Chapter 1

Operaciones matemáticas básicas

1.1 Introducción

Existen tres operaciones numéricas que son de primordial importancia en el modelado computacional de la mayor parte de los sistemas físicos. Estos son la diferenciación, la cuadratura y la búsqueda de raíces. Supongamos que somos capaces de calcular el valor de una función, digamos $f(x)$, para cualquier valor de la variable independiente x. En el caso de la diferenciación, buscamos una de las derivadas de f para un dado valor de x. En el caso de la cuadratura, que es en principio el proceso inverso de la diferenciación, deseamos obtener la integral definida de f entre dos límites especificados. Aclaramos aquí que nos reservamos el término de "integración " para el proceso de resolución de ecuaciones diferenciales ordinarias, que discutiremos más adelante. Finalmente, en el caso de la búsqueda de raíces, deseamos obtener el valor o los valores de x para los cuales la funcion f se anula.

Cuando f es conocida analíticamente, es casi siempre posible con el trabajo adecuado derivar fórmulas explícitas para las derivadas de f, y es a menudo posible calcular la integral definida. De todos modos, ocurre frecuentemente que no se puede evaluar o que no se puede encontrar un método analítico, aún cuando $f(x)$ pueda ser evaluada. Esto puede ocurrir porque se requiere algún procedimiento numérico complicado para evaluar f, y por lo tanto no tenemos una expresión analítica a la cual aplicar las reglas de derivación o de integración, o aún peor, porque sólo podemos tener a f eval-

uada en algunos puntos discretos de una red. El primero puede ser el caso de la obtención de la compresibilidad a partir de la energía de un metal, la cual se obtenga a partir del cálculo mecanocuántico de la energía de un sistema. El segundo puede ser el caso de la obtención de alguna propiedad integral a partir de un conjunto de datos experimentales. En ambas situaciones, debemos emplear fórmulas aproximadas que expresen las derivadas y las integrales en términos de los valores de f disponibles. Aún más, las raíces de la mayor parte de las funciones conocidas, con unas pocas excepciones, no se pueden calcular analíticamente y entonces los métodos numéricos son esenciales. Un ejemplo de ello lo encontramos en la búsqueda de raíces en la ecuación (1.1).

$$e^x - x = 0 \tag{1.1}$$

Este capítulo trata con la realización computacional de estas tres operaciones básica. La técnica central consiste en aproximar f por una función simple (tal como un polinomio de primero o segundo grado) sobre el cual estas operaciones se puedan realizar fácilmente. Vamos a derivar aquí solamente las fórmulas más simples y de uso corriente, las más complejas se pueden encontrar en textos más específicos de análisis numérico.

1.2 Diferenciación numérica

Supongamos que estamos interesados en la derivada de f en $x = 0$, que denotaremos $f'(0)$. Las fórmulas que derivaremos se pueden generalizar con facilidad para un x arbitrario por traslación. Supongamos que conocemos f sobre una red equiespaciada de valores de x, lo que podemos simbolizar como:

$$f_n = f(x_n);\ x_n = nh\ (n = 0, \pm 1, \pm 2, ...) \tag{1.2}$$

y que nuestro objetivo es computar un valor aproximado de $f'(0)$ en término de las f_n,tal como se muestra en la figura 1.1.

Comenzaremos usando una serie de Taylor para expandir a f en las proximidades de $x = 0$.

$$f(x) = f_0 + xf' + \frac{x^2}{2!}f'' + \frac{x^3}{3!}f''' + ... \tag{1.3}$$

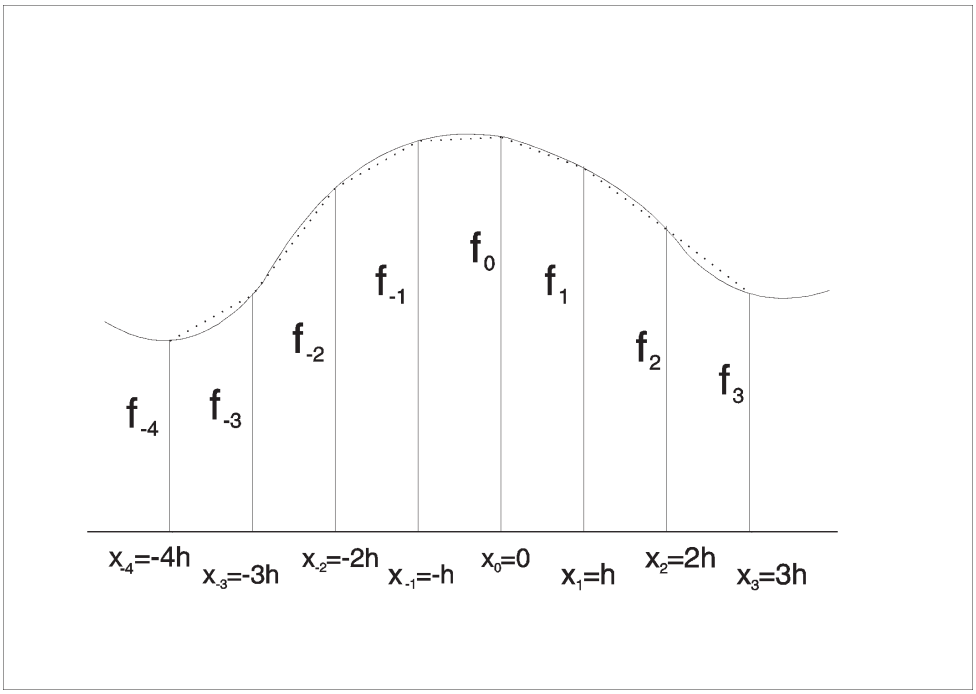

Figure 1.1: Valores de la función f para una red equiespaciada. La líneas discontinuas muestran una interpolación lineal entre los puntos de la red.

donde todas las derivadas están evaluadas en $x = 0$. Evaluemos esta expresión en $x = h$ y $x = -h$:

$$f_{+1} = f(x = h) = f_0 + hf' + \frac{h^2}{2!}f'' + \frac{h^3}{3!}f''' + \mathcal{O}(h^4) \tag{1.4}$$

$$f_{-1} = f(x = -h) = f_0 - hf' + \frac{h^2}{2!}f'' - \frac{h^3}{3!}f''' + \mathcal{O}(h^4) \tag{1.5}$$

donde $\mathcal{O}(h^4)$ significa términos del orden de h^4 o superior, de modo que podemos generalizar:

$$f_{\pm 1} = f(x = \pm h) = f_0 \pm hf' + \frac{h^2}{2!}f'' \pm \frac{h^3}{3!}f''' + \mathcal{O}(h^4) \tag{1.6}$$

y análogamente:

$$f_{\pm 2} = f(x = \pm 2h) = f_0 \pm 2hf' + 2h^2\ f'' \pm \frac{4h^3}{3}f''' + \mathcal{O}(h^4) \tag{1.7}$$

Para estimar la magnitud de los términos de orden superior, podemos suponer que f y sus derivadas son todas del mismo orden de magnitud, como es el caso de muchas magnitudes físicas de relevancia.

Después de sustraer la ecuación (1.5) de la ecuación (1.4) tenemos:

$$f_{+1} - f_{-1} = 2hf' + \frac{h^3}{3}f''' + \mathcal{O}(h^5) \tag{1.8}$$

de donde podemos despejar f':

$$f' = \frac{f_{+1} - f_{-1}}{2h} - \frac{h^2}{6}f''' + \mathcal{O}(h^4) \tag{1.9}$$

El término que contiene a f'' se vuelve despreciable cuando h se vuelve pequeño y es el error dominante asociado con la aproximación de diferencias finitas que retiene solamente al primer término:

$$f' \approx \frac{f_{+1} - f_{-1}}{2h} \tag{1.10}$$

Esta fórmula de "tres puntos" sería exacta si f fuese un polinomio de segundo grado en el intervalo de tres puntos $[-h, +h]$, dado que la derivada de tercer orden (y todas las de orden superior al tercero) se anularían. De este modo, la

esencia de la ecuación (1.10) es la suposición de la validez de una interpolación de f a través de un polinomio cuadrático en los tres puntos $x = 0, \pm h$

La ecuación (1.10) es un resultado muy natural, reminiscente de las fórmulas usadas para definir la derivada en los cursos elementales de cálculo. El término de error (del orden de h^2) se puede en principio hacer tan pequeño como se desee usando valores cada vez más pequeños de h. Nótese también que esta ecuación es la diferencia simétrica alrededor de $x = 0$, la cual es más precisa (por un orden de magnitud en h) que las diferencias finitas hacia adelante o hacia atrás, que pueden obtenerse despejando de las ecuaciones (1.4) o (1.5) para obtener:

$$f' = \frac{f_1 - f_0}{h} + \mathcal{O}(h) \tag{1.11}$$

y

$$f' = \frac{f_0 - f_{-1}}{h} + \mathcal{O}(h) \tag{1.12}$$

Esta fórmula de "dos puntos " están basadas en la suposición de que f está bien aproximada por una función lineal en el intervalo entre $x = 0$ y $x = \pm h$.

