$$

Los autovalores pueden ser un conjunto discreto o continuo de valores o una mezcla: valores discretos en un cierto rango y valores continuos en otro.

5.4.1. Problemas sencillos de autovalores

Una cuerda: Considérese el problema

$$\begin{aligned} y'' &= -k^2 y \\ y(0) &= y(1) = 0 \end{aligned} \tag{5.4.2}$$

En este tipo de problemas la normalización de la función es arbitrario y los valores posibles de k^2 deben ser determinados.

En este caso el algoritmo de Numerov es más sencillo porque k^2 es constante,

$$\left(1 + \frac{(hk)^2}{12}\right) y_{n+1} - 2\left(1 - \frac{5(hk)^2}{12}\right) y_n + \left(1 + \frac{(hk)^2}{12}\right) y_{n-1} = 0 \tag{5.4.3}$$

que se puede reescribir como

$$y_{n+1} = -y_{n-1} + \frac{24 - 10\,(hk)^2}{12 + (hk)^2} y_n \tag{5.4.4}$$

Usando esta última forma se integra desde $x = 0$ hacia la derecha obteniéndose en el extremo derecho un valor que depende del valor que se ha dado a k,

$$F(k) = y_N$$

y, usando algunas de las estrategias ya vistas, se busca los ceros de $F(k)$.

5.4.2. Ecuación de Schrödinger en una dimensión: estados ligados

El problema típico de autovalores con la ecuación de Schrödinger es

$$-\frac{\hbar^2}{2m}\psi'' + V(x)\,\psi = E\,\psi \qquad \int |\psi|^2\,dx = 1 \qquad (5.4.5)$$

La integral arriba es sobre el dominio sobre el cual la función ψ está definida. En lo que sigue se usa unidades tales que

$$2m = 1 \qquad y\hbar = 1\,. \qquad (5.4.6)$$

La ecuación entonces tiene la forma

$$\psi'' + k^2(x)\,\psi = 0 \quad \text{donde} \quad k^2 = E - V(x)$$

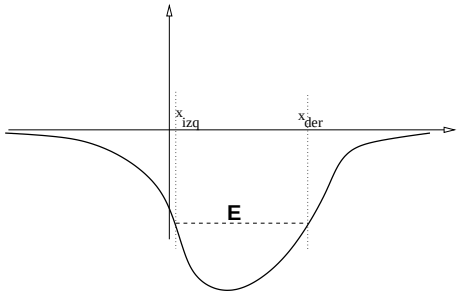


Figura 5.1: La función potencial V y el valor tentativo para la energía E definen valores x_{izq} y x_{der} que sirven de base para escoger $x_{\min}$ y $x_{\max}$.

Para el caso de estados ligados existe un rango finito $(x_{\text{izq}}, x_{\text{der}})$ fuera del cual $E < V(x)$, es decir $k^2(x) < 0$ y la solución de interés decae exponencialmente. Pero fuera de $(x_{\text{izq}}, x_{\text{der}})$ la ecuación también admite una solución exponencialmente creciente.

Los límites $(x_{\text{izq}}, x_{\text{der}})$ de esta zona son los puntos de retorno de mecánica clásica y el algoritmo debe abordar el hecho que la solución general crece exponencialmente fuera de las fronteras definidas por los puntos de retorno: $(x_{\text{izq}}, x_{\text{der}})$. Un problema de este tipo no puede ser simplemente integrado de izquierda a derecha o vice versa porque inevitablemente los errores de redondeo harían aparecer la solución exponencialmente creciente y ninguna rutina de reparación sería confiable.

Para evitar esto se integra desde algún punto $x_{\min}$ a la izquierda de x_{izq} hasta poco más allá de algún punto intermedio de *empalme* x_e. A esta función se la llama $\tilde{\psi}_{\text{Izq}}$. Además se obtiene ψ_{Der} integrando desde un punto $x_{\max}$ a la derecha de x_{der} hasta poco más a la izquierda del mismo punto de empalme.

Para hacer estas integraciones se escoge N y $h = \frac{1}{N}(x_{\max} - x_{\min})$, $x_n = x_{\min} + nh$ de tal modo que $x_e = x_{\min} + eh$. Supondremos que se integra ψ_{Der} desde la extrema derecha hasta $n = e - 1$ y ψ_{Izq} desde la extrema izquierda hasta $e + 1$.

Para integrar desde los extremos lo más sencillo es suponer que $\psi_{\text{Izq}}(x_{\min}) = 0$ y que $\psi_{\text{Der}}(x_{\max}) = 0$, pero esto puede ser insatisfactorio. En general es más apropiado determinar los comportamientos asintóticos de ψ a ambos lados y hacer uso de ellos.

Se debe exigir que tanto ψ como ψ' sean continuas en el punto de empalme x_e.

La primera condición se consigue renormalizando una de las funciones. Por ejemplo:

$$\left(\psi_{\mathsf{Izq}}[n] \leftarrow \frac{\psi_{\mathsf{Der}}[e]}{\tilde{\psi}_{\mathsf{Izq}}[e]}\,\tilde{\psi}_{\mathsf{Izq}}[n]\right)_{n=0,1,..,e+1} \tag{5.4.7}$$

Esto garantiza que en el punto de empalme $n=e$ haya un valor $\psi[e]$ común a ambas funciones.

La segunda condición se obtiene definiendo

$$F(E) = \frac{\psi_{\mathsf{Der}}[e+1]-\psi_{\mathsf{Der}}[e-1]}{h\,\psi[e]} - \frac{\psi_{\mathsf{Izq}}[e+1]-\psi_{\mathsf{Izq}}[e-1]}{h\,\psi[e]} \tag{5.4.8}$$

y con alguna estrategia ya estudiada se puede buscar los ceros de $F(E)$. Ellos son los autovalores E del problema de estados ligados.

La definición anterior de $F(E)$ corresponde a una definición particular de la derivada en x_e. Si se escoge utilizar una expresión más precisa para la derivada es necesario integrar ψ_{Izq} por más puntos a la derecha de $n=e$ y lo propio para ψ_{Der} hacia la izquierda.

En resumen:

1. Se escoge un valor semilla para E y un valor dE
2. Se calcula x_{izq} y x_{der}. Con ellos se escogen $(x_{\mathsf{min}}, x_{\mathsf{max}})$, se escoge el valor de N y se calcula $h=(x_{\mathsf{max}}-x_{\mathsf{min}})/N$. La asociación entre el índice discreto n y x es $x = x_{\mathsf{min}} + nh$. Se escoge también el punto del discreto que se usará para el empalme, $n=e$, el que debiera estar entre los puntos de retorno o coincidir con uno de ellos.
3. Se integra $\tilde{\psi}_{\mathsf{Izq}}$ desde $n=0$ hasta $n=e+1$; se integra ψ_{Der} desde $n=N$ hasta $e-1$.
4. Se define ψ_{Izq} por medio de una normalización de $\tilde{\psi}_{\mathsf{Izq}}$ utilizando (5.4.7); así se logra que ψ_{Izq} y ψ_{Der} tengan el mismo valor en el punto de empalme $n=e$.
5. Se calcula $F(E)$ por medio de (5.4.8)
6. Se define nuevo $E \leftarrow E+dE$ y se repite lo anterior hasta que $F(E)$ cambie de signo.
7. Se ingresa a la rutina `secante` que terminará por encontrar un valor de E tal que $F(E)$ sea nulo, es decir, un valor de E para el cual la derivada en x_e sea continua.

5.4.3. Ecuación de Schrödinger radial

5.4.3.1. La ecuación

En problemas con potencial central la ecuación de Schrödinger de función de onda $\Psi(\vec{r})$ se simplifica si se escribe con coordenadas esféricas en la forma

$$\Psi(\vec{r}) = \frac{1}{r}U(r)Y_{\ell m}(\theta,\phi)$$

donde los $Y_{\ell m}(\theta,\phi)$ son funciones conocidas llamadas *esféricos armónicos*.

La ecuación para la parte radial $U(r)$ es reducida en forma estándar a una ecuación donde la función de energía potencial $V(r)$ contiene un término de la barrera centrífuga

$$\frac{d^2U}{dr^2} + k^2(r)U = 0$$
$$k^2(r) = \left[E - \frac{\ell(\ell+1)}{r^2} - V(r)\right] \tag{5.4.9}$$

donde ℓ es un entero no negativo y se continua usando las unidades definidas por (5.4.6). El problema central es encontrar los autovalores E.

En lo que sigue se supondrá que la función $V(r)$ tiende a cero cuando r tiende a infinito.

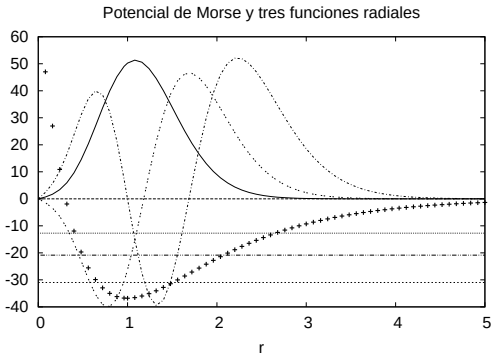


Figura 5.2: El potencial de Morse, $V = V_0(1 - \exp[-a(r - r_0)])^2$ en linea de cruces y tres de las la funciones U_0, U_1 y U_2 sin normalizar correspondientes a los estados más bajos (sin ceros, uno y dos ceros). Se aprecia que estas funciones se anulan en $r = 0$. Las líneas horizontales representan los correspondientes autovalores $E < 0$.

5.4.3.2. Comportamiento lejano

La función $U(r)$ para r muy grande debe tender a cero de modo que pueda ser normalizada:

$$\int_0^\infty |U(r)|^2 \, dr = 1$$

Si $E < 0$ la ecuación para r muy grande es $U''/U \approx E$ que implica que

$$U(r \sim \infty) \propto \exp\left[-\sqrt{|E|}\, r\right]$$

Si $E > 0$ hay problemas que no se verán.

5.4.3.3. El comportamiento cerca del origen

Primer caso: Si $V(r)$ cerca del origen es menos divergente que r^{-2} la ecuación en la vecindad de $r \sim 0$ es

$$\frac{d^2U}{dr^2} - \frac{\ell(\ell+1)}{r^2}U = 0$$

que tiene dos soluciones independientes, una proporsional a $r^{\ell+1}$ y la otra proporsional a $r^{-\ell}$. La segunda solución es inaceptable, de modo que

$$U(r \approx 0) \propto r^{\ell+1}$$

que se anula en el origen como una potencia de r. La integración numérica se puede hacer entonces comenzando con $U_{n=0} = 0$ y $U_{n=1} \propto h^{\ell+1}$ y además se debe integrar desde algún $r_{\max}$ y hacer un empalme para determinar los autovalores E.

Si $V(r)$ en el origen es más singular que r^{-2} se debe hacer un análisis diferente.

Segundo caso, un ejemplo: Por ejemplo el potencial de Lennard Jones 6-12 comúnmente usado en simulaciones moleculares es

$$V_{LJ} = V_0 \left[\left(\frac{\sigma}{r}\right)^{12} - 2\left(\frac{\sigma}{r}\right)^{6} \right] \tag{5.4.10}$$

Cerca de $r = 0$ se puede despreciar de $k^2(r)$ todos sus términos excepto aquel con r^{-12}. La solución no divergente en el origen satisface

$$U'' = V_0 \frac{\sigma^{12}}{r^{12}} U$$

y es

$$U(r \sim 0) = \exp\left[-\frac{\sqrt{V_0}\sigma^6}{5r^5} \right]$$

que se anula muy rápidamente al acercarse a $r = 0$. En la integración numérica se debe evitar el punto $r = 0$ y es usual comenzar a una pequeña distancia del origen tomando $0{,}5\sigma \leq r_{\min} \leq 0{,}8\sigma$. Entre este punto y el origen se debe usar la forma analítica dada en la última expresión.

Segundo caso, general: Más en general, si el potencial en el origen diverge como $V_0 \left(\frac{b}{r}\right)^n$ con $n > 2$, de modo que la ecuación cerca del origen es

$$U'' \approx V_0 \left(\frac{\sigma}{r}\right)^n U$$

puede verse que el comportamiento de U en el origen debe ser

$$U \sim \exp\left[-\frac{2\sqrt{V_0}\,\sigma}{n-2} \left(\frac{\sigma}{r}\right)^{\frac{n}{2}-1} \right]$$

En este caso general la función $U(r)$ también se anula exponencialmente en el origen.

La integración debe comenzar nuevamente desde un $r_{\min}$ que sea una fracción de σ y antes se debe usar la forma analítica.

La forma de integrar: Desde un $r_{\max}$ se integra hacia la izquierda usando el algoritmo de Numerov a partir de $U(N) \approx \exp[-\sqrt{|E|}\, r_{\max}]$ y $U(N-1) \approx \exp[-\sqrt{|E|}\,(r_{\max} - h)]$ hasta un punto menor al punto de empalme e. Desde un $r_{\min}$ se integra considerando la aproximación analítica cercana al origen según cuál sea el caso para iterar con el algoritmo de Numerov desde ese punto hacia la derecha. Luego se ejecuta el procedimiento de empalme.

5.5. Problemas

1. Obtenga numéricamente las funciones propias normalizadas y los valores propios asociados al potencial

$$V(x) = \frac{x^4 - 16}{1 + 4x^6}$$

en $-\infty \leq x \leq \infty$ correspondientes a los *primeros* cuatro autovalores.

2. Obtenga los tres primeros autovalores del problema de la ecuación de Schrödinger estacionaria adimensionalizada

$$\left(-\frac{d^2}{dx^2}+V\right)\Psi = E\,\Psi$$

con $0 \leq x \leq \infty$ y con $V = -\frac{5e^{-x}}{x}$.

3. El enlace covalente suele ser modelado por el potencial central de Lennard-Jones 9-6,

$$V = V_0\left(2\left(\frac{\sigma}{r}\right)^9 - 3\left(\frac{\sigma}{r}\right)^6\right)$$

donde σ es la distancia media entre átomos vecinos a temperatura muy baja.

Obtenga los niveles propios de energía y las correspondientes funciones radiales para los casos $\ell = 0$ y $\ell = 1$. Use $2m = 1$, $\sigma = 1$, $\hbar = 1$, $V_0 = 23{,}0$.

Capítulo 6

Integrales Monte Carlo y el algoritmo de Metropolis

La pregunta más persistente en este capítulo es cómo generar una secuencia de números que tengan una distribución W conocida.

6.1. Números aleatorios $r \leftarrow U(0,1)$

Desde los primeros tiempos en que se abordó problemas numéricos con los más primitivos computadares se vió la importancia de los números aleatorios. Hay problemas que, en el contexto de métodos numéricos, requieren de algún algoritmo para generar una secuencia de números (enteros o reales), $\{x_0, x_1, x_2, \ldots\}$ que pueda ser considerada aleatoria. El concepto de aleatoriedad, sin embargo, es sutil y en muchas de sus aplicaciones interesa poder decidir *cuán aleatorios* realmente es la secuencia que se construye. No existen medidores universales de aleatoriedad sino ciertos criterios reconocidos como condiciones necesarias que las secuencias deben satisfacer. Existe una amplia literatura sobre el tema. Para fines introductorios basta con estudiar los criterios que se encuentran en *Numerical Recipies*. En lo que sigue no nos ocuparemos de estos problemas y aceptaremos el generador de números aleatorios que nos da el compilador que estemos usando.

Se supondrá que se cuenta con un generador de números aleatorios r uniformemente distribuidos en el intervalo $(0,1)$. Para no tener que usar la expresión *números r uniformemente distribuidos* en el intervalo (a,b) se dirá

$$r \leftarrow U(a,b) \tag{6.1.1}$$

Si se tiene un generador $r \leftarrow U(0,1)$ se puede definir una variable

$$r' = (b-a)\,r + a \tag{6.1.2}$$

la que satisface $r' \leftarrow U(a,b)$. En C la función que usaremos es `drand48()`.

6.2. Densidades de probabilidad

6.2.1. Distribución y el promedio discreto

Las secuencias $\{x_j\}$ de números aleatorios reales usuales son las que están uniformemente distribuidas en el intervalo $(0,1)$, pero es de mucho interés saber generar secuencias que están distribuidas de acuerdo a una distribución escogida $W(x)$ en (a,b).

Se define una densidad de probabilidad $W(x)$ en (a,b) con una *función no negativa* en ese intervalo que está normalizada,

$$\int_a^b W(x)\,dx = 1 \tag{6.2.1}$$

Con ella se define el promedio de cualquier función $f(x)$ con $x \in (a,b)$,

$$\langle f\rangle_W = \int_a^b W(x)\,f(x)\,dx \tag{6.2.2}$$

La variancia σ_f^2 de h es

$$\sigma_f^2 = \left\langle f^2\right\rangle_W - \langle f\rangle_W^2 \tag{6.2.3}$$

y a σ_f, que se llama la *desviación estándar* de f, es una medida de cuánto se desvía f de su promedio.

Para calcular numéricamente (6.2.2) se debe utilizar una secuencia de números $\{x_i\}_{i=1..N}$ en el intervalo (a,b) que tengan distribución $W(x)$ y utilizar la forma aproximada

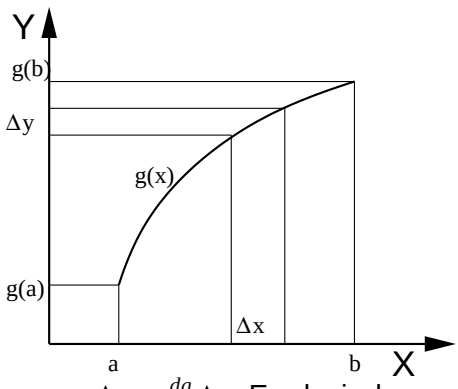
$$\boxed{\langle f\rangle_W \approx \frac{1}{N}\sum_{j=1}^{N} f(x_j) \qquad \text{con} \qquad x_j \leftarrow W} \tag{6.2.4}$$

Pero, ¿cómo generar la secuencia $x_j \leftarrow W$?

6.2.2. Distribuciones relacionadas

Supongamos que se tiene una variable aleatoria x en el intervalo (a,b) generada con distribución $W(x)$ y además sea $y = g(x)$ una relación monótona:

$$y = g(x)\,, \qquad dy = g'(x)\,dx \tag{6.2.5}$$



Es decir, se asocia a cada x_j de una secuencia con distribución W un valor $y_j = g(x_j)$ en el intervalo $(g(a),g(b))$. ¿Qué distribución se debe asociar a los y_j?

Puesto que la probabilidad asociada a un intervalo Δx en torno al punto x es $P = W(x)\Delta x$, la misma probabilidad tiene asociado el correspondiente intervalo Δy en torno a $y = g(x)$, siempre que $\Delta y = \frac{dg}{dx}\Delta x$. Es decir, la

probabilidad asociada a este intervalo Δy es $\widetilde{W}(y)\,\Delta y = \widetilde{W}(y)\,g'(x)\,\Delta x = W(x)\,\Delta x$. De la última igualdad se desprende que los y_j se distribuyen según

$$\widetilde{W}(y) = \left[\frac{W(x)}{g'(x)}\right]_{x=g^{-1}(y)} \tag{6.2.6}$$

> Si la variable $x = g^{-1}(y)$ tiene distribución $W(x)$ en el intervalo (a,b), la variable y tiene distribución $\widetilde{W}(y) = [W(x)/g'(x)]_{x=g^{-1}(y)}$ en el intervalo $[g(a),g(b)]$.

Es automático que $\widetilde{W}$ tiene la normalización correcta. En efecto, la relación $\int_a^b W\,dx = 1$ implica

$$\int_{g(a)}^{g(b)} \widetilde{W}(y)\,dy = 1$$

6.2.3. Obtención de secuencia $W(x)$ a partir de $U(0,1)$

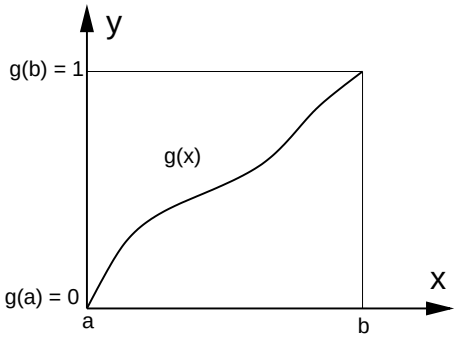
En particular, si los valores de y son generados por una distribución uniforme, $U(0,1)$, $y \leftarrow U(0,1)$, de modo que $g(a) = 0$, $g(b) = 1$ entonces $\widetilde{W}(y) = 1$ y los $x_j = g^{-1}(y_j)$ están distribuidos de acuerdo a

$$W(x) = g'(x) \tag{6.2.7}$$

en el intervalo $[a = g^{-1}(0),\ b = g^{-1}(1)]$. Esta última relación implica que si se conoce $W(x)$, la función $g(x)$ definida por

$$g(x) = \int_a^x W(x')\,dx' \tag{6.2.8}$$

permite definir el cambio de variable $y = g(x)$ tal que $y \in [0,1]$.



Si $y \leftarrow U(0,1)$ entonces los $x = g^{-1}(y)$ se distribuyen de acuerdo a $W(x)$ donde W y g se relacionan por (6.2.8).

Este método es de uso limitado porque para poder aplicarlo se requiere tanto poder calcular $g(x) = \int_a^x W(x')\,dx'$ como poder invertir la función g, para obtener $x_j = g^{-1}(y_j)$.

≫ Por ejemplo, si se escoge la distribucion de velocidades

$$W(c) = \frac{2c}{T}\,e^{-c^2/T} \qquad \text{con} \qquad 0 \le c < \infty \tag{6.2.9}$$

se obtiene

$$y = g(c) = \int_0^c W(c')\,dc' = 1 - e^{-c^2/T}$$

o equivalentemente

$$c = g^{-1}(y) = \sqrt{-T\,\ln(1-y)}$$

Puesto que si $y \leftarrow U(0,1)$ entonces $1-y \leftarrow U(0,1)$ y se puede usar y en lugar de $1-y$,

$$c = \sqrt{-T\,\ln y}$$

es decir, dada una secuencia de números $y_j \leftarrow U(0,1)$, la secuencia $c_j = \sqrt{-T\,\ln y_j}$ tiene la distribución (6.2.9).

6.2.3.1. El histograma asociado a un $W(x)$

Se tiene una secuencia x_j con distribución de probabilidad $W(x)$ en el intervalo $a \leq x \leq b$. Se divide este intervalo en M pequeños intervalos iguales de largo $\Delta = \frac{b-a}{M}$. Si la secuencia es de largo N suficientemente grande, el intervalo k-ésimo debiera contener H_k puntos de la secuencia con

$$H_k \approx \frac{(b-a)N}{M} w_k, \qquad k = 1, \ldots M$$

donde w_k es el valor de $W(x)$ en el punto medio del k-ésimo intervalo.

Si la igualdad anterior se divide por N y se suma sobre k, el lado izquierdo arroja necesariamente 1, muentras que el lado derecho arroja el valor de el valor de la integral $\int W(x)\,dx$ en la aproximación trapezoidal (con error $\mathcal{O}(M^{-2})$.

6.2.4. El caso de n variables

El razonamiento recién descrito puede ser generalizado a n variables. Se tiene un n-uplo de variables aleatorias, $X = \{x_1, \ldots, x_n\}$ en un dominio $\mathcal{D}_X$ con distribución $W(X)$. Además se tiene una función invertible, $Y = G(X)$, donde las nuevas variables Y están, en el dominio $\mathcal{D}_Y = G(\mathcal{D}_X)$. La transformación G induce una distribución $\widetilde{W}(Y)$ para las variables Y y ambas distribuciones se relacionan por

$$\widetilde{W}(Y) = \left[\frac{W(X)}{J(Y;X)}\right]_{X=G^{-1}(Y)}$$

donde $J(Y;X)$ es el Jacobiano $J = \det[\partial G(X)/\partial X]$. Esto es así porque

$$\widetilde{W}(Y)\,dY = \widetilde{W}(Y)\,J(Y;X)\,dX = W(X)\,dX$$

De especial interés es, al igual que en la subsección anterior, el caso en que los Y tienen distribución uniforme y el dominio $\mathcal{D}_Y$ es el hipercubo unitario, es decir $0 \leq y_a \leq 1$ para todo a. En tal caso $\widetilde{W}(Y) = 1$ en este hipercubo. Dada una función $G(X)$ se define secuencia de n-uplos $X = G^{-1}(Y)$ los cuales tienen una distribución

$$W(X) = J(Y;X) \tag{6.2.10}$$

Un ejemplo de esto se da a continuación.

6.2.5. Uso de $W(x_1, x_2)$ para generar gaussianas

El método visto en §6.2.3 no se puede aplicar en forma directa cuando se quiere generar una secuencia $\{x_j\}$ según la distribución gaussiana

$$W_\sigma(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-(x-a)^2/2\sigma^2}, \qquad -\infty \leq x \leq \infty$$

La distribución está trivialmente relacionada con la distribución centrada en el origen y con variancia unidad:

$$W(x) = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2} \tag{6.2.11}$$

que tampoco se puede resolver en forma directa porque la integral $\int_{-\infty}^{x} W(x')\,dx'$ no puede ser escrita con funciones elementales.

Sin embargo, como veremos, es fácil generar pares (x_1,x_2) con la distribución

$$W(x_1,x_2) = \frac{2}{\pi}\, e^{-(x_1^2+x_2^2)/2}, \qquad 0 \le x_i < \infty \quad \text{con} \quad i = 1,2 \tag{6.2.12}$$

que está correctamente normalizada,

$$\int_0^\infty \int_0^\infty W(x_1,x_2)\; dx_1\; dx_2 = 1$$

En efecto, si se hace el cambio de variable

$$\begin{aligned} y_1 &= \frac{2}{\pi}\arctan\frac{x_2}{x_1} & x_1 &= \sqrt{-2\ln y_2}\cos\frac{\pi y_1}{2} \\ y_2 &= e^{-(x_1^2+x_2^2)/2} & x_2 &= \sqrt{-2\ln y_2}\sin\frac{\pi y_1}{2} \end{aligned} \tag{6.2.13}$$

se obtiene que el Jacobiano de la transformación es precisamente

$$J(Y;X) = W(x_1,x_2) = \frac{2}{\pi}\, e^{-(x_1^2+x_2^2)/2} \tag{6.2.14}$$

Basta con generar $y_1 \leftarrow U(0,1)$ y también $y_2 \leftarrow U(0,1)$ para que (6.2.13) genere una secuencia de puntos en el plano (x_1,x_2) con la distribución (6.2.12).

6.3. Integración Monte Carlo

6.3.1. El problema

Se discutirá la forma de calcular

$$I = \int_a^b f(x)\,dx$$

utilizando secuencias de números aleatorios.

6.3.2. Primera forma

Sea $\{x_j\}_{j=1..N}$ una *muestra aleatoria* de valores de la variable aleatoria x con distribución W en (a,b). Si en (6.2.2) se toma $W(x) = \text{cte} = \frac{1}{b-a}$ y f es la función a promediar, entonces

$$\langle f\rangle = \frac{1}{b-a}\int_a^b f(x)\,dx$$

que, por (6.2.4) es $\frac{1}{N}\sum f_j$. Multiplicando por $(b-a)$ se obtiene que si x está uniformemente distribuida en (a,b),

$$I \equiv \int_a^b f(x)\,dx \approx \frac{b-a}{N}\sum_{j=1}^{N} f_j \quad \pm \quad \frac{b-a}{\sqrt{N}}\underbrace{\sqrt{\frac{1}{N}\sum_{j=1}^{N} f_j^2 - \left(\frac{1}{N}\sum_{j=1}^{N} f_j\right)^2}}_{\sigma_f} \tag{6.3.1}$$

donde los $x_i \leftarrow U(a,b)$ y $f_j = f(x_j)$.

La incertidumbre con que se evalua la integral depende tanto de la desviación estándar intrínseca de f en este intervalo, σ_f, como del tamaño N de la muestra. Nótese que es necesario cuadruplicar el valor de N para disminuir la incertidumbre a la mitad. Esto contrasta con la regla trapeziodal que es $\mathcal{O}(N^{-2})$. Sin embargo, como se verá en §6.3.3, en el cálculo de integrales sobre muchas variables, el método Monte Carlo es el más eficiente.

En resumen, la primera forma de integración Monte Carlo es

$$\int_a^b f(x)\,dx = \frac{b-a}{N}\sum_{j=1}^{N} f(x_j) + \mathcal{O}(\frac{\sigma_f}{\sqrt{N}}), \qquad x_j \leftarrow U(a,b) \qquad \textbf{MC1} \tag{6.3.2}$$

Este método da, por lo general, resultados más bien pobres, salvo que se use muestras de tamaño N muy grande o bien f varíe poco en el intervalo. Un punto a favor es que se usa un solo `drand48` y no hay ningún `if`.

```
n    = 0;
suma = 0.0;
do {  x     = a + (b-a)*drand48();
      suma += f(x);
      n++;
   }while(n<=N}
integral = (b-a)*suma/N;
```

Si $f = f_0$ es constante la integral es exactamente $(b-a)\,f_0$, y el lado derecho vale $\frac{(b-a)\,f_0}{N}N = (b-a)\,f_0$ para todo N. La forma (6.3.2) de calcular una integral es la más ingenua forma de integrar usando números aleatorios dentro de aquellas que pertenecen a la categoría *integración Monte Carlo*.

Por ejemplo el cálculo de $\int_0^{\pi}\sin(x)\,dx$:

```
res  = 0.0;
for(n=0; n<N; n++)
{ x    = PI*drand48();
  res += sin(x);
  if(n%50==1)  // cada 50 pasos escriba el resultado
  { intgr = PI*res/(1.0*n);
    printf(" %10d %12.8f\n",n,intgr);
  }
}
```

Como se acaba de comentar, si la función es constante el resultado que arroja MC1, (6.3.2), es exacto (en particular no depende de N) y si f varía muy poco este método da valores razonables. Pero en general se requiere hacer uso de formas más elaboradas de integración Monte Carlo, las que usan valores x_j que provienen de una distribución W escogida especialmente.

La otra obvia limitación del método recién descrito es que es aplicable tan solo si el dominio es acotado. Con dominios infinitos o con integrandos de alto contraste se procede siguiendo un camino emparentado al que se vió en §2.3

El algoritmo MC1 es generalizable a muchas dimensiones con la siguiente dificultad: si el dominio de integración se define como relaciones entre las variables (por ejemplo $x^2+y^2<1$ & $x-y>0$) se debe encontrar una forma económica de generar puntos uniformemente distribuidos sobre ese dominio y nada más. Normalmente la única solución razonable consiste en generar puntos uniformemente distribuidos en un dominio más grande pero sencillo (por ejemplo $-1\le x\le 1$ & $-1\le y\le 1$) y hacer uso tan solo de los puntos que caen dentro del dominio de interés.

Más adelante será útil tener presente que la propia fórmula MC1 sugiere que una forma de calcular $\langle f\rangle_W$ es

$$\langle f\rangle_W = \frac{b-a}{N}\sum_{j=1}^{N} W(x_j)\,f(x_j) \qquad \text{con} \qquad x_j \leftarrow U(a,b) \tag{6.3.3}$$

ya que la anterior no es sino MC1 donde se ha puesto $W\,f$ en lugar de f. Esta es una primera manera de resolver el problema que plantea (6.2.4).

Ya se ha comentado que si el integrando varía fuertemente en el dominio de integración este algoritmo puede dar resultados muy pobres. Una forma sencilla que—suele dar buenos resultados sin abandonar lo básico de MC1—consiste en encontrar una función integrable $\widetilde{f}(x)$ que sea parecida a $f(x)$. Es decir, si se conoce el valor $\mathbf{I}=\int_a^b \widetilde{f}(x)\,dx$ y además se cumple que $\overline{f}(x)\equiv f(x)-\widetilde{f}(x)$ es una función suficientemente plana en el dominio (de modo que MC1 en ella es satisfactoria) la integral se puede calcular usando

$$\int_a^b f(x)\,dx \approx \mathbf{I} + \underbrace{\int_a^b \overline{f}(x)\,dx}_{\text{con MC1}} \qquad \textbf{MC1b} \tag{6.3.4}$$

6.3.3. Aplicabilidad de los métodos Monte Carlo

Si se quiere calcular una integral con métodos tradicionales (trapezoidal, Simpson etc) en dimensón D tomando un total de N puntos, los intervalos en cada dimensión deben ser subdivididos en $N^{1/D}$ intervalos de tamaño $h\propto N^{1/D}$. La integral en cada dimensión arroja un error de order $\mathscr{O}(h^\nu)$ ($\nu=2$ con el método trapezoidal y $\nu=3$ con el método Simpson) y este mismo es el orden del error de la integral sobre todas las dimensiones: $\mathscr{O}(h^\nu)=\mathscr{O}(N^{-\nu/D})$. Si se desea calcular integrales en muchas dimensiones (por ejemplo, $D=10$) el error es bastante significativo salvo que N sea muy grande.

En cambio una integral Monte Carlo siempre tiene un error $\mathscr{O}(N^{-1/2})$ que normalmente resulta más conveniente cuando la dimensión es algo mayor que 4.

6.3.4. Método explícito

Se desea calcular $\int_a^b F(x)\,dx$ pero la variación de la función en este intervalo es muy grande o el intervalo es infinito o ambas cosas a la vez. En tal caso se debe seguir el siguiente método que se basa en escoger una distribución $w(x)$ tal que $f(x) = F(x)/w(x)$,

$$I = \int_a^b F(x)\,dx = \int_a^b w(x)\,f(x)\,dx \tag{6.3.5}$$

donde $w(x)$ debe satisfacer

$$w(x) \geq 0 \qquad \text{en} \quad (a,b) \qquad \text{y} \qquad \int_a^b w(x)\,dx = 1 \tag{6.3.6}$$

Haciendo el cambio de variable

$$y = g(x) = \int_a^x w(x')\,dx' \qquad \Rightarrow \qquad g(a) = 0\,, \quad g(b) = 1$$

se obtiene que

$$dy = g'(x)\,dx = w(x)\,dx \qquad \Rightarrow \qquad dx = \frac{dy}{w(x)}$$

que permite escribir

$$I = \int_0^1 f\left(g^{-1}(y)\right)\,dy = \int_0^1 \widetilde{f}(y)\,dy \approx \frac{1}{N}\sum_j \widetilde{f}(y_j) \qquad y_j \leftarrow U(0,1) \tag{6.3.7}$$

pero $\widetilde{f}(y_j) = f(g^{-1}(y_j)) = \left[\frac{F(x_j)}{w(x_j)}\right]_{x_j = g^{-1}(y_j)}$ por lo cual

$$I \approx \frac{1}{N}\sum_{j=1}^{N}\left[\frac{F(x_j)}{w(x_j)}\right]_{x_j = g^{-1}(y_j)} \qquad \text{con} \quad y_j \leftarrow U(0,1) \qquad \textbf{MC2} \tag{6.3.8}$$

En lugar de usar $x \leftarrow U(0,1)$ como en **MC1**, se usa una secuencia sesgada $x_j = g^{-1}(y_j)$ para sumar valores de la función $f = F/w$.

Este método es exitoso si la desviación estándar de los valores $\widetilde{f}(y_j) = [f(x_j)]_{x_j = g^{-1}(y_j)}$ es pequeña. Lo ideal sería escoger w proporcional a F, de tal modo que f sea tan solo una constante. Esto en general no es posible si se desea satisfacer las condiciones de hacer analíticamente la integral $g(x) = \int^x w\,dx$ y de conocer la función inversa g^{-1}.

En resumen

$$\begin{aligned} I &= \int_a^b F(x)\,dx \qquad \boxed{\textbf{MC2 explicíto}} \\ &= \int_0^1 \left[\frac{F(x)}{w(x)}\right]_{x=g^{-1}(y)} dy \\ &\approx \frac{1}{N}\sum_j \left[\frac{F(x_j)}{w(x_j)}\right]_{x_j=g^{-1}(y_j)} \qquad \text{con} \quad y_j \leftarrow U(0,1) \end{aligned} \tag{6.3.9}$$

donde $g=\int_0^x w(x')\,dx'$ y w se escoje para que la función $\widetilde{f}$ sea de poco contraste en el intervalo $[0,1]$

$$\widetilde{f}(y)=\left[\frac{F(x)}{w(x)}\right]_{x=g^{-1}(y)}, \quad w(a\le x\le b)\ge 0, \quad \int_a^b w(x)\,dx=1$$

El error en el caso MC2 continua siendo $\mathcal{O}(\sigma_f/N^{1/2})$. La diferencia está en que con MC2 se puede lograr una desviación estándar mucho menor.

Todo lo que se ha dicho en esta sección debe ser entendido aplicable al caso multidimensional.

$\gg$ Ejemplo. Se plantea calcular numéricamente la integral $I=\int_0^\infty \frac{2}{\sqrt{\pi}}e^{-x^2}\,dx$ que se sabe que vale 1 y se escoge $w(x)=\frac{1}{a}e^{-x/a}$. Con este w se obtiene que

$$y=g(x)=\int_0^x \frac{1}{a}e^{-x'/a}\,dx'=1-e^{-x/a}$$

relación que se puede invertir a

$$x=-a\ln(1-y)$$

Es decir, si la secuencia y_j proviene de $U(0,1)$ entonces $x_j=-a\ln(1-y)$ se distribuye de acuerdo a $w(x)=\frac{1}{a}e^{-x/a}$. Pero los y_j y los $1-y_j$ tienen la misma distribución lo que permite escribir

$$f(x)=\frac{2a}{\sqrt{\pi}}e^{-x^2+x/a}, \qquad \widetilde{f}(y)=\frac{2a}{\sqrt{\pi}}y^{-(1+a^2\ln y)}$$

y la integral I en la forma aproximada

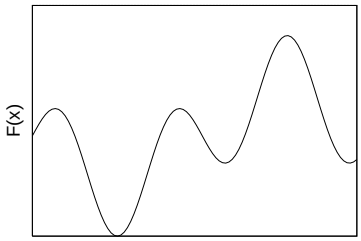
$$I\approx\frac{2a}{N\sqrt{\pi}}\sum_{j=1}^N\left[e^{-x^2+x/a}\right]_{x_j=-a\ln(y_j)}=\frac{2a}{N\sqrt{\pi}}\sum_{j=1}^N e^{-a^2\ln^2(y_j)-\ln(y_j)}, \qquad y_j\leftarrow U(0,1)$$

Un código posible para ejecutar la estrategia anterior, tomando $a=1$, es

```
res = 0.0;
for(n=0; n<N; n++)
{ y   = drand48();
  x   = -log(y);
  res+= exp(-x*x + x);
}
final = 2.0*res/(sqrt(PI)*N);
```

6.3.5. Estrategia de von Neumann

Consideremos una función $F(x)$ no negativa en el intervalo (a,b) y sea F_0 un valor mayor o igual al mayor valor de $F(x)$ en el intervalo. En la figura adjunta se ha dibujado la curva $F(x)$ en una caja de altura F_0 y base $(b-a)$. El área de la caja es $(b-a)F_0$ y el área bajo la curva es $I=\int F\,dx$. Si se lanza puntos al azar con distribución uniforme en la caja, la probabilidad que caigan bajo la curva $F(x)$ es $p=\frac{1}{(b-a)F_0}\int F\,dx$ y, computacionalemnte $p\approx n_1/n$ donde n es el total de puntos lanzados y n_1 son los que cayeron bajo la curva, lo que permite concluir que



$$I=\int_a^b F(x)\,dx\approx\frac{n_1}{n}(b-a)F_0 \tag{6.3.10}$$

Para hacer integrales de funciones que cambian de signo se debe integrar separadamente cada tramo donde no haya cambio de signo. Este método resulta muy pobre si $F(x)$ es de alto contraste. En general primero se debe hacer un cambio de variable.

Una rutina que ejecuta el cálculo de $\int F$ es

```
n    = 0;
n1   = 0;
do
{  x   = a + (b-a)*drand48();
   y   = F0*drand48();
   Fx  = F(x);
   n++;
   if(y<Fx)  n1++;
}while(n<=N}
integral = (b-a)*n1*F0/N;
```

Se genera un $x\leftarrow U(a,b)$ y un $y\leftarrow U(0,F_0)$. Con n se cuenta el total de puntos mientras que n_1 cuenta los puntos que caen bajo la curva $F(x)$. La secuancia que resulta tiene asociada una distribución $W(x)=\frac{F(x)}{(b-a)F_0}$.

Pruebe el método calculando $\int_0^\pi \sin x\,dx$.

En el método anterior n_1 cuenta los puntos (x,y) que satisfacen $y\le F(x)$. Es claro que la probabilidad de que un punto (x,y) sea aceptado es proporcional a la función $F(x)$. El costo de cada nuevo punto es llamar dos veces a la función `drand48()`, una vez a $F(x)$ y además hay un `if`. Compárese con MC1.

6.3.6. Integración Monte Carlo en dimensión D

La integración Monte Carlo en dimensión D toma las formas ya vistas. La fórmula básica es la generalización trivial de (6.3.1),

$$\int f(\vec{r})\,dV = V\,\langle f\rangle_{U(V)}\pm V\sqrt{\frac{\langle f^2\rangle-\langle f\rangle^2}{N}} \tag{6.3.11}$$

donde todos los promedios de arriba se refieren a los que se obtiene de

$$\langle A\rangle=\frac{1}{N}\sum_{i=1}^{N}A(\vec{r}_i)\qquad\qquad \vec{r}_i\leftarrow U(V) \tag{6.3.12}$$

y $U(V)$ designa un generador de puntos aleatorios y uniformemente distribuidos en el volumen V de integración. La integral se puede calcular usando una extención directa de MC1, (6.3.2),

$$\int f(\vec{r})\,dV = \frac{V}{N}\sum_{i=1}^{N} f(\vec{r}_i)\,, \qquad \vec{r}_i \leftarrow U(V) \tag{6.3.13}$$

Sin embargo el dominio sobre el cual se quiere hacer la integral puede ser suficientemente complejo para que no valga la pena calcular su volumen. En tal caso se define un volumen sencillo V_s que contiene a V y se genera puntos uniformemente distribuidos en V_s y se suma sólo los f_i que corresponden a puntos en el interior de V. En tal caso la integral se calcula como

$$\int f(\vec{r})\,dV = \frac{V_s}{N}\sum_{i=1}^{N} f(\vec{r}_i)\,, \qquad \vec{r}_i \leftarrow U(V_s) \tag{6.3.14}$$

y N son todos los puntos generados dentro de V_s. Si f fuese una constante $f(x) = f_0$ entonces el resultado anterior sería $I = \frac{n_1}{N} V_s\, f_0$ donde n_1 son los puntos que caen dentro de V, pero $n_1/N \approx V/V_s$ y entonces se obtiene $I \approx V\, f_0$ que es lo que debe ser.

Suele tener que hacerse cambios de variable (por ejemplo si el dominio es infinito o la función tiene variancia grande). El cambio que se haga tiene que ser explícito.

6.4. La estrategia Metropolis para calcular promedios

6.4.1. El algoritmo de Metropolis

El algoritmo de Metropolis fue concebido originalmente en el contexto de Mecánica Estadística para calcular promedios asociados a sistemas estadísticos en equilibrio. Esta estrategia sin embargo es de gran generalidad y primero será presentada sin hacer uso de los detalles que Mecánica Estadística require. Luego se verá cómo se aplica, en particular, en esa área.

Reducido a su esencia, el algoritmo representa una estrategia para generar una secuencia $X \leftarrow W(X)$ para cualquier distribución de probabilidad W en un un espacio de *puntos* X sobre los que se pueda asociar una distribución de probabilidad $W(X)$. En un sentido muy poderoso, Metropolis resuelve el problema planteado en (6.2.4) con un código eficiente y breve. Un promedio, como señala (6.2.4), se logra con el simple promedio aritmético de los valores generados.

Primero se va a definir el algoritmo de Metropolis y luego se va a argumentar que tiene las propiedades deseadas.

El algoritmo. Se escoge una semilla X_0 y se entra en el ciclo que sigue donde, a partir del último punto X_n, se define un nuevo punto X_p el cual puede ser aceptado o rechado. El esquema es el siguiente:

a) Se usa alguna regla, cuyas propiedasdes se discuten más adelante, para generar, a partir de X_n, un X_p (el subíndice p es por "prueba"). Esta regla debe ser simétrica, en el sentido que sea igualmente problable generar X_p si se proviene de X_n que vice versa.

b) Se usa un criterio para aceptar o rechazar el punto de prueba X_p.

c) Si se acepta entonces $X_{n+1} = X_p$ y si se rechaza no se hace otro intento, sino que entonces $X_{n+1} = X_n$.

El algoritmo de Metropolis consiste en definir la acción conjunta (b) y (c) tomando un $r \leftarrow U(0,1)$ y aplicar

$$\boxed{\text{if} \quad \frac{W(X_p)}{W(X_n)} > r \quad \text{then} \quad X_{n+1} = X_p \quad \text{else} \quad X_{n+1} = X_n} \tag{6.4.1}$$

Si $W_p > W_n$ entonces W_p/W_n es mayor que cualquier r y el nuevo X_p es aceptado. Es decir, X_p es aceptado incondicionalmente si es más probable que X_n. Si, por el contrario, X_p es menos probable que X_n solo a veces X_p es aceptado. Se subraya que con este método la secuencia $\{X_n\}$ tiene puntos repetidos por cada vez que un X_p no es aceptado. Esto debe ser así por razones estadísticas.

> Dado un número x en el intervalo $(0,1)$, ¿cuál es la probabilidad $P(x)$ de que un número aleatorio r en $(0,1)$ satisfaga $x > r$? La respuesta es $P(x) = x$. El algoritmo de Metropolis hace uso de esto.

Un problema que se presenta en algoritmos como estos es que X_j y X_{j+1} no sean independientes. Por la forma como se generan los puntos, ellos normalmente son cercanos (en algún sentido) lo que implica una *correlación* entre ellos. Una forma de resolver esto consiste en no utilizar todos los X_j generados, sino que se toma en cuenta sólo uno de cada v de ellos. El valor de v depende del problema que se esté resolviendo.

En muchos casos interesantes en física la probabilidad W es proporcional a la exponencial $e^{-E/kT}$: menos la energía dividida por kT (k es la constante de Boltzmann y T la temperatura en grados Kelvin). Si la temparatura se mide en unidades de energía se puede sustituir kT por T y $W \propto e^{-E/T}$.

En tales casos la condición $\frac{W(X_p)}{W(X_n)} > r$ es idéntica a $e^{(E_n - E_p)/T} > r$ y a su vez esta condición es idéntica a

$$E_n - E_p > T \ln r \tag{6.4.2}$$

que, computacionalmente, puede ser más rápida y manejable.

6.4.2. Por qué funciona

Supongamos que se han generado N secuencias de largo n a partir de N semillas $X_j^{(i)}$ con $i = 1...N$ y $j = 1...n$.

Sea $D_n(X)$ la densidad de probabilidad de presencia de los puntos $X_n^{(i)}$ en la vecindad de X y $D_n(Y)$ es la densidad en torno a Y. Y es el punto, antes llamado X_p, que será aceptado o rechazado. Al iterar una vez más hay un cambio en D_n que se debe a traspaso neto de puntos de

la vecindad de X a la vecindad de todos los puntos Y. Este cambio que sufre $D_n(X)$ formalmente es $\delta D_n(X) = D_{n+1}(X) - D_n(X)$ y se puede expresar como

$$\begin{aligned}\delta D_n(X) &= \underbrace{\sum_Y D_n(Y)\,P(Y\to X)}_{\text{ganancia}} - \underbrace{\sum_Y D_n(X)\,P(X\to Y)}_{\text{pérdida}} \\ &= \sum_Y D_n(Y)\,P(X\to Y)\left[\frac{P(Y\to X)}{P(X\to Y)} - \frac{D_n(X)}{D_n(Y)}\right] \qquad (6.4.3)\end{aligned}$$

donde $P(X\to Y)$ es la probabilidad de que del punto X se pase al punto Y. Asintóticamente se alcanza un estado de *equilibrio* en el sentido que $D(X)$ ya no evoluciona más, esto es, δD se anula. La solución general que anula el lado derecho es muy difícil de encontrar. No demostraremos que la solución requiere que el corchete sea nulo para todos los pares (X,Y).

• Se dice que se alcanza la densidad de equilibrio $D_{\rm eq}$ cuando $D(X)$ deja de evolucionar en todos los puntos X y se satisface

$$\frac{D_{\rm eq}(X)}{D_{\rm eq}(Y)} = \frac{P(Y\to X)}{P(X\to Y)} \qquad (6.4.4)$$

es decir, $\delta D = 0$, lo que hace a $D_{\rm eq}$ un punto fijo.

• Si D está muy cerca de $D_{\rm eq}$ y ocurre que

$$\frac{D_n(X)}{D_n(Y)} > \frac{P(Y\to X)}{P(X\to Y)} \qquad (6.4.5)$$

el factor en el corchete en (6.4.3) es negativo y hay traspaso neto hacia Y, lo que acerca a $D(X)$ al equilibrio. Las dos últimas propiedades muestran que el mecanismo anterior conduce a un equilibrio estable. El punto fijo es un punto de equilibrio estable.

Si se escribe

$$P(X\to Y) = p_{XY}\,A_{XY} \qquad (6.4.6)$$

donde p_{XY} es la probabilidad de que si la semilla es X se intente Y, y A_{XY} es la probabilidad de que (6.4.1) acepte a Y si la semilla es X. Por definición de la estrategia de Metropolis p_{XY} es simétrica: $p_{XY} = p_{YX}$. Entonces (6.4.4) se puede escribir

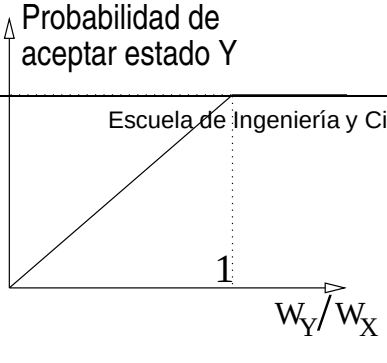
$$\frac{D_{\rm eq}(X)}{D_{\rm eq}(Y)} = \frac{A_{YX}}{A_{XY}} \qquad (6.4.7)$$

Ahora analicemos (6.4.1) a la luz de (6.4.7). Teniendo X e Y fijos, y X como semilla, se sortea un $r \leftarrow U(0,1)$.

	$W_X < W_Y$	$W_Y < W_X$
A_{XY}	1	W_Y/W_X
A_{YX}	W_X/W_Y	1

Valores de A en los distintos casos.

- Si $W_X < W_Y$, se ve que $\frac{W_Y}{W_X} > 1$ es mayor que cualquier r y $X_{n+1} = Y$. Esto se traduce en el 1 del primer casillero de la tabla.
- Si $W_X > W_Y$, es decir $\frac{W_Y}{W_X} < 1$, la probabilidad de aceptar a Y es $\frac{W_Y}{W_X}$, que es el valor superior derecho de la tabla.



En forma similar se obtiene los valores que están abajo en la tabla para el caso en que Y es la semilla y X es el que se acepta o rechaza. Se comprueba inmediatamente que el valor del cuociente A_{YX}/A_{XY} es siempre el mismo y es W_X/W_Y, es decir,

$$\frac{D_{\rm eq}(X)}{D_{\rm eq}(Y)} = \frac{W_X}{W_Y} \tag{6.4.8}$$

y usando los casilleros de la derecha se obtiene lo mismo, lo que muestra que $\boxed{D_{\rm eq} = W}$ y por tanto queda aclarado que (6.4.1) finalmente conduce a $W(X)$.

En el análisis anterior se hizo uso de la ya cidada propiedad elemental: dado un x en el intervalo $(0,1)$, la probabilidad de que un $r \leftarrow U(0,1)$ esté entre 0 y x es exactamente $P = x$.

6.4.3. Metropolis en mecánica estadística

Es muy típico considerar un sistema de $\mathscr{N}$ partículas cuyo hamiltoniano es de la forma $H = \sum_a \frac{p_a^2}{2m} + \sum_{a<b} V_{ab}$ y querer calcular, por ejemplo, un promedio estadístico canónico de una cantidad que solo depende de las posiciones de las partículas, $A(\vec{r}_a)$ como

$$\langle A \rangle = \frac{\sum A(\vec{r}^{\,\mathscr{N}})\, e^{-V/(kT)}}{\sum e^{-V/(kT)}} \tag{6.4.9}$$

donde $V = \sum_{a<b} V_{ab}$ quiere decir la energía potencial total del sistema y la suma (integral) es sobre los estados configuracionales del sistema. Normalmente lo anterior es una integral $d^{\mathscr{N}}\vec{r}$ sobre todas las posiciones posibles de las partículas del sistema. En la práctica la cantidad de estados es un continuo que no puede ser integrado o es un discreto gigantezco, de modo que lo que se hace es un muestreo del espacio de estados, tal como se hace en la integral Monte Carlo. El método de Metropolis genera estados (configuracionales) directamente con la distribución W requerida, por ejemplo,

$$W = \frac{e^{-\beta V}}{\sum e^{-\beta V}} \tag{6.4.10}$$

Cuando, por ejemplo, se calcula en mecánica estadística el promedio de una cantidad que depende tan solo de las coordenadas de las partículas del sistemas (por ejemplo de un líquido), se utiliza una distribución W como en (6.4.10). Más en detalle, si se tiene un potencial interpartícula, $V_{ab} = V(r_{ab})$, el exponente simbolizado como $\beta \sum V$ es

$$\beta \mathbf{V} = \beta \sum_{a<b} V_{ab} \tag{6.4.11}$$

Esta es una suma sobre todos los pares posibles de partículas del sistema. Si el sistema tiene N partículas, la suma tiene $\mathscr{O}(N^2)$ sumandos. Para sistemas medianamente grandes tal suma sería un inconveniente prohibitivo. Pero lo que interesa es el cuociente W_p/W_n, es decir, interesa calcular $\Delta\mathbf{V} = \mathbf{V}_n - \mathbf{V}_p$.

Si en la iteración $X_n \to X_p$ se cambia las coordenadas de una sola partícula—la partícula k-esima—$\vec{r}_k \to \vec{r}\,'_k$ entonces en la diferencia (6.4.11) la mayoría de los términos se cancela idénticamente y queda tan solo aquellos que involucran a la partícula k,

$$\Delta \mathbf{V} = \sum_a \left[V(\vec{r}_a - \vec{r}\,'_k) - V(\vec{r}_a - \vec{r}_k) \right] \tag{6.4.12}$$

Esto hace que ahora la suma tenga $\mathscr{O}(N)$ sumandos. Aun esto es demasiado cuando se desea hacer cálculos sobre sistemas muy grandes.

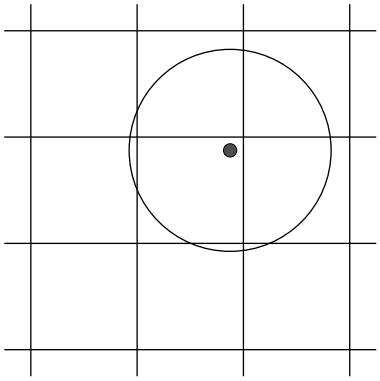


Figura 6.1: El espacio se divide en celdas de tal tamaño que cada partícula interactua tan solo con aquellas que están en su propia celda o en las celdas vecinas.

La solución viene de una aproximación que solo puede hacerse si el potencial V_{ab} decae suficientemente rápido. Si decae rápido se opta por aproximar a cero el potencial más allá de una distancia R_0, es decir, se hace la aproximación $V(r_{ab} \geq R_0) = 0$. Se dice que R_0 es el radio de influencia de cada partícula. Una vez que se ha escogido el valor de R_0 el sistema se divide en celdas cúbicas de tamaño mayor o igual a R_0. De esta manera se logra que cada partícula solo interactue con otras que están en su propia celda o en alguna de las celdas vecinas. Por cada celda c el programa mantiene una lista L_c con el nombre de las partículas que hay en c y cada vez que una partícula cambia de celda el programa actualiza borrando a esa partícula de la lista que deja y anotándola en la nueva. La suma (6.4.12) se hace tan solo sobre las partículas de las celdas que corresponda. Esta suma ya no dependen del tamaño N del sistema y se dice que es $\mathscr{O}(1)$ precisamente porque no depende de N.

La generación del estado de prueba X_p no es un asunto enteramente trivial. Si es muy cercano a X_n la probabilidad de aceptación puede ser muy alta y puede haber una gran correlación entre los estados sucesivos. Por otro lado, si X_p es muy lejano a X_n la probabilidad de rechazo puede ser muy alta lo que hace muy probable que $X_{n+1} = X_n$ lo que implica correlación máxima.

En la literatura en el tema suele decirse que el algoritmo de generación de los X_p sea ajustado en una corrida en blanco para lograr una tasa de aceptación de alrededor de 50 %. Y además para promediar no se tome todos los estados de la secuencia $\{X_n\}$ sino uno de cada K estados, para desminuir los efectos de correlación entre estados consecutivos. Mi experiencia es que hay casos en que lo anterior es claramente inconveniente.

• Para saber más sobre los temas de este capítulo se recomienda

- *Monte Carlo Methods in Statistical Physics*, M.E.J. Newman & G.T. Barkena, Clarendon Press, Oxford, 1999.

- *A guide to Monte Carlo Simulations in Statistical Physics*, D.P. Landau, K. Binder, Cambridge University Press, 2000.

6.4.4. Propiedades necesarias

Un algoritmo como el de Metropolis genera una secuencia de puntos X distribuidos según $W(X)$. El algoritmo debe cumplir con una serie de propiedades para ser satisfactorio a los propósitos de Mecánica Estadística.

Estas propiedades no necesariamente se refieren a toda la historia de la secuencia generada, sino a las propiedades que la secuencia alcanza una vez que *ha relajado al equilibrio*. Es necesaria esta aclaración porque si el punto de partida es muy improbable, la primera parte de la secuencia puede ocurrir en una zona muy poco representativa de la distribución $W(X)$. Pero tarde o temprano la secuencia va a alcanzar las zonas importantes de W y es a partir de ahí que se dice que se tiene propiedades de equilibrio.

Proceso de Markov. Para los fines de este capítulo, un *proceso* de Markov genera aleatoriamente una secuencia de puntos X en algún espacio cumpliendo ciertas propiedades. La probabilidad $P(X,Y)$ de generar el punto Y si se proviene del punto X solo depende de estos dos puntos. En particular no depende de la historia anterior de la secuencia y debe satisfacer $\sum_Y P(X,Y) = 1$, es decir, dado un X siempre se produce un Y. La probabilidad $P(X,X)$ puede ser no nula (en el algoritmo de Metropolis claramente es no nula).

Ergodicidad. Se desea que, sin importar el punto X_0 de partida, la secuencia debe en algún momento alcanzar todo X que tenga probabilidad no nula. Más aun, la frecuencia con que la secuencia visita al punto X debe ser proporcional a $W(X)$.

Balance detallado. Si se considera la parte de la secuencia que se tiene después de haber *llegado al equilibrio*, la probabilidad de alcanzar un punto X debe ser igual a la probabilida de salir de ese punto, en el siguiente sentido: $\sum_Y W(X)P(X,Y) = \sum_Y W(Y)P(Y,X)$.

El balance detallado exige algo más restrictivo:

$$W(X)P(X,Y) = W(Y)P(Y,X)$$

Esta última exigencia—aunque no se justificará—garantiza que la secuencia efectivamente alcanza un equilibrio y no es atrapada en algún tipo de ciclo límite. En el caso de Metropolis esta propiedad quedó establecida en (6.4.7).

6.5. Problemas

1. Determine un mínimo de la energía potencial E asociada a un sistema bidimensional de 20 partículas puntuales que interactuan de a pares con el potencial

$$V_{ab} = 4\left(\frac{1}{r_{ab}^8} - \frac{1}{r_{ab}^4}\right)$$

donde r_{ab} es la distancia entre a y b. Las partículas se pueden mover tan solo en el plano XY con coordenadas $0 \le x_a \le 10$ y también $0 \le y_a \le 10$. Suponga que una de las partículas está fija en $x = 5$, $y = 5$ y que otra solo puede moverse en el eje X con $y = 5$. Un sistema de este tipo está en su mínimo de energía tan solo si está a temperatura cero. Más adelante se verá cómo determinar estados representativos asociados a una cierta temperatura. Como configuración inicial coloque a las partículas desordenadamente dentro de la caja de 10×10. La "evolución" del sistema se hace en forma aleatoria intentando modificar una sola coordenada a la vez en la forma $z_{\text{nueva}} = z_{\text{actual}} + (0{,}5 - \texttt{drand48}())\,\delta$ y el nuevo valor se acepta tan solo si la energía disminuye. Conviene aumentar el valor de δ (por ejemplo en

un 50 %) cuando z_{nueva} es aceptado y disminuirlo levemente cuando z_{nueva} es rechazado, ¿por qué? [a] Explique muy claramente el procedimiento seguido. [b] Haga un gráfico (o una tabla) con la evolución de la energía como función del número de ciclos, donde se define como ciclo el conjunto de las 37 iteraciones que intentan modificar cada una de las 37 coordenadas del sistema libres de variar. [c] Una vez que haya obtenido un mínimo de E dibuje la posición de las partículas. Interprete el valor de E en base: a este dibujo y a la expresión para V_{ab}.

2. Calcule la integral $\int_0^\pi \sin x\,dx$, $f = \sin x$, usando MC2, factorizandolo en la forma $f = wh$ con

$$w = \frac{3}{2\pi}\left[1 - \frac{4}{\pi^2}\left(x - \frac{\pi}{2}\right)^2\right]$$

y compare la velocidad de convergencia con el resultado de integrar en forma directa usando MC1.

3. Compruebe que $I = \int_1^\infty \frac{\sin x}{x^3}\,dx$ puede calcularse usando $y = \frac{1}{x}$ en la forma $I \approx \frac{1}{N}\sum_j y_j \sin(1/y_j)$ con $y_j \leftarrow U(0,1)$.

4. Obtenga el valor de la integral 9-dimensional

$$\int \sqrt{x_1^4 + 2x_2^4 + 3x_3^4 + .. + 9x_9^4}\; dx_1..dx_9$$

dentro de una hiperesfera de radio 2 sin hacer cambio de variables. Estime el valor numérico del error de su resultado. El resultado muy aproximadamente es $6{,}3 \times 10^3$.

5. Escriba y ejecute un programa inteligente que genere una secuencia aleatoria $\{x_k\}$ de 20 millones de valores que se distribuyan según

$$W(x) = \frac{2x}{(1+x^2)^2} \qquad \text{con} \quad 0 \le x \le \infty$$

Haga un histograma $H[k]$ que registre la frecuencia de ocurrencia de los valores $0 \le x \le 20$ en celdas de largo 0.1, es decir, el histograma tiene 200 componentes. ¿Qué porciento de la secuencia está en este intervalo? En un mismo gráfico superponga los valores de $W(x_j)$ y del histograma normalizado, es decir los valores $\mathcal{N}H[j]$, ¿cómo debe hacerse la comparación realmente? ¿cómo se escoge $\mathcal{N}$? Explique en detalle. Obtenga además el promedio Monte Carlo de e^{-x} con respecto a W usando los primeros n millones de valores de la secuencia, con $n = 1$, $n = 2$ .. hasta $n = 20$.