Como ejemplo concreto, consideremos la evaluación $f'(x = 1)$ cuando $f(x) = \sin\ x$. La respuesta exacta es $f'(x = 1) = \cos 1 = 0.540302...$ El siguiente programa FORTRAN evalúa la ecuación (1.10), para el valor de h que se de como entrada:

```
      IMPLICIT REAL*8 (A-H,O-Z)
      X=1.D0
      EXACT=DCOS(X)
10    PRINT *, 'ENTRE VALOR DE H (.LE. 0 TO STOP) '
      READ *, H
      IF(H .LE. 0)STOP
      FPRIME=(DSIN(X+H)-DSIN(X-H))/(2*H)
      DIFF=EXACT-FPRIME
      PRINT 20,H,DIFF
20    FORMAT( ' H= ', E15.8, 5X, 'ERROR=', E15.8)
      GOTO 10
      END
```

Como puede observarse, el valor de H se pide del teclado. En el caso de que se entre un valor negativo de H, se detendrá la ejecución del programa. La

elección de las variables sigue la intuición natural, y la fórmula matemática dada en la ecuación (1.10) se transcribe usando la función SIN en la sexta línea del programa. La cantidad de dígitos significativos para la escritura de H y del errror se especifica en el formato señalado en 20. En la penúltima línea del programa, se realiza un salto a la línea marcada con 10 mediante la instrucción GOTO.

En la siguiente tabla se muestran los resultados generados con este programa, y algunos similares evaluando las fórmulas de las diferencias hacia atrás y hacia adelante.

h	Simétrica 3 puntos	Adelante 2 puntos	Atrás 2 puntos
1.0000000000000	0.0856535924553	0.4724758638504	-.3011686789398
0.1000000000000	0.0009000536984	0.0429385533328	-.0411384459360
0.0100000000000	0.0000090049934	0.0042163248563	-.0041983148695
0.0010000000000	0.0000000900504	0.0004208255077	-.0004206454070
0.0001000000000	0.0000000009005	0.0000420744497	-.0000420726487
0.0000100000000	0.0000000000090	0.0000042073639	-.0000042073459
0.0000010000000	0.0000000000001	0.0000004207356	-.0000004207354
0.0000001000000	0.0000000000000	0.0000000420738	-.0000000420738
0.0000000100000	-.0000000000008	0.0000000042059	-.0000000042075
0.0000000010000	-.0000000000117	0.0000000004220	-.0000000004454
0.0000000001000	-.0000000000659	-.0000000000659	-.0000000000659
0.0000000000100	0.0000000004762	0.0000000031867	-.0000000022343
0.0000000000010	0.0000000194498	0.0000000194498	0.0000000194498
0.0000000000001	0.0000001820801	0.0000004531336	-.0000000889705

Tabla 1.1 Error al evaluar $d \sin x/dx \mid_{x=1}= 0.540302...$

Vale la pena señalar que el resultado mejora cuando cuando decrecemos h, pero solamente hasta cierto punto, luego empeora. Eso es así porque la aritmética se lleva a cabo con una cierta precisión, de manera que cuando se forman las diferencias en el numerador, se tienen grandes errores de redondeo si h es pequeño y f_1y f_{-1} difieren en muy poco.

Supongamos por ejemplo que $h = 10^{-6}$, y que trabajamos con una precisión de 6 cifras después de la coma. Tenemos:

$$f_1 = \sin(1.000001) = 0.841472; \qquad f_{-1} = \sin(0.999999) = 0.841470$$

de manera que tenemos $f_1 - f_{-1} = 0.000002$ dentro de las seis cifras significativas. Cuando sustituímos en (1.10) encontramos $f' \approx 1.000000$, que es un resultado muy malo. Si en cambio usamos diez cifras significativas, tenemos:

$$f_1 = 0.8414715251; \qquad f_{-1} = 0.8414704445$$

lo que arroja el resultado $f' \approx 0.540330$, que es considerablemente mejor. De este modo, vemos que la diferenciación numérica es un proceso intrínsecamente inestable, ya que no se encuentra bien definido en el límite de $h \to 0$, y por lo tanto debe realizarse con sumo cuidado.

La fórmula de tres puntos se puede mejorar, relacionando f' a puntos más alejados de $x = 0$. Por ejemplo, podemos usar las ecuaciones (1.6) y (1.7) para obtener una fórmula de 5 puntos. En efecto, a partir de estas ecuaciones podemos obtener:

$$f_{+1} - f_{-1} = 2hf' + \frac{1}{3}h^3 f''' + \mathcal{O}(h^5) \tag{1.13}$$

y

$$f_{+2} - f_{-2} = 4hf' + \frac{8}{3}h^3 f''' + \mathcal{O}(h^5) \tag{1.14}$$

Si ahora multiplicamos a (1.13) por 8 y a (1.14) por (-1) y sumamos tendremos:

$$8(f_{+1} - f_{-1}) - (f_{+2} - f_{-2}) = 12hf' + \mathcal{O}(h^5) \tag{1.15}$$

de donde podemos despejar f' para obtener:

$$f' \approx \frac{1}{12h}\left[f_{-2} - 8f_{-1} + 8f_1 - f_2\right] + \mathcal{O}(h^4) \tag{1.16}$$

De este modo, si se usa la primera parte del segundo miembro de esta ecuación, se cometen errores de cuarto orden en h. Como hemos usado en la expansión de Taylor los términos hasta h^4, el cálculo de la derivada por esta ecuación supone que f se encuentra bien aproximada por un polinomio de cuarto orden sobre el intervalo de cinco puntos en $[-2h, 2h]$. A pesar de que requiere un mayor esfuerzo computacional, esta aproximación es considerablemente más precisa. Si se realizan algunas estimaciones de la precisión numérica de esta aproximación, se encuentra que con esta fórmula se obtiene una precisión comparable a la de la ecuación (1.10) con un intervalo h 10

veces mayor. Esto puede ser de importancia cuando hay que guardar muchos valores de f en la computadora, ya que la mayor precisión implica la tabulación de f para un menor número de valores de x, con lo que se ahorra memoria. De todos modos, la ecuación (1.16) requiere un mayor número de operaciones que (1.10), asi que en el caso de que interese una aplicación más rápida, esta última ecuación es preferible a expensas de un mayor gasto de memoria.

También se pueden construir fórmulas para derivadas de orden superior, tomando las apropiadas combinaciones lineales de las ecuaciones (1.6) y (1.7). Así, sumando (1.4) y (1.5) tenemos:

$$f_1 + f_{-1} = 2f_0 + h^2 f'' + \mathcal{O}(h^4) f_1 + f_{-1} \tag{1.17}$$

de donde podemos despejar:

$$\begin{aligned} f'' &= \frac{1}{h^2}\left[f_1 - 2f_0 + f_{-1}\right] + \mathcal{O}(h^2) \\ &\approx \frac{1}{h^2}\left[f_1 - 2f_0 + f_{-1}\right] \end{aligned} \tag{1.18}$$

en forma análoga se pueden derivar fórmulas para otras derivadas de f con distinto orden en h. La tabla 1.2 muestra algunas de estas aproximaciones:

	4 puntos	5 puntos
$h\ f'$	$\pm\frac{1}{6}(-2f_{\mp 1} - 3f_0 + 6f_{\pm 1} - f_{\pm 2})$	$\frac{1}{12}(f_{-2} - 8f_{-1} + 8f_1 - f_2)$
$h^2\ f''$	$f_{-1} - 2f_0 + f_1$	$\frac{1}{12}(-f_{-2} + 16f_{-1} - 30f_0 + 16f_1 - f_2)$
$h^3\ f'''$	$\pm(-f_{\mp 1} + 3f_0 - 3f_{\pm 1} + f_{\pm 2})$	$\frac{1}{2}(-f_{-2} + 2f_{-1} - 2f_1 + f_2)$
$h^4\ f^{iv}$	...	$f_{-2} - 4f_{-1} + 6f_0 - 4f_1 + f_2$

Tabla 1.2. Fórmulas de diferencias para derivas en las aproximaciones de 4 y cinco puntos

1.3 Cuadratura numérica

En el caso de la cuadratura numérica, estamos interesados en calcular la integral definida de f entre dos límites, digamos $a < b$. Con el fin del cálculo numérico, supondremos que estos dos puntos se encuentran separados por un

número N par de intervalos, que constituyen una red, que se relacionará con estos límites según:

$$N = \frac{(b-a)}{h}; \quad N = 2M; \quad M \text{ entero} \tag{1.19}$$

La evaluación de la integral en el intervalo a, b puede entoces visualizarse como una suma de integrales en cada par de intervalos de acuerdo a:

$$\int_a^b f(x)dx = \int_a^{a+2h} f(x)dx + \int_{a+2h}^{a+4h} f(x)dx + ... + \int_{b-2h}^{b} f(x)dx \tag{1.20}$$