6. Use la función

$$W(x) = \frac{1}{5\sqrt{\pi}}\left(2e^{-(x-1)^2} + 3e^{-(x+1)^2}\right)$$

para generar con el algoritmo de Metropolis una secuencia $\{x_n\}$ partiendo de x_0 escogido a gusto entre -2 y 2 y definiendo $x_p = x_n + \delta\Delta$ donde $\Delta \leftarrow U(-1,1)$ y $\delta \approx 6{,}0$ (sí, dice 6.0). La literatura dice que un buen δ es aquel que implica que aproximadamente la mitad de los x_p son aceptados. Use la función `drand48()` de C inicializada una sola vez con `srand48(M)`, y M es un entero cualquiera. Haga un histograma h_k ($k = 0.,799$) de los valores de la secuencia que quepan dentro de $-4 \le x \le 4$, y dibuje $100 * h_j/n_1$ comparando con $W(x)$ para 50 mil y 5 millones de iteraciones, donde n_1 el el número de puntos de la secuencia x_n que cayeron dentro de $(-4,4)$. Por cada punto x que está en tal intervalo se puede hacer:

```
j = (int) (100.0*(4.0 + x));
h[j]++;
n1++;
```

Si el histograma fuese $h[0 \leq j < N]$ puede comprobar que la instrucción de más arriba sería `j = (int) 0.125*N*(4.0+x);` ya que (4.0+x) puede alcanzar el valor 8.0. Se debe dibujar $0{,}125 * N * h[j]/n1$.

7. Considere un sistema unidimensional de 11 varas de largo 1 dentro de una "caja" de largo 14. Suponga que la interacción entre ellas es de energía potencial $V_0 = 1$ si se penetran y nula si no se tocan. Por medio del algoritmo de Metropolis obtenga la densidad $n(x)$ media del sistema. Para ello divida la caja en 200 intervalos y determine numéricamente la probabilidad de ocupación del centro de cada vara en cada uno de los 200 intervalos. Es claro que $\int n(x)\,dx = 1$. También determine la probabilidad $g(x)$ de que la partícula 6 (la central) tenga distancia relativa x con alguna otra partícula. La función $g(x)$ se llama *función de correlación de pares*. Haga variar esta distancia x relativa en el intervalo $(0,5)$ y haga una normalización arbitrara a $g(x)$. Para unas tres temperaturas diferentes obténga tanto la densidad como $g(x)$ usando $T = 0{,}01$, $T = 0{,}4$, $T = 10$. Explique claramente todo lo que haga.

8. Considere un sistema 2D de 30 partículas interactuando con un potencial

$$V_{ab} = 4\left(\frac{1}{r_{ab}^{12}} - \frac{1}{r_{ab}^{6}}\right)$$

inicializadas en orden cristalográfico triangular (hexagonal de cara centrada) en tres filas de 10 partículas. Utilizando el algoritmo de Metropolis estudie el coeficiente de expansión lineal del sistema (cambio de la longitud media del sistema) como función de la temperatura.

Comience con una temperatura muy baja $(T = 0{,}001)$. Con cada nueva temperatura se comienza de la configuración final que se obtuvo con la temperatura anterior. Al cambiar de temperatura el sistema debe ser relajado (unas 30 mil iteraciones) antes de comenzar a "medir".

9. Considere el modelo de Ising ferromagnético en dos dimensiones definido sobre un reticulado cuadrado, descrito por el hamiltoniano

$$H = -J \sum_{(i,j)} \sum_{(k,l)} S_{ij} S_{kl} - B \sum_{(i,j)} S_{ij}$$

donde las variables dinámicas S_{ij} solo pueden tomar los valores ± 1, J es la constante de acoplamiento positiva y B es el campo magnético multiplicado por el momento magnético de cada spin. Los índices (i,j) recorren la red cuadrada y los índices (k,l) recorren los cuatro vecinos del nodo (i,j), esto es,

$$(k,l) = \{(i-1,j);(i+1,j);(i,j-1);(i,j+1)\}$$

Se debe usar una red cuadrada de $N \times N$ spines con condiciones de borde periódicas. Cada producto $S_{ij}S_{kl}$ tiene asociado en forma natural un trazo elemental de la red y el término $-J\sum_{(i,j)}\sum_{(k,l)} S_{ij}S_{kl}$ debe entenderse como una suma sobre todos los trazos elementales de la red, sumando una sola vez cada trazo.

Para usar las condiciones de borde periódicas sin complicaciones, la lista de spines vecinos se llama de la siguiente forma

$$(k,l) = \{(i-1 \bmod N, j); (i+1 \bmod N, j)(i, j-1 \bmod N); (i, j+1 \bmod N)\}$$

donde `mod` es la función resto, que en **C** corresponde al operador `%`.

Se define la magnetización instantánea como

$$M = \sum_{(i,j)} S_{ij}$$

Se desea medir la curva $M(T)$ para $B = 0$ y la curva $M(B)$ para $T = T_0 < T_c$. Si se usa el sistema de unidades en que $J = 1$, mida la primera curva desde $T = 0$ a $T = 3{,}0$, y la segunda curva desde $B = -2$ a $B = 2$, para $T = 1{,}0$. Debe hacer ambas curvas en los dos sentidos, esto es, una vez aumentando y la otra disminuyendo el parámetro de control. Use una red de tamaño $N = 50$. Además, dé una pequeña interpretación de los resultados que obtenga e identifique la temperatura crítica.

Para construir cada curva, se propone usar el siguiente método. Se genera primero una condición al azar (cada spin toma al azar un valor 1 ó -1). Luego, se relaja el sistema para una temperatura y campo magnético iguales al punto inicial de la curva. Después de relajar se promedia la magnetización con esos parámetros. Se varían los parámetros y se usa como condición inicial el estado final que resultó de la simulación anterior. Después de una breve relajación (porque como los parámetros son similares, el estado de equilibrio debe ser similar) se promedia la magnetización, y así sucesivamente.

10. Considere el siguiente modelo para la molécula H_2. La molécula está compuesta por dos núcleos de hidrógeno (protones) a distancia L. En torno a cada núcleo hay un electrón que supondremos que está en un órbital s. La expresión para la función de onda de un electrón en ese orbital en torno a un núcleo ubicado en $\vec{R}$ es

$$\Psi(\vec{r};\vec{R}) = \frac{1}{\sqrt{\pi a_0^3}}\, e^{-|\vec{r}-\vec{R}|/a_0} \tag{6.5.1}$$

donde a_0 es el radio de Bohr.

El operador de energía electrostática de la molécula, que incluye la interacción entre electrones y núcleos, está dado por

$$U(\vec{r}_1,\vec{r}_2;\vec{R}_1,\vec{R}_2) = \frac{e^2}{|\vec{r}_1-\vec{r}_2|} - \frac{e^2}{|\vec{r}_1-\vec{R}_1|} - \frac{e^2}{|\vec{r}_1-\vec{R}_2|} - \frac{e^2}{|\vec{r}_2-\vec{R}_1|} - \frac{e^2}{|\vec{r}_2-\vec{R}_2|} + \frac{e^2}{|\vec{R}_1-\vec{R}_2|} \tag{6.5.2}$$

donde $\vec{r}_1$ y $\vec{r}_2$ son las posiciones de los electrones, $\vec{R}_1$ y $\vec{R}_2$ las posiciones de los núcleos y e es la carga electrónica.

Luego, sin considerar antisimetrización por spin de los electrones, el valor medio de la energía electrostática es

$$\langle U \rangle = \int d^3r_1\, d^3r_2\, \Psi(\vec{r}_1;\vec{R}_1)^2\, \Psi(\vec{r}_2;\vec{R}_2)^2\, U(\vec{r}_1,\vec{r}_2;\vec{R}_1,\vec{R}_2) \tag{6.5.3}$$

Use un sistema de unidades donde $a_0 = 1$ y $e = 1$ y coloque a los núcleos en $\vec{R}_1 = L/2\hat{\imath}$ y $\vec{R}_2 = -L/2\hat{\imath}$. Se pide graficar $\langle U \rangle$ para separaciones L que van desde $L = 1{,}5$ hasta $L = 2{,}5$. Integre usando el algoritmo de Metropolis consideando que las dos distribuciones $p(\vec{r}_i) = \Psi(\vec{r}_i; \vec{R}_i)^2$ $(i = 1, 2)$ dan la probabilidades con que se distribuyen $\vec{r}_1$ y $\vec{r}_2$ respectivamente. No es necesario preocuparse por las divergencias en U, pues las integrales son finitas.

Tal vez le convenga separar en coordenadas centro de masa y relativas.

Capítulo 7

Ecuaciones elípticas

En este capítulo y los que siguen se verá tan solo los métodos más sencillos para integrar ecuaciones a derivadas parciales. Existe una amplia variedad de métodos más refinados.

7.1. Ecuación y condiciones de borde

Considérese el sencillo caso bidimensional de una ecuación de Poisson dentro de un dominio en el plano XY, con borde dado por una curva cerrada Γ

$$\nabla^2\Phi = G(x,y) \tag{7.1.1}$$

Como condiciones de borde se puede condiciones rígidas, o de Dirichlet

$$[\Phi]_\Gamma = [g_1(x,y)]_{(x,y)\in\Gamma} \tag{7.1.2}$$

o bien algún tipo de condición sobre las derivadas, por ejemplo,

$$\left(\frac{\partial\Phi}{\partial n}\right)_\Gamma = [g_2(x,y)]_{(x,y)\in\Gamma} \tag{7.1.3}$$

donde $\partial/\partial n = \hat{n}\cdot\nabla$ es la derivada normal al borde Γ del dominio de integración y se llama condición de borde tipo Neumann. Cuando en todo el borde se tiene este tipo de condición la solución del problema no es única porque a la solución que se tenga se le puede agregar una constante arbitraria y sigue siendo una solución del mismo problema. Pero una ecuación como $\nabla^2\Phi = G(\Phi,x,y)$ puede tener condiciones de borde tipo Neumann en todos lados y tiene solución única excepto que sea una ecuación de autovalores ($G = k(x,y)\,\Phi$).

Más en general se puede tener condiciones de borde mixtas,

$$\left(\frac{\partial\Phi}{\partial n}\right)_\Gamma + \gamma(x,y)\,\Phi(x,y) = [g(x,y)]_{(x,y)\in\Gamma} \tag{7.1.4}$$

La derivada $\left(\frac{\partial\Phi}{\partial n}\right)_\Gamma$ se debe entender como $\hat{n}\cdot\nabla\Phi$ donde $\hat{n}$ es la normal al borde.

En todo lo que sigue se puede rehacer los cálculos considerando ecuaciones elípticas más generales, tales como

$$\nabla\left(p(x,y)\,\nabla\Phi(x,y)\right)+q(x,y)\,\Phi(x,y)=G(x,y) \tag{7.1.5}$$

pero en nada sustancial cambian los métodos de aquellos requeridos para integrar (7.1.1).

7.1.1. Integral de acción

Definamos la integral de acción $S[\Phi]$

$$S[\Phi]=\int dxdy\,\left[\frac{1}{2}\left(\nabla\Phi\right)^2+\Phi G\right] \tag{7.1.6}$$

busca la condición bajo la cual S permanece estacionaria—esto es $\delta S=0$ — y que sea compatible con las condiciones de borde rígidas,

$$\delta S=\int dxdy\,\left[\nabla\Phi\cdot\nabla(\delta\Phi)+\delta\Phi G\right] \tag{7.1.7}$$

pero

$$\int\nabla\Phi\cdot\nabla(\delta\Phi)\,dxdy=\int\nabla\cdot(\delta\Phi\nabla\Phi)\,dxdy-\int\delta\Phi\nabla^2\Phi dxdy \tag{7.1.8}$$

La primera de las dos integrales de la derecha es equivalente a una integral sobre el borde Γ de la zona de integración,

$$\int_\Gamma\delta\Phi\nabla\Phi\cdot\hat{n}\,ds$$

donde $\hat{n}$ es el vector normal al borde Γ y ds es el elemento de arco. Puesto que Φ no varía sobre el borde ($(\delta\Phi)_\Gamma=0$) la primera integral es nula concluyéndose que

$$\delta S=\int dxdy\,\left[-\nabla^2\Phi+G\right]\,\delta\Phi \tag{7.1.9}$$

que debe ser nula para cualquier variación $\delta\Phi$. Esto implica (7.1.1).

El hecho que la ecuación provenga del mínimo de la integral de acción reduce el problema al de encontrar el mínimo de S compatible con las condiciones de borde.

Para una ecuación como (7.1.5) basta con considerar la integral de acción

$$S[\Phi]=\int dxdy\,\left[\frac{1}{2}\,p\,\left(\nabla\Phi\right)^2+F(\Phi)\right]$$

que implica la ecuación

$$\nabla\left(p\nabla\Phi\right)-\frac{\delta F}{\delta\Phi}=0$$

$F(\Phi)$ puede, por ejemplo, ser $-\frac{1}{2}q\Phi^2-\Phi G$.

7.2. Discretización

7.2.1. Discretización en el volumen

La ecuación (7.1.1) puede ser discretizada en la forma

$$\frac{\Phi_{i+1,k}-2\Phi_{i,k}+\Phi_{i-1,k}}{h^2}+\frac{\Phi_{i,k+1}-2\Phi_{i,k}+\Phi_{i,k-1}}{h^2}=G_{ik} \tag{7.2.1}$$

De esta ecuación se despeja $\Phi_{i,k}$ definiéndose una relación de recurrencia para ir actualizando los valores de los $\Phi_{i,k}$

$$\Phi_{i,k}\leftarrow\frac{1}{4}\left[\Phi_{i+1,k}+\Phi_{i-1,k}+\Phi_{i,k+1}+\Phi_{i,k-1}-h^2G_{ik}\right] \tag{7.2.2}$$

Esta relación de recurrencia converge prácticamente a partir de cualquier conjunto inicial de valores para $\Phi_{i,k}$.

Algo más en general se define, en lugar de (7.2.2)

$$\Phi_{i,k}\leftarrow(1-\omega)\,\Phi_{i,k}+\frac{\omega}{4}\left[\Phi_{i+1,k}+\Phi_{i-1,k}+\Phi_{i,k+1}+\Phi_{i,k-1}-h^2G_{ik}\right] \tag{7.2.3}$$

que se reduce a (7.2.2) cuando $\omega=1$. Esta regla debe manejarse usando un solo arreglo Φ_{ik}. Si se intenta actualizar el valor de Φ en todos los sitios usando (7.2.3) con el arreglo previo, el método, en general, no es convergente.

Si las condiciones de borde son tipo Dirichlet se puede usar directamente (7.2.3) sin más consideraciones tomando cuidado de no alterar los valores colocados al comienzo en los bordes.

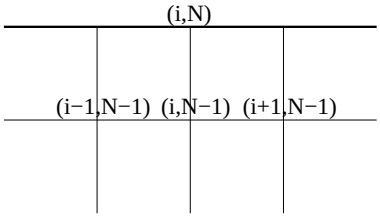


Figura 7.1: Detalle de la discretización en el borde superior del dominio de integración.

7.2.2. Discretización en los bordes en un caso tipo Neumann

La condición (7.1.3) discretizada es

$$\Phi_{iN}-\Phi_{iN-1}=hg_i \tag{7.2.4}$$

Al considerar (7.2.2) con $k=N-1$, a la derecha aparece un Φ_{iN} que se reemplaza, usando la última expresion, y se obtiene

$$4\Phi_{i,N-1}=\Phi_{i+1,N-1}+\Phi_{i-1,N-1}+\{\Phi_{i,N-1}+hg_i\}+\Phi_{i,N-2}-h^2G_{i,N-1} \tag{7.2.5}$$

Entre llaves aparece la expresión usada en lugar de $\Phi_{i,N}$. De esta nueva expresión se deduce inmediatamente la relación de recurrencia específica para los puntos vecinos al borde,

$$\Phi_{i,N-1}\leftarrow(1-\omega)\,\Phi_{i,N-1}+\frac{\omega}{3}\left[\Phi_{i+1,N-1}+\Phi_{i-1,N-1}+\Phi_{i,N-2}-h^2G_{i,N-1}+hg_i\right] \tag{7.2.6}$$

Obtenidos los valores anteriores se actualiza los puntos del borde mismo con

$$\Phi_{iN}=\Phi_{iN-1}+hg_i \tag{7.2.7}$$

En resumen: lejos de los bordes se itera con (7.2.3), al lado de los bordes se itera con (7.2.6) y los puntos del borde se iteran con (7.2.7).

7.2.3. Convergencia

7.2.3.1. Iteración en el volumen

Para estudiar la convergencia de (7.2.3) se utilizará la regla de iteración lejos de los bordes.

La versión discreta de la integral de acción es

$$S=\sum_{i,k}\left[\frac{1}{2}\left(\frac{\Phi_{ik}-\Phi_{i-1k}}{h}\right)^2+\frac{1}{2}\left(\frac{\Phi_{ik}-\Phi_{ik-1}}{h}\right)^2+\Phi_{ik}G_{ik}\right]h^2 \tag{7.2.8}$$

y, puesto que interesará la condición $dS/d\Phi_{ik}=0$ para cada Φ_{ik} separadamente, basta con tomar en cuenta de todas las contribuciones a S solo aquellas con términos que tienen un Φ_{ik} con índices fijos:

$$\begin{aligned}
S &= \frac{1}{2}\left(\Phi_{ik}-\Phi_{i-1k}\right)^2++\frac{1}{2}\left(\Phi_{i+1k}-\Phi_{ik}\right)^2+\frac{1}{2}\left(\Phi_{ik}-\Phi_{ik-1}\right)^2\\
&\quad+\frac{1}{2}\left(\Phi_{ik+1}-\Phi_{ik}\right)^2+h^2\Phi_{ik}G_{ik}+\text{términos sin }\Phi_{ik}\\
&= 2\Phi_{ik}^2-\Phi_{ik}\Phi_{i-1k}-\Phi_{ik}\Phi_{i+1k}-\Phi_{ik}\Phi_{ik-1}-\Phi_{ik}\Phi_{ik+1}\\
&\quad+h^2\Phi_{ik}G_{ik}+\text{términos sin }\Phi_{ik}\\
&= \Phi_{ik}\left(\underbrace{2\Phi_{ik}-\Phi_{i-1k}-\Phi_{i+1k}-\Phi_{ik-1}-\Phi_{ik+1}+h^2G_{ik}}_{B}\right)+\text{términos sin }\Phi_{ik}\\
&= \Phi_{ik}B+\text{términos sin }\Phi_{ik}
\end{aligned} \tag{7.2.9}$$

En esta expresión para S se va a reemplazar Φ_{ik} por la expresión dada en (7.2.3), definiendo así un nuevo valor S'. Pero primero (7.2.3) será reescrita para las necesidades actuales,

$$\begin{aligned}
\Phi_{i,k} &\leftarrow \Phi_{ik}+\underbrace{\frac{\omega}{4}\left[-4\Phi_{ik}+\Phi_{i+1,k}+\Phi_{i-1,k}+\Phi_{i,k+1}+\Phi_{i,k-1}-h^2G_{ik}\right]}_{A=\omega\Psi/4}\\
&\leftarrow \Phi_{ik}+A
\end{aligned} \tag{7.2.10}$$

Nótese que A y B se relacionan por

$$2\Phi_{ik}+B=-\frac{4}{\omega}A \tag{7.2.11}$$

Al hacer el reemplazo (7.2.10) en la expresión para S, tanto en el factor explícito Φ_{ik} como en B se obtiene

$$S'=\left(\Phi_{ik}+A\right)\left(2A+B\right)$$

que permite calcular la variación de S como

$$\begin{aligned}
\Delta &= S'-S\\
&= 2A^2+A\left(2\Phi_{ik}+B\right)\\
&= 2A^2+A\left(-\frac{4A}{\omega}\right)\\
&= 2\frac{\omega-2}{\omega}A^2\\
&= \frac{\omega\left(\omega-2\right)}{8}\Psi^2
\end{aligned} \tag{7.2.12}$$

Si $\omega < 0$ el último factor es positivo, lo que no se quiere. Para que S disminuya se necesita $\Delta < 0$ lo que requiere que ω sea positivo y menor que dos,

$$0 \leq \omega \leq 2 \tag{7.2.13}$$

De este sencillo análisis parece desprenderse que el valor óptimo es $\omega = 1$, sin embargo puede comprobarse que valores de ω más cerca de 2 son los que dan una convergencia mucho más rápida. Debido a errores numéricos, si ω se acerca mucho a 2, el algoritmo se desestabiliza y diverge.

7.2.3.2. Iteración con condición de borde tipo Neumann

Se presenta en forma muy esquemática la forma de obtener la condición cuando se itera cerca de un borde cuando se tiene las condición de borde de tipo analizado en §7.2.2.

- Se define Ψ imponiendo que la expresión a la derecha en (7.2.6) sea $\Phi_{i,N-1} + \omega\Psi$. De esta ecuación se despeja $h^2 G_{i,N-1}$.
- Se define $S_{i,N-1}$ como la suma de todos los términos de S que contienen $\Phi_{i,N-1}$. Una vez que se tiene esta expresión se elimina $\Phi_{i,N}$ usando (7.2.7), lo que da una nueva expresión para $S_{i,N-1}$.
- En $S_{i,N-1}$ se sustituye $\Phi_{i,N-1}$ usando la regla de iteración (7.2.6), lo que da una expresión que designamos $S'_{i,N-1}$.
- Se define $\delta = S'_{i,N-1} - S_{i,N-1}$. En ella se sustituye $h^2 G_{i,N-1}$ por la expresión que se obtuvo en el primer paso.

El resultado de este procedimiento es

$$\delta = \frac{3\omega(\omega-2)}{2}\Psi^2$$

que nuevamente garantiza convergencia cuando se satisface (7.2.13).

Si el problema que se resuelve es (7.1.1) y la condición de borde es tipo Neumann en todos los bordes, entonces la solución no es única, porque si Ψ es solución, también lo es $\Psi + \text{cte}$. Basta con fijar arbitrariamente el valor de Ψ en un punto del dominio para que el método garantice unicidad.

7.3. Fluidos incompresibles estacionarios

7.3.1. Las ecuaciones

Como aplicación de lo anterior se estudiará el caso *bidimensional* que resulta de comenzar con las ecuaciones hidrodinámicas

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \vec{v}) = 0 \tag{7.3.1}$$

$$\rho\left(\frac{\partial\vec{v}}{\partial t}+(\vec{v}\cdot\nabla)\vec{v}\right) = -\nabla p+\eta\nabla^2\vec{v}+\rho\,\vec{g} \tag{7.3.2}$$

$$\frac{\partial T}{\partial t}+(\vec{v}\cdot\nabla)\,T = \kappa\nabla^2 T \tag{7.3.3}$$

donde

ρ	densidad	$\vec{v}$	velocidad hidrodinámica
p	presión	T	temperatura
η	viscosidad	$\vec{g}$	aceleración de gravedad
κ	conductividad térmica	α	coeficiente de expansión térmica lineal
ψ	función corriente	ζ	vorticidad
$\nu=\eta/\rho_0$	viscosidad cinemática		

Se va a hacer dos simplificaciones: nos restringiremos a estados en los que no hay variación temporal y en los que la densidad pueda ser reemplazada por la densidad media ρ_0 excepto en el término con gravedad donde se coloca

$$\rho=(1-\alpha(T-T_0))\,\rho_0\,. \tag{7.3.4}$$

T_0 es la temperatura media del sistema. En el caso de líquidos la densidad varía un poco, no así en el caso de gases.

Existe una sencilla solución hidrostática a las ecuaciones anteriores si se supone que el fluido está entre una base a temperatura fija T_b en $y=0$ y un borde superior a temperatura T_t en $y=y_1$, suponiendo que $\vec{g}=(0,-g)$

$$\begin{aligned}\vec{v} &= 0\\ T &= T_b+(T_t-T_b)\,\frac{y}{y_1} \quad \text{la temperatura cambia linealmente con la altura}\\ p &= g\rho_0\left[\frac{\alpha}{2}\frac{T_t-T_b}{y_1}y^2-y\right]+p_0\end{aligned} \tag{7.3.5}$$

7.3.2. Ecuaciones estacionarias para la función corriente, la vorticidad y la temperatura

Sin tiempo y densidad uniforme (7.3.1) se convierte en

$$\nabla\cdot\vec{v}=0 \tag{7.3.6}$$

pero todo campo con divergencia nula puede ser expresado como el rotor de otro campo que, en el caso presente, se denomina *función corriente*, ψ,

$$v_i=\varepsilon_{ij}\partial_j\psi \qquad \text{esto es en 2D} \tag{7.3.7}$$

Donde ε_{ij} es antisimétrico y $\varepsilon_{12}=1$ mientras que ψ es un seudovector, lo que bidimensionalmente lo hace un seudoescalar,

$$\vec{v}=\begin{pmatrix}u\\ v\end{pmatrix}=\begin{pmatrix}\frac{\partial\psi}{\partial y}\\ -\frac{\partial\psi}{\partial x}\end{pmatrix}=\begin{pmatrix}\psi_y\\ -\psi_x\end{pmatrix} \tag{7.3.8}$$

Nótese que ψ está definido salvo por una constante aditiva.

Adicionalmente se define la vorticidad, esencialmente como el rotor de la velocidad

$$\begin{aligned} \zeta &= -\varepsilon_{ij}\partial_i v_j \\ &= \nabla^2\psi \end{aligned} \tag{7.3.9}$$

Las ecuaciones hidrodinámicas originales se reducen a

$$\nu\nabla^2\zeta = \psi_y\zeta_x - \psi_x\zeta_y + \alpha g T_x \tag{7.3.10}$$

$$\nabla^2\psi = \zeta \tag{7.3.11}$$

$$\kappa\nabla^2 T = \psi_y T_x - \psi_x T_y \tag{7.3.12}$$

que es un sistema acoplado de tres ecuaciones diferenciales para los tres campos ψ, ζ y T.

En problemas con $g = 0$, la ecuación (7.3.12) se desacopla de las otras de modo que basta con resolver solamente (7.3.10) acoplada con (7.3.11).

Las dimensiones de estas cantidades son

$$[\psi] = [\ell^2/t], \qquad [\zeta] = [1/t], \qquad [T] = [m\ell^2/t^2], \qquad [\kappa] = [\nu] = [\ell^2/t] \tag{7.3.13}$$

EJERCICIO: Demostrar que

$$\nabla^2 p = 2\rho_0 \left[\frac{\partial^2\psi}{\partial x^2}\frac{\partial^2\psi}{\partial y^2} - \left(\frac{\partial^2\psi}{\partial x\partial y}\right)^2 \right] - g\frac{\partial\rho}{\partial y} \tag{7.3.14}$$

7.3.3. Líneas de corriente

Si la curva Γ, definida por $[x(s), y(s)]$, es una curva en el plano XY sobre la cual la función Ψ es constante se puede deducir que

$$\begin{aligned} 0 &= \frac{d\Psi}{ds} \\ &= \frac{\partial\Psi}{\partial x}\frac{dx}{ds} + \frac{\partial\Psi}{\partial y}\frac{dy}{ds} \\ &= -v\dot{x} + u\dot{y} \end{aligned}$$

de donde

$$\frac{dy}{dx} = \frac{v}{u}$$

lo que implica que la velocidad $\vec{v}$ es tangente a la curva Γ. De aquí que si se desea imponer que en alguna parte la velocidad siga una línea específica, se debe imponer que Ψ sea constante sobre esa línea.

7.3.4. Versión discreta de ψ y ζ

Las componentes de la velocidad son u y v. En términos discretos estas componentes se asocian a los trazos horizonatales y verticales del reticulado, como lo muestra la figura y

$$u_{ik} = \frac{\psi_{i,k} - \psi_{i,k-1}}{h} \qquad\qquad v_{ik} = -\frac{\psi_{i,k} - \psi_{i-1,k}}{h} \tag{7.3.15}$$

Se puede pensar que en cada cuadrilátero elemental hay una *corriente* ψ_{ik} en el sentido que indica la figura, y las componentes de la velocidad resultan de sumar las corrientes que impone cada celda.

Por otro lado, la vorticidad es el rotor de la velocidad,

$$\begin{aligned} \zeta_{ik} &= (u_y - v_x)_{ik} \\ &= \frac{u_{i,k+1} - u_{i,k}}{h} - \frac{v_{i+1,k} - v_{i,k}}{h} \\ &= \frac{u_{i,k+1} - v_{i+1,k} - u_{i,k} + v_{i,k}}{h} \end{aligned} \tag{7.3.16}$$

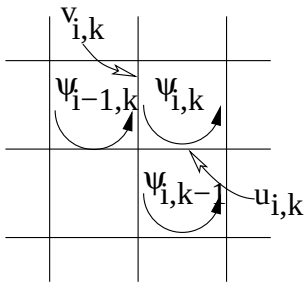


Figura 7.2: Esta figura ayuda a tener una imagen pictórica del significado "local" de ψ.

En el nunmerador está la suma de las componentes de la velocidad asociadas a cada uno de los cuatro lados de la celda, tomando en cuenta el signo según la forma que indica la figura.

En general se llama circulacin de un campo vectorial $\vec{A}$ por un camino cerrado Γ a la integral,

$$C = \oint_{\Gamma} \vec{A} \cdot d\vec{r}. \tag{7.3.17}$$

El signo de la circulacin est ligado al signo con que se escoja recorrer a la curva cerrada Γ. La expresión (7.3.16) es proporcional a la circulación de la velocidad $\vec{v}$ en una celda de integración.

7.4. Primer ejemplo: convección térmica

Esta vez se estudiará la dinámica de un fluido 2D en una caja rectangular $ABCD$ de $L_x \times L_y$ con pared inferior AB mantenida a temperatura fija $T = T_b$ y pared superior CD mantenida a temperatura fija $T = T_t$, $\quad T_b \geq T_t$. Hay gravedad que apunta hacia abajo como lo indica la figura 7.3. Las paredes laterales AD y BC son perfectamente aislantes (flujo de calor nulo) por lo que $\frac{\partial T}{\partial x} = 0$ en los bordes verticales. Se debe resolver las tres ecuaciones acopladas (7.3.10), (7.3.11) y (7.3.12), que con la notación actual son,

$$\nu \nabla^2 \overline{\zeta} = \overline{\psi}_y \overline{\zeta}_x - \overline{\psi}_x \overline{\zeta}_y + g\,\alpha\, \overline{T}_x \tag{7.4.1}$$

$$\nabla^2 \overline{\psi} = \overline{\zeta} \tag{7.4.2}$$

$$\kappa \nabla^2 \overline{T} = \overline{\psi}_y \overline{T}_x - \overline{\psi}_x \overline{T}_y \tag{7.4.3}$$

Puesto que las paredes son sólidas el campo de velocidad se anula en ellas lo que hace que el campo $\overline{\psi}$ sea constante en todo el perímetro y se escoge nulo. La temperatura aparece solo derivada, lo que deja la libertad $T \to c_0 + T'$. Por ejemplo, se puede tomar $T_{\text{top}} = 0$ y $T_{\text{bot}} = \Delta$.

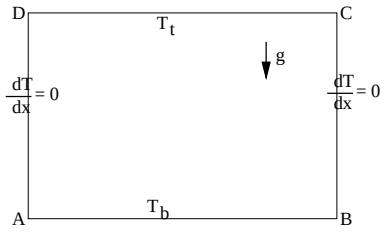


Figura 7.3: *Note que los vértices tienen nombres diferentes que en la figura del flujo con obstáculo. La función corriente es nula en todo el perímetro.*

El problema se discretiza en un reticulado de $N_x \times N_y$ y, para simplificar la notación, supondremos que se logra tener celdas cuadradas de $h \times h$.

En cuanto a la vorticidad $\overline{\zeta}$ se razona en forma parecida al caso visto anteriormente. Se comienza por hacer una expansión de ψ en potencias de h y hasta segundo orden en un punto $(1,k)$ a distancia h del borde izquierdo,

$$\overline{\psi}_{1k} = \overline{\psi}_{0k} + h\left(\frac{\partial \overline{\psi}}{\partial x}\right)_{0k} + \frac{h^2}{2}\left(\frac{\partial^2 \overline{\psi}}{\partial x^2}\right)_{0k} \tag{7.4.4}$$

El primer término es nulo porque $\overline{\psi}$ es nulo en el borde. El segundo también es nulo porque es la segunda componente de la velocidad evaluada en el borde. Por otro lado, $\overline{\zeta} = \partial_y u - \partial_x v$ pero a lo largo del borde izquierdo $u = 0$, es decir, $\partial_y u = 0$, de donde, $\overline{\zeta}_{AD} = -\partial_x v = +\partial_{xx}\overline{\psi} = \overline{\psi}_{xx}$. Un punto sobre AD es un punto $(0,k)$ por lo cual

$$\overline{\zeta}_{0,k} = \left(\frac{\partial^2 \overline{\psi}}{\partial x^2}\right)_{0k} = \frac{2}{h^2}\,\overline{\psi}_{1k} \tag{7.4.5}$$

La última igualdad viene de (7.4.4). En forma semejante se obtiene las condiciones de borde para $\overline{\zeta}$ en los otros tres bordes. Todas ellas son:

$$\overline{\zeta}_{0,k} = \frac{2}{h^2}\,\overline{\psi}_{1,k}\,, \quad \overline{\zeta}_{i,N_2} = \frac{2}{h^2}\,\overline{\psi}_{i,N_2-1}\,, \quad \overline{\zeta}_{i,0} = \frac{2}{h^2}\,\overline{\psi}_{i,1}\,, \quad \overline{\zeta}_{N_1,k} = \frac{2}{h^2}\,\overline{\psi}_{N_1-1,k}\,. \tag{7.4.6}$$

Una forma interesante de adimensionalizar cuando se hace un análisis de las ecuaciones continuas es: $x_k = L_y \hat{x}_k$, $\overline{\psi} = \nu\,\psi$ y $\overline{\zeta} = \frac{\kappa}{L_y^2}\,\zeta$.

Las ecuaciones van a ser adimensionalizadas primero en forma genérica para luego buscar la forma específica más conveniente. Se va a usar

$$\overline{\psi} = \lambda_1 \psi\,, \quad \overline{\zeta} = \lambda_2 \zeta\,, \quad \overline{T} = \Delta T \tag{7.4.7}$$

donde $\Delta = T_b - T_t$. Además se usará una forma compacta para los operadores diferenciales discretos

$$\nabla^2 = \frac{1}{h^2}\,\delta_2 \qquad \frac{\partial}{\partial x_k} = \frac{1}{2h}\delta_k \tag{7.4.8}$$

de tal modo que

$$\delta_2 f = f_{i+1,k} + f_{i-1,k} + f_{i,k+1} + f_{i,k-1} - 4 f_{ik} \tag{7.4.9}$$

$$\delta_k f = f_{i,k+1} - f_{i,k-1} \tag{7.4.10}$$

las tres ecuaciones hidrodinámicas entonces son

$$\begin{aligned}
\frac{\nu\lambda_2}{h^2}\delta_2\zeta &= \frac{\lambda_1\lambda_2}{4h^2}\left(\delta_k\psi\,\delta_i\zeta - \delta_i\psi\delta_k\zeta\right) + \frac{\alpha g\Delta}{2h}\delta_i T \\
\frac{\lambda_1}{h^2}\delta_2\psi &= \lambda_2\zeta \\
\frac{\kappa}{h^2}\delta_2 T &= \frac{\lambda_1}{4h^2}\left(\delta_k\psi\delta_i T - \delta_i\psi\delta_k T\right)
\end{aligned} \tag{7.4.11}$$

Escogiendo $\lambda_2 = \lambda_1/h^2$ se logra que la segunda ecuación sea

$$\delta_2\psi = \zeta$$

y ahora las otras dos ecuaciones quedan

$$\begin{aligned}
\delta_2\zeta &= \frac{\lambda_1}{4\nu}\left(\delta_k\psi\,\delta_i\zeta - \delta_i\psi\delta_k\zeta\right) + \frac{\alpha g h^3\Delta}{2\nu\lambda_1}\delta_i T \\
\delta_2 T &= \frac{\lambda_1}{4\kappa}\left(\delta_k\psi\delta_i T - \delta_i\psi\delta_k T\right)
\end{aligned}$$

Aparecen tres coeficientes numéricos. Se escoge $\lambda_1 = 4\nu$ para que el primer coeficiente sea uno. Automáticamente el tercer coeficiente toma el valor

$$\mathrm{Pr} \equiv \frac{\nu}{\kappa}$$

que se conoce como *número de Prandtl*. El segundo coeficiente puede ser escrito como

$$\mu \equiv \frac{\mathrm{Ra}}{8\,\mathrm{Pr}\,N_y^3}$$

donde el *número de Rayleigh* es

$$\mathrm{Ra} \equiv \frac{\Delta L_y^3\,\alpha\,g}{\nu\,\kappa} \tag{7.4.12}$$

En resumen, las ecuaciones de iteración en puntos del interior son (seguro que he cometido errores: rechequear)

$$\begin{aligned}
\psi_{ik} &= \frac{1}{4}\left[\psi_{i+1,k} + \psi_{i-1,k} + \psi_{i,k+1} + \psi_{i,k-1} - \zeta_{ik}\right] \\
\zeta_{ik} &= \frac{1}{4}\Big[\zeta_{i+1,k} + \zeta_{i-1,k} + \zeta_{i,k+1} + \zeta_{i,k-1} - \left(\psi_{i,k+1} - \psi_{i,k-1}\right)\left(\zeta_{i+1,k} - \zeta_{i-1,k}\right) \\
&\quad + \left(\zeta_{i,k+1} - \zeta_{i,k-1}\right)\left(\psi_{i+1,k} - \psi_{i-1,k}\right) - \mu\,\left(T_{i+1,k} - T_{i-1,k}\right)\Big] \\
T_{ik} &= \frac{1}{4}\Big[T_{i+1,k} + T_{i-1,k} + T_{i,k+1} + T_{i,k-1} \\
&\quad + \mathrm{Pr}\left\{\left(\psi_{i+1,k} - \psi_{i-1,k}\right)\left(T_{i,k+1} - T_{i,k-1}\right) - \left(\psi_{i,k+1} - \psi_{i,k-1}\right)\left(T_{i+1,k} - T_{i-1,k}\right)\right\}\Big]
\end{aligned} \tag{7.4.13}$$

A estas ecuaciones aun se les debe agregar el parámetro ω de sobrerelajación para acelerar la convergencia. El problema aparece con dos parámetros de control en las ecuaciones mismas, μ y Pr aparte de los que puedan entrar a través de las condiciones de borde.

Puesto que el campo de velocidad en los bordes debe anularse, debe cumplirse que ψ sea constante en los bordes, y no hay pérdida de generalidad tomando nula tal constante:

$$\psi_{i,0} = 0\,, \quad \psi_{i,N_2} = 0\,, \quad \psi_{0,k} = 0\,, \quad \psi_{N_1,k} = 0$$

De modo que ψ se itera en los puntos interiores sin restricciones.

Arriba y abajo la temperatura debe ser fija y como aparecen sus derivadas no hay pérdida tomando

$$T_{i,0} = 1\,, \quad T_{i,N_2} = 0$$

mientras que a los costados la derivada de T debe ser nula, esto es, $T_{1,k} - T_{0,k} = 0$ y también $T_{N_1,k} - T_{N_1-1,k} = 0$. Al usar la fórmula con que se itera T para $T_{1,k}$ se obtiene al lado derecho un par de $T_{0,k}$ que son reemplazados por $T_{1,k}$. Esto conduce a la ley de iteración

$$\begin{aligned} T_{1k} &= \frac{1}{3}\Big[T_{2,k} + T_{1,k+1} + T_{1,k-1} \qquad (7.4.14)\\ &\quad + \mathrm{Pr}\left\{\left(\psi_{2,k} - \psi_{0,k}\right)\left(T_{1,k+1} - T_{1,k-1}\right) - \left(\psi_{1,k+1} - \psi_{1,k-1}\right)\left(T_{2,k} - T_{0,k}\right)\right\}\Big] \\ T_{0k} &= T_{1k} \qquad (7.4.15) \end{aligned}$$

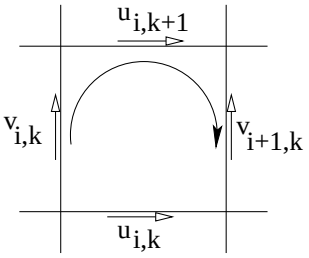


Figura 7.4: La vorticidad puede ser interpretada como la circulación local del campo de velocidad.

Algo semejante debe hacerse al lado derecho.

En el caso del campo de vorticidad ζ, se debe cumplir las condiciones de borde (7.4.6) que en el caso adimensionalizado son

$$\zeta_{0,k} = 2\,\psi_{1,k}\,, \quad \zeta_{i,N_2} = 2\,\psi_{i,N_2-1}\,, \quad \zeta_{i,0} = 2\,\psi_{i,1}\,, \quad \zeta_{N_1,k} = 2\,\psi_{N_1-1,k}\,.$$

de modo que ζ debe ser relajado en los puntos interiores y a continuación deben imponerse estas condiones de borde.

7.5. Segundo ejemplo: flujo y obstáculo

El sistema se discretiza en N_1 intervalos en la dirección X y en N_2 intervalos en la dirección Y de tal modo que el intervalo elemental en ambas direcciones es h (tan solo para que la notación sea sencilla),

$$x = ih \qquad y = kh \qquad i = 0,..N_1\,, \quad k = 0,..N_2 \qquad (7.5.1)$$

Las coordenadas enteras del obstáculo son $E \leftrightarrow (i_1, k_1)$ y $G \leftrightarrow (i_2, k_2)$

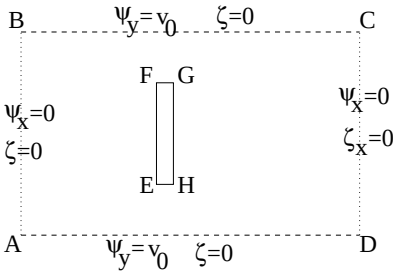


Figura 7.5: Con línea punteada se señala un “borde” de origen puramente algorítmico para integrar en una zona finita. El obstáculo es el rectángulo $EFGH$.

7.5.1. Las ecuaciones discretas en el volumen

Este es un problema que se plantea sin gravedad por lo que bastará con resolver (7.3.10) acoplada con (7.3.11). El campo de temperatura se puede evaluar al final si se desea.

Se verá el caso de un fluido 2D que pasa más allá de un obstáculo rectangular $EFGH$. Para resolver este problema se impondrá ciertas condiciones de borde de naturaleza física y otras que son más bien hipótesis simplificatorias. Se escoge resolver el problema dentro de un rectángulo ficticio $ABCD$ suficientemente lejos del obstáculo. Se supondrá que el fluido fluye laminarmente desde la izquierda de modo que en los bordes izquierdo y derecho

$$-v = \frac{\partial \psi}{\partial x} = 0 \qquad \text{a izquierda y derecha la velocidad es solo horizontal}$$

Se supondrá que la componente horizontal de la velocidad hidrodinámica arriba $\overline{BC}$ y abajo $\overline{AD}$ vale v_0. Es decir, arriba y abajo se tomará

$$u = \psi_y = v_0 \qquad \text{arriba y abajo la componente horizontal de la velocidad está fija}$$

Respecto a la vorticidad, sin mayores argumentos se supondrá que es cero abajo, arriba y a la izquierda. En cambio a la derecha satisfará $\partial\zeta/\partial x = 0$: ley de Kelvin sobre la conservación de la vorticidad que establece que en fluidos sin viscosidad la vorticidad se conserva.

Todas las condiciones de borde impuestas hasta aquí se hacen sobre un borde ficticio y son simplificaciones del problema. La condición física típica de hidrodinámica es que la velocidad es cero en los puntos de contacto con un sólido, es decir, ψ es constante en el perímetro $EFGH$. Puesto que ψ está definido salvo por una constante aditiva, se toma

$$\psi = 0 \qquad \text{en } EFGH$$

La condición sobre ζ en este perímetro será discutida más adelante.

Adimensionalización: Sea L una distancia caractarística L del problema y se define los campos adimensionales como sigue:

$$\psi = v_0 L\, \hat{\psi} \qquad \zeta = \frac{v_0}{L}\, \hat{\zeta} \tag{7.5.2}$$

Los campos $\hat{X}$ son adimensionales. Puesto que en todo lo que sigue de este problema se trata solo con los campos adimensionales no se pondrá el acento. En cambio las pocas veces que se haga referencia a los campos con dimensiones se les sobrepondrá una barra: $\overline{\psi}$ y $\overline{\zeta}$

La ecuación (7.3.11) se escribe

$$v_0 L \left(\frac{\psi_{i+1,k} - 2\psi_{i,k} + \psi_{i-1,k}}{h^2} + \frac{\psi_{i,k+1} - 2\psi_{i,k} + \psi_{i,k-1}}{h^2} \right) = \frac{v_0}{L}\, \zeta_{i,k}$$

Los factores v_0 se cancelan y, si se escoge $L = h$, no aparece ninguna constante en esta ecuación, reduciéndose a

$$\psi_{i,k} = \frac{1}{4} \left[\psi_{i+1,k} + \psi_{i-1,k} + \psi_{i,k+1} + \psi_{i,k-1} - \zeta_{i,k} \right] \tag{7.5.3}$$

que es la primera ecuación que se usará para iterar.

En forma similar (7.3.10) puede ser convertida en

$$\begin{aligned}\zeta_{ik} &= \frac{1}{4}\left[\zeta_{i+1,k}+\zeta_{i-1,k}+\zeta_{i,k+1}+\zeta_{i,k-1}+\frac{R}{4}\left\{(\psi_{i+1,k}-\psi_{i-1,k})(\zeta_{i,k+1}-\zeta_{i,k-1})\right.\right.\\ &\left.\left.-(\psi_{i,k+1}-\psi_{i,k-1})(\zeta_{i+1,k}-\zeta_{i-1,k})\right\}\right] \qquad (7.5.4)\end{aligned}$$

en esta expresión

$$R=\frac{h v_0}{\nu}$$

es una espacie de número de Reynolds. Un número de Reynolds con significado físico es $\overline{EF}\, v_0/\nu$.

7.5.2. Las ecuaciones en los bordes

Ecuaciones para ψ: Ya se ha dicho que $\psi = 0$ en todo el borde con el obstáculo.

En $\boxed{AB}$ se tiene $\psi_x = 0$, que se expresa como

$$\psi_{0,k} = \psi_{1,k} \qquad (7.5.5)$$

Usando esta relación en (7.5.3) tomada con $i = 1$ da

$$4\psi_{1,k} = \psi_{2,k} + \underbrace{\psi_{0,k}}_{\psi_{1,k}} + \psi_{1,k+1} + \psi_{1,k-1} - \zeta_{1,k}$$

que permite escribir, ya con ω incorporada,

$$\psi_{1,k} = (1-\omega)\psi_{1,k} + \frac{\omega}{3}\left(\psi_{2,k}+\psi_{1,k+1}+\psi_{1,k-1}-\zeta_{1,k}\right) \qquad (7.5.6)$$

En $\boxed{AD}$ se tiene $\overline{\psi}_y = v_0$, es decir, $\psi_{i,1} - \psi_{i,0} = 1$, esto es

$$\psi_{i,0} = \psi_{i,1} - 1 \qquad (7.5.7)$$

que se usa en (7.5.3) y se obtiene, en forma similar que en el caso anterior,

$$\psi_{i,1} = (1-\omega)\psi_{i,1} + \frac{\omega}{3}\left(\psi_{i+1,1}+\psi_{i-1,1}+\psi_{i,2}-1-\zeta_{i,1}\right) \qquad (7.5.8)$$

En forma entéramente análoga, en $\boxed{BC}$ se satisface

$$\psi_{i,N_2} = \psi_{i,N_2-1} + 1 \qquad (7.5.9)$$

y también

$$\psi_{i,N_2-1} = (1-\omega)\psi_{i,N_2-1} + \frac{\omega}{3}\left(\psi_{i+1,N_2-1}+\psi_{i-1,N_2-1}+1+\psi_{i,N_2-2}-\zeta_{i,N_2-1}\right) \qquad (7.5.10)$$

Y también en forma entéramente análoga, en $\boxed{CD}$ se satisface

$$\psi_{N_1,k} = \psi_{N_1-1,k} \qquad (7.5.11)$$

$$\psi_{N_1-1,k} = (1-\omega)\psi_{N_1-1,k} + \frac{\omega}{3}\left(\psi_{N_1-2,k}+\psi_{N_1-1,k+1}+\psi_{N_1-1,k-1}-\zeta_{N_1-1,k}\right) \qquad (7.5.12)$$

Ecuaciones para ζ en bordes exteriores: Las condiciones sobre ζ en $ABCD$ son sencillas,

$$\begin{aligned} \zeta_{0,k} &= 0 \quad \text{izquierda} \\ \zeta_{i,0} &= 0 \quad \text{abajo} \\ \zeta_{i,N_2} &= 0 \quad \text{arriba} \end{aligned} \tag{7.5.13}$$

A la derecha se exige $\zeta_x = 0$ (ley de Kelvin de conservación de la vorticidad) lo que conduce a

$$\zeta_{N_1,k} = \zeta_{N_1-1,k} \tag{7.5.14}$$

Tomando (7.5.4) con $i = N_1 - 1$ se obtiene

$$\begin{aligned} 4\zeta_{N_1-1,k} &= \overbrace{\zeta_{N_1,k}}^{\zeta_{N_1-1,k}} + \zeta_{N_1-2,k} + \zeta_{N_1-1,k+1} + \zeta_{N_1-1,k-1} \\ &\quad + \frac{R}{4}\Big\{(\psi_{N_1,k} - \psi_{N_1-2,k})(\zeta_{N_1-1,k+1} - \zeta_{N_1-1,k-1}) \\ &\quad - (\psi_{N_1-1,k+1} - \psi_{N_1-1,k-1})(\zeta_{N_1,k} - \zeta_{N_1-2,k})\Big\} \end{aligned}$$

que trivialmente se convierte en

$$\begin{aligned} \zeta_{N_1-1,k} &= (1-\omega)\zeta_{N_1-1,k} + \frac{\omega}{3}\Big(\zeta_{N_1-2,k} + \zeta_{N_1-1,k+1} + \zeta_{N_1-1,k-1} \\ &\quad + \frac{R}{4}\Big\{(\psi_{N_1,k} - \psi_{N_1-2,k})(\zeta_{N_1-1,k+1} - \zeta_{N_1-1,k-1}) \\ &\quad - (\psi_{N_1-1,k+1} - \psi_{N_1-1,k-1})(\zeta_{N_1,k} - \zeta_{N_1-2,k})\Big\}\Big) \end{aligned} \tag{7.5.15}$$

Ecuaciones para ζ en bordes del obstáculo: Para obtener las condiciones de borde para ζ en torno al obstáculo se hace una expansión de Taylor de ψ para un punto (i, k_2+1) con $i_1 < i < i_2$,

$$\overline{\psi}_{i,k_2+1} = \overline{\psi}_{i,k_2} + h\left(\frac{\partial\overline{\psi}}{\partial y}\right)_{i,k_2} + \frac{h^2}{2}\left(\frac{\partial^2\overline{\psi}}{\partial y^2}\right)_{i,k_2} + \ldots \tag{7.5.16}$$

El primer término de la derecha es nulo porque ψ es nulo alrededor de todo el obstáculo. El segundo es cero porque corresponde a la componente tangencial de la velocidad en contacto con un sólido. Ambas componentes de la velocidad son cero en los puntos de contacto con un sólido. En particular v_y es cero a lo largo de FG, es decir, $\frac{\partial v_y}{\partial x} = 0$ sobre FG, lo que implica que $\partial^2\overline{\psi}/\partial x^2 = 0$ sobre FG. Pero en general,

$$\overline{\zeta} = \frac{\partial^2\overline{\psi}}{\partial x^2} + \frac{\partial^2\overline{\psi}}{\partial y^2}$$

entonces sobre FG es

$$\overline{\zeta} = \left[\frac{\partial^2 \overline{\psi}}{\partial y^2}\right]_{FG}$$

que se puede reemplazar en (7.5.16) obteniéndose

$$\overline{\psi}_{i,k_2+1} = \frac{h^2}{2}\overline{\zeta}_{i,k_2}$$

que, al pasar a campos adimensionales se reduce (y las otras se obtienen por métodos semejantes):

$$\begin{aligned}
\zeta_{i,k_2} &= 2\,\psi_{i,k_2+1} \qquad FG \\
\zeta_{i_1,k} &= 2\,\psi_{i_1-1,k} \qquad EF \\
\zeta_{i_2,k} &= 2\,\psi_{i_2+1,k} \qquad GH \qquad (7.5.17) \\
\zeta_{i,k_1} &= 2\,\psi_{i,k_1-1} \qquad HE
\end{aligned}$$

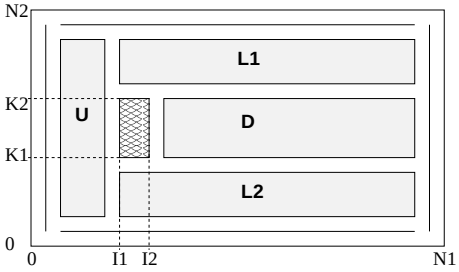


Figura 7.6: Como explica el texto, conviene definir vaeias zonas de integración.

Las rutinas de integración deben tomar en cuenta con cuidado el rango de las variables enteras (i,k). En la figura se muestra un caso en que $0 \le i \le N_1$ y $0 \le k \le N_2$. El obstáculo es un rectángulo con vértices opuestos (i_1,k_1) y (i_2,k_2). Debe tenerse rutinas que aplican todas las condiciones de borde del perímetro. En estas rutinas debe primero actualizarse los puntos inmediatos al perímetro y sólo entonces los del perímetro mismo. Debe haber otra rutina que aplica las relaciones asociadas al contacto con el obstáculo y finalmente rutinas que integran en los "puntos interiores" de las zonas que en la figura aparecen como U (up wind), $L1$ y $L2$ que son laterales y D (down wind). Naturalmente que puede escogerse integrar en forma algo diferente, pero es importante escoger correctamente los rangos de integración de cada rutina para no deshacer, por ejemplo, las condiciones de borde.

Problemas

1. **Ecuación de Poisson**: (a) Integre la ecuación de Poisson 2D para el potencial electrostático

   $$\nabla^2 V(x,y) = -\frac{1}{\varepsilon_0}\rho(x,y) \equiv G(x,y)$$

   dentro de una caja cuadrada de $L \times L$ con $L = 10$. El potencial V en el perímetro vale: $V(x,0) = V(x,L) = x/10$, $\quad V(0,y) = 0 \quad$ y $\quad V(L,y) = 1$. Utilice un coeficiente de relajación $\omega \ge 1{,}5$.

   Dentro de la caja hay una zona rectangular con vértices opuestos en $(2{,}4;\ 2{,}4)$ y $(4{,}4;\ 4{,}4)$ dentro de la cual $G(x,y) = -1$. Vale cero fuera de esa zona.

2. **Hidrodinámica 2D**: Integre las ecuaciones hidrodinámicas 2D definidas en este capítulo para las funciones corriente, vorticidad y temperatura usando el método de relajación para el caso de un fluido en una caja rectangular de 3×1 tal que la pared de abajo es más caliente que la de arriba. Considere además que paredes laterales son perfectamente aislantes. La admimensionalización definida en el capítulo condujo a ecuaciones discretas que

tienen solo dos parámetros: μ y Pr. Fije $Pr = 1$ y varíe μ desde $0{,}001$ con $d\mu = 0{,}0001$. Si bien la diferencia de temperatura arriba y abajo va cambiando, por defínción la diferencia adimensionalizada es siempre la unidad.

Mientras está iterando con un μ fijo se desea saber si hay convergencia y cuándo. Una vez se ha convergido guarde en archivo el valor de μ, el valor de dT y el número de iteraciones que fueron necesarias. Nunca barra menos de 50 mil veces antes de comenzar a observar los cambios que sufre dT. Esto debe hacerse así porque hay que dar la oportunidad para que la solución se desestabilice hacia una nueva forma más estable. Siempre tome los campos del caso anterior como los valores iniciales.

3. **Flujo en 2D**: Integre las ecuaciones 2D que determinan las funciones corriente y vorticidad en la geometría que indica la figura 7.5. Un fluido viene de la izquierda y hay un obstaculo rectangular EFGH. Para efectos prácticos se escoge un dominio finito ABCD que no corresponde límites físico sino al borde del dominio que se usa para integrar y al cual se le asocia condiciones de borde sencillas.

   Antes de adimensionalizar se escoge las siguientes condiciones de borde: $\psi_x = 0$ en AB y CD, $\psi_y = v_0$ en BC y AD; $\zeta = 0$ en DA, AB y BC, en cambio $\zeta_x = 0$ en CD. En el perímetro EFGH se toma $\psi = 0$ mientras que las condiciones para ζ son más complicadas y son las que se discuten en clases. Se adimensionaliza usando $\psi \to hv_0\psi$ y $\zeta \to \frac{v_0}{h}\zeta$. Esta forma de proceder hace que aparezca un único parámetro en las ecuaciones, $R \equiv \frac{hv_0}{\nu}$, donde ν es la viscosidad cinemática. Aunque no es totalmente correcto se lo denomina número de Reynolds.

   Utilice una grilla definida por coordenadas enteras (N_x, N_y): A=$(0,0)$, C=$(1000,120)$, E=$(60,40)$, G=$(80,80)$. Conviene que los campos inicialmente valgan 0.0 excepto si alguna condición de borde exigiera otra cosa.

   a) Integre el caso R =0.8. usando al menos dos valores mayores que la unidad del coeficiente de sobrerelajación ω. Indique cuántas iteraciones fueron necesarias. Presente en un solo gráfico una familia de curvas iso-ψ y en otro una familia de curvas iso-$\log|\zeta|$. Tabule los valores de ψ y ζ en los puntos (N_x, N_y) de la grilla que corresponden a $N_y = 70$ y N_x es múltiplo de 50, esto es N_x es 0, 50, 100, 150, ... hasta 1000.

   b) Integre el caso R =4.5. Luego de haber iterado al menos 50 mil veces guarde los valores de ψ y ζ y presente gráficos de las curvas iso-ψ y las curvas iso-$\log|\zeta|$ y la misma tabla definida en (a). Una vez alcanzada la iteración 50 mil, guarde los valores de $\psi_{90,80}$ y de $\zeta_{90,80}$ y coloque *tic*=0. Luego cada 5000 iteraciones aumenta *tic* en uno y vuelva a guardar $\psi_{90,80}$ y de $\zeta_{90,80}$ hasta tener tantos puntos que se vea una gráfico interesante (por sobre 200 puntos). Haga gráficos de $\psi_{90,80}$ y de $\zeta_{90,80}$ versus *tic*.

Capítulo 8

Ecuaciones parabólicas

8.1. Ecuación general

Los métodos que se presenta a continuación son generalizables a ecuaciones de la forma

$$\frac{\partial F}{\partial t} + \frac{\partial}{\partial x}\left(B(t,x,F)\,\frac{\partial F}{\partial x}\right) = S(t,x,F) \tag{8.1.1}$$

tanto con condiciones de borde rígidas como derivativas.

La ecuación propiamente parabólica más general tiene la forma

$$aF_{tt} + bF_{xx} + \sqrt{ab}F_{tx} + \ldots = 0 \tag{8.1.2}$$

donde los subíndices señalan derivadas, los coeficientes a y b representan funciones de (t,x) y los puntos suspensivos representan términos con tan solo primeras derivadas o sin derivadas. La ecuación (8.1.1) es un caso con $a = 0$.

Suele ocurrir que estas ecuaciones admitan separación de variables. Eso las convierte en ecuaciones diferenciales ordinarias lo que no es de interés en este capítulo.

8.2. Ecuaciones típicas

8.2.1. Ecuación de calor

Las ecuaciones parabólicas más conocidas en física posiblemente son la ecuación de difusión de calor que en su forma tridimensional es

$$\frac{\partial T}{\partial t} = \kappa \nabla^2 T \tag{8.2.1}$$

pero que normalmente veremos en una sola dimensión espacial

$$\frac{\partial T}{\partial t} = \kappa \frac{\partial^2 T}{\partial x^2} \tag{8.2.2}$$

Puede pensarse que se trata del problema de una barra muy larga con temperatura $T(t,x)$ que varía en el tiempo. Pero también puede verse un caso tridimensional con simetría esférica. La ecuación no tiene exactamente la forma (8.2.2) pero es igualmente tratable.

Más en general se puede estudiar ecuaciones como

$$\frac{\partial \Phi}{\partial t} = \frac{\partial^2 \Phi}{\partial x^2} + S(x,t) \tag{8.2.3}$$

8.2.2. Ecuación de Schrödinger

Un caso interesante es el de la ecuación de Schrödinger dependiente del tiempo, que podría ser

$$\begin{aligned} \frac{\partial \Psi}{\partial t} &= -iH\Psi \\ H &= -\nabla^2 + V \\ \Psi(t=0) &= \Psi_0(x), \qquad \Psi(t,\pm\infty) = 0 \end{aligned} \tag{8.2.4}$$

8.2.3. Otros ejemplos de ecuaciones parabólicas

La ecuación de Burgers:

$$\frac{\partial U}{\partial t} = -U\frac{\partial U}{\partial x} + \frac{1}{\mathrm{Re}}\frac{\partial^2 U}{\partial x^2} \tag{8.2.5}$$

proviene de hidrodinámica. Sin el último término se la llama *ecuación de Burgers* con viscosidad nula y Re es el número de Reynolds.

La ecuación de Swift-Hohenberg:

$$\frac{\partial U}{\partial t} = \lambda U\,(1-U^2) - (1+\nabla^2)^2 U \tag{8.2.6}$$

Esta ecuación tiene derivadas de cuarto orden en las coordenadas, pero es comúnmente considerada como parabólica y se integra con los mismos métodos.

Los autores se inspiraron en el estudio del fenómeno de convección térmica. Si se resuelve con una sola dimensión espacial y condiciones iniciales arbitrarias lo típico es que la solución evoluciones hacia un conjunto de “patrones” que se desplazan a velocidad constante.

Hidrodinámica incompresible 2D dependiente del tiempo: Si en las ecuaciones con que se presenta §7.3.1 se impone $\rho = \rho_0$ estrictamente en todas partes, la primera ecuación es $\nabla \cdot \vec{v} = 0$ que, como ya se sabe, se puede reducir en 2D a la introducción de la función corriente ψ tal que $v_x = \psi_y$ y $v_y = -\psi_x$. Si se define $\zeta = \nabla^2\psi$ se encuentra una ecuación dinámica para ζ, reemplazando a (7.3.10) con lo que las ecuaciones son

$$\frac{\partial \zeta}{\partial t} = \psi_x\zeta_y - \psi_y\zeta_x + \nu\nabla^2\zeta \tag{8.2.7}$$

$$0 = \nabla^2\psi - \zeta \tag{8.2.8}$$

Esto es, una ecuación parabólica acoplada con una ecuación elíptica.

8.3. Adimensionalización de la ecuación de difusión de calor 1D

Si en la ecuación (8.2.2) se hace el cambio de variables y función a cantidades prima adimensionales:

$$x = L x', \qquad t = \frac{L^2 t'}{\kappa}, \qquad T = T_0 T' \tag{8.3.1}$$

donde L normalmente es el largo del intervalo, por lo cual ahora $0 \leq x' \leq 1$ y T_0 es alguna temperatura característica, mientras que T' es adminesional.