A los efectos de derivar algunas ecuaciones, será más útil considerar integrales de la forma

$$\int_{-h}^{h} g(x)dx \tag{1.21}$$

Veamos como podemos llevar las integrales de arriba a esta forma. Tomemos la primera integral $\int_a^{a+2h} f(x)dx$. Es evidente que para que de alguna forma nos quede en el límite inferior $-h$, de alguna forma debemos sustraer la cantidad $(a+h)$. Por este motivo, realicemos el cambio de variables $x' = x-a-h$, con lo que tendremos $x = x'+a+h$ y $dx = dx'$. Esto nos lleva a la igualdad:

$$\int_{x_1=a}^{x_2=a+2h} f(x)dx = \int_{x'_1=-h}^{x'_2=+h} f(x'+a+h)dx \tag{1.22}$$

que tiene precisamente la forma (1.21) con $g(x) = f(x'+a+h)$. Pasando al caso más genérico de una integral del tipo $\int_{a+nh}^{a+(n+2)h} f(x)dx$, con n par, nos conviene sustituir $x' = x-a-nh-h$, con lo que tendremos $x = x'+a+nh+h$ y $dx = dx'$. La sustitución nos lleva a la igualdad:

$$\int_{x_1=a+nh}^{x_2=a+(n+2)h} f(x)dx = \int_{x'_1=-h}^{x'_2=+h} f(x'+a+nh+h)dx \tag{1.23}$$

donde vemos ahora que nuevamente tenemos la (1.21) con $g(x) = f(x + a + (n+1)h)$. En lo que sigue, cuando veamos $f(x)$, lo que en realidad estaremos considerando es entonces $f(x' + a + (n+1)h)$.

Pasemos entonces a considerar la integral $\int_{-h}^{h} f(x)dx$. La idea fundamental dentro de las fórmulas de cuadratura que discutiremos aquí es aproximar a f entre $-h$ y $+h$ por una función que sea integrable exactamente en este intervalo. Por ejemplo, la aproximación más simple que podemos tener es considerar los intervalos $[-h, 0]$ y $[0, h]$ separadamente, y suponer que f es lineal dentro de cada uno de estos intervalos. De este modo tenemos:

$$\int_{-h}^{h} f(x)dx = \int_{-h}^{0} f(x)dx + \int_{0}^{h} f(x)dx \tag{1.24}$$

Veamos la primera de estas integrales, aproximando hasta orden x con $f = f_0 + xf' + \mathcal{O}(x^2)$ tenemos:

$$\begin{aligned} \int_{-h}^{0} f(x)dx &= \int_{-h}^{0} \left(f_0 + xf' + \mathcal{O}(x^2)\right) dx \\ &= f_0 h - f' \frac{h^2}{2} + \mathcal{O}(h^3) \end{aligned} \tag{1.25}$$

y si aproximamos f' por la derivada hacia atrás $\frac{f_0 - f_{-1}}{h} + \mathcal{O}(h)$ llegamos a:

$$\int_{-h}^{0} f(x)dx = f_0 h - (f_0 - f_{-1})\frac{h}{2} + \mathcal{O}(h^3) \tag{1.26}$$

Note que hemos usado una aproximación para la derivada que tiene un error de orden h, que al multiplicarse por el factor x^2 da después de integrar un error del orden de h^3, que es del mismo orden que el error que da la integración del factor $\mathcal{O}(x^2)$.

Análogamente, considerando la segunda integral en la ec. (1.24) y reemplazando f por su desarrollo hasta primer orden en x y f' por la derivada hacia adelante $\frac{f_1 - f_0}{h}$ tenemos:

$$\int_0^h f(x)dx = f_0 h + (f_1 - f_0)\frac{h}{2} + \mathcal{O}(h^3) \tag{1.27}$$

de manera que (1.24) queda:

$$\int_{-h}^h f(x)dx = \frac{h}{2}\left(f_{-1} + 2f_0 + f_1\right) + \mathcal{O}(h^3) \tag{1.28}$$

Esta es la que se conoce como **fórmula de los trapecios.**

Una mejor aproximación puede realizarse expandiendo la función f hasta tercer orden:

$$f = f_0 + f'x + f''\frac{x^2}{2} + f'''\frac{x^3}{3!} + \mathcal{O}(x^4) \tag{1.29}$$

y reemplazando las derivadas primera y segunda por $f' = \frac{f_{+1}-f_{-1}}{2h} + \mathcal{O}(h^2)$ y $f'' = \frac{1}{h^2}\left[f_1 - 2f_0 + f_{-1}\right] + \mathcal{O}(h^2)$ para tener:

$$\begin{aligned}
\int_{-h}^h f(x)dx &= f_0 x \,|_{-h}^{+h} \\
&+\frac{(f_1 - f_{-1})}{2h}\frac{x^2}{2} \quad | \quad {}_{-h}^{+h} + \mathcal{O}(h^2)\frac{x^2}{2}\,|_{-h}^{+h} \\
+\frac{1}{2h^2}\left[f_1 - 2f_0 + f_{-1}\right]\frac{x^3}{3} &\quad | \quad {}_{-h}^{+h} + \mathcal{O}(h^2)\frac{x^3}{3}\,|_{-h}^{+h} \\
&+\frac{f'''}{3!}\frac{x^4}{4} \quad | \quad {}_{-h}^{+h} \\
&+\mathcal{O}(x^5) \quad | \quad {}_{-h}^{+h}
\end{aligned} \tag{1.30}$$

donde en esta última ecuación hemos separado en diferentes renglones las contribuciones de las derivadas de distinto orden en (1.29). La evaluación de los límites de integración da:

$$\begin{aligned}\int_{-h}^{h} f(x)dx &= 2f_0h + \frac{1}{h^2}\left[f_1 - 2f_0 + f_{-1}\right]\frac{h^3}{3} + \mathcal{O}(h^5) \qquad (1.31)\\ &= \frac{h}{3}[f_{-1} + 4f_0 + f_1] + \mathcal{O}(h^5)\end{aligned}$$

Que es la **regla de Simpson**. A pesar de que aumentamos en un orden la precisión de la expansión, hemos mejorado en dos órdenes la precisión en la integral. Esto ocurre debido a la anulación de los términos que contienen x^2.

Para usar esta fórmula en el intevalo $[a, b]$ debemos componer la integral:

$$\begin{aligned}\int_{a}^{b} f(x)dx &= \int_{a}^{a+2h} f(x)dx + \int_{a+2h}^{a+4h} f(x)dx + ... + \int_{b-2h}^{b} f(x)dx \qquad (1.32)\\ &= \int_{-h}^{+h} f(x' + a + h)dx + \int_{-h}^{+h} f(x' + a + 3h)dx + \int_{-h}^{+h} f(x' + a + 5h)dx\\ &\quad + ... + \int_{-h}^{+h} f(x' + b - h)dx\end{aligned}$$

Si ahora aplicamos a esta última ecuación el resultado de la ec. (1.31) tendremos:

$$\begin{aligned}\int_{a}^{b} f(x)dx &= \frac{h}{3}\begin{bmatrix} f(a) + 4f(a+h) + f(a+2h) \\ +f(a+2h) + 4f(a+3h) + f(a+4h) \\ +f(a+4h) + 4f(a+5h) + f(a+6h) \\ +..... + f(b-2h) + 4f(b-h) + f(b)\end{bmatrix} = \qquad (1.33)\\ &\quad \frac{h}{3}\begin{bmatrix} f(a) + 4f(a+h) + 2f(a+2h) + 4f(a+3h) + 2f(a+4h) \\ +..... + 4f(b-h) + f(b)\end{bmatrix}\end{aligned}$$

Como ejemplo, consideraremos el siguiente programa que efectúa la integración:

$$\int_{0}^{1} e^x dx = e - 1 = 1.718282... \qquad (1.34)$$

```
      FUNC(X)=EXP(X)
      EXACT=EXP(1.)-1.
30    PRINT *,'ENTER N EVEN (.LT. 2 TO STOP)'
      READ *, N IF (N .LT. 2) STOP
      IF (MOD(N,2) .NE. 0) N=N+1
      H=1./N
      SUM=FUNC(0.)
      FAC=2
      DO 10 I=1,N-1
      IF (FAC .EQ. 2.) THEN
      FAC=4
      ELSE
      FAC=2.
      END IF
      X=I*H
      SUM=SUM+FAC*FUNC(X)
10    CONTINUE
      SUM=SUM+FUNC(1.)
      XINT=SUM*H/3.
      DIFF=EXACT-XINT
      PRINT 20,N,DIFF
20    FORMAT (5X,'N=',I5,5X,'ERROR=',E15.8)
      GOTO 30
      END
```

Los resultados se muestran en la Tabla 1.3 para varios valores de N, comparados con resultados de la regla de los trapecios.

N	h	Trapecios	Simpson	Bode
4	0.25	-0.008940	-0.000037	-0.000001
8	0.125	-0.002237	0.000002	0.000000
16	0.0625	-0.000559	0.000000	0.000000
32	0.03125	-0.000140	0.000000	0.000000
64	0.0156250	-0.000035	0.000000	0.000000
128	0.0078125	-0.000008	0.000000	0.000000

Tabla 1.3 . Errores al evaluar $\int_0^1 e^x dx = e - 1 = 1.718282...$

La mejora que introduce el método de mayor orden es evidente. Nótese que el método de integración es estable, en el sentido que se obtiene un límite bien definido cuando N se vuelve muy grande y el espaciamiento h se vuelve muy pequeño. Los errores de redondeo no son importantes ya que todos los valores de f participan en la ecuación tienen el mismo signo, a diferencia de lo que ocurre con la diferenciación.