De esta manera, si además se elimina las primas, la ecuación queda

$$\frac{\partial T}{\partial t} = \frac{\partial^2 T}{\partial x^2} \qquad 0 \leq x \leq 1 \tag{8.3.2}$$

8.4. Integración explícita directa

El problema adimensionalizado anterior se puede escribir en forma discreta en la forma

$$\frac{T_k^{n+1} - T_k^n}{\varepsilon} = \frac{T_{k+1}^n - 2T_k^n + T_{k-1}^n}{h^2} \tag{8.4.1}$$

que conduce a

$$T_k^{n+1} = r\, T_{k-1}^n + (1-2r)\, T_k^n + r\, T_{k+1}^n \tag{8.4.2}$$

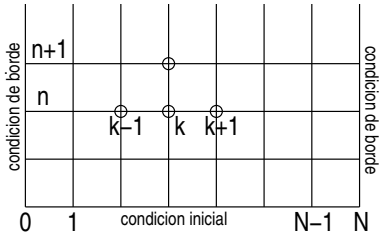


Figura 8.1: *Las variables (x,t) se discretizan en la forma $x = kh$, con $k = 0,1,\ldots N$ y $t = n\varepsilon$, con $n = 0,1,\ldots$ Los valores $k = 0$ y $k = N$ corresponden a los bordes. Se subraya que el intervalo X de integración está dividido en N celdas de largo h.*

donde $r = \dfrac{\varepsilon}{h^2}$. Esta forma de integrar puede dar buenos resultados. En la literatura se ha demostrado que si $0 < r < \frac{1}{2}$ esta regla de iteración no diverge. Sin embargo dentro del rango permitido no da resultados muy precisos salvo que r sea bastante pequeño, pero esto implica ε pequeño, es decir, la integración avanza lentamente en el tiempo.

8.4.1. Condiciones de borde rígidas

La integración (8.4.2) con condiciones de borde rígidas es trivial ya que basta con usar la fórmula iterativa (8.4.2) sucesivamente con $n = 1$, $n = 2$ ... recorriendo cada vez de $k = 1$ hasta $k = N - 1$.

8.4.2. Condiciones de borde con derivada

Puede haber varias condiciones de borde con derivada. Algunas son:

$$\frac{\partial T}{\partial n} = 0 \qquad \text{borde aislante perfecto} \tag{8.4.3}$$

o bien un borde que tiene asociada una tasa de absorcióno emisión de energía. En cada punta la versión adimensionalizada de condiciones de borde derivativas se puede escribirse en la forma

$$\left[\frac{\partial T}{\partial x}\right]_{\text{izq}} = \mu\,(T - T_{\text{izq}}), \qquad \left[\frac{\partial T}{\partial x}\right]_{\text{der}} = -\mu\,(T - T_{\text{der}}) \tag{8.4.4}$$

Estas condiciones se podrían discretizar en la forma

$$\frac{T_1^n - T_0^n}{h} = \mu\,(T_0^n - T_{\text{izq}}) \qquad \frac{T_N^n - T_{N-1}^n}{h} = -\mu\,(T_0^n - T_{\text{der}}) \tag{8.4.5}$$

pero es más conveniente agregar puntos ficticios más allá de los extremos e imponer

$$\frac{T_1^n - T_{-1}^n}{2h} = \mu\,(T_0^n - T_{\text{izq}}) \qquad \frac{T_{N+1}^n - T_{N-1}^n}{2h} = -\mu\,(T_N^n - T_{\text{der}}) \tag{8.4.6}$$

Este método se usa a continuación en un caso concreto.

Se va a considerar el caso

$$\begin{array}{rll} \frac{\partial T}{\partial t} & = \frac{\partial^2 T}{\partial x^2} & \\ T(0,x) & = 1 & \text{condición inicial} \\ T(t,x=0) & = 1 & \text{condición de borde rígida} \\ T'(t,x=1) & = \mu\,(T - T_{\text{der}}) & \text{condición de borde derivativa} \end{array} \tag{8.4.7}$$

La última condición describe una vara cuyo extremo derecho dinámicamente ajusta su temperatura acercándola al valor T_{der}.

Al discretizar el método explícito (8.4.2) exige imponer $T_0^n = 1$. El el extremo derecho, como se ha dicho, conviene agregar un punto ficticio $k = N+1$ e imponer

$$\frac{T_{N+1}^n - T_{N-1}^n}{2h} = -\mu\,(T_N^n - T_{\text{der}}) \qquad \Longrightarrow \qquad T_{N+1}^n = T_{N-1}^n - 2\mu h\,(T_N^n - T_{\text{der}})$$

Puesto que en la expresión (8.4.2) para $k = N$ aparece T_{N+1}^n se debe hacer uso de la expresión anterior para obtener la regla correcta para T_N^{n+1} sin que aparezca el punto ficticio $k = N+1$.

Como se ha visto, el hecho que una de las condiciones de borde sea derivativa implica que existe una variable dinámica adicional que en este caso es T_N.

Si ambas condiciones de borde hubiesen sido derivativa se tendría $N+1$ variables dinámicas: $T_0,\ T_1, \ldots T_N$, esto es, todas ellas tendrían que ser iteradas.

8.4.3. Condiciones de borde periódicas

Si las condiciones de borde son periódicas se puede considerar la misma malla discreta que muestra la figura 8.1 tan solo que se debe considerar que los puntos correspondientes de $k = 0$ e $k = N$ deben considerarse idénticos. Así entonces la regla de iteración (8.4.2) sigue válida pero se debe tomar en cuenta que

$$T_0^{n+1} = r\,T_{N-1}^n + (1-2r)\,T_0^n + r\,T_1^n, \qquad T_{N-1}^{n+1} = r\,T_{N-2}^n + (1-2r)\,T_{N-1}^n + r\,T_0^n \tag{8.4.8}$$

8.5. El método de Du Fort-Frankel

En el caso de la ecuación

$$\frac{\partial U}{\partial t} = \kappa \frac{d^2 U}{dt^2}$$

los autores proponen

$$\frac{U_k^{n+1} - U_k^{n-1}}{2\varepsilon} = \kappa \frac{U_{k+1}^n + U_k^{n+1} + U_k^{n-1} + U_{k-1}^n}{h^2}$$

que lleva al algoritmo explícito

$$U_k^{n+1} = \frac{1-\alpha}{1+\alpha} U_k^{n-1} + \frac{\alpha}{1+\alpha}\left(U_{k+1}^n + U_{k-1}^n\right) \qquad \text{donde} \quad \alpha = \frac{2\kappa\varepsilon}{h^2} \tag{8.5.1}$$

Un defecto del método es que requiere condiciones iniciales en dos tiempos consecutivos. Una condición necesaria para que este método dé la solución correcta es que $\frac{\varepsilon}{h} \to 0$ lo que es automático si se exige que $\varepsilon = \mathscr{O}(h^2)$ y $h \to 0$.

8.6. El método tridiagonal

8.6.1. La ecuación de calor

A continuación se verá un método que no está limitado a que $r = \dfrac{\varepsilon}{h^2}$ sea pequeño y que es aplicable a una gran variedad de ecuaciones parabólicas.

Se comienza planteando nuevamente la ecuación

$$\frac{\partial T}{\partial t} = \frac{\partial^2 T}{\partial x^2}$$

con condiciones de borde que por el momento serán rígidas:

$$\begin{aligned} T(t,0) &= T_{\text{izq}}\,, \qquad T(t,1) = T_{\text{der}} \\ T(0,x) &= g(x) \qquad \text{con} \quad g(0) = T_{\text{izq}}\,, \quad g(1) = T_{\text{der}} \end{aligned} \tag{8.6.1}$$

y $g(x)$ es alguna condición inicial dada.

En forma discreta las condiciones de borde son

$$T_0^n = T_{\text{izq}}\,, \qquad T_N^n = T_{\text{der}} \tag{8.6.2}$$

y la condición inicial es la función $g(x)$ que se abreviará, en notación discreta, como g_k,

$$T_k^0 = g_k, \quad \text{con} \quad g_0 = T_{\text{izq}}\,, \quad g_N = T_{\text{der}} \tag{8.6.3}$$

Esta vez se plantea la ecuación discretizada

$$\frac{T_k^{n+1} - T_k^n}{\varepsilon} = \frac{a}{2}\left[\frac{T_{k+1}^{n+1} - 2T_k^{n+1} + T_{k-1}^{n+1}}{h^2}\right] + \frac{2-a}{2}\left[\frac{T_{k+1}^n - 2T_k^n + T_{k-1}^n}{h^2}\right] \tag{8.6.4}$$

donde $1 \leq k \leq N-1$, lo que implica que en particular son necesarios los valores $T_0 = T_{\text{izq}}$ y $T_N = T_{\text{der}}$.

Escogiendo $a = 0$ se tendría el método explícito y queda excluido en el contexto actual. El caso $a = 1$ conduce al método conocido como de Crank-Nicolson y con $a = 2$ se lo llama el *método implícito*, pero a puede tomar un continuo de valores. El caso $a = 1$ de Crank-Nicolson es especial porque corresponde a

$$\left(\frac{\partial T}{\partial t}\right)_k^{n+\frac{1}{2}} = \frac{1}{2}\left[\left(\nabla^2 T\right)_k^{n+1} + \left(\nabla^2 T\right)_k^{n}\right]$$

Para a cualquiera la ecuación (8.6.4) puede ser reescrita con las cantidades con $n+1$ a la izquierda y con n a la derecha

$$-ar\,T_{k-1}^{n+1} + 2(1+ar)T_k^{n+1} - ar\,T_{k+1}^{n+1} = (2-a)\,rT_{k-1}^{n} + 2\,(1-(2-a)\,r)\,T_k^n + (2-a)\,rT_{k+1}^n \tag{8.6.5}$$

El problema consiste en obtener los T^{n+1} suponiendo que se conoce los T^n. La ecuación anterior puede ser vista de la forma

$$A_k^-\,\phi_{k-1} + A_k^0\,\phi_k + A_k^+\,\phi_{k+1} = b_k \qquad \text{con} \quad 1 \leq k \leq N-1 \tag{8.6.6}$$

De arriba se ve que $\phi_k = T_k^{n+1}$, $A_k^+ = -ar$, $A_k^0 = 2\,(1+ar)$, $A_k^- = -ar$ mientras que

$$b_k^n = (2-a)\,rT_{k-1}^n + 2(1-(2-a)\,r)\,T_k^n + (2-a)\,rT_{k+1}^n \tag{8.6.7}$$

que es un valor conocido cuando se está por obtener los T_k^{n+1}.

Más en general, el problema es resolver un problema de la forma

$$M\vec{\phi} = \vec{b} \tag{8.6.8}$$

donde M es una matriz rectangular tridiagonal:

$$M = \begin{bmatrix} A_1^- & A_1^0 & A_1^+ & 0 & .. & & 0 \\ 0 & A_2^- & A_2^0 & A_2^+ & 0 & & 0 \\ 0 & 0 & A_3^- & A_3^0 & A_3^+ & & 0 \\ 0 & 0 & 0 & & & & 0 \\ 0 & 0 & .. & .. & .. & .. & 0 \\ 0 & 0 & & .. & 0 & A_{N-1}^- \;\; A_{N-1}^0 & A_{N-1}^+ \end{bmatrix}$$

Se ve que M tiene $N+1$ columnas (numeradas del 0 al N) y $N-1$ filas. Esta matriz multiplica al vector $\vec{\phi} = \{\phi_0, \phi_1, ..., \phi_{N-1}, \phi_N\}$ donde las componentes ϕ_0 y ϕ_N están definidas por las condiciones rígidas. El vector $\vec{b}$ tiene $N-1$ componentes. La ecuación (8.6.8) conduce, entonces a $N-1$ ecuaciones para las $N-1$ incógnitas $\vec{\varphi} = \{\phi_1, ..., \phi_{N-1}\}$. Este problema es idéntico al problema

$$\begin{aligned} A_1^-\,\phi_0 + (\mathbf{A}\vec{\varphi})_1 &= b_1 \qquad \Rightarrow \qquad (\mathbf{A}\vec{\varphi})_1 = b_1 - A_1^-\,\phi_0 \\ (\mathbf{A}\vec{\varphi})_k &= b_k \qquad 2 \leq k \leq N-2 \\ (\mathbf{A}\vec{\varphi})_{N-1} + A_{N-1}^+\,\phi_N &= b_{N-1} \qquad \Rightarrow \qquad (\mathbf{A}\vec{\varphi})_{N-1} = b_{N-1} - A_{N-1}^+\,\phi_N \end{aligned} \tag{8.6.9}$$

donde **A** es una matriz tridiagonal de $(N-1)\times(N-1)$. De esta manera el problema se reduce auténticamente el de una matriz cuadrada tridiagonal **A** de $(N-1)\times(N-1)$ y un vector $\vec{\tilde{b}}$ modificado por las condiciones de borde:

$$\mathbf{A}\vec{\varphi}=\vec{\tilde{b}} \qquad \text{cuya solución formal es} \qquad \vec{\varphi}=\mathbf{A}^{-1}\vec{\tilde{b}} \tag{8.6.10}$$

Nótese que si se tuviera que $\phi_0=\phi_N=0$ el problema descrito en (8.6.9) es sencillamente

$$\mathbf{A}\vec{\varphi}=\vec{b}$$

con el vector $\vec{b}$ original.

8.6.2. El algoritmo para el caso rígido

El método que se explica a continuación se usa para resolver ecuaciones parabólicas lineales.

Se plantea resolver para $\vec{\phi}$ la ecuación

$$\mathbf{A}\,\vec{\phi}=\vec{b} \tag{8.6.11}$$

donde **A** es una matriz de $(N-1)\times(N-1)$ tridiagonal, es decir, (8.6.11) es

$$A_k^-\,\phi_{k-1}+A_k^0\,\phi_k+A_k^+\,\phi_{k+1}=b_k \quad \text{con} \begin{cases} k & =1,\ldots,N-1 \\ \phi_0 & =U_0 \\ \phi_N & =U_1 \end{cases} \tag{8.6.12}$$

Se destaca que si bien en la ecuación anterior el rango de k es de 1 a $N-1$, los ϕ_k están definidos también con $k=0$ y $k=N$. El problema que se plantea implica, de alguna forma, invertir la matriz **A**. Se va a encontrar un algoritmo que permite encontrar $\vec{\phi}$ en pocos pasos.

Ecuaciones lineales de recurrencia como estas siempre tienen solución de la forma

$$\phi_{k+1}=\alpha_k\phi_k+\beta_k \tag{8.6.13}$$

Reemplazando este ϕ_{k+1} en (8.6.12) se obtiene

$$A_k^+\,(\alpha_k\,\phi_k+\beta_k)+A_k^0\phi_k+A_k^-\,\phi_{k-1}=b_k$$

que lleva a

$$\phi_k=\frac{-A_k^-}{A_k^+\,\alpha_k+A_k^0}\,\phi_{k-1}+\frac{b_k-A_k^+\beta_k}{A_k^+\,\alpha_k+A_k^0}$$

que tiene la forma de (8.6.13) y por lo tanto se debe identificar

$$\alpha_{k-1} = \frac{-A_k^-}{A_k^+\,\alpha_k+A_k^0} \tag{8.6.14}$$

$$\beta_{k-1} = \frac{b_k-A_k^+\beta_k}{A_k^+\,\alpha_k+A_k^0} \tag{8.6.15}$$

que son ecuaciones de recurrencia para los α_k y β_k.

Para que todo sea consistente se debe cuidar los puntos del borde. La ecuación (8.6.12) para $k = N-1$ debe coincidir con (8.6.13) con $k = N-2$. Pero ellas son

$$A^+_{N-1}U_1 + A^0_1\,\phi_{N-1} + A^-_{N-1}\,\phi_{N-2} = b_{N-1}\,, \qquad \phi_{N-1} = \alpha_{N-2}\phi_{N-2} + \beta_{N-2} \tag{8.6.16}$$

Comparándolas se obtiene $\alpha_{N-1} = -A^-_N/A^0_N$ y $\beta_{N-1} = (b_N - A^+_N U_1)/A^0_N$.

Juntando estos resultados con las relaciones de recurrencia (8.6.14) y (8.6.15) con $k = N$ se obtiene que

$$\alpha_{N-1} = 0\,, \qquad \beta_{N-1} = U_1 \equiv \phi_N \tag{8.6.17}$$

Con esto se usa las relaciones de recurrencia (8.6.14) y (8.6.15) para obtener en forma descendente todos los α_k y todos los β_k. Una vez que estos coeficientes se conocen se usa (8.6.13) en forma ascendente, sabiendo que $\phi_0 = U_0$ para obtener todos los ϕ_k. El problema ha sido resuelto.

8.6.3. Ecuación de calor con conductividad variable

La ecuación de calor en espacio unidimensional para $T(t,x)$ es

$$\begin{aligned} \frac{\partial T}{\partial t} &= \frac{\partial}{\partial x}\left(K(x)\frac{\partial T}{\partial x}\right) \\ &= \frac{dK}{dx}\frac{\partial T}{\partial x} + K\frac{\partial^2 T}{\partial x^2} \end{aligned} \tag{8.6.18}$$

Discretizando con igual peso en los instante $n+1$ y n se obtiene

$$\begin{aligned} \frac{T^{n+1}_k - T^n_k}{\varepsilon} &= \frac{1}{2}\left[\frac{K_{k+1}-K_{k-1}}{2h}\frac{T^{n+1}_{k+1}-T^{n+1}_{k-1}}{2h} + K_k\frac{T^{n+1}_{k+1}-2T^{n+1}_k+T^{n+1}_{k-1}}{h^2}\right] \\ &\quad + \frac{1}{2}\left[\frac{K_{k+1}-K_{k-1}}{2h}\frac{T^{n}_{k+1}-T^{n}_{k-1}}{2h} + K_k\frac{T^{n}_{k+1}-2T^{n}_k+T^{n}_{k-1}}{h^2}\right] \end{aligned} \tag{8.6.19}$$

Definiendo $r \equiv \varepsilon/h^2$ la ecuación puede ser llevada a la forma

$$r(K_{k+1} - 4K_k - K_{k-1})\,T^{n+1}_{k-1} + 8(1+rK_k)\,T^{n+1}_k + r(-K_{k+1} - 4K_k + K_{k-1})\,T^{n+1}_{k+1} = B^n_k \tag{8.6.20}$$

donde

$$B^n_k = r(-K_{k+1} + 4K_k + K_{k-1})\,T^n_{k-1} + 8(1-rK_k)\,T^n_k + r(K_{k+1} + 4K_k - K_{k-1})\,T^n_{k+1}$$

vinéndose que (8.6.20) tiene la forma (8.6.6) ya estudiada.

8.6.4. El caso con condiciones de borde periódicas

Esta vez consideraremos la forma de aplicar le método tridiagonal para la ecuación

$$A^-_k\,\phi_{k-1} + A^0_k\,\phi_k + A^+_k\,\phi_{k+1} = b_k \qquad \text{válida con} \quad 1 \le k \le N-2 \tag{8.6.21}$$

y, debido a la periodicida, la ecuación anterior toma formas especiales en dos puntos

$$\begin{array}{llll} A^-_0\,\phi_{N-1} & +A^0_0\,\phi_0 & +A^+_0\,\phi_1 & = b_0 \\ A^-_{N-1}\,\phi_{N-2} & +A^0_{N-1}\,\phi_{N-1} & +A^+_{N-1}\,\phi_0 & = b_{N-1} \end{array} \tag{8.6.22}$$

lo que da un conjunto de N ecuaciones lineales para las N incógnitas $\phi_0, \phi_1, \ldots, \phi_{N-1}$.

El problema anterior corresponde a enfrentar un problema de la forma

$$\mathbf{M}\,\vec{\phi} = \vec{b} \qquad \Longleftrightarrow \qquad M_{ij}\phi_j = b_i \tag{8.6.23}$$

donde la matriz M es de $N \times N$ y tiene la forma

$$\mathbf{M} = \begin{bmatrix} A_0^0 & A_0^+ & 0 & .. & 0 & A_0^- \\ A_1^- & A_1^0 & A_1^+ & 0 & .. & 0 \\ 0 & A_2^- & A_2^0 & A_2^+ & & 0 \\ 0 & 0 & & & & 0 \\ 0 & .. & 0 & A_{N-1}^- & A_{N-2}^0 & A_{N-2}^+ \\ A_{N-1}^+ & 0 & .. & 0 & A_{N-1}^- & A_{N-1}^0 \end{bmatrix} \tag{8.6.24}$$

Los elementos de esta matriz los denotamos M_{ij} y tanto i como j varían entre 0 y $N-1$. La matriz M tiene no nulas las tres diaginales centrales y dos elementos de vértice como se aprecia en (8.6.24).

La matriz **M** puede escribirse como una matriz tridiagonal **A** más valores en los vértices:

$$M_{ij} = A_{ij} + A_0^-\,\delta_{i0}\delta_{j(N-1)} + A_{N-1}^+\,\delta_{i(N-1)}\delta_{0j} \tag{8.6.25}$$

Si se definen los vectores $\vec{u} = \{a, 0, .., 0, b\}$ y $\vec{w} = \{c, 0, .., 0, d\}$, esto es $u_i = a\delta_{i0} + b\delta_{i(N-1)}$ y $w_j = c\delta_{j0} + d\delta_{j(N-1)}$, se obtiene que $u_i w_j = ac\delta_{i0}\delta_{0j} + ad\delta_{i0}\delta_{(N-1)j} + bc\delta_{i(N-1)}\delta_{0j} + bd\delta_{i(N-1)}\delta_{(N-1)j}$. Se necesita que $ad = A_0^-$ y que $bc = A_{N-1}^+$. La matriz M escrita en la forma (8.6.25), como una nueva matriz tridiagonal **A** más una matriz $\vec{u} \otimes \vec{w}$ cuyos únicos elementos no nulos están en los cuatro vértices matricialmente se puede representar por

$$\mathbf{M} = \mathbf{A} + \vec{u} \otimes \vec{v}$$

Escogiendo $w_0 = \frac{A_{N-1}^+}{b}$ y $w_{N-1} = \frac{A_0^-}{a}$ se obtiene

$$\vec{u} \otimes \vec{w} = \begin{bmatrix} \frac{a}{b}A_{N-1}^+ & 0 & 0 & .. & 0 & A_0^- \\ 0 & 0 & 0 & 0 & .. & 0 \\ 0 & 0 & 0 & 0 & & 0 \\ 0 & 0 & & & & 0 \\ 0 & .. & 0 & 0 & 0 & 0 \\ A_{N-1}^+ & 0 & .. & 0 & 0 & \frac{b}{a}A_{N-1}^- \end{bmatrix} \tag{8.6.26}$$

viéndose que $\vec{u}$ queda indeterminado.

Se verá que la respuesta al problema (8.6.23) se obtiene resolviendo dos problemas tridiagonales estándar para vectores auxiliares $\vec{\eta}$ y $\vec{\xi}$,

$$\mathbf{A}\vec{\eta} = \vec{b}\,, \qquad \mathbf{A}\vec{\xi} = \vec{u} \tag{8.6.27}$$

y la solución es

$$\vec{\phi} = \vec{\eta} - \frac{\vec{w}\cdot\vec{\eta}}{1+\vec{w}\cdot\vec{\xi}}\,\vec{\xi} \tag{8.6.28}$$

como se comprueba a continuación.

Primero se multiplica el lado derecho de (8.6.28) por la matriz **A** lo que da

$$b_i - \frac{w_k\eta_k u_i}{1+w_n\xi_n} \tag{8.6.29}$$

y segundo, la acción de $\vec{u}\otimes\vec{w}$ sobre el lado derecho de (8.6.28) da

$$u_i w_j \eta_j - \frac{w_k\eta_k}{1+w_n\xi_n} u_i w_j \xi_j = u_i\,\frac{w_j\eta_j + w_j\eta_j w_n\xi_n - w_k\eta_k w_j\xi_j}{1+w_n\xi_n} = u_i\,\frac{w_j\eta_j}{1+w_n\xi_n} \tag{8.6.30}$$

Sumando (8.6.29) y (8.6.30) efectivamente se obtiene b_i.

En resumen, para resolver el problema $\mathbf{M}\vec{\phi} = \vec{b}$ con **M** de la forma (8.6.24) se definen $\vec{u} = \{a,0,..,0,b\}$ y $\vec{w} = \{\frac{A^+_{N-1}}{b},0,..,0,\frac{A^-_0}{a}\}$ y deben primero resolverse dos problemas tridiagonales planteados en (8.6.27) siguiendo lo aprendido en §8.6.2 para luego platear la solución del problema (8.6.23) en la forma (8.6.28).

Desde el punto de vista analítico, la solución $\vec{\phi}$ no depende de los valores escogidos a y b, pero numéricamente se deben escoger con cuidado para que la precisión sea óptima.

8.7. Un caso parabólico en 1+2 dimensiones

A continuación se comenta brevemente una forma de integrar una ecuación parabólica en (t,x,y) de la forma

$$\frac{\partial F}{\partial t} = -\vec{u}\cdot\nabla F + \nu\nabla^2 F \tag{8.7.1}$$

Paso 1: En un primer paso se calcula los $F^{n+\frac{1}{2}}$ a partir de los F^n planteando un problema tridiagonal en el índice i, mientras que en j el método es explícito.

$$\begin{aligned}\frac{F^{n+\frac{1}{2}}_{i,j} - F^n_{i,j}}{\frac{\varepsilon}{2}} &= -u_1\frac{F^{n+\frac{1}{2}}_{i+1,j} - F^{n+\frac{1}{2}}_{i-1,j}}{2h} - u_2\frac{F^n_{i,j+1} - F^n_{i,j-1}}{2h} \\ &\quad +\nu\left[\frac{F^{n+\frac{1}{2}}_{i+1,j} - 2F^{n+\frac{1}{2}}_{i,j} + F^{n+\frac{1}{2}}_{i-1,j}}{h^2} + \frac{F^n_{i,j+1} - 2F^n_{i,j} + F^n_{i,j-1}}{h^2}\right]\end{aligned} \tag{8.7.2}$$

Paso 2: En un segundo paso se calcula los F^{n+1} a partir de los $F^{n+\frac{1}{2}}$ planteando un problema tridiagonal en el índice j, mientras que en i el método es explícito.

$$\begin{aligned}\frac{F_{i,j}^{n+1}-F_{i,j}^{n+\frac{1}{2}}}{\frac{\varepsilon}{2}} &= -u_1\frac{F_{i+1,j}^{n+\frac{1}{2}}-F_{i-1,j}^{n+\frac{1}{2}}}{2h}-u_2\frac{F_{i,j+1}^{n+1}-F_{i,j-1}^{n+1}}{2h}\\ &\quad +\nu\left[\frac{F_{i+1,j}^{n+\frac{1}{2}}-2F_{i,j}^{n+\frac{1}{2}}+F_{i-1,j}^{n+\frac{1}{2}}}{h^2}+\frac{F_{i,j+1}^{n+1}-2F_{i,j}^{n+1}+F_{i,j-1}^{n+1}}{h^2}\right]\end{aligned} \tag{8.7.3}$$

Este método puede ser aplicado, por ejemplo, a las ecuaciones de fluidos incompresibles vistas en el capítulo anterior. Sería especialmente sencillo en el caso de un flujo con obstáculo, sin gravedad y paredes laterales rígidas.

8.8. Dos métodos adicionales

8.8.1. Método de Richtmayer

Se estudia resolver ecuación

$$\frac{\partial U}{\partial t}=\frac{\partial^2 U^m}{\partial x^2} \qquad\qquad m \text{ entero mayor o igual a } 2$$

Se hace la siguiente deducción del método (donde δ_2 fue definido en (7.4.9))

$$\frac{U_k^{n+1}-U_k^n}{\varepsilon} = \frac{\theta\delta_2\left(U^m\right)_k^{n+1}+(1-\theta)\delta_2\left(U^m\right)_k^n}{h^2} \tag{8.8.1}$$

pero

$$\begin{aligned}\left(U^m\right)_k^{n+1} &= \left(U^m\right)_k^n+\varepsilon\left(\frac{\partial U^m}{\partial t}\right)_k^n\\ &= \left(U^m\right)_k^n+\varepsilon\left(\frac{\partial U^m}{\partial u}\right)_k^n\left(\frac{\partial U}{\partial t}\right)_k^n\\ &= \left(U^m\right)_k^n+m\left(U^{m-1}\right)_k^n\left(U_k^{n+1}-U_k^n\right)\end{aligned} \tag{8.8.2}$$

que se reemplaza en (8.8.1) usando la notacion

$$\Delta_k \equiv U_k^{n+1}-U_k^n$$

lo que da

$$\begin{aligned}\frac{\Delta_k}{\varepsilon} &= \frac{1}{h^2}\left[\theta\ \delta_2\left\{U_k^{m,n}+mU_k^{m-1,n}\Delta_k\right\}+(1-\theta)\delta_2U_k^{m,n}\right]\\ &= \frac{1}{h^2}\left[m\theta\delta_2\left\{U_k^{m-1,n}\Delta_k\right\}+\delta_2U_k^{m,n}\right]\\ &= \frac{1}{h^2}\left[m\theta\left\{U_{k+1}^{m-1,n}\Delta_{k+1}-2U_k^{m-1,n}\Delta_k+U_{k-1}^{m-1,n}\Delta_{k-1}\right\}+U_{k+1}^{m,n}-2U_k^{m,n}+U_{k-1}^{m,n}\right]\end{aligned}$$

La última expresión convierte al problema en uno tridiagonal.

8.8.2. Método de Lees

El método de Lees (M.Lees, Math. Comp. 20 516 (1966)) aborda ecuaciones parabólicas de la forma

$$b(U)\,\frac{\partial U}{\partial t} = \frac{\partial}{\partial x}\left(a(U)\,\frac{\partial U}{\partial x}\right), \qquad a(U) > 0\,, \quad b(U) > 0$$

La derivada espacial se expresa en la forma

$$\delta U_k^n \equiv U_{k+\frac{1}{2}}^n - U_{k-\frac{1}{2}}^n$$

La ecuación se discretiza

$$\begin{aligned} b_k^n \frac{U_k^{n+1} - U_k^{n-1}}{2\varepsilon} &= \frac{1}{h}\delta\left\{a_k^n \delta U_k^n\right\} = \frac{1}{h^2}\delta\left\{a_k^n\left(U_{k+\frac{1}{2}}^n - U_{k-\frac{1}{2}}^n\right)\right\} \\ &= \frac{1}{h^2}\left[a_{k+\frac{1}{2}}^n\left(U_{k+1}^n - U_k^n\right) - a_{k-\frac{1}{2}}^n\left(U_k^n - U_{k-1}^n\right)\right] \end{aligned} \tag{8.8.3}$$

Esta expresión es inestable si $a = b = 1$. Además, hasta aquí, es explícito. La estabilidad se resuelve si se hace los reemplazos que siguen, los que convierten al problema en uno que ya no es explícito,

$$U_{k'}^n \quad \to \frac{1}{3}\left(U_{k'}^{n+1} + U_{k'}^n + U_{k'}^{n-1}\right) \quad \text{donde } k' \text{ puede se} \quad k' = k+1,\ \ k,\ \ k-1$$

$$a_{k+\frac{1}{2}}^n \to \frac{1}{2}\left(a_{k+1}^n + a_k^n\right) \qquad a_{k-\frac{1}{2}}^n \to \frac{1}{2}\left(a_k^n + a_{k-1}^n\right)$$

Con lo cual el problema es tridiagonal.

8.9. Ecuación de Schrödinger dependiente del tiempo

8.9.1. Usando el método de Crank Nicolson

Se estudiará un método para integrar (8.2.4) con el método de Crank Nicolson. En el caso actual la ecuación de Schrödinger discretizada formalmente se plantea en la forma

$$\frac{\psi^{n+1} - \psi^n}{\varepsilon} = -\frac{i}{2}\left[H\,\psi^{n+1} + H\,\psi^n\right] \tag{8.9.1}$$

que se puede reordenar en la forma

$$\psi^{n+1} + \frac{i\varepsilon}{2} H\,\psi^{n+1} = \psi^n - \frac{i\varepsilon}{2} H\,\psi^n \tag{8.9.2}$$

o bien

$$\psi^{n+1} = \frac{1 - \frac{i\varepsilon}{2} H}{1 + \frac{i\varepsilon}{2} H}\,\psi^n \tag{8.9.3}$$

Este método es bueno porque el operador que actua sobre ψ^n tiene la forma $F = \frac{1-iz}{1+iz}$ y $|F| = 1$, lo que sugiere que se preserva la norma de ψ en su evolución temporal. También se observa que

$$\frac{1-iz}{1+iz} = \frac{2}{1+iz} - 1$$

y basado en esto se escribe

$$\psi^{n+1} = \left[\frac{2}{1+\frac{i\varepsilon}{2}H} - 1 \right] \psi^n = \chi - \psi^n \tag{8.9.4}$$

donde se ha definido,

$$\left(1 + \frac{i\,\varepsilon}{2} H\right) \chi = 2\,\psi^n \tag{8.9.5}$$

Se quiere resolver (8.9.5), o más bien su versión discreta,

$$\chi_k + \frac{i\,\varepsilon}{2} \left[-\frac{\chi_{k+1} - 2\chi_k + \chi_{k-1}}{h^2} + V_k \chi_k \right] = 2\psi_k^n$$

que se reescribe como

$$\chi_{k+1} + \left(-2 + \frac{2ih^2}{\varepsilon} - h^2 V_k \right) \chi_k + \chi_{k-1} = \frac{4h^2}{-i\varepsilon}\, \psi_k^n \tag{8.9.6}$$

que es un problema tridiagonal y, por lo tanto, en principio ya se sabe resolver. Una vez que se obtiene χ, se inserta en (8.9.4) y de ahí se sigue. Se ha demostrado que este método es estable.

8.9.2. El método explícito de Visscher

Este método apareció presentado en Computational Physics, **5** 596 (1991). La idea es separar a la función de onda en sus partes real e imaginaria,

$$\psi = \psi_R + i\,\psi_I$$

que lleva a escribir la ecuación de Schrödinger como un par de ecuaciones reales,

$$\dot{\psi}_R = H\,\psi_I\,, \qquad\qquad \dot{\psi}_I = -H\,\psi_R \tag{8.9.7}$$

que se discretizan evaluando a ψ_R en tiempos enteros ψ_R^n y la parte imaginaria en tiempos semi-enteros

$$\frac{\psi_R^n - \psi_R^{n-1}}{\varepsilon} = H\,\psi_I^{n-\frac{1}{2}}\,, \qquad\qquad \frac{\psi_I^{n+\frac{1}{2}} - \psi_I^{n-\frac{1}{2}}}{\varepsilon} = -H\,\psi_R^n \tag{8.9.8}$$

y se despeja

$$\psi_R^n = \psi_R^{n-1} + \varepsilon\, H\,\psi_I^{n-\frac{1}{2}}\,, \qquad\qquad \psi_I^{n+\frac{1}{2}} = \psi_I^{n-\frac{1}{2}} - \varepsilon\, H\,\psi_R^n \tag{8.9.9}$$

Con estas ecuaciones se itera en forma explícita. Antes, claro, se debe usar que

$$H\,f = -\frac{f_{k+1} - 2\,f_k + f_{k-1}}{h^2} + V_k\,f_k = -\frac{f_{k+1} + f_{k-1}}{h^2} + \left(V_k + \frac{2}{h^2} \right)\, f_k \tag{8.9.10}$$

8.9.2.1. Conservación de la norma

Se verá que la norma de la función de onda se conserva en el tiempo. Para ello primero se define la densidad de probabilidad en tiempos enteros y semienteros,

$$P^n = (\psi_R^n)^2 + \psi_I^{n+\frac{1}{2}}\,\psi_I^{n-\frac{1}{2}}, \qquad P^{n+\frac{1}{2}} = \psi_R^{n+1}\,\psi_R^n + \left(\psi_I^{n+\frac{1}{2}}\right)^2 \tag{8.9.11}$$

A continuación se calculará la diferencia entre es estas dos densidades y finalmente se va a demostrar que al sumar sobre k (integrar sobre x) se obtiene cero.

$$\begin{aligned}\Delta = P^{n+\frac{1}{2}} - P^n &= \left(\psi_R^n + \varepsilon H\,\psi_I^{n+\frac{1}{2}}\right)\psi_R^n + \left(\psi_I^{n+\frac{1}{2}}\right)^2 - (\psi_R^n)^2 - \psi_I^{n+\frac{1}{2}}\left(\psi_I^{n+\frac{1}{2}} + \varepsilon H\,\psi_R^n\right)\\ &= \varepsilon\left(\psi_R^n H\,\psi_I^{n+\frac{1}{2}} - \psi_I^{n+\frac{1}{2}} H\,\psi_R^n\right)\end{aligned} \tag{8.9.12}$$

Al reemplazar H en la última expresión se obtiene que la contribución de V se anula trivialmente. La contribución de $-d^2/dx^2$ es menos trivial. Se escribe omitiendo los índices de tiempo que ya se sabe cuanto valen para ψ_R y ψ_I,

$$\Delta = -\frac{\varepsilon}{h^2}\left(\psi_{Rk}\left\{\psi_{Ik+1} - 2\psi_{Ik} + \psi_{Ik-1}\right\} - \psi_{Ik}\left\{\psi_{Rk+1} - 2\psi_{Rk} + \psi_{Rk-1}\right\}\right)$$

y es fácil ver que hay términos que por pares se cancelan, por ejemplo

$$\sum_k \psi_{Rk}\psi_{Ik+1} - \sum_k \psi_{Ik}\,\psi_{Rk-1} = 0$$

Podría haber contribución de términos de borde, pero en los bordes la función de onda es nula. Habiéndose obtenido cero al sumar sobre k, se ha demostrado que $\int |\psi|^2\,dx$ no cambia en el tiempo y por lo tanto la norma se preserva. Este es un gran mérito del método de Vissher.

8.9.2.2. Estabilidad

El análisis de estabilidad se hará en forma bastante limitada y a pesar de ellos la experiencia muestra que da criterios que funcionan. Se comienza replanteando las ecuaciones (8.9.9) en forma matricial,

$$\begin{pmatrix} \psi_R^n \\ \psi_I^{n-1/2} \end{pmatrix} = \begin{pmatrix} 1 & \varepsilon H \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \psi_R^{n-1} \\ \psi_I^{n-1/2} \end{pmatrix}, \qquad \begin{pmatrix} \psi_R^{n-1} \\ \psi_I^{n-1/2} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -\varepsilon H & 1 \end{pmatrix}\begin{pmatrix} \psi_R^{n-1} \\ \psi_I^{n-3/2} \end{pmatrix}$$

En los pasos que sigue se usa la abreviación $A \equiv \varepsilon H$ y las dos ecuaciones anteriores se pueden escribir como una sola,

$$\begin{pmatrix} \psi_R^n \\ \psi_I^{n-1/2} \end{pmatrix} = \begin{pmatrix} 1-A^2 & A \\ -A & 1 \end{pmatrix}\begin{pmatrix} \psi_R^{n-1} \\ \psi_I^{n-3/2} \end{pmatrix}$$

Esta ecuación tiene una estrutura del tipo $\psi^n = M\,\psi^{n-1}$ lo que implica que $\psi^n = M^n\,\psi^0$. Un método de iteración con esta estructura explota si al menos uno de los autovalores de M tiene módulo

mayor que 1 y produce una función nula si el mayor autovalor es menor que 1. Para que el método funcione es necesario que los autovalores tengan módulo 1.

Los autovalores de la ecuación matriz de arriba son

$$\lambda_{\pm} = 1 - \frac{A^2}{2} \pm \frac{A}{2}\sqrt{A^2-4} \tag{8.9.13}$$

• Si $|A| > 2$ la primera parte $1 - A^2/2$ es menor que -1 y por lo tanto $\lambda_- < -1$ lo que es inaceptable.

• Si $|A| \leq 2$ los autovalores se pueden escribir en la forma

$$\lambda_{\pm} = 1 - \frac{A^2}{2} \pm \frac{iA}{2}\sqrt{4-A^2} \tag{8.9.14}$$

y trivialmente se comprueba que $|\lambda_{\pm}| = 1$ que es lo que se necesita para la estabilidad. Ahora se verá las implicasiones que tiene esto sobre los valores de ε y h.

Con este objetivo se usará un potencial esencialmente plano, $V_k = V_0$ y funciones de onda estacionarias planas,

$$\psi_\nu = T_\nu(t)\, e^{i\frac{\nu k\pi}{N}}$$

La acción de H sobre ψ_ν da

$$H\,\psi_\nu = T_\nu\left[-\frac{e^{i\frac{\nu(k+1)\pi}{N}} - 2\,e^{i\frac{\nu k\pi}{N}} + e^{i\frac{\nu(k-1)\pi}{N}}}{h^2} + V_0\, e^{i\frac{\nu k\pi}{N}}\right]$$

El numerador del primer término se puede escribir,

$$\begin{aligned}\left(e^{i\frac{\nu\pi}{N}} - 2 + e^{-i\frac{\nu\pi}{N}}\right) e^{i\frac{\nu k\pi}{N}} &= \left(e^{i\frac{\nu\pi}{2N}} - e^{-i\frac{\nu\pi}{2N}}\right)^2 e^{i\frac{\nu k\pi}{N}} \\ &= -4\sin^2\left(\frac{\nu\,\pi}{2N}\right) e^{i\frac{\nu k\pi}{N}}\end{aligned}$$

lo que muestra que los autovalores son

$$E_\nu = \frac{4}{h^2}\sin^2\left(\frac{\nu\,\pi}{2N}\right) + V_0$$

por lo cual

$$A = \frac{4\varepsilon}{h^2}\sin^2\left(\frac{\nu\,\pi}{2N}\right) + \varepsilon V_0$$

Los valores extremos que toma A son εV_0 y $\varepsilon V_0 + \frac{4\varepsilon}{h^2}$ lo que lleva a

$$-2 < \varepsilon V_0 < 2 \qquad \text{y también} \qquad -2 < \varepsilon V_0 + \frac{4\,\varepsilon}{h^2} < 2$$

Escogiendo las más exigentes se obtiene

$$-2 < \varepsilon V_0 < 2 - \frac{4\varepsilon}{h^2} \tag{8.9.15}$$

8.10. Método implícito

Se considerará en $0 \leq x \leq 1$ la ecuación

$$\begin{aligned} \frac{\partial \phi}{\partial t} &= \frac{\partial^2 \phi}{\partial x^2} + S(x) \\ \phi(t,0) &= a\,, \qquad \phi(t,1) = b \\ \phi(0,x) &= f(x) \end{aligned} \tag{8.10.1}$$

Sea $\Phi(x)$ una solución estática, es decir,

$$\begin{aligned} \Phi'' &= -S(x) \\ \Phi(0) &= a\,, \qquad \Phi(1) = b \end{aligned} \tag{8.10.2}$$

Si se define $u(t,x)$ tal que $\phi(t,x) = \Phi(x) + u(x,t)$ entonces se cumple que $u(t,0) = u(t,1) = 0$ y el problema se reduce a

$$\begin{aligned} \dot{u} &= u'' \\ u(t,0) &= 0\,, \qquad u(t,1) = 0 \end{aligned} \tag{8.10.3}$$

Por el método de separación de variable este problema es trivial.

Si se hace la expansión

$$u(t,x) = \sum_{r=1}^{\infty} a_r\, e^{\varepsilon_r t}\, \sin(r\pi x) \tag{8.10.4}$$

las ecuaciones de borde nulas en $x = 0$ y $x = 1$ implican

$$\varepsilon_r = -(\pi r)^2 \tag{8.10.5}$$

lo que asegura que todos los sumandos van a cero cuando $t \to \infty$. Esto muestra que en ecuaciones parabólicas del tipo (8.10.1) tienen soluciones estáticas estables: las oscilaciones mueren como lo muestra (8.10.4). Se hace notar que la suma en (8.10.4) comienza en $r = 1$ porque si hubiera una componente $r = 0$ ésta sería estática y la parte estática ya fue considerada.

La ecuación (8.10.1) se discretiza ahora en la forma

$$\frac{\phi_k^{n+1} - \phi_k^n}{\varepsilon} = \frac{\phi_{k+1}^{n+1} - 2\phi_k^{n+1} + \phi_{k-1}^{n+1}}{h^2} + S_k \tag{8.10.6}$$

que corresponde a escoger $a = 2$ en (8.6.4) y arroja

$$\phi_k^{n+1} = \phi_k^n + \varepsilon \underbrace{\frac{1}{h^2}\left(\phi_{k+1}^{n+1} - 2\phi_k^{n+1} + \phi_{k-1}^{n+1}\right)}_{-\mathbf{H}\vec{\phi}} + \varepsilon S_k \tag{8.10.7}$$

Estabilidad: La ecuación anterior se escribe en forma vectorial

$$\vec{\phi}^{n+1} = \vec{\phi}^{n} - \varepsilon \mathbf{H} \vec{\phi}^{n+1} + \varepsilon \vec{S} \tag{8.10.8}$$

que formalmente es

$$\vec{\phi}^{n+1} = (1+\varepsilon \mathbf{H})^{-1} \left(\vec{\phi}^{n} + \varepsilon \vec{S} \right) \tag{8.10.9}$$

Puesto que **H** es la versión discreta de $-\frac{\partial^2}{\partial x^2}$, sus autovalores E_ν son positivos implicando que los autovalores de $(1+\varepsilon \mathbf{H})^{-1}$ son

$$\frac{1}{1+\varepsilon E_\nu}$$

que necesariamente son menores que la unidad si $\varepsilon > 0$. Esto asegura que el método es estable.

Integración numérica: Para integrar el problema (8.10.7) se utiliza un método tridiagonal semejante al descrito en sec.8.6.1 usando $a = 2$. La ecuación discreta puede ser puesta en la forma estándar

$$-r\,\phi_{k+1}^{n+1} + (1+2r)\,\phi_k^{n+1} - r\,\phi_{k-1}^{n+1} = \phi_k^n + \varepsilon S_k \tag{8.10.10}$$

que tiene forma tridiagonal con coeficientes A que no dependen de k

$$A^+ = -r \qquad A^0 = 1+2r \qquad A^- = -r \qquad b_k^n = \phi_k^n + \varepsilon S_k \tag{8.10.11}$$

Las condiciones de borde e inicial son

$$\begin{aligned} \phi_0^n &= a\,, \quad \phi_N^0 = b \\ \phi_k^0 &= f_k\,, \qquad\qquad \text{con} \quad f_0 = a\,, \quad f_N = b \end{aligned} \tag{8.10.12}$$

Tal como en el caso de la ecuación de calor el problema se resuelve por medio de

$$\phi_k^{n+1} = \alpha_k\, \phi_{k+1}^{n+1} + \beta_k^n \tag{8.10.13}$$

Para comenzar, se sabe que $\alpha_{N-1} = 0$ y se puede obtener todos los α_k de una sola vez. Y todos los β a partir de las αs y de $\beta_{N-1} = b$ se obtiene todos los βs. De estos dos conjuntos y la condición inicial para ϕ se obtiene todos los ϕ_k^n.

Problemas

1. Una esfera de radio R es sacada de un horno con distribución casi uniforme, $T(t=0,r) = T_0$, de temperatura y es colocada en un ambiente que se mantiene a temperatura T_A. La esfera se irá enfriando en forma esféricamente simétrica de acuerdo a

$$\begin{aligned} \frac{\partial T}{\partial t} &= \kappa \nabla^2 T \\ \left[\frac{\partial T}{\partial r} \right]_{r=R} &= -\mu \left(T(t,R) - T_A \right) \end{aligned}$$

Para efectos de hacer compatible la condición de borde en $r = R$ con la condición inicial, suponga que en $t = 0$ la temperatura vale T_0 en todo el volumen excepto en la capa externa $0{,}95R \leq r \leq R$ donde la temperatura, en $t = 0$, decrece linealmente desde el valor T_0 en $r = 0{,}95R$. Use que $dT/dr|_{r=0} = 0$.

1 Escriba la ecuación anterior, aun en forma continua, en la forma de la ecuación parabólica que va a integrar numéricamente y también escriba las condiciones de borde que va a utilizar para tal ecuación (todo en lenguaje continuo).

2 Escriba todas las ecuaciones del algoritmo tridiagonal que va utilizar, en particular, deje muy claro la forma en que va manejar los bordes y la forma en que va a escoger α_0 y β_0.

3 Use los siguientes valores: $R = 100$, $T_0 = 400$, $\kappa = 0{,}01$, $\mu = 1$, $T_A = 20$. Integre la evolución de $T(t,r)$ escogiendo $N = 5000$, $dr = h = R/N$ y tres valores para $dt = \varepsilon$ en torno a $\varepsilon = 1$. Entrege un gráfico del perfil de temperatura inicial y el perfil de T cada vez que en el centro de la esfera la temperatura tome los valores 350, 300, 250, 200, 150 y 100. Indique el valor del tiempo t para los que el centro toma tales valores. Compare los resultados que se obtienen con los tres $dt = \varepsilon$ escogidos. Si las eleccciones han sido razonables debieran ser parecidos.

2. Integre la ecuación de Schrödinger adimensional

$$\frac{\partial \psi}{\partial t} = -i\left(-\frac{\partial^2}{\partial x^2} + V(x)\right)\psi$$

usando el algoritmo de Visscher. Tome como condición inicial

$$\psi(x,0) = A e^{-\mu(x-x_0)^2} e^{ibx}$$

con $\mu = 0{,}006$, $b = 0{,}333$. El potencial es un pozo o una barrera cuadrada de ancho 10 y magnitud $0{,}1$, es decir, $V(x) = \pm 0{,}1$ en un rango $(x_1, x_2 = x_1 + 10)$.

Se recomienda reticular el espacio en 4000 segmentos que abarquen desde $x = 0$ hasta $x = 2000$, $(h = 0{,}5)$, $x_1 = 1000$ y el paquete inicial está muy cerca de la zona con pontencial no nulo, por ejemplo, $x_0 = 920$. Para el incremento del tiempo parece ser bueno $\varepsilon = 0{,}05$ (discútalo).

Para simplificar el problema puede aproximar $\psi_I^{1/2}$ a la parte imaginaria de $\psi(x,0)$. Para evitar problemas lógicos no actualice los valores de ψ en los extremos (piense que corresponden a $x = \pm\infty$) y detenga el cálculo cuando el paquete llegue en forma significativa a alguno de los extremos. Parece que esto es $t \approx 890$.

Dibuje $P(t) = \int |\psi|^2 dx$ como función del tiempo y para algunos tiempos t_k interesantes dibuje $\rho_k = |\psi(x,t_k)|^2$.

Resuelva tanto para $V_0 = 0{,}1$ como para $V_0 = -0{,}1$.

Capítulo 9

Ecuaciones hiperbólicas

Las ecuaciones hiperbólicas requieren de métodos especiales porque la información contenida en las condiciones iniciales y las condiciones de borde normalmente no pueden alcanzar a todo el dominio de integración. La influencia de tales condiciones impuestas se propaga a lo largo de las llamadas curvas características y es difícil lograr un método de integración correcto sin tomar en cuenta esta limitación en la debida forma. Es probable que sólo cuando las características son líneas rectas un programador inadvertido puede dar con una forma satisfactoria de integrar ecuaciones hiperbólicas. Para hacer fácil comprender el papel jugado por las características es conveniente comenzar estudiando un problema análogo que se presenta con ecuaciones de primer orden.

Referencia: G.D. Smith, *Numerical Solution of Partial Differencial Equations: Finite Difference Methods*, Clarendon Press, Oxford, third edition, 1984.

9.1. Ecuaciones de primer orden y sus curvas características

Consideremos la ecuación

$$a\frac{\partial U}{\partial x}+b\frac{\partial U}{\partial y}=c \tag{9.1.1}$$

donde a, b y c pueden ser funciones de (x,y,U). Se va a ver que en cada punto del dominio de solución pasa una curva C a lo largo de la cual hay que resolver una ecuación diferencial ordinaria.

Se usará la notación,

$$p\equiv\frac{\partial U}{\partial x}\qquad\qquad q\equiv\frac{\partial U}{\partial y} \tag{9.1.2}$$

Con esta notación la ecuación original es

$$a\,p+b\,q=c \tag{9.1.3}$$

Un pequeño desplazamiento produce un cambio en el valor de U

$$dU=p\,dx+q\,dy$$

pero de (9.1.3) se sabe que $p = \frac{c-bq}{a}$ lo que da

$$dU = \frac{c-bq}{a}\,dx + q\,dy$$

que se reordena como

$$q\,(a\,dy - b\,dx) + c\,dx - a\,dU = 0 \tag{9.1.4}$$

La ecuación anterior es independiente de $p = \frac{\partial U}{\partial x}$ porque a, b y c solo dependen de (x,y,U). Además (9.1.4) se hace independiente de q si se escoge que dx y dy no sean independientes, es decir, si se define que el desplazamiento sea a lo largo de una curva C cuya pendiente satisface

$$a\,dy - b\,dx = 0 \tag{9.1.5}$$

con lo que (9.1.4) se reduce a

$$c\,dx - a\,dU = 0 \tag{9.1.6}$$

y se ve que (9.1.5) es una ecuación para C y (9.1.6) es una ecuación para U sobre C. Ellas se pueden resumir, además de definir un elemento de arco ds, planteando

$$\frac{dx}{a} = \frac{dy}{b} = \frac{dU}{c} = ds \tag{9.1.7}$$

De esta igualdad doble, la primera es la ecuación para la característica C mientras que la igualdad del último miembro con cualquiera de las primeras es la ecuación para U sobre C. Planteada la ecuación de la característica en la forma (9.1.5) pareciera que se debe obtener $x(y)$ o bien $y(x)$, sin embargo puede ocurrir que ninguna de las dos formas pueda dar una función univaluada. Es más general formular (9.1.5) como una ecuación en términos del parámetro s ya mencionado,

$$\frac{dx}{ds} = a\,, \qquad \frac{dy}{ds} = b \tag{9.1.8}$$

La ecuación para la función U sobre la característica es

$$\frac{dU}{ds} = c \tag{9.1.9}$$

Reiterando, todo lo anterior puede ser presentado planteando la ecuación original (9.1.1) sobre una curva parametrizada por un parámetro s. Sobre la curva se tiene que

$$\frac{dU}{ds} = \frac{dx}{ds}\frac{\partial U}{\partial x} + \frac{dy}{ds}\frac{\partial U}{\partial y} \tag{9.1.10}$$

y se escoge que la curva sea tal que el lado izquierdo de (9.1.1) coincida con dU/ds, es decir, se exige (9.1.8) y por lo tanto $dU/ds = c$. Un par de ejemplos sencillos debieran ayudar a asimilar lo anterior.

9.2. El método de las características

9.2.1. Ejemplos para ilustrar los conceptos básicos

9.2.1.1. Ejemplo muy sencillo

$$\begin{aligned} y\frac{\partial U}{\partial x}+\frac{\partial U}{\partial y} &= 2 \\ U(0\le x\le 1, y=0) &= \phi(x) \qquad \text{condición de borde} \end{aligned} \tag{9.2.1}$$

En esta ecuación se tiene $a=y$, $b=1$ y $c=2$ por lo que las ecuaciones paramétricas para las características son

$$\frac{dx}{ds}=y(s), \qquad \frac{dy}{ds}=1 \tag{9.2.2}$$

Se resuelve la segunda escogiendo $y(0)=0$ lo que da $y=s$. La primera ecuación se convierte en $dx/ds=s$ que da $x(s)=s^2/2+x_R$, es decir $x=\frac{1}{2}y^2+x_R$ lo que permite escribir la ecuación de las características,

$$y^2=2x-2x_R \tag{9.2.3}$$

estas son parábolas con el eje X como eje de simetría parametrizadas por x_R, que es el punto sobre el eje X por el que pasa la parábola descrita en (9.2.3).

Para los propósitos del problema planteado se necesita solo las parábolas para las que

$$0\le x_R\le 1$$

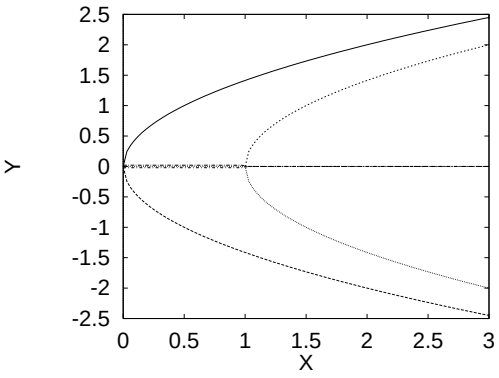


Figura 9.1: *Las características asociadas a los extremos del dominio de las condiciones iniciales del problema.*

De (9.1.9) se desprende que $dy=dU/2$, que implica $U=2y+U_R$, pero como $U(x_R,0)=\phi(x_R)$ entonces $U_R=\phi(x_R)$. De (9.2.3) formalmente se despeja que $x_R=\frac{1}{2}(2x-y^2)$ lo que finalmente permite escribir

$$U(x,y)=2y+\phi\left(x-\frac{y^2}{2}\right)$$

El término ϕ en la expresión anterior es constante a lo largo de cada caractarística. Para verificar que esta solución es correcta notamos que

$$\frac{\partial U}{\partial x}=\phi' \qquad \frac{\partial U}{\partial y}=2-y\phi'$$

donde la prima indica la derivada de ϕ con respecto a su argumento.

En resumen, lo que se ha logrado es determinar el valor de la función $U(x,y)$ en el dominio comprendido entre las dos parábolas [$y^2=2x$; $y^2=2(x-1)$]. El método analítico debiera dejar claro que con la condición inicial dada no es posible conocer a la función U fuera del dominio descrito. Un método numérico ingenuo podría intentar integrar la ecuación haciendo un reticulado rectangular regular con celdas de $(h_x\times h_y)$ en el plano (x,y), lo que llevaría a una solución incorrecta tan pronto el incremento h_y sobrepase el valor $\frac{h_x}{y}$.

9.2.1.2. Ejemplo algo más elaborado

Se busca integrar la ecuación

$$\frac{\partial U}{\partial x} + U\frac{\partial U}{\partial y} = -U^2 \qquad \text{con}$$

$$U(x=0, y>0) = 1+e^{-y} \qquad U(x>0, y=0) = \frac{2}{1+x} \tag{9.2.4}$$

Las condiciones de borde coinciden en el origen. las ecuaciones para las características son

$$\frac{dx}{ds} = 1, \qquad \frac{dy}{ds} = U \tag{9.2.5}$$

y la ecuación para U sobre las características es

$$\frac{dU}{ds} = -U^2 \tag{9.2.6}$$

La primera de las ecuaciones (9.2.5) permite escoger la parametrización $s = x$. De (9.2.6) se obtiene inmediatamente que

$$U(x,y) = \frac{1}{A(x,y)+x} \tag{9.2.7}$$

pero aun se debe determinar $A(x,y)$, que es constante a lo largo de cada característica, pero que depende de x e y.

La segunda de las ecuaciones (9.2.5), y (9.2.7) se obtiene que sobre la característica

$$dy = \frac{dx}{A(x,y)+x}$$

que se puede resolver de dos maneras equivalentes. La primera es directa y da

$$y = y_R + \ln\frac{A(x,y)+x}{A(x,y)} \tag{9.2.8}$$

Si de aquí se despeja x da $x = A(e^{y-y_R}-1)$. Definiendo $x_0 \equiv x(y=0)$ se despeja $y_R = -\ln((A+x_0)/A)$ que finalmente da

$$x = A(x,y)\,(e^y - 1) + x_0\, e^y \tag{9.2.9}$$

Estas dos formas equivalentes que definen a las características. La primera forma es útil para describir las característica que nacen en $(x=0, y_R)$ y la segunda para las que nacen de $(x_0, y=0)$. Se comprobará que en el primer caso se debe tomar $A(0,y_R) = 1/(1+e^{-y_R})$ y en el segundo caso $A(x_0,0) = \frac{1-x_0}{2}$.

A partir de la primera condición de borde: Se usa $U = 1/(A+x)$ en $x=0$ junto a la condición de borde en $x=0$, lo que permite deducir que $A = 1/(1+e^{-y_R})$, valor que se reemplaza en (9.2.8) para despejar formalmente que

$$y_R = \ln \frac{e^y - x}{1+x}$$

Esta expresión permite escribir A como función de (x,y), lo que lleva a

$$U(x,y) = \frac{1+e^{-y}}{1+x}$$

En este caso las características son parametrizadas por el punto $y_R > 0$ del que nacen sobre el eje Y y se definen por

$$y_{\text{carac}} = y_R + \ln\left(1 + (1+e^{-y_R})\,x\right)$$

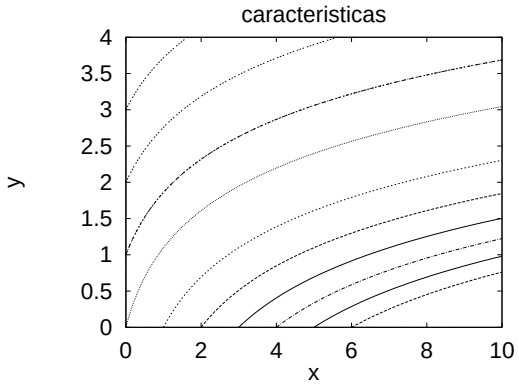


Figura 9.2: Aquí se dibujan características pertenecientes a las dos familias del ejemplo.

A partir de la segunda condición de borde: Se usa $U = 1/(A+x)$ en $y = 0$ junto a la condición de borde en $y = 0$, lo que permite deducir que $A = (1-x_0)/2$, valor que se reemplaza en (9.2.9) para despejar formalmente que

$$x_0 = \frac{1+2x-e^y}{1+e^y}$$

Esta expresión permite escribir A como función de (x,y), lo que lleva a

$$U(x,y) = \frac{1+e^{-y}}{1+x}$$

En este segundo caso las características, parametrizadas por el punto x_0 en que nacen sobre el eje X satisfacen

$$y_{\text{carac}} = \ln\left(\frac{1+2x-x_0}{1+x_0}\right)$$

9.2.2. Integración numérica a lo largo de una característica

La integración numérica general debe simultáneamente encontrar la forma de la característica como la función U sobre ella.

Supongamos que se nos ha dado los valores de U sobre una curva Γ (que no puede ser una característica). Sea P un punto sobre Γ: $P = (x_P, y_P) \in \Gamma$ y sea C la característica que pasa por P. Al punto P se le asocia $s = 0$.

Para integrar se escoge un paso h para el parámetro s. Las ecuaciones (9.1.8) y (9.1.9) al más bajo orden se pueden escribir

$$\begin{aligned} \frac{x_R^{(1)} - x_P}{h} &= a_P \\ \frac{y_R^{(1)} - y_P}{h} &= b_P \\ \frac{U_R^{(1)} - U_P}{h} &= c_P \end{aligned} \qquad (9.2.10)$$

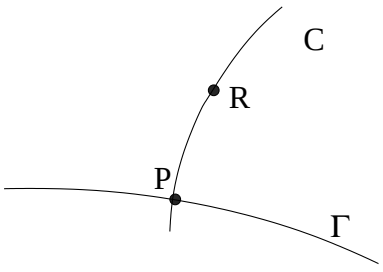


Figura 9.3: *La curva Γ corta a la característica C en $P = (x_P, y_P)$. Otro punto cercano sobre C es $R = (x_R, y_R)$.*

de las que se obtiene trivialmente las tres cantidades de orden 1, $(x_R^{(1)}, y_R^{(1)}, U_R^{(1)})$. Estos tres valores permiten calcular valores $(a_R^{(1)}, b_R^{(1)}, c_R^{(1)})$.