Un problema importante es cuán pequeño tiene que ser h para calcular la integral con una dada precisión. A pesar de que se puede después de un análisis cuidadoso encontrar una cota superior para los errores, en la práctica lo que se hace es correr un dado cálculo con varios h y observar cómo varía la precisión.

Se pueden desarrollar fórmulas de mayor cuadratura reteniendo más términos en la expansión de Taylor (1.4) que se usa para interpolar entre puntos de red, empleando las correspondientes diferencias finitas para las derivadas. La generalización de la regla de Simpson usando polinomios cúbico y cuárticos para interpolar son:

$$\int_{x_0}^{x_3} f(x)dx = \frac{3h}{8}\left[f_0 + 3f_1 + 3f_2 + f_3\right] + \mathcal{O}(h^5) \tag{1.35}$$

(Fórmula de Simpson 3/8)

$$\int_{x_0}^{x_4} f(x)dx = \frac{2h}{45}\left[7f_0 + 32f_1 + 12f_2 + 32f_3 + 7f_4\right] + \mathcal{O}(h^7) \tag{1.36}$$

(Fórmula de Bode)

En estos casos, N debe ser múltiplo de 3 y de 4 respectivamente.

Fórmulas que empleen polinomios de interpolación más altos no son adecuadas, ya que éstos tienden a oscilar y no dan una buena aproximación a la función. Los coeficientes que acompañan a las evaluaciones de la función pueden además tener signos alternados, y los errores de redondeos pueden crear problemas, como pasaba en el caso de las derivadas. De todos modos, se pueden desarrollar métodos que emplean polinomios interpolantes de mayor grado que funcionan bien, siempre y cuando se renuncie a la condición de equiespaciamiento que pusimos aca ($h = cte$).

Aparte de la integración numérica, un análisis de la función a integrar puede también ayudar a tener resultados más precisos. Por ejemplo, en el caso de que el límite superior de integración sea grande, conviene realizar un cambio de variable. Para dar un ejemplo, consideremos la integral:

$$\int_1^b dx\ x^{-2} g(x) \tag{1.37}$$

donde $g(x)$ tiende a tomar un valor constante para grandes valores de x. En este límite el argumento de la integral se va a comportar como $1/x^2$ y por lo tanto la integral va a converger muy lentamente, para lo cual tendremos que usar un N muy grande para grandes valores de b. Si en cambio hacemos un cambio de variables, poniendo $t = x^{-1}$, con lo que tendremos $dx = -dt/t^2$, tendremos:

$$\int_{x_1=1}^{x_2=b} dx\ x^{-2} g(x) = \int_{t_1=1}^{t_2=1/b} -\frac{dt}{t^2}\ t^2 g(1/t) = \int_{1/b}^{1} g(t^{-1}) dt \tag{1.38}$$

que se puede evaluar por algunos de los métodos que hemos visto.

1.4 Búsqueda de raíces

La tercera de las operaciones fundamentales que nos faltaba era el hallazgo de una raíz de una función $f(x)$ que uno puede calcular para un x arbitrario. Un método seguro, aunque algo primitivo, consiste en lo que se llama método de la bisección. Si uno conoce la localización aproximada de la raíz (digamos $x = x_r$), consiste en elegir un valor para comenzar que seguramente es menor que el de la raiz, digamos $x_0 < x_r$, y luego comenzar a dar pequeños pasos buscando el cambio de signo, reduciendo a la mitad el paso, cada vez que se detecta el cambio de signo. Los valores de x generados por este procedimiento inevitablemente convergen a x_r, de manera que la búsqueda se puede interrumpir al lograr la precisión deseada. El siguiente programa de FORTRAN encuentra la raíz de la función $f(x) = x^2 - 5$, $x_r = \sqrt{5} = 2.236068...$, con una tolerancia de 10^{-6} usando $x = 1$ como valor inicial con un paso inicial de 0.5.

```
FUNC(X)=X*X-5.
TOLX=1.E-06
X=1.
FOLD=FUNC(X)
DX=.5
```

```
      ITER=0
10    CONTINUE
      ITER=ITER+1
      X=X+DX
      PRINT *,ITER,X,SQRT(5.)-X
      IF ((FOLD*FUNC(X)) .LT. 0) THEN
      X=X-DX
      DX=DX/2
      END IF
      IF (ABS(DX) .GT. TOLX) GOTO 10
      STOP
      END
```

Los resultados para la iteración se muestran en la tabla 1.4.

Iteración	Bisección	Newton	Secante
0	1.236076	1.236076	1.236076
1	0.736068	-0.763932	-1.430599
2	0.236068	-0.097265	0.378925
3	-0.263932	-0.002027	0.0981327
4	-0.013932	-0.000001	-0.009308
5	0.111068	0.000000	0.000008
6	-0.013932	0.000000	0.000000
⋮	⋮	⋮	⋮
33	0.000001	0.000000	0.000000

Tabla 1.4. Error en hallar la raíz positiva de la ecuación $f(x) = x^2 - 5$

Como se puede apreciar, este método converge a la respuesta correcta, aunque tarda para ello del orden de 33 iteraciones.

Hay que tener además cuidado si la ecuación presenta raíces múltiples, ya que si el paso elegido es demasiado grande,se puede "pasar de largo" una raíz.

Un método más eficiente para la búsqueda de raíces lo constituye el método de Newton-Raphson. Este método, según se puede apreciar en la Figura 1.2, calcula una aproximación x_{i+1} a la raíz a partir del valor de la función para un dado x y de su derivada para ese mismo valor de x.

De la figura 1.2 vemos que:

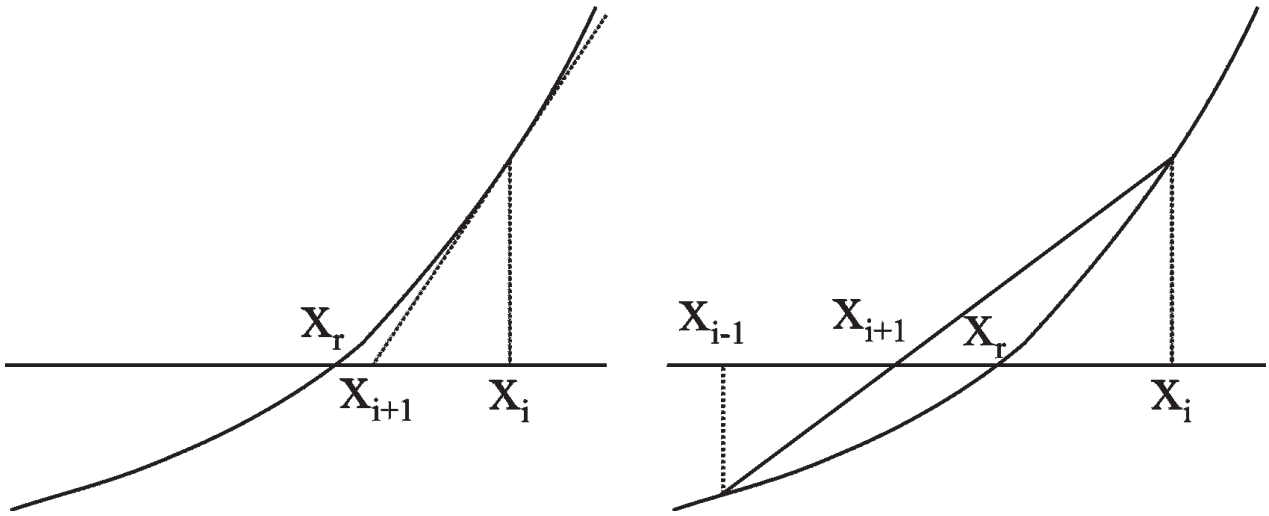

Figure 1.2: Fundamento geométrico del los métodos de Newton-Raphson(izquierda) y de la secante(derecha).

$$f'(x_i) = \frac{f(x_i)}{x_i - x_{i+1}} \tag{1.39}$$

de donde obtenemos la relación para iterar:

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)} \tag{1.40}$$

En la tabla 1.4 vemos la veloz convergencia de este método comparado con el anterior. Este es el algoritmo que usan las computadoras para calcular la raíz cuadrada de un número.