A continuación las mismas ecuaciones se escriben

$$\frac{x_R^{(\nu+1)} - x_P}{h} = \frac{a_P + a_R^{(\nu)}}{2}, \qquad \frac{y_R^{(\nu+1)} - y_P}{h} = \frac{b_P + b_R^{(\nu)}}{2}, \qquad \frac{U_R^{(\nu+1)} - U_P}{h} = \frac{c_P + c_R^{(\nu)}}{2} \tag{9.2.11}$$

Las que se iteran desde $\nu = 1$ hasta el orden que se considere necesario.

Una vez que se tiene el valor aceptado para (x_R, y_R, U_R) que corresponde a $s = h$ se procede a usar estos valores para avanzar, sobre la misma característica hasta un punto que corresponde a $s = 2h$. El procedimiento es el mismo. De esta manera se logra obtener toda la característica que nace en P y todos los valores de U sobre esa característica.

♠ *Usando el método anterior integrar*

$$\sqrt{x}\,\frac{\partial U}{\partial x} + U\,\frac{\partial U}{\partial y} = -U^2 \qquad \textit{con la condición de borde} \qquad U(x > 0, 0) = 1$$

9.3. Sistema de ecuaciones hiperbólicas de primer orden

Si se tiene el sistema de M ecuaciones para $\vec{U}(x,t)$,

$$\frac{\partial \vec{U}}{\partial t} + \mathbf{A}(x,t)\frac{\partial \vec{U}}{\partial x} + \mathbf{B}(x,t)\vec{U} = \vec{F}(x,t) \tag{9.3.1}$$

este sistema es hiperbólico si la matriz **A** de $M \times M$ es diagonalizable y sus autovalores α_k son reales. En general se supondrá que ellos son funciones de (x,t) y que no dependen de $\vec{U}$.

Para resolver este sistema se comienza escribiendo

$$\begin{aligned} d\vec{U} &= dx\,\frac{\partial \vec{U}}{\partial x} + dt\,\frac{\partial \vec{U}}{\partial t} \\ &= dx\,\frac{\partial \vec{U}}{\partial x} + dt\,\left(\vec{F} - \mathbf{A}\frac{\partial}{\partial x}\vec{U} - \mathbf{B}\vec{U}\right) \\ &= (dx - dt\,\mathbf{A})\,\frac{\partial \vec{U}}{\partial x} + dt\,\left(\vec{F} - \mathbf{B}\vec{U}\right) \end{aligned}$$

En el primer paso se eliminó la derivada parcial con respecto al tiempo. A continuación se encuentra la condición para que la derivada con respecto a x también desaparezca. Se escribe $\mathbf{A} = P^{-1}\Lambda P$, donde Λ es una matriz diagonal (en la diagonal están los autovalores α_k de **A**) y P es la matriz para diagonalizar a **A**. Al multiplicar a la ecuación por P se obtiene

$$P\,d\vec{U} = (dx - dt\,\Lambda)\,P\,\frac{\partial \vec{U}}{\partial x} + dt\,P\,\left(\vec{F} - \mathbf{B}\vec{U}\right) \tag{9.3.2}$$

que por componentes es un conjunto de M ecuaciones

$$\sum_{j=0}^{M-1} P_{kj}\, dU_j = (dx - dt\,\alpha_k) \sum_{j=0}^{M-1} P_{kj} \frac{\partial U_j}{\partial x} + dt \sum_{j=0}^{M-1} P_{kj}\, (F_j - B_{ji} U_i)$$

donde los índices j y k varías de 1 a M o de 0 a $M-1$ según lo que se prefiera. Por cada uno de los valores del índice k se define una característica $\mathscr{C}_k$ como la curva $x^{(k)}(t)$ que satisface la ecuación

$$\frac{dx^{(k)}}{dt} = \alpha_k(x,t) \tag{9.3.3}$$

A lo largo de $\mathscr{C}_k$ se satisface

$$\sum_{j=0}^{M-1} P_{kj} dU_j = dt \sum_{j=0}^{M-1} P_{kj} \left(F_j - \sum_{i=0}^{M-1} B_{ji} U_i \right) \tag{9.3.4}$$

Lo anterior se ilustra a continuación con un problema unidimensional de fluidos.

9.3.1. Fluido compresible sencillo

> Esta parte sigue de cerca el modelo planteado en "*Numerical solution of partial differential equations: finite difference methods*" de G.D. Smith

Consideremos las ecuaciones (7.3.1) y (7.3.2) para el caso en que sólo hay velocidad hidronámica en la dirección x, los campos sólo dependen de (x,t), se desprecia el término con viscosidad ($\eta = 0$), no hay gravedad y la temperatura es uniforme y constante en el tiempo. En tal caso las ecuaciones se reducen a

$$\begin{aligned} \frac{\partial \rho}{\partial t} + \frac{\partial \rho v}{\partial x} &= 0 \\ \rho \frac{\partial v}{\partial t} + \rho\, v \frac{\partial v}{\partial x} &= -\frac{\partial p}{\partial x} \end{aligned} \tag{9.3.5}$$

Adicionalmente se supondrá que vale la ecuación de estado $p = A\rho^\gamma$, lo que implica que $\partial p/\partial \rho = A\gamma\rho^{\gamma-1} = \gamma p/\rho$. Se define

$$c^2 = \frac{dp}{d\rho} = \frac{\gamma p}{\rho} \tag{9.3.6}$$

que permite reescribir la segunda de las ecuaciones (9.3.5) en la forma

$$\frac{\partial v}{\partial t} + v\frac{\partial v}{\partial x} + \frac{c^2}{\rho}\frac{\partial \rho}{\partial x} = 0 \tag{9.3.7}$$

La cantidad c no es una constante; es la velocidad del sonido y depende de la densidad.

El sistema de ecuaciones (9.3.5a) y (9.3.7) puede ser escrito como

$$\frac{\partial}{\partial t}\begin{pmatrix} \rho \\ v \end{pmatrix} + \begin{pmatrix} v & \rho \\ \frac{c^2}{\rho} & v \end{pmatrix} \frac{\partial}{\partial x}\begin{pmatrix} \rho \\ v \end{pmatrix} = 0 \tag{9.3.8}$$

Es fácil comprobar que los autovalores de la matriz son $\lambda_1 = v - c$ y $\lambda_2 = v + c$. Sin embargo, reobtendremos este resultado avanzando por un camino más explícito.

Para determinar curvas características escribimos la variación de ambos campos en la forma $dF = F_x\,dx + F_t\,dt$ pero inmediatamente reemplazamos F_t por lo que dan las ecuaciones dinámicas que tenemos más arriba:

$$\begin{aligned} dv &= v_x\,dx - \left(v v_x + \tfrac{c^2}{\rho}\rho_x\right) dt \\ d\rho &= \rho_x\,dx - (v\rho_x + \rho v_x)\,dt \end{aligned} \tag{9.3.9}$$

De la segunda ecuación se obtiene que

$$\rho_x = \frac{d\rho + \rho v_x\,dt}{dx - v\,dt}$$

que se reemplaza en la primera, obteniéndose

$$\rho\,(dx - v\,dt)\,dv = \rho\left[(dx - v\,dt)^2 - c^2\,dt^2\right] v_x - c^2\,d\rho\,dt \tag{9.3.10}$$

En el paso anterior se eliminó ρ_x. La última relación es independiente de v_x si se escoge que la integración se haga en curvas que satisfagan

$$(dx - v\,dt)^2 = c^2\,dt^2$$

lo que se cumple en dos familias de curvas características:

$$\frac{dx}{dt} = v + c \qquad \text{curvas } f \tag{9.3.11}$$

$$\frac{dx}{dt} = v - c \qquad \text{curvas } g \tag{9.3.12}$$

No se trata de dos curvas, sino de dos *familias* de curvas. Cada miembro de estas dos familias se caracteriza por una constante de integración.

La evolución sobre estas características reduce la ecuación (9.3.10) a la forma

$$\begin{aligned} \rho\,(dx - v\,dt)\,dv &= -c^2\,d\rho\,dt \\ \rho\,((v \pm c)\,dt - v\,dt)\,dv &= -c^2\,d\rho\,dt \\ c\,d\rho \pm \rho\,dv &= 0 \qquad \pm \text{ sobre curvas } f/g \end{aligned} \tag{9.3.13}$$

Las ecuaciones que hay que integrar numéricamente se pueden resumir en

$$\mathscr{C}_+ : \quad dx = (v + c)\,dt \quad \text{combinada con} \quad c\,d\rho + \rho\,dv = 0 \tag{9.3.14}$$

$$\mathscr{C}_- : \quad dx = (v - c)\,dt \quad \text{combinada con} \quad c\,d\rho - \rho\,dv = 0 \tag{9.3.15}$$

Los diferenciales que aparecen en estas ecuaciones están definidos sobre las características. Por ejemplo, en (9.3.14), $d\rho$ es la variación de la densidad cuando hay un desplazamiento sobre la característica f, en cambio en (9.3.15) aparece el cambio de la densidad sobre la característica g. Ambos $d\rho$ llevan el mismo nombre pero son objetos diferentes.

A continuación se define un método para integrar estas ecuaciones. Se hará en dos etapas. En la primera etapa, ver la figura 9.4, se define las ecuaciones

$$\begin{aligned} x_R - x_P &= (v_P + c_P)\,(t_R - t_P) \\ x_R - x_Q &= (v_Q - c_Q)\,(t_R - t_Q) \end{aligned} \tag{9.3.16}$$

que permiten despejar x_R y t_R. Luego las ecuaciones $c\,d\rho \pm \rho\,dv$ se discretizan,

$$\begin{aligned} c_P\,(\rho_R-\rho_P)+(v_R-v_P)\,\rho_P &= 0 \\ c_Q\,(\rho_R-\rho_Q)-(v_R-v_Q)\,\rho_Q &= 0 \end{aligned} \tag{9.3.17}$$

que permite despejar ρ_R y v_R. De ρ_R se obtiene c_R.

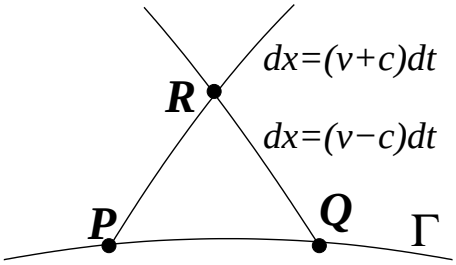


Figura 9.4: *Las curvas PR y QR son curvas características.*

Las cantidades recién obtenidas son una primera aproximación a x_R, t_R, ρ_R y v_R. Ahora se puede escribir una forma más precisa para los dos pares de ecuaciones:

$$\begin{aligned} x_R^{n+1}-x_P &= \tfrac{1}{2}(v_P+v_R^n+c_P+c_R^n)\,(t_R^{n+1}-t_P) \\ x_R^{n+1}-x_Q &= \tfrac{1}{2}(v_Q+v_R^n-c_Q-c_R^n)\,(t_R^{n+1}-t_Q) \end{aligned} \tag{9.3.18}$$

Se usa las primeras dos ecuaciones con $n=1$ para obtener valores x_R y t_R de segundo orden e inmediatamente se usa las otras dos, también con $n=1$:

$$\begin{aligned} (c_P+c_R^n)\,(\rho_R^{n+1}-\rho_P) &= -(v_R^{n+1}-v_P)\,(\rho_P+\rho_R^n) \\ (c_Q+c_R^n)\,(\rho_R^{n+1}-\rho_Q) &= (v_R^{n+1}-v_Q)\,(\rho_Q+\rho_R^n) \end{aligned} \tag{9.3.19}$$

para obtener valores ρ_R y v_R de segundo orden. Estas cuatro ecuaciones se puede usar reiteradamente para lograr convergencia a valores finales (x_R,t_R,ρ_R,v_R).

Para integrar, entonces, se parte de la base que se tiene curvas Γ sobre las que se ha definido condiciones iniciales y/o de borde. Esto significa que sobre Γ se conoce las cuatro cantidades (x,t,ρ,v). Se escoge puntos sobre estas curvas, que van a jugar los papeles de P y Q de más arriba.

La forma típica de integrar en este caso consiste en obtener toda la secuencia de "puntos R" a partir de puntos consecutivos de Γ que juegan el papel de P y Q. La secuencia de puntos R así obtenidos definen una curva Γ_1. que será la nueva curva a partir de la cual se construirá una curva Γ_2.

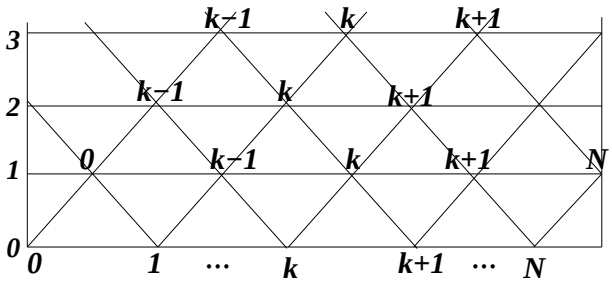


Figura 9.5: *Las dos faimilias de características de este problema forman un reticulardo no regular.*

La iteración de este problema es distinta al pasar de tiempos pares a tiempos impares que al revés. En el primer caso las identificaciones son $P=(x_k^{2\nu},t_k^{2\nu})$, $Q=(x_{k+1}^{2\nu},t_{k+1}^{2\nu})$ y $R=(x_k^{2\nu+1},t_k^{2\nu+1})$ mientras que en el segundo las identificaciones son $P=(x_{k-1}^{2\nu-1},t_{k-1}^{2\nu-1})$, $Q=(x_k^{2\nu-1},t_k^{2\nu-1})$ y $R=(x_k^{2\nu},t_k^{2\nu})$.

Condiciones de borde. Supongamos que en la malla de caractarísticas se tiene un punto Q en una característica g (crece hacia la izquierda) y el próximo punto, R, está en el borde izquierdo. De (9.3.17) y (9.3.18) las ecuación que fijan las condiciones de borde a la izquierda en una primera iteración son

$$\begin{aligned} x_R-x_Q-(v_Q+c_Q)\,(t_R-t_Q) &= 0 \\ c_Q\,(\rho_R-\rho_Q)-(v_R-v_Q)\,\rho_) &= 0 \end{aligned} \tag{9.3.20}$$

Por ejemplo, se puede presentar el caso que $x_R = 0$ y que las condiciones de borde del problema sean que la velocidad en el borde es nula, $v_R = 0$. Se tiene dos ecuaciones y dos incógnitas: t_R y ρ_R. Algo similar debe ser planteado al lado derecho.

9.3.2. Fluido compresible con entalpía variable

Esta parte sigue de cerca el modelo planteado en el libro "*Computational methods in engineering and science" de S. Nakamura.*

Esta vez se considerará las ecuaciones

$$\begin{aligned} \frac{\partial \rho}{\partial t} + v\frac{\partial \rho}{\partial x} + \rho\frac{\partial v}{\partial x} &= 0 \\ \frac{\partial \rho v}{\partial t} + \frac{\partial \rho v^2}{\partial x} + \frac{\partial p}{\partial x} &= f \end{aligned} \tag{9.3.21}$$

donde f, reemplazando al término $\eta\nabla^2\vec{v}$, es una forma sencilla de representar la resistencia del flujo. Si en la segunda ecuación se elimina $\frac{\partial \rho}{\partial t}$, usando la primera, se obtiene la forma

$$\rho\left(\frac{\partial v}{\partial t} + v\frac{\partial v}{\partial x}\right) + \frac{\partial p}{\partial x} = f \tag{9.3.22}$$

y la ecuación para la entalpía H,

$$\rho\left(\frac{\partial H}{\partial t} + v\frac{\partial H}{\partial x}\right) - \frac{1}{J}\left(\frac{\partial p}{\partial t} + v\frac{\partial p}{\partial x}\right) = Q \tag{9.3.23}$$

donde Q es la tasa de generación de calor y J es una constante que hace calzar las unidades. Puesto que

$$\begin{aligned} \frac{\partial H}{\partial t} &= \frac{\partial H}{\partial \rho}\frac{\partial \rho}{\partial t} + \frac{\partial H}{\partial p}\frac{\partial p}{\partial t} \\ \frac{\partial H}{\partial x} &= \frac{\partial H}{\partial \rho}\frac{\partial \rho}{\partial x} + \frac{\partial H}{\partial p}\frac{\partial p}{\partial x} \end{aligned}$$

se puede reescribir (9.3.23) en la forma

$$\rho\left(\frac{\partial H}{\partial \rho}\frac{\partial \rho}{\partial t} + \frac{\partial H}{\partial p}\frac{\partial p}{\partial t}\right) + \rho v\left(\frac{\partial H}{\partial \rho}\frac{\partial \rho}{\partial x} + \frac{\partial H}{\partial p}\frac{\partial p}{\partial x}\right) - \frac{1}{J}\left(\frac{\partial p}{\partial t} + v\frac{\partial p}{\partial x}\right) = Q \tag{9.3.24}$$

el cuadrado de la velocidad del sonido es

$$c^2 = -\frac{\partial H/\partial \rho}{\partial H/\partial p - 1/(\rho J)} \tag{9.3.25}$$

lo que permite obtener

$$\frac{\partial p}{\partial t} + \rho c^2\frac{\partial v}{\partial x} + v\frac{\partial p}{\partial x} = -\frac{c^2 Q}{\rho \partial H \partial \rho} \tag{9.3.26}$$

Si ahora en (9.3.23) se elimina $\frac{\partial p}{\partial t}$ usando la ecuación recién escrita, se obtiene

$$\frac{\partial H}{\partial t}+v\frac{\partial H}{\partial x}+\frac{c^2}{J}\frac{\partial v}{\partial x}=\frac{Q}{\rho}-\frac{c^2Q}{\rho^2 J\frac{\partial H}{\partial \rho}} \tag{9.3.27}$$

Las ecuaciones (9.3.22), (9.3.26) y (9.3.27) se pueden resumir en la forma

$$\frac{\partial}{\partial t}\begin{pmatrix} v\\ p\\ H\end{pmatrix}+\begin{pmatrix} v & 1/\rho & 0\\ \rho c^2 & v & 0\\ c^2/J & 0 & v\end{pmatrix}\frac{\partial}{\partial x}\begin{pmatrix} v\\ p\\ H\end{pmatrix}=\begin{pmatrix} f/\rho\\ -c^2Q/(\rho\partial H/\partial\rho)\\ Q/\rho-c^2Q/(J\rho^2\partial H/\partial\rho)\end{pmatrix} \tag{9.3.28}$$

La matriz A en este caso tiene autovalores

$$\alpha_1=v+c\,,\qquad \alpha_2=v-c\,,\qquad \alpha_3=v$$

Si se define

$$P=\begin{pmatrix} 1 & -1/\rho c & 0\\ 0 & c & -\rho Jc\\ 1/2J & 1/(2\rho Jc) & 0\end{pmatrix}\quad\Longleftrightarrow\quad P^{-1}=\begin{pmatrix} 1/2 & 0 & J\\ -\rho c/2 & 0 & \rho Jc\\ -c/(2J) & 1/(\rho Jc) & c\end{pmatrix} \tag{9.3.29}$$

se obtiene que $PAP^{-1}=$diag$\{v-c,v,v+c\}$.

En base a estos resultados se puede llegar a las siguientes ecuaciones que deben ser integradas a lo largo de las respectivas características.

Ecuaciones para las características y v, p, H .

Característica $\mathscr{C}_+$:

$$\frac{dx}{dt} = v+c \tag{9.3.30}$$

$$\rho c\frac{dv}{dt}+\frac{dp}{dt} = c\,f-\frac{Qc^2}{\rho\partial H/\partial\rho} \tag{9.3.31}$$

Característica $\mathscr{C}_-$:

$$\frac{dx}{dt} = v-c \tag{9.3.32}$$

$$-\rho c\frac{dv}{dt}+\frac{dp}{dt} = -c\,f-\frac{Qc^2}{\rho\partial H/\partial\rho} \tag{9.3.33}$$

En ambos casos las derivadas son sobre la respectiva características. Estas dos características se denominan *sónicas*.

Característica $\mathscr{C}_3$ o característica *material*:

$$\frac{dx}{dt} = v \tag{9.3.34}$$

$$\rho\frac{dH}{dt}-\frac{1}{J}\frac{dp}{dt} = Q \tag{9.3.35}$$

Método explícito parcialmente basado en las características .

Un método basado en la integranción a lo largo de las características, tal como se hizo en §9.3.1 sería mucho más satisfactorio.

Este método calcula p^{n+1}, v^{n+1} y H^{n+1} a partir de p^n, v^n y H^n. En lo que sigue $h = x_{k+1} - x_k = x_k - x_{k-1}$ y los intervalos temporales valen ε. Se usa además que

$$\begin{aligned} p_A &= \frac{x_k - x_A}{h} p^n_{k-1} + \frac{x_A - x_{k-1}}{h} p^n_k \\ v_A &= \frac{x_k - x_A}{h} v^n_{k-1} + \frac{x_A - x_{k-1}}{h} v^n_k \\ p_B &= \frac{x_{k+1} - x_B}{h} p^n_k + \frac{x_B - x_k}{h} p^n_{k+1} \\ v_B &= \frac{x_{k+1} - x_B}{h} v^n_k + \frac{x_B - x_k}{h} v^n_{k+1} \end{aligned}$$

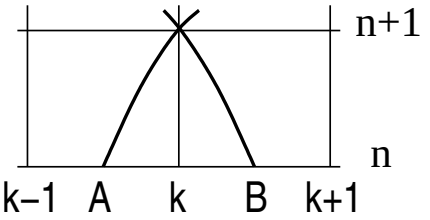


Figura 9.6: *Los puntos A y B son las intersecciones de las características $\mathscr{C}_+$ y $\mathscr{C}_-$ que pasan por el punto $(n+1,k)$ con el eje n.*

Se plantean las ecuaciones

$$\frac{\partial p^{n+1}_k - p_A}{\partial \varepsilon} + \langle \rho c \rangle_+ \frac{\partial v^{n+1}_k - v_A}{\partial \varepsilon} = R_+ \equiv cf - \frac{Qc^2}{\rho \partial H/\partial \rho}$$

$$\frac{\partial p^{n+1}_k - p_B}{\partial \varepsilon} - \langle \rho c \rangle_- \frac{\partial v^{n+1}_k - v_B}{\partial \varepsilon} = R_- \equiv -cf - \frac{Qc^2}{\rho \partial H/\partial \rho}$$

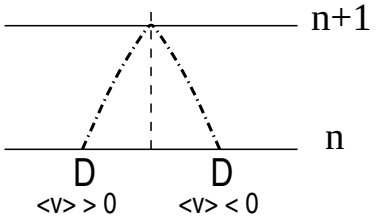


Figura 9.7: *La característica material define un punto D que está a la izquierda o derecha según el signo de promedio $\langle v \rangle_m$ de v en $\mathscr{C}_3$.*

donde debe entenderse que $\langle \rho c \rangle_\pm$ es el promedio sobre la característica $\mathscr{C}_\pm$. De lo anterior se obtiene

$$p^{n+1}_k + \langle \rho c \rangle_+ v^{n+1}_k = \varepsilon R_+ + p_A + \langle \rho c \rangle_+ v_A \qquad (9.3.36)$$

$$p^{n+1}_k - \langle \rho c \rangle_- v^{n+1}_k = \varepsilon R_- + p_B - \langle \rho c \rangle_- v_B \qquad (9.3.37)$$

Los factores que paracen en los dos lados derechos se conocen, de modo que estas dos ecuaciones permiten obtener p^{n+1}_k y v^{n+1}_k.

Estas ecuaciones pueden no ser válidas en los bordes que, típicamente, corresponden a $k = 0$ y $k = N$.

Para determinar la entalpía se usa (9.3.35) en la forma

$$\langle \rho \rangle_m \frac{H^{n+1}_k - H_D}{\varepsilon} - \frac{1}{J} \frac{p^{n+1}_k - p_D}{\varepsilon} = \langle Q \rangle_m$$

donde H_D es el valor de H en el punto D que es la intersección de $\mathscr{C}_3$ con el eje n y se cumple que

$$\begin{aligned} x_D > x_k &\Leftrightarrow \langle v \rangle_m < 0 \\ x_D < x_k &\Leftrightarrow \langle v \rangle_m > 0 \end{aligned}$$

como muestra la figura 9.7 donde H_D se define por medio de

$$H_D = \begin{cases} \langle v \rangle_m \frac{\varepsilon}{h} H^n_{k-1} + (1 - \langle v \rangle_m \frac{\varepsilon}{h}) H^n_k & \text{con} \quad \langle v \rangle_m > 0 \\ |\langle v \rangle_m| \frac{\varepsilon}{h} H^n_{k+1} + (1 - |\langle v \rangle_m| \frac{\varepsilon}{h}) H^n_k & \text{con} \quad \langle v \rangle_m < 0 \end{cases} \qquad (9.3.38)$$

y expresiones similares para p_D. Pero p_k^{n+1} ya se conoce, de modo que

$$H_k^{n+1} = \frac{1}{\langle\rho\rangle_m}\left[\varepsilon\langle Q\rangle_m + \frac{1}{J}\left(p_k^{n+1} - p_D\right)\right] + H_D \tag{9.3.39}$$

Conociendo H_k^{n+1} y p_k^{n+1}, se usa la ecuación de estado para determina la densidad ρ y otras propiedades que pudieran interesar.

9.4. Ecuaciones de segundo orden cuasilineales

Se considerará la ecuación

$$a\frac{\partial^2 U}{\partial t^2} + b\frac{\partial^2 U}{\partial t\,\partial x} + c\frac{\partial^2 U}{\partial x^2} + e = 0 \tag{9.4.1}$$

donde los coeficientes a, b, c, e son, en general, función de t, x , U, $\partial_t U$ y $\partial_x U$.

Se usará la notación

$$\begin{gathered} p = \frac{\partial U}{\partial t}, \quad q = \frac{\partial U}{\partial x}, \\ R = \frac{\partial^2 U}{\partial t^2}, \quad W = \frac{\partial^2 U}{\partial x^2}, \quad S = \frac{\partial^2 U}{\partial t\,\partial x}. \end{gathered} \tag{9.4.2}$$

Tal como en el caso de las ecuaciones de primer orden, se busca las lineas *características* en el plano (x,t), parametrizadas por un parámetro s, a lo largo de las cuales la ecuación diferencial se transforma en la expresión de la diferencial dU cuando varía s, es decir, $dU = p\,dt + q\,dx$ donde dt y dx son las variaciones a lo largo de las características.

La notación anterior permite que la ecuación original se escriba

$$aR + bS + cW + e = 0 \tag{9.4.3}$$

A continuación se manipulará para que desaparezcan las segundas derivadas, lo que conduce a la definición de las características.

Como se ha dicho, la ecuación original anterior será convertida en ecuaciones para curvas características f y g y una forma diferencial para U sobre las características. Las curvas características quedarán definidas por funciones $(x_f(s), t_f(s))$ y $(x_g(s), t_g(s))$ de un parámetro s. Usando un punto para indicar derivada con respecto al parámetro s, se puede escribir

$$\begin{aligned} \dot{p} &= R\dot{t} + S\dot{x} \\ \dot{q} &= S\dot{t} + W\dot{x} \end{aligned} \tag{9.4.4}$$

de donde se despeja

$$\begin{aligned} R &= \frac{\dot{p} - S\dot{x}}{\dot{t}} \\ W &= \frac{\dot{q} - S\dot{t}}{\dot{x}} \end{aligned} \tag{9.4.5}$$

que permite reescribir (9.4.3) en la forma

$$a\frac{\dot{p} - S\dot{x}}{\dot{t}} + bS + c\frac{\dot{q} - S\dot{t}}{\dot{x}} + e = 0$$

que reordenada y multiplicada por $\dot{x}$ da

$$-\dot{t}\,S\left(a\left(\frac{\dot{x}}{\dot{t}}\right)^2-b\frac{\dot{x}}{\dot{t}}+c\right)+a\,\dot{p}\frac{\dot{x}}{\dot{t}}+c\,\dot{q}+e\,\dot{x}=0$$

La ecuación anterior ya no depende ni de R ni de W. El próximo paso es lograr que tampoco dependa de S. El cuociente $\frac{\dot{x}}{\dot{t}}$ será llamado $X(s)$ y, para que la ecuación no dependa de S se exige que el paréntesis grande sea nulo:

$$aX^2-bX+c=0 \tag{9.4.6}$$

la que da, en cada punto $(x(s),t(s))$ dos soluciones para X llamadas $f(s)$ y $g(s)$,

$$\begin{aligned} f &= \frac{b+\sqrt{b^2-4ac}}{2a} \\ g &= \frac{b-\sqrt{b^2-4ac}}{2a} \end{aligned} \tag{9.4.7}$$

Esto quiere decir que (9.4.3) permite dos direcciones en cada punto (t,x)—definidas por $X=\frac{dx}{dt}$—que resuelven (9.4.6). A lo largo de tales curvas dp y dq satisfacen,

$$aX\,dp+c\,dq+e\,dx=0 \tag{9.4.8}$$

Las direcciones que emergen de (9.4.6) se llaman *direcciones características* y las ecuaciones de tipo (9.4.1) son clasificables de acuerdo al tipo de soluciones que emergen de (9.4.6), lo que se decide con el signo/valor de

$$b^2-4\,a\,c \tag{9.4.9}$$

como lo resume la tabla que sigue:

tipo de ecuación	raíces	signo
hiperbólica	reales y diferentes	$b^2-4ac>0$
parabólica	reales e iguales	$b^2-4ac=0$
elíptica	complejas	$b^2-4ac<0$

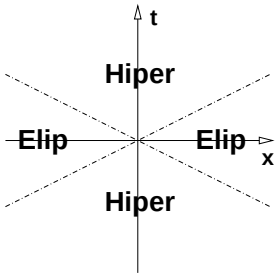


Figura 9.8: *La ecuación (9.4.10) es hiperbólica, parabólica o elíptica dependiendo de la zona del plano (x,t) que se considere.*

Una ecuación puede ser de distinta naturaleza en distintas zonas del espacio (t,x), como por ejemplo

$$xU_{tt}+tU_{tx}+xU_{xx}=F(t,x,U,U_t,U_x) \tag{9.4.10}$$

En este caso $a=x$, $b=t$, $c=x$ y la naturaleza de esta ecuación es hiperbólica, parabólica o elíptica según si $b^2-4ac=t^2-4x^2$ es positivo, nulo o negativo. En la figura se ve los dominios donde es hiperbólica y elíptica. En las fronteras la ecuación es parabólica.

9.5. Ecuaciones hiperbólicas

9.5.1. Planteamiento del problema

En el caso hiperbólico pasan dos características diferentes por cada punto. Ellas tienen pendientes

$$\frac{dx}{dt}=f\,,\qquad \frac{dx}{dt}=g \tag{9.5.1}$$

que se denominarán las f-características y las g-características.

Por lo anterior, la ecuación

$$a\frac{\partial^2 U}{\partial t^2}+b\frac{\partial^2 U}{\partial t\,\partial x}+c\frac{\partial^2 U}{\partial x^2}+e=0 \tag{9.5.2}$$

con coeficientes a, b, c y e que son funciones conocidas de (t,x,U,p,q), tiene características cuyas pendientes se obtienen resolviendo la ecuación algebraica

$$a\left(\frac{dx}{dt}\right)^2-b\frac{dx}{dt}+c=0 \tag{9.5.3}$$

Las raíces se denominan f y g.

A lo largo de las características los diferenciales dp y dq están relacionados por

$$a\frac{dp}{dt}\frac{dx}{dt}+c\frac{dq}{dt}+e\frac{dx}{dt}=0$$

que se puede escribir

$$a\frac{dx}{dt}\,dp+c\,dq+e\,dx=0$$

que toma dos formas, según de qué característica se trate,

$$\begin{aligned} a\,f\,dp+c\,dq+e\,dx&=0\\ a\,g\,dp+c\,dq+e\,dx&=0 \end{aligned} \tag{9.5.4}$$

En las dos relaciones anteriores los diferenciales, a pesar de tener el mismo nombre tienen valores distintos; dp, dq y dx en la cada una de estas ecuaciones representan las variacion de p, q y x en la característica correspondiente, f en la primera y g en la segunda.

Finalmente se usa además

$$dU=p\,dt+q\,dx \tag{9.5.5}$$

que puede evaluarse a lo largo de cualquiera de las dos características.

9.5.2. Integración explícita

Existe una variedad de problemas para los cuales las características son integrables en forma explícita antes de iniciar el método numérico. En lo que sigue se describe el caso en que las características se van obteniendo numéricamente punto a punto.

Sea Γ una curva que no es una característica y supongamos que sobre ella se conocen U, p y q. Sean P, Q puntos muy cercanos en Γ, es decir los valores u_P, u_Q, p_P, p_Q, q_P y q_Q son conocidos.

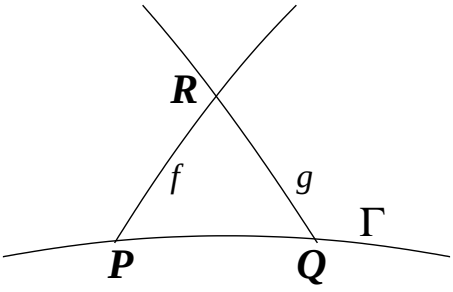


Figura 9.9: *Naturalmente en las ecuaciones hiperbólicas de segundo orden existen dos familias de características.*

El primer paso consiste en resolver (9.4.6)

$$\begin{aligned} a_P\,f^2-b_P\,f+c_P&=0\\ a_Q\,g^2-b_Q\,g+c_Q&=0 \end{aligned} \tag{9.5.6}$$

para escoger un valor para f_P y otro para g_Q consistentes para que exista un punto cercano R donde se corta la f-característica que pasa por P con la g-característica que pasa por Q. En una primera aproximación se toman los arcos PR y QR como rectas con pendientes f_P y g_Q por lo que se puede escribir

$$\begin{aligned} x_R^{(1)} - x_P &= f_P\,(t_R^{(1)} - t_P) \\ x_R^{(1)} - x_Q &= g_Q\,(t_R^{(1)} - t_Q) \end{aligned} \tag{9.5.7}$$

De estas dos ecuaciones se obtiene valores aproximados para $(t_R^{(1)}, x_R^{(1)})$.

A las ecuaciones (9.5.4) se les puede dar la forma aproximada

$$\begin{aligned} a_P\,f_P\,(p_R^{(1)} - p_P) + c_P\,(q_R^{(1)} - q_P) + e_P\,(x_R^{(1)} - x_P) &= 0 \\ a_Q\,g_Q\,(p_R^{(1)} - p_Q) + c_Q\,(q_R^{(1)} - q_Q) + e_Q\,(x_R^{(1)} - x_Q) &= 0 \end{aligned} \tag{9.5.8}$$

que dan primeras expresiones $(p_R^{(1)}, q_R^{(1)})$.

La ecuación para determinar U

$$dU = \frac{\partial U}{\partial t}\,dt + \frac{\partial U}{\partial x}\,dx = p\,dt + q\,dx$$

en forma aproximada es

$$u_R^{(1)} - u_P = \frac{1}{2}\left[(p_P + p_R^{(1)})\,(t_R^{(1)} - t_P) + (q_R^{(1)} + q_P)\,(x_R^{(1)} - x_P)\right] \tag{9.5.9}$$

que determina un valor para $u_R^{(1)}$. También se podría usar la otra característica para aproximarse al valor de $u_R^{(1)}$, o ambas y luego utilizar el promedio.

De esta manera se obtiene primeros valores para las cinco cantidades $\left(t_R^{(1)}, x_R^{(1)}, p_R^{(1)}, q_R^{(1)}, u_R^{(1)}\right)$. Con estos valores se determinan $\left(a_R^{(1)}, b_R^{(1)}, c_R^{(1)}, e_R^{(1)}\right)$.

A partir de todo lo anterior se define un método iterativo para mejorar los valores en R. Este método iterativo permite obtener valores $X^{(n+1)}$ a partir de valores $X^{(n)}$. Se ingresa al ciclo iterativo siguiente ya conociendo todo sobre las cantidades en el punto R a primer orden. Es decir, en lo que sigue, el primer valor de n es 1 y se determinan cantidades $X^{(2)}$.

1. Se usa el conocimiento de la iteración n para resolver la ecuación

$$a_R^{(n)} X^2 - b_R^{(n)} X + c_R^{(n)} = 0 \tag{9.5.10}$$

   cuyas raices son $f_R^{(n)}$ y $g_R^{(n)}$.

2. Se reemplaza (9.5.7) por

$$x_R^{(n+1)} - x_P = \frac{f_P + f_R^{(n)}}{2}\,(t_R^{(n+1)} - t_P) \qquad x_R^{(n+1)} - x_Q = \frac{g_Q + g_R^{(n)}}{2}\,(t_R^{(n+1)} - t_Q) \tag{9.5.11}$$

   para obtener $(t_R^{(n+1)}, x_R^{(n+1)})$.

3. Se reemplaza (9.5.8) por

$$\begin{aligned}
&\frac{a_P+a_R^{(n)}}{2}\,\frac{f_P+f_R^{(n)}}{2}\,(p_R^{(n+1)}-p_P)+\frac{c_P+c_R^{(n)}}{2}\,(q_R^{(n+1)}-q_P)+\frac{e_P+e_R^{(n)}}{2}\,(x_R^{(n+1)}-x_P)=0\\
&\frac{a_Q+a_R^{(n)}}{2}\,\frac{g_Q+g_R^{(n)}}{2}\,(p_R^{(n+1)}-p_Q)+\frac{c_Q+c_R^{(n)}}{2}\,(q_R^{(n+1)}-q_Q)+\frac{e_Q+e_R^{(n)}}{2}\,(x_R^{(n+1)}-x_Q)=0
\end{aligned} \tag{9.5.12}$$

para obtener $(p_R^{(n+1)},q_R^{(n+1)})$.

4. Hecho todo lo anterior se usa (9.5.9) en la forma

$$u_R^{(n+1)}-u_P=\frac{p_P+p_R^{(n+1)}}{2}\,(t_R^{(n+1)}-t_P)+\frac{q_R^{(n+1)}+q_P}{2}\,(x_R^{(n+1)}-x_P) \tag{9.5.13}$$

para tener $u_R^{(n+1)}$.

5. Habiéndose obtenido en primera aproximación las cantidades: t_R, x_R, p_R, q_R y u_R se calculan, en esta aproximación, los coeficientes $a_R^{(n+1)},b_R^{(n+1)},c_R^{(n+1)},e_R^{(n+1)}$.

Los pasos recién descritos se repiten hasta tener convergencia para las cinco cantidades que se quiere evaluar en el punto R. Si P, Q y R están suficientemente cerca se requiere de pocas iteraciones para que las interaciones anteriores converjan.

Recorriendo la curva Γ con pares cercanos (P,Q), se obtiene una nueva curva (esto es, nuevos puntos (x,t)), cercana a Γ sobre la cual se determina la función y sus derivadas. Esta nueva curva se usa para seguir avanzando.

9.6. Condiciones de borde

En lo que sigue, se analiza la forma de determinar los valores en los puntos R en el borde izquierdo a partir de los valores que ya se han determinado en puntos Q como lo ilustra la figura 9.10.

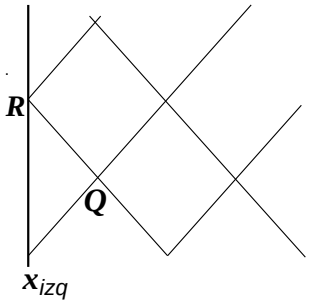


Figura 9.10: *En el texto se analiza la forma como se determina las cantidades asociadas a un punto R en el borde izquiero cuando se tiene toda la información en un punto Q sobre la misma característica.*

(a) Si la condición $U(x,t)$ en $x=x_{\text{izq}}$ es rígida entonces es dato

$$U_R=U(x_{\text{izq}},t) \quad \text{que implica conocer} \quad p_R=p(x_{\text{izq}},t) \tag{9.6.14}$$

(b) Si la condición $U(x,t)$ en $x=x_{\text{izq}}$ es derivativa entonces es dato

$$U'_R=\left[\frac{\partial U(x,t)}{\partial x}\right]_{x=x_{\text{izq}}}=q_R \tag{9.6.15}$$

De modo que según si la condición de borde a la izquierda es rígida o derivativa se conoce p_R o q_R.

Puesto que se conoce todo en el punto Q de la figura 9.10, se determina que

$$g_Q = \frac{b_Q - \sqrt{b_Q^2 - 4a_Q c_Q}}{2a_Q}$$

que se usa en la forma discreta de $dx = g\,dt$,

$$x_{\text{izq}} - x_Q = g_G\,(t_R - t_Q) \tag{9.6.16}$$

para determinar t_R. Seguídamente se una la forma discreta de $a\,g\,dp + c\,dq + e\,dx$,

$$a_Q g_Q\,(p_R - p_Q) + c_Q\,(q_R - q_Q) + e_Q\,(x_{\text{izq}} - x_Q) \tag{9.6.17}$$

En esta expresión se conocen todos los factores excepto por uno. Si la condición de borde a la izquierda es rígida se conoce p_R y (9.6.17) permite determinar q_R. Si condición de borde a la izquierda es derivativa se conoce q_R y (9.6.17) determina p_R.

Finalmente de (9.5.9) se tiene que

$$U_R = U_Q + \frac{1}{2}\left[(p_R + p_Q)(t_R - t_Q) + (q_R + q_Q)(x_{\text{izq}} - x_Q)\right]$$

Todos los factores que hay en el lado derecho ya han sido determinados de modo que se ha integrado correctamente desde Q hasta R.

Un razonamiento similar se pude usar para tratar las condiciones de borde al lado derecho.

9.7. Problemas

1. Integre el problema de un fluido compresible unidimensional sin viscosidad que obedece las ecuaciones

$$\begin{aligned} \frac{\partial \rho}{\partial t} + \frac{\partial \rho\, v}{\partial x} &= 0 \\ \rho\,\frac{\partial v}{\partial t} + \rho\, v \frac{\partial v}{\partial x} &= -\frac{\partial p}{\partial x} \end{aligned} \tag{9.7.1}$$

   y ecuación de estado $p = A\rho^\gamma$. Considere $0 \le x \le 1$, condiciones de borde $v(0,t) = 0$ y $v(1,t) = 0$ y condición inicial $\rho(x,0) = 1$ en todo el domino excepto que en el intervalo $0 \le x \le 0{,}1$ vale

$$\rho(x,0) = 1 + 0{,}1\cos(5\pi x)$$

   Escoja $A = 2$, $\gamma = \frac{5}{3}$ y divida el intervalo $(0,1)$ en 10 mil trazos iguales. Guarde los cuatro campos en un archivo a 4 columnas (x,t,ρ,v), unos 100 valores cada vez. Por ejemplo la primera columna contiene x_0, x_{100}, x_{200} ..., la segunda tiene t_0, t_{100}, t_{200} ... etc. Su archivo debe registrar estos bloques de altura 101 a intervalos regulares (tal vez cada 400 barridos de $0 \le x \le 1$) para tener un visión de la evolución del sistema. Itere su sistema al menos unas 25 mil veces.

   Una de las muchas formas de mostrar lo que ha obtenido puede ser, por ejemplo, mostrar la evolución de la densidad usando gnuplot con las instrucciones

```
set nokey
set nosurface
set contour base
set view 0, 0, 1, 1
splot "mis.datos" u 2:1:3 w d
```

En el caso de la evolución de $v(x,t)$ también interesa ver el signo.

2. Integre la ecuación

$$\frac{\partial^2 U}{\partial t^2} - \frac{\partial^2 U}{\partial x^2} + \alpha \frac{\partial U}{\partial t} = 0$$

usando el método de las características en el intervalo $0 \leq x \leq 1$ con las condiciones de borde $U(0,t) = 2\sin \omega t$, $U(1,t) = 0$, y con las condiciones iniciales $U(x,0) = 0$ y $\partial U(x,0)/\partial t = 0$. Estudie los casos con $\omega = 14\pi$ y $\omega = 15\pi$, y $\alpha = 2{,}0$ y $\alpha = 8{,}0$. Divida el intervalo $(0,1)$ en $N + \frac{1}{2}$ trazos de longitud $h = 1/(N + \frac{1}{2})$ usando $N = 5000$. Necesitará construir dos rutinas de iteración, las pares y las impares.

a) Dibuje $U(x, t = \text{fijo})$ para diversos valores k de iteraciones: para $k = 5000$, 10500 x 13400. (Para $k = 5000$ la función U debiera ser cero sobre la mitad del intervalo).

b) Grafique $U(x = \frac{1}{7}, t)$ y $U(x = \frac{1}{2}, t)$ desde $t = 0$ hasta un tiempo suficientemente largo para que se vea que el sistema ha alcanzado un estado de régimen.

c) Una vez en el estado de régimen dibuje $U(x,t_0)$ para un instante t_0 para el cual $U(0,t_0) = 2{,}0$

La ecuación planteada es la que satisfacen las componentes del campo eléctrico y magnético cuando se propaga una onda electromagnética en un medio algo conductor. El coeficiente α es proporcional a la conductividad del medio. La condición de borde en $x = 0$ puede pensarse como la que impone una onda que llega desde el vacío al medio conductor que comienza en $x = 0$. La condición de borde en $x = 1$ corresponde a la presencia de un conductor perfecto de ahí en adelante. Ese borde actua como un espejo.

Capítulo 10

Transformada rápida de Fourier

Este capítulo sigue de cerca la primera parte del correspondiente capítulo de Numerical Recipes y aun está muy incompleto.

10.1. La transformada continua

10.1.1. La delta de Dirac

$$\delta(f) = \int_{-\infty}^{\infty} e^{2\pi i f t}\, dt \qquad \text{y} \qquad \delta(t) = \int_{-\infty}^{\infty} e^{-2\pi i f t}\, df \tag{10.1.1}$$

donde $\delta(y \neq 0) = 0$ y

$$\int_{-\infty}^{\infty} \delta(x)\, h(x)\, dx = h(0) \tag{10.1.2}$$

para cualquier función $h(x)$ continua en $x = 0$.

10.1.2. Relación entre una función y su transformada

Si $g(t)$ es una función continua y $G(f)$ es su *transformada continua* de Fourier se satisface

$$G(f) = \int_{-\infty}^{\infty} g(t)\, e^{2\pi i f t}\, dt \qquad \Longleftrightarrow \qquad g(t) = \int_{-\infty}^{\infty} G(f)\, e^{-2\pi i f t}\, df \tag{10.1.3}$$

10.2. Transformada de Fourier discreta

En lugar de considerar la función continua $g(t)$ con $-\infty < t < \infty$ se utilizará ahora la muestra discreta de datos g_k,

$$g_k = g(\Delta k), \qquad k = 0, \pm 1, \pm 2, \pm 3, ... \tag{10.2.1}$$

y Δ se llama la tasa de muestreo.

Asociada a Δ se tiene la *frecuencia crítica de Nyquist*,

$$f_c \equiv \frac{1}{2\Delta} \tag{10.2.2}$$

Hay razones que permiten apreciar la importancia de la frecuencia de Nyquist:

a) Si se muestrea una función continua $g(t)$ a intervalos Δ y $g(t)$ resulta tener un ancho de banda limitado por frecuencias de magnitudes menores que f_c, esto es,

$$G(f) = 0 \quad \text{para todo} \quad |f| \geq f_c \qquad \text{ancho de banda limitado} \tag{10.2.3}$$

entonces la función $g(t)$ queda totalmente determinada por el muestreo g_k. En efecto, si $g(t)$ está dado por el *teorema del muestreo* o "*sampling theorem*",

$$g(t) = \Delta \sum_{k=-\infty}^{\infty} g_k \frac{\sin(2\pi f_c (t - k\Delta))}{\pi (t - k\Delta)} \tag{10.2.4}$$

La expresión anterior muestra que el contenido de información de una función de ancho de banda limitado es, en cierto sentido, infinitamente menor que el de una función continua general. En tal caso basta una muestra discreta y no continua.

A menudo se debe trabajar con una señal que—por razones físicas—tiene un ancho de banda limitado, al menos en un sentido aproximado. Por ejemplo, la señal puede haber pasado por un amplificador que responde a un ancho de banda limitado. En tal caso, el teorema dice que toda la información de la señal puede ser registrada muestreando a una tasa Δ^{-1}, igual al doble de la frecuencia máxima, (10.2.2), que entrega el amplificador.

b) Si el muestrea una función continua no está limitada a un ancho de banda ($-f_c < f < f_c$) frecuencia menor que f_c ocurre que toda la potencia espectral que está fuera del rango $-f_c < f < f_c$ es erróneamente trasladada a ese rango. Este fenómeno se denomina *traducción falsa* (en inglés se dice *aliasing*). Se reitera: cualquier componente de frecuencia fuera del rango $(-f_c, f_c)$ llevada erradamanete al rango $(-f_c, f_c)$ por el mero hecho de tomar muestras discretamente.

c) La forma de superar la traducción falsa consiste en primero tener claro cuál es el ancho de banda de la señal, o bien imponer un límite conocido por medio de un filtro analógico de la señal continua y, segundo, muestrear a una tasa suficientemente fina (Δ pequeño) para dar al menos dos puntos por ciclo para la frecuencia más alta.

d) Para reparar el efecto de la traducción falsa conviene estimar la transformada de Fourier $G(f)$ imponiendo que ella sea nula fuera del rango $(-f_c, f_c)$. Si, al hacer esto, se observa que $G(f)$ se acerca a cero en sus extremos ($f \to f_c$ y $f \to -f_c$) se puede suponer que se tendrá un resultado razonablemente bueno, pero si esto no ocurre podemos sospechar fuertemente que componentes fuera del rango se han colado al rango crítico.

Se tiene entonces una muestra discreta y de tamaño finito N,

$$g_k \equiv g(t_k), \qquad t_k = k\Delta, \qquad k = 0, 1, ..., N-1 \tag{10.2.5}$$

Supongamos además que N es par.

Ya que se tiene un número finito de datos se quiere calcular igual número N de transformadas $G(f)$ en el rango $(-f_c, f_c)$. Se escoge hacerlo para las frecuencias discretas:

$$f_n = \frac{n}{\Delta N}, \qquad n = -\frac{N}{2}, ..., \frac{N}{2} \tag{10.2.6}$$

Se acaba de definir $N+1$ frecuencias, pero en lo que sigue se verá que solo se toman en cuenta N. La exponencial $\exp[2\pi i f t]$ en el caso discreto (usando $f = f_n$ y $t = k\Delta$) es

$$e^{2\pi i f t} = e^{2\pi i n k/N} \tag{10.2.7}$$

Esta exponencial no cambia si se reemplaza $n \to n \pm N$ porque

$$e^{2\pi i (n\pm N)k/N} = e^{2\pi i n k/N} e^{\pm 2\pi i k} = e^{2\pi i n k/N} \tag{10.2.8}$$

ya que $\exp[\pm 2\pi i k] = \cos(2\pi i k) = 1$. Esto es, $e^{2\pi i n k/N}$ es periódica bajo el reemplazo $n \to n \pm N$. En particular el caso $n = 0$ es equivalente al caso $n = N$ por lo que a ambos se les asocia la frecuencia nula.

Consideremos los casos extremos $n = \pm\frac{N}{2}$ a los que se asocia las frecuencias $\pm f_c$. Si el caso es $n = -\frac{N}{2}$, usando el reemplazo $n \to n+N$, se obtiene el caso $n = \frac{N}{2}$, lo que muestra que los casos $f = f_c$ y $f = -f_c$ son el mismo resolviendo así lo que se comentó bajo (10.2.6), esto es, solo N valores de n son necesarios.

Supongamos ahora que se tiene un término $e^{2\pi i f t} = e^{2\pi i n k/N}$ con $-\frac{N}{2} < n < 0$, lo que se va a describir como $n = -\frac{N}{2} + j$ con $0 < j < \frac{N}{2}$ y que corresponde a la frecuencia

$$f_{-\frac{N}{2}+j} = -f_c + \frac{j}{\Delta N} < 0$$

Pero, dada la periodicidad, se puede hacer el reemplazo $n \to n+N$, lo que conduce a una exponencial que, en lugar del n original, tiene $n' = j + \frac{N}{2} > \frac{N}{2}$, esto es, se tiene la equivalencia

$$f_{j-N/2} \equiv -f_c + \frac{j}{\Delta N} \qquad \Longleftrightarrow \qquad f_{j+N/2} \equiv f_c + \frac{j}{\Delta N} \tag{10.2.9}$$

Esta es una equivalencia entre una frecuencia negativa y una positiva mayor que f_c. En lo sucesivo las frecuencias mayores que f_c serán reemplazadas por la frecuencia negativa correspondiente. Ésta última tiene módulo menor que f_c.

Se ve entonces que las frecuencias que entran en este formalismo tiene siempre una magnitud que no sobrepasa f_c.

De todo lo anterior se ve que el rango para n dado en (10.2.6) puede ser reemplazado por el rango $n = 0, 1, ..N-1$ de tal manera que para $0 < n \leq \frac{N}{2}$ se obtiene freciencias f_n tales que

$$0 < n \leq \frac{N}{2} \qquad \Longleftrightarrow \qquad 0 \leq f_n < f_c$$

Mientras que con n tal que $\frac{N}{2} \leq n < N$ se asocia frecuencias negativas.

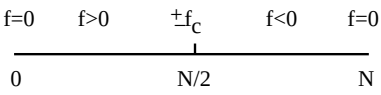


Figura 10.1: *Esta figura ilustra la relación entre los valores de n abajo con las frecuencias que se le asocia, arriba. Ver también la tabla (10.2.10)*

La tabla que sigue muestra la relación entre los valores de n y la frecuencia asociada

$$\begin{array}{|c|l|}\hline 0 \le n < \dfrac{N}{2} & f_n = \dfrac{n}{\Delta N} = \dfrac{n}{(N/2)}\, f_c \\ n = \dfrac{N}{2} & f_{N/2} = \pm\tfrac{1}{2\Delta} = \pm f_c \\ \dfrac{N}{2} < n \le N & f_n = -\dfrac{N-n}{\Delta N} = -\dfrac{N-n}{(N/2)}\, f_c \\ \hline \end{array} \tag{10.2.10}$$

Nótese que las frecuencias de mayor magnitud absoluta son precisamente $\pm f_c$.

Con lo anterior se define la transformada

$$G(f_n) = \int_{-\infty}^{\infty} g(t)\, e^{2\pi i f_n t}\, dt \approx \sum_{k=0}^{N-1} g_k\, e^{2\pi i f_n t_k}\, \Delta = \Delta \sum_{k=0}^{N-1} g_k\, e^{2\pi i k n/N} \tag{10.2.11}$$

Pero la transformada de Fourier discreta G_n se define omitiendo el factor Δ,

$$G_n \equiv \sum_{k=0}^{N-1} g_k\, e^{2\pi i k n/N} \tag{10.2.12}$$

que gráficamente se representa por la figura 10.1. Por ejemplo, en el caso $N = 16$, esta figura arriba tiene las 17 frecuencias:

$$0,\ \ \frac{1}{8}f_c,\ \ \frac{2}{8}f_c,\ \ \frac{3}{8}f_c,\ \ \ldots\ \ \frac{7}{8}f_c,\ \ \pm f_c,\ \ -\frac{7}{8}f_c,\ \ -\frac{6}{8}f_c\ \ \ldots\ \ -\frac{1}{8}f_c,\ \ 0$$

y abajo los valores de n

$$0 \quad 1 \quad 2 \quad \ldots 7 \quad 8 \quad 9 \quad \ldots 14 \quad 15 \quad 16$$

Puesto que la primera y última frecuencias son iguales (y nulas) efectivamente hay tan solo 16 frecuencias diferentes.

La transformada discreta asocia a los N números complejos g_k los N números complejos G_n. Ella no depende de ningún parámetro dimensional tal como Δ.

$$G(f_n) \approx \Delta\, G_n \tag{10.2.13}$$

La transformada inversa que—a partir de G_n—recupera la función original g_k, es

$$g_k = \frac{1}{N} \sum_{n=0}^{N-1} G_n\, e^{-2\pi i k n/N} \tag{10.2.14}$$

y el teorema de Parseval (**??**) toma la forma

$$\sum_{k=0}^{N-1} |g_k|^2 = \frac{1}{N} \sum_{n=0}^{N-1} |G_n|^2 \tag{10.2.15}$$

También se puede escribir formas discretas para la convolución y la correlación.

10.3. La transformada rápida de Fourier (FFT)

Se define la matriz $\mathbb{W}$ de componentes W^{nk} como

$$W^{nk} = e^{2\pi ikn/N} \tag{10.3.1}$$

con lo cual (10.2.12) se puede escribir

$$G_n = \sum_{k=0}^{N-1} W^{nk} g_k \tag{10.3.2}$$

que puede mirarse como el producto de la matriz $\mathbb{W}$ multiplicando al vector $\vec{g}$. Según la expresión anterior, cada uno de los G_n requiere de N multiplicaciones y $n = 0, 1, \ldots N-1$ toma N valores, de modo que hasta aquí se debe hacer $\mathcal{O}(N^2)$ multiplicaciones. En lo que sigue se verá cómo reducir el número de operaciones.

En efecto, en lo que sigue se ve que la transformada discreta de Fourier se puede reescribir como la suma de dos transformadas de largo $N/2$, una usando los índices pares y otra los índices impares:

$$\begin{aligned}
G_n &= \sum_{k=0}^{N-1} e^{2\pi ikn/N} g_k \\
&= \sum_{k=0}^{N/2-1} e^{2\pi in(2k)/N} g_{2k} + \sum_{k=0}^{N/2-1} e^{2\pi in(2k+1)/N} g_{2k+1} \\
&= \sum_{k=0}^{N/2-1} e^{2\pi ink/(N/2)} g_{2k} + W^n \sum_{k=0}^{N/2-1} e^{2\pi ink/(N/2)} g_{2k+1}\,, \qquad n = 0, 1, \ldots N-1 \quad (10.3.3) \\
&= G_n^+ + W^n G_n^- \quad (10.3.4)
\end{aligned}$$

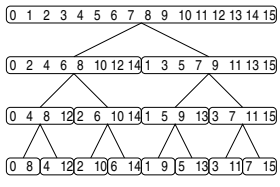
donde $\pm$ quiere decir parte par o impar. Se puede comprobar que G_n^+ y G_n^- son periódicas en n con período $N/2$ debido a lo siguiente:

$$e^{2\pi i(n+N/2)k/(N/2)} = e^{2\pi ink/(N/2)}\, e^{2\pi ik} = e^{2\pi ink/(N/2)} \tag{10.3.5}$$

Con esto debiera ser claro que el costo de calcular G_n^+ es $N/2$ productos complejos y lo mismo ocurre con G_n^-.

Dado un n tal que $0 \le n < \frac{N}{2}$ para calcular $G_{n+N/2}$ se debe considerar sumas como en (10.3.3). En ambas sumas aparecen exponenciales como (10.3.5) y además y, puesto que

$$W^{n+N/2} = W^n e^{\pi i} = -W^n\,, \tag{10.3.6}$$



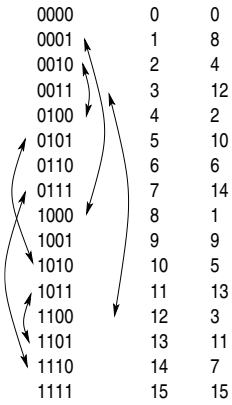


Figura 10.2: *A la izquierda se muestra como las sucesivas separaciones en términos pares e impares lleva al orden final. A la derecha se ilustran el reordenamiento de los índices de 16 datos comprobándose que corresponde a la inversión de los bits.*

se cumple que

$$G_{n+\frac{N}{2}} = G_n^+ - W^n G_n^- \tag{10.3.7}$$

lo que muestra que si se calcula los G_n^+ y G_n^- con $0 \leq n < \frac{N}{2}$, se puede calcular no tan solo los G_n sino además los $G_{n+\frac{N}{2}}$, esto es, hay que hacer la mitad de los cálculos para obtener todos los G_n.

Habiendo reducido el cálculo de G_n al cálculo de G_n^+ y G_n^-, con la mitad del esfuerzo, en forma semejante se puede reducir el problema de calcular G_n^+ al problema de calcular sus $N/4$ datos pares y sus $N/4$ datos impares, lo que podríamos llamar el cálculo de G_n^{++} y G_n^{+-}. Y algo similar se hace con G_n^- dando G_n^{-+} G_n^{--}. Esto hace que el caso más sencillo se presente cuando N es una potencia de 2, $N = 2^q$.

Teniendo $N = 2^q$ diversos autores argumentan que la transformada de Fourier estándar requiere de $N^2 = 2^{2q}$ operaciones, mientras que la FFT (transformada de Fourier rápida) requiere de $q2^q = N \ln N$ operaciones. Por ejemplo, si $q = 12$ se reduce de 2^{24} =16.777.216 a 12×2^{12} =49.152 operaciones.

Se ve de lo anterior que existe una relación de recurrencia donde las componentes llegan a tener una secuencia de índices "par" e "impar". Se hace uso de la representación binaria de n, esto es, n es expresado con ceros y unos, de modo que a 0 se asocia "par" y a 1 se asocia "impar", como se ilustra en las figura 10.2 y en (**??**). Las transformadas cada vez más gruesas (menos puntos) deben ser vistas, en la representación binaria de n, como la eliminación secuencial de los bits menos relevantes.

La idea básica es la siguiente. Se toma el vector original $\vec{g}$ de componentes g_j y se lo reordena según el orden natural que resulta una vez que se cambia cada índice j por el que corresponda en una inversión de bits como se ilustra en la Fig. 10.2. Esto corresponde a tomar objetos como $G_n^{+--+-+-++-\cdots}$ e invertir el orden de estos índices superiores.

Lo anterior no es tan sencillo y debe explicarse en más detalles. Si se tiene tan solo 16 datos, en la primera separación, tipo (10.3.4), se suman $(0,2,4,6,8,10,12,14)$ y luego los datos $(1,3,5,7,9,11,\ 13,15)$; en la segunda separación se debe sumar $(0,4,8,12)$ y $(2,6,10,14)$ seguido de $(1,5,9,13)$ y $(3,7,11,15)$; en la tercera separacion se tiene $(0,8)$, $(4,12)$, $(2,10)$, $(6,14)$, $(1,9)$, $(5,13)$, $(3,11)$, $(7,15)$. Si se compara los pares anteriores con la columna de la derecha en la Fig. **??** se comprueba que coinciden. Esto hace conveniente que se reordene los datos colocando el dato 0 seguido del dato 8, seguido de 4, seguido del 12 etc.

Cooley y Tukey demostraron en 1965 que si el índice de los datos se coloca en forma binaria, el reordenamiento de los datos corresponde a una inversión de los *bits* que es lo que señalan las dobles flechas en la figura, por ejemplo, $[0101]$ intercambia lugar con $[1010]$: el quinto dato es renombrado como el décimo y el décimo como el quinto.