Uno de los inconvenientes del método de Newton, es la necesidad de tener que calcular la derivada. Esto se evita con el método de la secante, donde la derivada se aproxima por la diferencia finita hacia atrás:

$$f'(x_i) \approx \frac{f(x_i) - f(x_{i-1})}{x_i - x_{i-1}} \tag{1.41}$$

De modo que reemplazando (1.41) en (1.40) tendremos:

$$x_{i+1} = x_i - f(x_i)\frac{x_i - x_{i-1}}{f(x_i) - f(x_{i-1})} \tag{1.42}$$

El significado geométrico es esta ecuación se ilustra en la Figura 1.2. Desde el punto de vista geométrico, la ecuación (1.42) puede derivarse planteando la igualdad que se obtiene a partir de la semejanza de los triángulos mostrados en esta figura:

$$-\frac{f(x_{i-1})}{(x_{i+1}-x_{i-1})} = \frac{f(x_i)}{(x_i - x_{i+1})} \tag{1.43}$$

de donde simplemente hay que despejar x_{i+1} para llegar a (1.42).

Se pueden usar valores aproximados de x_0 y x_1 para iniciar el algoritmo, que se concluye al obtener la tolerancia deseada. Usando para el problema que consideramos $x_0 = 0.5$ y $x_1 = 1.0$, tenemos los resultados de la tabla 1.4. Vemos que la convergencia es casi tan rápida como la del método de Newton-Raphson.

Cuando la función no está bien corportada cerca de la raíz, por ejemplo cuando tiene puntos de inflexión, este método puede fallar si el valor inicial no es adecuado. En estos casos conviene usar el método de la bisección para aproximar la raíz, y de allí empalmar con el método de Newton-Raphson.

1.5 Ejemplo:Cuantización semiclásica de las vibraciones moleculares.

1.5.1 El método WKB

El método WKB debe su nombre a Wenzel, Kramers, Brillouin, que fueron quienes lo desarrollaron. Este consiste en un procedimiento para resolver la ecuación de Schrödinger para una partícula *pesada* en un potencial que varía lentamente con la posición, y una descripción detallada de este método se puede consultar por ejemplo en el libro Quantum Mechanics, de Donald Rapp, pág. 145 y subsiguientes. Consideremos la ecuación:

$$-\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2} + V(x)\psi = E\psi \tag{1.44}$$

que se puede reordenar para dar:

$$\frac{d^2\psi}{dx^2} + \frac{2m}{\hbar^2}\left[E - V(x)\right]\psi = 0 \tag{1.45}$$

Si ahora definimos $k^2(x) = \frac{2m}{\hbar^2}\left[E - V(x)\right]$ obtenemos:

$$\frac{d^2\psi}{dx^2} + k^2(x)\psi = 0 \tag{1.46}$$

La solución de la ecuación (1.46) se puede escribir de la forma:

$$\psi = F(x)e^{iS(x)} \tag{1.47}$$

donde $F(x)$ y $S(x)$ son funciones reales de x. Si reemplazamos (1.47) en (1.46) y agrupamos las partes reales e imaginarias llegamos a:

$$\left[F'' - F(S')^2 + k^2F\right] + i\left[2F'S' + FS''\right] = 0$$

Como en esta ecuación las partes reales e imaginaria deben ser cero llegamos a:

$$2F'S' + FS'' = 0 \tag{1.48}$$

$$F'' - F(S')^2 + k^2F = 0 \tag{1.49}$$

Si multiplicamos a (1.48)por F tenemos:

$$2FF'S' + F^2S'' = 0 \tag{1.50}$$

lo que es equivalente a:

$$\frac{d}{dx}(F^2S') = 0 \tag{1.51}$$

lo que implica que

$$F^2\frac{dS}{dx} = cte = C^2$$

donde definimos el cuadrado por conveniencia. Despejando F tenemos:

$$F = \frac{C}{(dS/dx)^{1/2}}$$

de modo que la ecuación para la función de onda (1.47) queda:

$$\psi(x) = \frac{C}{(dS/dx)^{1/2}}e^{iS(x)} \tag{1.52}$$

ecuación que es formalmente exacta porque no hemos realizado ninguna aproximación.

Consideremos ahora las soluciones de la ecuación (1.47) para el caso en que $E > V(x)$, con $V(x) \approx$constante. Sabemos que esta es del tipo oscilatorio:

$$\psi(x) \propto e^{\pm ikx}$$

por lo que comparando con (1.52) concluimos que $S(x) = \pm kx$ y que $dS/dx = \pm k$.

Que ocurre ahora si $V(x)$ varía, aunque lentamente comparado con la variación de $\psi(x)$? Podemos suponer que la fase $S(x) = \pm kx$ varía lentamente, y que su derivada $dS/dx = \pm k$ varía todavía más lentamente, así como sus derivadas superiores $d^nS/dx^n \approx 0$. Veamos qué relevancia tiene esto para las ecuaciones de arriba.

A partir de (1.48) y usando la relación $F = \frac{C}{(S')^{1/2}}$tenemos:

$$F' = -\frac{1}{2}F\frac{S''}{S'} = -\frac{1}{2}\frac{C}{(S')^{1/2}}\frac{S''}{S'} = \frac{1}{2}\frac{C}{(S')^{3/2}}S'' \tag{1.53}$$

Derivando nuevamente tendremos:

$$F'' = -\frac{1}{2}\frac{C}{(S')^{3/2}}S''' - \frac{1}{2}CS''(-\frac{3}{2})\frac{S''}{(S')^{5/2}} \tag{1.54}$$

Si en esta ecuación reemplazamos $\frac{C}{(S')^{1/2}}$ por F y los sacamos como factor común tendremos:

$$F'' = F\left[-\frac{1}{2}\frac{S'''}{S'} + \frac{3}{4}\frac{(S'')^2}{(S')^2}\right] = F(S')^2\left[-\frac{1}{2}\frac{S'''}{(S')^3} + \frac{3}{4}\left[\frac{S''}{(S')^2}\right]^2\right] \tag{1.55}$$

Volvamos a:

$$F'' - F(S')^2 + k^2F = 0$$

utilizando la penúltima ecuación para reemplazar F'', con lo que tendremos:

$$F(S')^2\left[-\frac{1}{2}\frac{S'''}{(S')^3} + \frac{3}{4}\left[\frac{S''}{(S')^2}\right]^2 - 1\right] + k^2F = 0 \tag{1.56}$$

Si despreciamos ahora los dos primeros términos en el corchete tenemos:

$$-F(S')^2 + k^2 F \approx 0 \tag{1.57}$$

o bien:

$$\left(\frac{dS}{dx}\right)^2 \approx k^2(x) \tag{1.58}$$

con lo que tenemos:

$$\left(\frac{dS}{dx}\right) \approx \pm k(x) \tag{1.59}$$

y

$$S \approx \pm \int k(x)dx \tag{1.60}$$

en base a esto, la función F resultará:

$$F = \frac{C}{(S')^{1/2}} \approx \frac{C}{\sqrt{k(x)}} \tag{1.61}$$

con lo que la solución general para ψ será:

$$\psi(x) \approx \frac{C_-}{\sqrt{k(x)}} e^{i\int k(x)dx} + \frac{C_+}{\sqrt{k(x)}} e^{-i\int k(x)dx} \tag{1.62}$$

donde C_- y C_+ se derivan a partir de las condiciones de contorno.

En el caso de que $V(x) - E > 0$, se obtendrían las soluciones correspondientes reemplazando $k = \frac{\{2m[V(x)-E]/\hbar^2\}^{1/2}}{2}$ donde estaba ik, para obtener:

$$\psi(x) \approx \frac{C_-}{\sqrt{\kappa(x)}} e^{\int \kappa(x)dx} + \frac{C_+}{\sqrt{\kappa(x)}} e^{\int \kappa(x)dx} \tag{1.63}$$

Estas formas de las funciones de onda son lo que se conoce como aproximación WKB. Se puede demostrar que la aproximación que hicimos arriba de que:

$$\left| -\frac{1}{2}\frac{S'''}{(S')^3} + \frac{3}{4}\left[\frac{S''}{(S')^2}\right]^2 \right| << 1$$

implica :

$$\frac{|\Delta\lambda|}{\lambda} << 2\pi$$

Es decir, que el cambio fraccional de la longitud de onda de De Broglie es pequeño comparado con 2π.

De las ecuaciones (1.62) y (1.63) queda claro que la función de onda aproximada diverge en el punto de retorno clásico, por lo que en esta zona se deben reemplazar por otra expresión. Como resultado del empalme de la ecuaciones, aparece la condición:

$$\int_a^b k(x)dx = (n + \frac{1}{2})\pi \tag{1.64}$$

1.5.2 Aplicación del método WKB

Como ejemplo para combinar las operaciones matemáticas básicas que hemos aprendido a manejar numéricamente, consideraremos el problema de describir el movimiento de una molécula diatómica, como por ejemplo el N_2, que consiste de dos núcleos ligados por electrones. Dado que los núcleos son mucho más pesados que los electrones, podemos suponer que éstos se mueven suficientemente rápido como para ajustarse instantáneamente a la posición de los núcleos. Esto es lo que se conoce en mecánica cuántica como aproximación de Born-Oppenheimer. El problema se reduce entonces al del movimiento de dos núcleos en un potencial que designaremos como V, que dependerá de la distancia entre los mismos, que designaremos r. En principio, V se debería obtener a partir de un cálculo de primeros principios, pero en términos generales se puede decir que el potencial será atractivo a distancias largas, debido a fuerzas de dispersión del tipo dipolo inducido-dipolo inducido, y repulsivo a distancias cortas, debido a la interacción de los núcleos y a la repulsión de Pauli de los electrones. Lo que generalmente se hace es emplear un potencial que contiene una serie de parámetros, que se ajustan en base a alguna consideración teórica o semi empírica. Un potencial de uso corriente con estos fines es el potencial de Lennard-Jones o potencial 12-6, que tiene la forma:

$$V(r) = 4V_0 \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right] \tag{1.65}$$

el cual tiene la forma que se muestra en la figura 1.3.

La masa relativamente grande de los núcleos permite que el problema se simplifique todavía más desacoplando la rotación lenta de los núcleos de los cambios rápidos de su separación. Mientras que la rotación se puede describir con el modelo cuántico del rotor rígido, a los estados vibracionales,

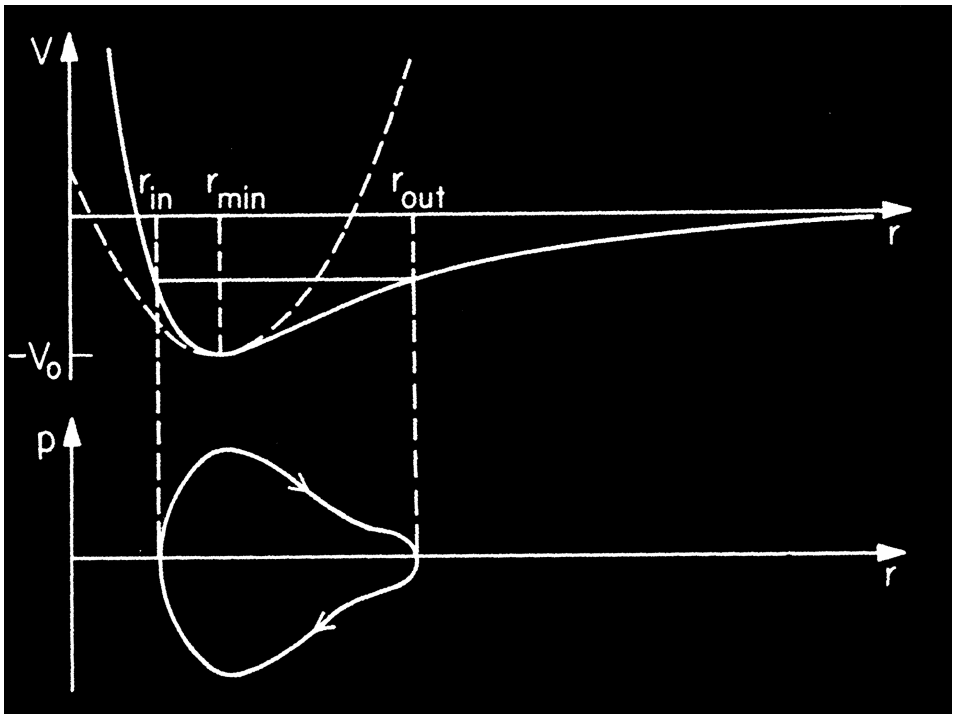

Figure 1.3: Figura 1.3: Parte superior: potencial de Lennard-Jones, y los puntos de retorno clásico interno y externo para un valor de enrgía negativo. La línea punteada muestra la aproximación parabólica a este potencial. Parte inferior: Trayectoria en el espacio de las fases. Este potencial presenta un mínimo para r = $\sigma 2^{1/6}$ una profundidad V_0

a los que les asignaremos una energía E_n, los describiremos por los estados ligado $\psi_n(r)$ de una ecuación de Scrhrödinger monodimensional:

$$\left[-\frac{\hbar^2}{2m}\frac{d^2}{dr^2}+V(r)\right]\psi_n(r)=E_n\ \psi_n(r) \tag{1.66}$$

donde m denota la masa reducida de los núcleos.

El objetivo del presente proyecto será encontrar las energías E_n, para un dado potencial. Esto se puede hacer resolviendo la ecuación de autovalores (1.66), para lo que veremos más adelante métodos numéricos. De todos modos, la gran masa de los núcleos hace que su movimiento sea cuasi clásico, de manera que valores aproximados de la energía vibracional E_n pueden ser obtenidos considerando el movimiento clásico de los núcleos en V y aplicando las llamadas "reglas de cuantización " para determinar las energías. Estas reglas de cuantización, originalmente postuladas por Bohr, Sommerfeld y Wilson, fueron la base de la vieja teoría cuántica de la cual evolucionó la teoría cuántica moderna. De todos modos, también se pueden obtener rigurosamente de la aproximación WKB de la ecuación de onda.

El movimiento clásico de los núcleos puede tener lugar en el potencial $V(r)$ para energía comprendidas en el rango $V_0 < V < 0$. La distancia entre los núcleos oscila periódicamente(aunque no armónicamente) entre los puntos de retorno clásico interno y externos r_{in} y r_{out} mostrados en la Figura 1.3. Durante estas oscilaciones, se intercambia energía potencial con energía cinética de manera que la energía total E permanezca constante de acuerdo a:

$$E=\frac{p^2}{2m}+V(r) \tag{1.67}$$

donde p representa el momento de los núcleos. De este modo, el sistema describe una trayectoria cerrada en el espacio de las fases, en la cual se satisface la ecuación (1.67), tal como se muestra en la porción inferior de la Figura 1.3. Podemos despejar de (1.67) el momento para obtener:

$$p(r)=\pm\left[2m\left(E-V(r)\right)\right]^{1/2} \tag{1.68}$$

El movimiento clásico que indicamos, ocurre para cualquier valor de energía E comprendido entre V_0 y 0. Para cuantizar este movimiento, debemos obtener de algún modo valores aproximados a los autovalores E_n que aparecen en la ecuación (1.66). Para ello consideraremos la siguiente integral:

$$S(E) = \oint \frac{p(r)}{\hbar} dr = \oint k(r) dr \tag{1.69}$$

que suele denominarse "acción" en analogía con lo empleado en la mecánica lagrangiana. $k(r)$ es lo que se denomina número de onda de De Broglie y la integral se realiza sobre un período completo. Esta integral es proporcional al área encerrada por la trayectoria en el espacio de las fases. Las reglas de cuantización implican que, para las energías permitidas E_n, la acción es un múltiplo semientero de 2π. Si tenemos en cuenta que la oscilación pasa por cada valor de r dos veces (una vez con valor positivo y otra con valor negativo) tenemos la condición:

$$S(E_n) = 2\left(\frac{2m}{\hbar^2}\right)^{1/2} \int_{r_{in}}^{r_{out}} \left[E_n - V(r)\right]^{1/2} dr = \left(n + \frac{1}{2}\right) 2\pi \tag{1.70}$$

donde n es un entero no negativo. En los límites de la integral, el integrando se anula.

En el caso particular del potencial de Lennard-Jones, conviene definir cantidades adimensionales:

$$\epsilon = \frac{E}{V_0}, \quad x = \frac{r}{\sigma}, \quad \gamma = \left(\frac{2m\sigma^2 V_0}{\hbar^2}\right)^{1/2} \tag{1.71}$$

de manera que la ecuación (1.70) queda:

$$s(\epsilon_n) \equiv \frac{1}{2} S(\epsilon_n V_0) = \gamma \int_{x_{in}}^{x_{out}} \left[\epsilon_n - v(x)\right]^{1/2} dx = \left(n + \frac{1}{2}\right) \pi \tag{1.72}$$

donde hemos definido:

$$v(x) = 4\left(\frac{1}{x^{12}} - \frac{1}{x^6}\right) \tag{1.73}$$

que es el potencial "escaleado". Nuestra tarea sera entonces, por una parte realizar una integración numérica para obtener la acción en la ecuación (1.72) y encontrar los valores ϵ_n para los diferentes n, lo que implica una búsqueda de raíces.

La cantidad γ es un parámetro que mide la nautraleza cuántica del problema. En el límite clásico ($\hbar$ pequeño y m grande), γ se vuelve grande. El conocimiento del momento de inercia de una molécula (a partir de las energías rotacionales) y de su energía de disociación permite obtener entonces a partir de observaciones experimentales a los parámetros σ y V_0, y por lo tanto γ. Algunos valores de γ se muestran en la siguiente tabla.

Molécula	γ
H_2	21.7
HD	24.8
O_2	150

Tabla 1.5. Valores del parámetro γ para algunas moléculas.

1.6 Introducción a la minimización y maximización numérica de funciones

Consideremos el siguiente problema. Dada una función $f(\vec{x})$ que depende de una o más variables independientes, se desea encontrar al valor de aquellas variables $\vec{x}_i$ donde f toma el valor máximo o mínimo(extremo). Los problemas de minimización y de maximización se encuentran relacionados en forma trivial, ya que si desea considerar el problema de maximizar f, basta con considerar el problema de minimizar -f. En general, nos referiremos a este último caso.

Un mínimo puede ser global (cuando la función toma el valor más bajo que puede tomar en todo su dominio) o local (cuando la función toma el valor más bajo en un dado entorno finito, y no en la frontera de ese entorno).El hallazgo de un mínimo global es un problema extremadamente difícil comparado con el hallazgo de mínimos locales. En este último caso son usualmente aplicados métodos que llamaremos *deterministas*, en donde se evalúa la función y a veces sus derivadas primeras y segundas, con el objeto de dirigirse al mínimo más cercano a partir de un punto de partida dado. En el primer caso, se emplean fundamentalmente dos tipos de enfoque. En algunos casos se parte de un conjunto de coordenadas iniciales, digamos $\{\vec{x}_i\}$, y se realiza una selección de las que dan los valores de f más bajos, aunque sin rechazar necesariamente los valores de $\vec{x}_i$ que incrementan el valor de la función. Dentro de

estos métodos, que denominaremos *estocásticos*, se encuentran los llamados de templado simulado, que han demostrado recientemente lograr una muy buena performance en problemas de gran complejidad.

En esta sección consideraremos solamente un método determinista muy sencillo que se emplea para encontrar mínimos locales, que es el llamado método **simplex**. Este método no implica el cálculo de gradientes ni derivadas superiores de la función, por lo que puede ser muy útil aún en el caso de que la derivada de la función no esté definida en algunos puntos del espacio de búsqueda.

1.6.1 El método simplex

Un simplex es una figura geométrica que tiene un vértice más que la dimensión del espacio en el cual se encuentra definido. Así, en una dimensión un simplex es un segmento, en dos dimensiones es un triángulo y en tres dimensiones es un tetrahedro, en general irregular. Este método fue propuesto en 1965 por J. A. Nedler y R. Mead /Computer Journal, 7(1965)308/ y ha sido analizado y aplicado por S. Caceci and W. P. Cacheris /Byte, (1984)340 y a los efectos de alcanzar el mínimo el programa mueve el simplex cuesta abajo, acelerando o frenando su movimiento según convenga. El criterio adoptado es el siguiente: Primero se encuentran los vértices $\vec{R}_a$ y $\vec{R}_b$ que originan el valor más alto y el valor más bajo de la función respectivamente. Después se reemplaza al que tiene el valor más alto por otro según se explica a continuación, y se comienza nuevamente.El programa computa al nuevo vértice de acuerdo a uno de los siguientes mecanismos: reflexión, expansión y contracción del vector que contiene al valor más alto de la función. Las operaciones de reflexión, expansión y contracción se realizan respecto al centro de masa $\vec{R}_m$ de los puntos del simplex, excluyendo al punto que tiene el valor más alto. De este modo, tendremos:

$$\vec{R}_m = \frac{\sum\limits_{i \neq a} \vec{R}_i}{m} \tag{1.74}$$

Consideremos ahora las diferentes operaciones mencionadas en el contexto de la búsqueda

Reflexión: El punto $\vec{R}_a$ se refleja según la ecuación:

$$\vec{R}^* = \vec{R}_m + (\vec{R}_a - \vec{R}_m)(-\alpha) \tag{1.75}$$

con $\alpha > 0$, o lo que es equivalente:

$$\vec{R}^* = \vec{R}_m(1+\alpha) - \alpha\vec{R}_a$$

Si el valor $f(\vec{R}^*)$ es menor que $f(\vec{R}_a)$ pero mayor que $f(\vec{R}_b)$, entonces el programa hace $f(\vec{R}_a) = f(\vec{R}^*)$ y vuelve a comenzar. Es decir que reemplaza al vector que originaba el "peor " de f por el vector reflejado.

Si el valor $f(\vec{R}^*)$ resulta menor que $f(\vec{R}_b)$, el programa intenta una **expansión** en la dirección de $\vec{R}^*$:

$$\vec{R}^{**} = \vec{R}^* + \beta(\vec{R}^* - \vec{R}_m) \tag{1.76}$$

donde vale $0 < \beta$, lo que también se puede escribir:

$$\vec{R}^{**} = \vec{R}^*(1+\beta) - \beta\vec{R}_m$$

Dependiendo del valor de $f(\vec{R}^{**})$ el programa tiene dos opciones.

Si el valor de $f(\vec{R}^{**}) < f(\vec{R}_a)$, entonces $\vec{R}^{**}$ es aceptado para reemplazar a $\vec{R}_a$. De lo contrario se emplea a $\vec{R}^*$ (el valor reflejado) para reemplazar a $\vec{R}_a$.

Si el valor $f(\vec{R}^*)$ resulta mayor que $f(\vec{R}_a)$, el programa intenta una **contracción** del peor vector $\vec{R}_a$ en la dirección de $\vec{R}_m$:

$$\vec{R}^{**} = \vec{R}_m + \gamma(\vec{R}_a - \vec{R}_m) \tag{1.77}$$

donde se cumple $0 < \gamma < 1$. Esta ecuación puede también escribirse:

$$\vec{R}^{**} = \gamma\vec{R}_a + (1-\gamma)\vec{R}_m$$

Si $f(\vec{R}^{**}) < f(\vec{R}_a)$, entonces $\vec{R}^{**}$ es aceptado, haciendo $\vec{R}_a = \vec{R}^{**}$. En el caso mas bien improbable de que $f(\vec{R}^{**}) > f(\vec{R}_a)$, entonces se contraen todos los vértices (**contracción múltiple**) excepto $\vec{R}_b$ en la dirección de $\vec{R}_b$, de acuerdo a:

$$\vec{R}_i^{\#} = \vec{R}_b + \frac{\vec{R}_i - \vec{R}_b}{2} = \frac{\vec{R}_i + \vec{R}_b}{2} \qquad , \qquad i \neq b \tag{1.78}$$

Nedler y R. Mead aconsejan tomar $\alpha = 1$, $\beta = 1$ y $\gamma = 0.5$.

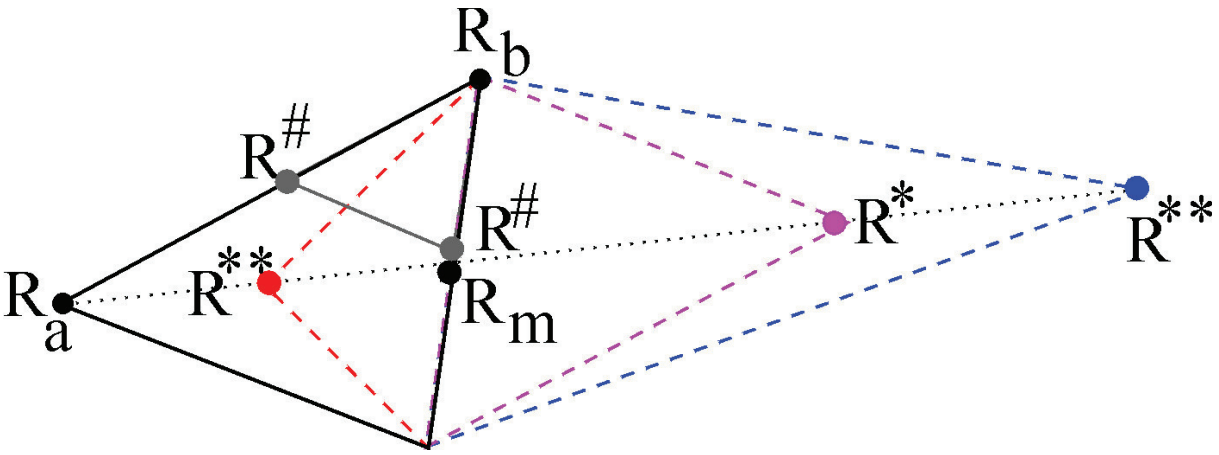

Figure 1.4: Figura 1.4: Representación de un simplex y los posibles cambios que puede experimentar. R_a da el valor mas alto, R_b da el valor más bajo, R_m es el centro de masa, R^* es la reflexión, R^{**}(azul) es la expansión, R^{**}(rojo) es la contracción, $R^{\#}$ es la contracción múltiple.

Todos los tipos de movimientos posibles se resumen en la figura **??**

y un programa para calcular el mínimo de una función por este método se da a continuación:

```
c fun es la función a minimizar
c p son los puntos que conforman el simplex
IMPLICIT REAL*8 (a-h,o-z)
PARAMETER (ndim=2)
DIMENSION p(ndim+1,ndim)
DIMENSION r(ndim),rm(ndim)
DIMENSION rstar(ndim),rstar2(ndim)
COMMON/niter/niter
EXTERNAL fun niter=0
OPEN(7,file='output.txt')
p(1,1)=-1.d0
p(1,2)=1.d0
p(2,1)=-1.d0
p(2,2)=-1.d0
```

```
p(3,1)=2.d0
p(3,2)=0.d0
tol=1.d-8
CALL simplex(fun,p,r,rm,rstar,rstar2,ndim,tol)
STOP
END

SUBROUTINE simplex(fun,p,r,rm,rstar,rstar2,ndim,tol)
IMPLICIT REAL*8 (a-h,o-z)
PARAMETER (alfa=1.d0,beta=1.d0,gama=0.5d0)
c el vector p contiene en las filas los vectores de los puntos
DIMENSION p(ndim+1,ndim)
c r se usa para poner a cada uno de los vectores r que se usan
c para evaluar la funcion fun
DIMENSION r(ndim),rm(ndim)
DIMENSION rstar(ndim),rstar2(ndim)
c determine el vector que corresponde al valor mas alto de la funcion
c primero debe transformar la matriz p en cada uno de los vectores
c r
COMMON/niter/niter
100      CONTINUE
fmax=-1.d20
fmin=+1.d20
niter=niter+1
WRITE(7,*)niter
WRITE(7,*)p
IF(niter.eq.100)
STOP
DO10 i=1,ndim+1
     DO 20 j=1,ndim
          r(j)=p(i,j)
20      END DO
          fval=fun(r)
c WRITE(7,*)'r'
c WRITE(7,*)r
WRITE(7,*)'r(',i,')','=',r
WRITE(7,*)'fval(',i,')','=',fval
c encuentre los vectores que tienen a los valores minimos y
```

```
c maximos de f
          IF(fval.lt.fmin)THEN
                    fmin=fval
                    imin=i
          END IF
          IF(fval.ge.fmax)THEN
                    fmax=fval
                    imax=i
          END IF
10      END DO
IF((fmax-fmin).lt.tol)THEN
WRITE(7,*)fmin
WRITE(7,*)r
STOP
END IF
c calcule el vector del centro de masa considerando todos los puntos
c excepto el de mayor energia
          DO 30j=1,ndim
                    sum=0.d0
                    DO 40i=1,ndim+1
                              sum=sum+p(i,j)
40                    END DO
                    rm(j)=(sum-p(imax,j))/dfloat(ndim)
30            END DO
WRITE(*,*)'masscenter'
WRITE(*,*)rm
c STOP
c refleje al vector que da el valor mas grande
               DO 50j=1,ndim
                    rstar(j)=rm(j)*(1.d0+alfa)-alfa*p(imax,j)
50             END DO
c
cwrite(*,*)'rstar'
cwrite(*,*)rstar
cstop
                    funstar=fun(rstar)
c si el valor nuevo esta entre el maximo y el minimo
c lo acepta, cambiandolo por el punto que antes daba el
```

```
c valor mas grande
IF(funstar.lt.fmax.and.funstar.gt.fmin)THEN
                    DO 60j=1,ndim
                    p(imax,j)=rstar(j)
60                  END DO
                    WRITE(7,*)'reflexionaceptada'
                    GO TO100
c si el valor nuevo es menor que el que daba el  valor mas bajo
c se intenta una nueva expansion en ese sentido
ELSE IF(funstar.le.fmin)THEN
          WRITE(7,*)'intentaexpansion'
                    DO70j=1,ndim
                    rstar2(j)=rstar(j)*(1.d0+beta)-beta*rm(j)
70                  END DO
                    funstar2=fun(rstar2)
                IF(funstar2.lt.fmin)THEN
                    DO 80j=1,ndim
                    p(imax,j)=rstar2(j)
80            END DO
                    WRITE(7,*)'expansionaceptada'
                    GOTO100
                    ELSE
                    DO 90j=1,ndim
                    p(imax,j)=rstar(j)
90                  END DO
                              WRITE(7,*)'expasionrechazada'
                              WRITE(7,*)'semantienelareflexion'
                    GO TO100
                    END IF
c si el valor nuevo es mayor que el que daba el valor mas alto,
c se intenta una contraccion
                    ELSE IF(funstar.ge.fmax)THEN
                    WRITE(7,*)'intentacontraccion'
                    DO110j=1,ndim
                              rstar2(j)=gama*p(imax,j)+(1.d0-gama)*rm(j)
110                 END DO
                    funstar2=fun(rstar2)
c si la contraccion resulta,
```

```
                    IF(funstar2.lt.fmax)THEN
                    DO120j=1,ndim
                    p(imax,j)=rstar2(j)
120                 END DO
                    WRITE(7,*)'contraccionaceptada'
                    GOTO100
                    ELSE
c si la contraccion no resulta, contraer todos los vectores
c hacia el mejor
                    DO130i=1,ndim+1
                            IF(i.ne.imin)THEN
                            DO140j=1,ndim
                            p(i,j)=(p(i,j)+p(imin,j))/2.d0
140                         END DO
                             END IF
130                         END DO
                    WRITE(7,*)'contraccion rechazada'
                    WRITE(7,*)'se contrae todo el simplex'
                    GO TO100
                    END IF
                    END IF
                    RETURN
                    END
c funcion
REAL*8 function fun(r)
IMPLICIT REAL*8 (a-h,o-z)
DIMENSION r(2)
fun=r(1)**2+6*r(2)**4
RETURN
END
```

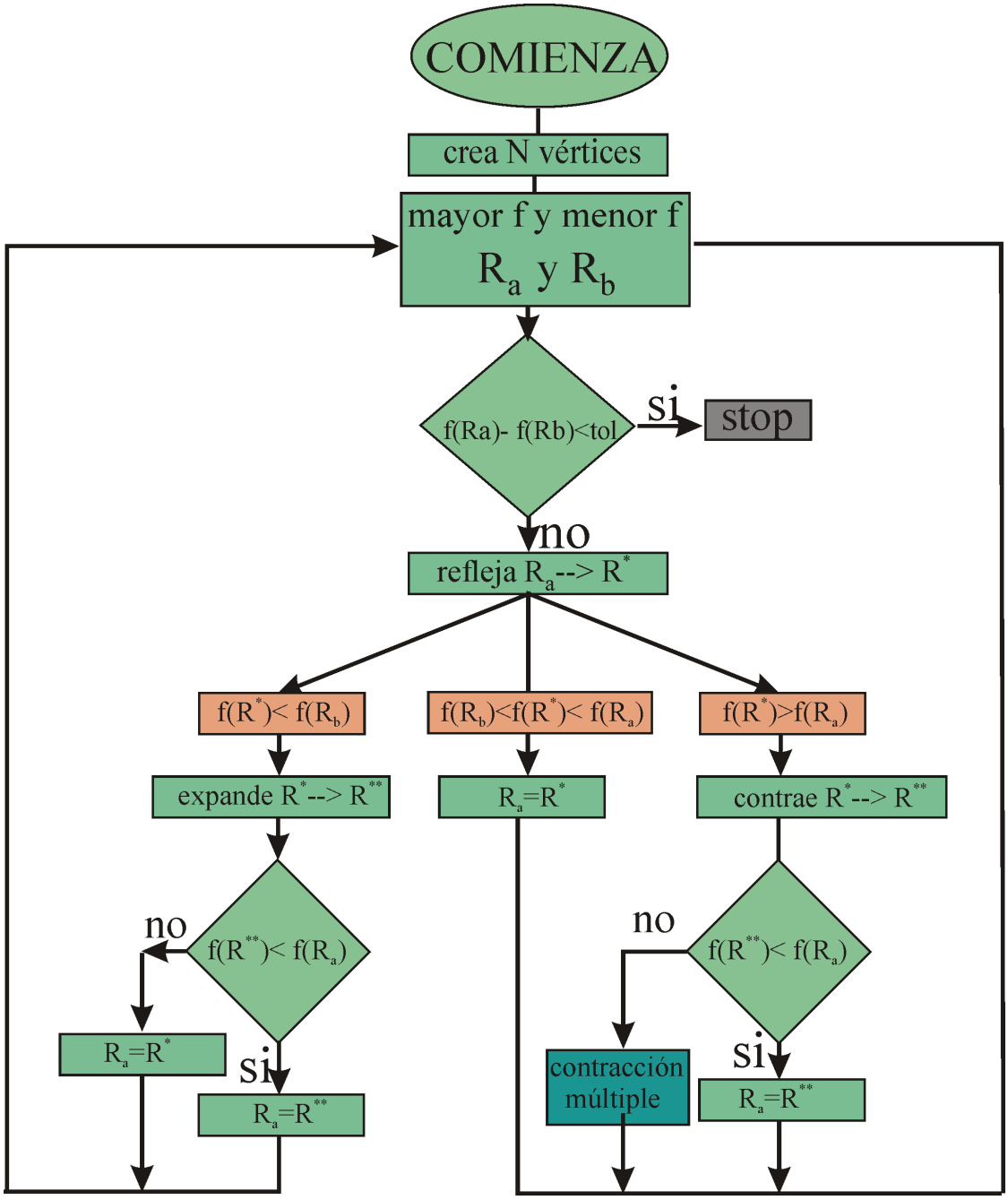
COMIENZA
crea N vértices
mayor f y menor f
R_a y R_b
f(Ra)- f(Rb)<tol
si
stop
no
refleja R_a--> R^*
$f(R^*)< f(R_b)$
$f(R_b)<f(R^*)< f(R_a)$
$f(R^*)>f(R_a)$
expande R^*--> R^{**}
$R_a=R^*$
contrae R^*--> R^{**}
no
$f(R^{**})< f(R_a)$
no
$f(R^{**})< f(R_a)$
$R_a=R^*$
si
$R_a=R^{**}$
contracción múltiple
si
$R_a=R^{**}$

Figure 1.5: