

7 Ecuaciones no lineales y raíces de polinomios	169
7.1 Ecuaciones no lineales en problemas físicos	169
7.1.1 La ecuación de Kepler	169
7.1.2 La trayectoria de un proyectil	172
7.2 Dificultades del problema	173
7.3 Ecuaciones no lineales simples	176
7.3.1 Método de bisección	178
7.3.2 Método de la posición falsa o regla falsi	179
7.3.3 Método de Newton-Raphson	181
7.3.4 Convergencia y fractales ¡NO ENTRA EXAMEN!	182
7.3.5 Métodos iterativos de punto fijo	183
7.3.6 Métodos de punto fijo más generales	188
7.3.7 Orden de convergencia de un método de punto fijo	189
7.3.8 Convergencia del método de Newton-Raphson	189
7.3.9 Método de la secante	191
7.3.10 Aproximación cuadrática, métodos de Newton y de la secante	193
7.3.11 Método de Steffensen	195

7.3.12	Iteración inversa y método de Newton	196
7.3.13	Aceleración de Aitken	197
7.4	Recomendaciones generales	198
7.5	Sistemas de ecuaciones no lineales	199
7.5.1	Método de Picard á la Gauss-Jacobi	199
7.5.2	Método de Picard á la Gauss-Seidel	200
7.5.3	Método de Newton-Raphson y sus modificaciones	201
7.5.4	Métodos de continuación	203
7.5.5	Optimización y descenso más rápido	204
7.6	Raíces de polinomios	206
7.6.1	Acotación de raíces. Reglas elementales y sucesiones de Sturm	206
7.6.2	Método de Horner para la evaluación de polinomios	208
7.6.3	Método de Newton-Raphson para raíces reales	209
7.6.4	Deflación (reducción) de Maehly	211
7.6.5	Método de Müller para raíces complejas	211
7.6.6	Método de Bairstow	213
7.6.7	Método de Graeffe	215
7.6.8	Método de Bernouilli	217
7.7	Aplicaciones a ecuaciones diferenciales ordinarias	218

Bibliografía

221

5 de febrero de 2002

© Francisco R. Villatoro, Carmen M. García, Juan I. Ramos. Estas notas están protegidas por derechos de copyright y pueden ser distribuidas libremente sólo con propósitos educativos sin ánimo de lucro. *These notes are copyright-protected but may be freely distributed for instructional nonprofit purposes.*

CAPÍTULO 7

ECUACIONES NO LINEALES Y RAÍCES DE POLINOMIOS

7.1 Ecuaciones no lineales en problemas físicos

7.1.1 La ecuación de Kepler

Una de las primeras ecuaciones que se resolvieron mediante métodos numéricos fue la ecuación de Kepler, que permite determinar donde se encuentra un planeta en un momento dado a partir de donde se encontraba en un momento anterior. Es una ecuación de importancia fundamental en mecánica celeste, pero no se puede invertir directamente en término de funciones elementales.

Las ley de Kepler de las áreas dice que el área subtendida por el radio vector de un planeta en órbita alrededor del sol es constante para intervalos fijos de tiempo. Esta ley se cumple si el planeta describe un órbita con forma de elipse, es decir, si despreciamos el efecto del resto de los planetas.

Localizar en su órbita elíptica un planeta que vemos en el cielo requiere determinar la anomalía verdadera, ν , en función de anomalía media, M . La figura 7.1.1 (izquierda) muestra gráficamente estos conceptos en una elipse con centro O , foco F , semieje mayor de longitud a y excentricidad e . La figura también muestra la circunferencia que circunscribe a la elipse de radio a y centro O . Dado un punto P de dicha elipse, dibujemos una línea perpendicular al semieje mayor que pase por P , y que lo interseca en el punto A . Se denomina excentricidad anómala al ángulo E de la figura. En mecánica celeste también su utiliza la media anómala, M , que corresponde con el área sombreada en la figura 7.1.1 (derecha). La media anómala (M) es una manera de medir el tiempo transcurrido y la excentricidad anómala (E) es una parametrización del ángulo en coordenadas polares. La ecuación de una elipse se puede escribir

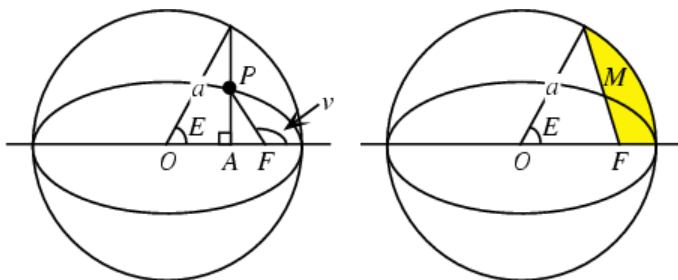


Figura 7.1. Dibujo ilustrando conceptos básicos sobre elipses utilizados en mecánica celeste, reproducido de Eric Weissteins's World of Mathematics.

como

$$M = E - e \sin(E),$$

que se denomina ecuación de Kepler, quien también introdujo la ecuación

$$\tan \frac{\nu}{2} = \sqrt{\frac{1+e}{1-e}} \tan \frac{E}{2},$$

que es más fácil de evaluar si se conoce la excentricidad anómala (E).

A partir de un almanaque astronómico es fácil determinar la anomalía media de un planeta en un día determinado del año. Si M no es un múltiplo de π , la ecuación de Kepler tiene una solución única, pero requiere un método numérico iterativo para determinarla. Este método parte de una estimación inicial, E_0 y obtiene sucesivas aproximaciones $E_{i+1} = g(E_i)$, donde $g(E)$ es una función a determinar.

El método más simple es el de Picard, que conduce al proceso iterativo

$$E_{i+1} = M + e \sin E_i, \quad E_0 = M.$$

Este método funciona bastante bien para este problema porque la excentricidad de los planetas, e , suele ser pequeña. Para Mercurio, $e = 0.205635$, si un almanaque astronómico nos da una anomalía media, $M = 1.2$ radianes, en 5 iteraciones obtenemos $E = 1.402738$ radianes, con 7 dígitos exactos, que nos permite localizar fácilmente dicho planeta en la eclíptica. De hecho para los nueve planetas del sistema solar, 8 iteraciones bastan para obtener un error menor de una millonésima en todos los casos.

Sin embargo, el método de Picard converge muy lentamente cuando la excentricidad es muy grande. Por ejemplo, para $e = 0.99$ y $M = 150$ grados = 2.618 radianes, el método requiere 287 iteraciones para obtener la respuesta $E = 2.876949$.

Desde los tiempos de Newton se prefiere utilizar un método más rápido, llamado método de Newton-Raphson, que conduce a la iteración

$$E_{i+1} = E_i - \frac{f(E_i)}{f'(E_i)}, \quad E_0 = M,$$

con $f(E) = E - M - e \sin E$. Para Mercurio este método requiere 3 iteraciones, en lugar de 8, para obtener 7 dígitos exactos, y para nuestro hipotético planeta con $e = 0.99$, son necesarias sólo 41 iteraciones, en lugar de 287. Por tanto, podemos concluir que el método de Newton es mucho más rápido que el de Picard.

En 1983, Danby [4] introdujo un nuevo método numérico que utiliza como valor inicial

$$E_0 = M + 0.85 \operatorname{signo}(\sin M) e,$$

y una iteración con derivadas de la función hasta de tercer grado,

$$\begin{aligned} k_1(E_i) &= -\frac{f}{f'}, \\ k_2(E_i) &= -\frac{f}{f' + \frac{k_1}{2} f''}, \\ k_3(E_i) &= -\frac{f}{f' + \frac{k_1}{2} f'' + \frac{(k_2)^2}{6} f'''}, \\ E_{i+1} &= E_i + k_3(E_i), \end{aligned}$$

donde hemos omitido los argumentos de f y sus derivadas, que son E_i . Recuerde que las derivadas de f son muy fáciles de calcular y poco costosas de evaluar numéricamente,

$$f'(E) = 1 - \cos E, \quad f''(E) = e \sin E, \quad f'''(E) = e \cos E.$$

Este método es mucho más rápido que los dos anteriores, ya que para Mercurio sólo requiere 1 iteración, y para $e = 0.99$, sólo 2. Además, el coste por iteración sólo es un poco más grande.

En este tema estudiaremos por qué el método de Danby (que es de cuarto orden) es más rápido que el de Newton (de segundo), y éste que el de Picard (sólo de primer orden). También estudiaremos diferentes condiciones de parada, que nos indican cuando el método ha convergido. En estos ejemplos hemos utilizado $|f(E_i)| < 10^{-7}$, lo que garantiza en nuestro ejemplo unos 7 dígitos exactos.

7.1.2 La trayectoria de un proyectil

Consideremos el problema del disparo (o lanzamiento) de un proyectil como se muestra en la Figura 7.1.2. Considerando el movimiento como plano, y despreciando la variación de masa del proyectil por el consumo de combustible, la segunda ley de Newton

$$\frac{d}{dt}(m v) = F,$$

donde m es la masa del proyectil, v su velocidad y F la resultante de las fuerzas a las que está sometido, aplicada a nuestro problema da,

$$\frac{d}{dt}(m \dot{x}) = (T - D) \cos \theta,$$

$$\frac{d}{dt}(m \dot{y}) = (T - D) \sin \theta - m g,$$

donde v es la velocidad del proyectil, $\dot{x} = v \cos \theta$ e $\dot{y} = v \sin \theta$, g es la aceleración de la gravedad, T el empuje producido por el motor del proyectil y D es la fuerza de fricción (*drag*). Para un fluido a altos números de Reynolds (cuando su viscosidad es despreciable), la fuerza de fricción toma la forma

$$D = \frac{1}{2} \rho v^2 S C_f$$

donde C_f es el coeficiente de fricción, S la sección transversal del proyectil y ρ la densidad del aire. Además, como hemos despreciado la pérdida de masa por consumo de combustible, $\dot{m} = 0$, podemos simplificar la expresión a

$$m \ddot{x} = (T - D) \cos \theta,$$

$$m \ddot{y} = (T - D) \sin \theta - m g,$$

donde D es la fuerza de fricción (*drag*). Como condiciones iniciales supondremos

$$x(0) = y(0) = 0, \quad v_0 = (\dot{x}(0), \dot{y}(0)).$$

Es más fácil trabajar utilizando la velocidad y el ángulo como variables a integrar. Sustituyendo en las ecuaciones anteriores obtenemos

$$m \dot{v} \cos \theta - m v \sin \theta \dot{\theta} = (T - D) \cos \theta,$$

$$m \dot{v} \sin \theta + m v \cos \theta \dot{\theta} = (T - D) \sin \theta - m g,$$

que es un sistema lineal en \dot{v} y $\dot{\theta}$ cuya solución conduce a

$$\dot{v} = \frac{T - D}{m} - \frac{g}{v} \sin \theta, \quad \dot{\theta} = -\frac{g}{v} \cos \theta.$$

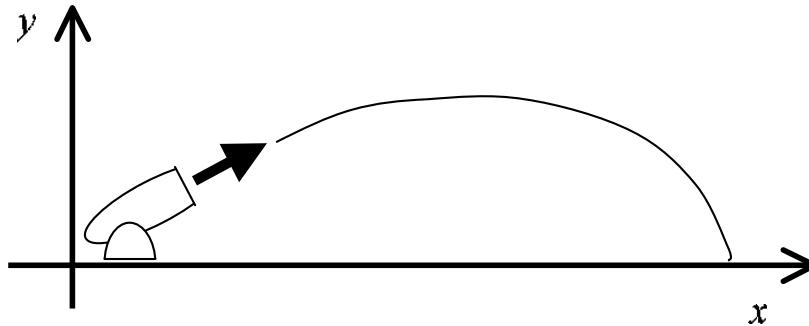


Figura 7.2. Dibujo esquemático del lanzamiento de un proyectil y de su trayectoria de vuelo.

Ahora utilizaremos como condiciones iniciales

$$\theta(0) = 0, \quad v(0) = \sqrt{\dot{x}(0)^2 + \dot{y}(0)^2}.$$

La solución del problema anterior se puede reducir a cuadraturas. Sin embargo, no presentaremos dicha solución.

Calcular el momento t_g cuando el proyectil alcance la tierra después de haber sido disparado es un problema que requiere la solución del problema de valores iniciales de estos sistemas de ecuaciones diferenciales ordinarias. Una vez obtenida la solución hay que determinar el cero de la función $y(t) = 0$, con $t > 0$. Otro problema, de gran interés, es dado el momento en que se alcanza la tierra, t_g , encontrar la velocidad inicial y el ángulo inicial $\theta(0)$ que hace que $y(t_g) = 0$.

Éstos son ejemplos de la necesidad de calcular los ceros de funciones a la hora de resolver determinados problemas propios de la ingeniería, por ejemplo, en mecánica de fluidos a altos números de Reynolds, elasticidad no lineal donde hay funciones no lineales de las deformaciones, en electrónica cuando se usan dispositivos activos como diodos que tienen características voltaje-corriente no lineales, en estructuras mecánicas que usan elementos no lineales, etc.

7.2 Dificultades del problema

Antes de estudiar algunos métodos numéricos para resolver ecuaciones no lineales hemos de indicar un punto que nunca se debe olvidar, el problema de calcular las raíces o ceros de una función $f(x)$, general, es muy complejo por no decir casi imposible. ¿Pero por qué?

El problema fundamental es la propia evaluación de $f(x)$ como nos muestran los siguientes ejemplos de Kahan [2]. Consideremos la función $g(x) = x + 2(x-5)$, cuya raíz exacta es $x = 10/3$, pero que evaluada numéricamente es aproximadamente del orden del cuadrado del épsilon de la máquina. Por ejemplo, en aritmética IEEE-754 de doble precisión es $c = g(10/3) \approx 2 \times 10^{-31}$. Si definimos la función $f(x) = 1 - 2 \exp(-g(x)^2/c^2)$, ésta tiene dos raíces analíticas cerca de $x = 10/3$, que como es fácil comprobar son

$$x_{\pm} = 10/3 \pm c \frac{\sqrt{\ln 2}}{3}.$$

Sin embargo, la evaluación numérica de $f(x)$, debido a los errores de redondeo, siempre da exactamente $f(x) = 1$, para todo x . Por ello, es imposible numéricamente, en aritmética IEEE-754, descubrir estos ceros de la función.

Peor aún, cualquier algoritmo numérico para calcular raíces puede encontrar raíces que en realidad no lo son. Consideremos las dos funciones siguientes

$$f_1(x) = 1/g(x), \quad f_2(x) = 1/(g(x) + c^2/g(x)),$$

donde $g(x)$ y c se definen como antes. Estas dos funciones tienen valores idénticos en aritmética flotante IEEE-754, sin embargo, en $x = 10/3$ la primera tiene un polo, $f_1(10/3) = \infty$, y la segunda un cero, $f_2(10/3) = 0$.

Kahan [2] ha demostrado que el problema fundamental de la resolución de ecuaciones es la imposibilidad de decidir si una función siempre vale cero. Sea $f(x) \equiv 1$, que no tiene ninguna raíz. Cualquier método numérico que pretenda calcular las raíces de $f(x)$ tendrá que evaluarla en una serie de puntos, sean $\{x_i\}$, $i = 1, 2, \dots, n$. Para estos puntos la función $f(x_i) = 1$. Sin embargo, consideremos la función

$$\hat{f}(x) = (x - x_1)(x - x_2) \cdots (x - x_n) + 1,$$

que es un polinomio de grado n y por el teorema fundamental del álgebra tiene n raíces. ¿Cómo podemos saber si $f(x)$ y $\hat{f}(x)$ son distintas si valen lo mismo en todos los puntos que el algoritmo numérico ha considerado? Este problema es equivalente al problema de la parada en teoría de la calculabilidad (Turing (1933) [2]) del que no daremos más detalles.

Los ejemplos anteriores pueden parecer forzados, pero muestran claramente que resolver ecuaciones en forma general es imposible, aunque en la mayoría de los casos de interés práctico sí es posible. Lo importante es que un conocimiento detallado de las funciones con las que estamos trabajando es fundamental para el éxito del procedimiento.

A la hora de calcular una raíz real x de una ecuación $f(x) = 0$ es necesario tener en cuenta seis puntos importantes [3]:

1. ¿Qué ecuación resolver? La función f se puede reescribir de diferentes formas manteniendo sus raíces reales. Las dos ecuaciones $f(x) = 0$, y $\exp(f(x)) - 1 = 0$, tienen las mismas raíces reales, pero casi siempre una de ellas es más fácil de resolver que la otra. Para $f(x) = 6x - x^4 - 1$, la primera ecuación es más fácil, pero para $f(x) = \ln(6x - x^4)$, lo es la segunda [2]. En general, cada ecuación es un representante de una familia infinita de funciones que tienen las mismas raíces reales, y cualquier método numérico funcionará mejor con algunas de ellas que con otras.

2. ¿Qué método elegir? Normalmente se utiliza un método numérico iterativo, como

$$x^{(k+1)} = g(x^{(k)}, x^{(k-1)}, \dots),$$

que parte de una estimación inicial $x^{(0)}$ a la raíz y obtiene un sucesión de aproximaciones $\{x^{(k)}\}$. Si esta sucesión converge y si $x = g(x, x, \dots)$ (x es un punto fijo de g) implica que $f(x) = 0$, entonces el método convergerá a una raíz de la ecuación. Hay infinitos métodos numéricos que se pueden aplicar una función dada. Unos serán de convergencia más rápida, otros de menor costo por iteración, o más fáciles de implementar numéricamente.

3. ¿Por dónde debemos empezar a buscar la raíz? Parece obvio que cuando conocemos un buen valor aproximado para la raíz, un buen método numérico debería encontrar fácilmente dicha raíz. Sin embargo, si este valor inicial no es próximo a la raíz, muchos métodos no convergerán a ella. Un buen valor inicial depende siempre de nuestro conocimiento de las propiedades de la ecuación a resolver. Para intentar responder a esta cuestión se ha desarrollado la teoría matemática de la convergencia global de métodos numéricos.

4. ¿Cuán rápida es la convergencia del método? Como hemos visto para la ecuación de Kepler, hay métodos numéricos convergentes que se acercan a la raíz más rápidamente que otros. La eficiencia computacional de un método numérico depende no sólo del número de iteraciones, si no también del coste de cada una de ellas. El concepto de orden de convergencia nos permite estimar el número de iteraciones del método, aunque sólo localmente, o sea, cerca de la raíz. Normalmente el coste de un iteración se mide con el número de evaluaciones de la función $f(x)$, y sus derivadas, por iteración.

5. ¿Cuándo debemos parar la iteración? El chequeo de la convergencia de un método iterativo, como de cualquier sucesión, se basa en un análisis de los errores del método, aunque se debe tener en cuenta la posibilidad de que el cero (o punto fijo del método) no exista. Normalmente se especifica una tolerancia de error, de hecho, a varias maneras de hacerlo. Es muy importante tener en cuenta que los errores de redondeo pueden impedir que se

alcance esta tolerancia de error, sobre todo si es próxima al ϵ de la máquina, ya que la evaluación de la función f , o del método iterativo g , puede incurrir en errores de redondeo mucho mayores. De hecho, a veces, el método iterativo se comporta de una manera aleatoria muy cerca de la raíz (fenómeno llamado *dithering*).

6. ¿Cómo podemos garantizar la precisión de la raíz? Una vez garantizada la convergencia del método numérico debemos determinar de alguna manera cuántos dígitos de precisión tiene la aproximación a la raíz que hemos determinado. Aquí, el análisis de los errores del método es de fundamental importancia.

Estas seis cuestiones ni son completamente independientes entre sí, ni pueden contestarse siempre en dicho orden. Además, todavía no existe una teoría de la convergencia de métodos numéricos iterativos lo suficientemente potente como para contestar todas estas cuestiones de forma completamente general. De hecho, la interpretación de los métodos numéricos iterativos en el marco de la teoría de los sistemas dinámicos discretos, llamados también aplicaciones (*maps*), que pueden tener comportamiento de gran complejidad (soluciones casi-periódicas, caos, ...), permite asegurar que nunca existirá una teoría general que conteste completamente a estas seis cuestiones. Sin embargo, ello no quita que en la práctica tratemos de contestarlas, al menos de manera aproximada.

El uso de programas matemáticos de propósito general, como Mathematica o Matlab, puede hacer pensar a algunos que el problema de determinar los ceros de una función se reduce a utilizar cualquiera de los algoritmos de propósito general que se incluyen en estos paquetes informáticos. Sin embargo, como ya hemos indicado y como veremos en este tema, es imposible desarrollar un programa automático de resolución general de cualquier ecuación (*general-purpose root-finder*), luego estos programas no pueden tener uno, aunque para determinadas familias de funciones, como los polinomios, ya hay programas muy buenos. En general, la ventaja fundamental de los algoritmos de propósito específico es la eficiencia, la velocidad, que es especialmente importante cuando la resolución de una ecuación es parte de un problema numérico de mayor magnitud, como la resolución de una ecuación diferencial no lineal.

7.3 Ecuaciones no lineales simples

Para la resolución de ecuaciones no lineales de una variable real podemos utilizar dos tipos de métodos, que llamaremos métodos de convergencia por intervalos y puntual. Los métodos de convergencia por intervalos parten de un intervalo que acota la raíz, está garantizado que

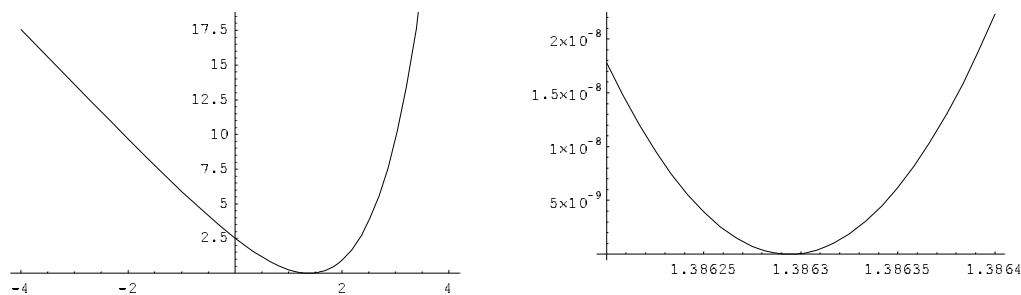


Figura 7.3. Representación gráfica de la función $\exp(x) - 4x - (4 - 4\ln(4))$ (izquierda) y cerca de su raíz doble $\ln 4 \approx 1.38629$.

ésta se encuentra en su interior, y durante todas las iteraciones mantienen esta propiedad. Los métodos puntuales, parten de una estimación inicial de la raíz y tratan de aproximarse a ella. La desventaja de los primeros está en la dificultad de encontrar un intervalo que acote a la raíz, por un lado, y en que suelen ser más lentos, por otro, aunque su convergencia está garantizada. La desventaja de los segundos está en que tampoco es fácil encontrar una buena estimación inicial de una raíz y su convergencia no está garantizada, aunque son muchos más rápidos que los anteriores cuando están cerca de la raíz.

Los métodos de convergencia por intervalos se suelen basar en el teorema de Bolzano, que garantiza para toda función continua en un intervalo a, b en el que cambia de signo, $f(x) \in \mathbf{C}^0[a, b]$ con $f(a)f(b) < 0$, que existe al menos una raíz $\xi \in (a, b)$ de $f(\xi) = 0$. La demostración de este teorema, aunque no es difícil, requiere una definición rigurosa del concepto de continuidad y es propio de un primer curso de Cálculo.

Es importante notar que este teorema no funciona para las raíces que son un mínimo o un máximo de la función, raíces dobles. Por ejemplo, la función $f(x) = \exp(x) - 4x - (4 - 4\ln(4))$, tiene una raíz doble en $x = \ln(4)$, por lo que no existe ningún entorno de esta raíz que cumpla las condiciones del teorema de Bolzano, ya que es no negativa para todo x como se muestra gráficamente en la figura 7.3.

Estudiaremos sólo dos métodos de convergencia por intervalos, los más sencillos, el método de bisección y el de la posición falsa o regula falsi. Nos centraremos en los métodos de convergencia puntual, presentando el método de Newton y el de la secante, y la teoría general de los métodos iterativos de punto fijo. Hemos de indicar que la aritmética basada en intervalos [5, 6] permite obtener algoritmos de convergencia por intervalos a partir de cualquier método de convergencia puntual. No trataremos este interesante tópico en este curso.

En la práctica, es habitual tomar un método de convergencia por intervalos para las primeras iteraciones, y cuando estamos suficientemente cerca de la raíz pasar a un método de convergencia puntual más rápido.

7.3.1 Método de bisección

El método de bisección se basa en la aplicación directa del teorema de Bolzano. Se parte de un intervalo que cumpla las condiciones de este teorema. Este intervalo se divide en dos partes iguales y se elige, para continuar, el subintervalo que sigue cumpliendo la condición de Bolzano que garantiza que en su interior existe una raíz. Repitiendo este proceso para dicho sub-intervalo se obtiene una sucesión de intervalos encajados que contienen a la raíz (o una de las raíces, si es que hay un número impar).

El algoritmo de bisección se puede escribir:

1. Encontrar un intervalo $I = [a, b]$ tal que $f(a)f(b) < 0$.
2. Determinar el punto medio $\mu = (a + b)/2$ del intervalo I .
3. Si $f(\mu) = 0$, hemos acabado.
4. Si $f(a)f(\mu) > 0$ y $f(b)f(\mu) < 0$ elegir como intervalo $I = [\mu, b]$, es decir, $a = \mu$.
5. Si $f(b)f(\mu) > 0$ y $f(a)f(\mu) < 0$ elegir como intervalo $I = [a, \mu]$, es decir, $b = \mu$.
6. Condición de parada. Si $|\mu - a|/2 > \epsilon_1$ entonces repetir el paso (2), si no μ es una aproximación a la raíz con un error absoluto acotado por ϵ_1 .

Como en cada paso la longitud del intervalo se reduce a la mitad, el teorema de Bolzano garantiza la convergencia del método. Sin embargo, este método es de convergencia muy lenta debido a que sólo usa información relativa al intervalo $[a, b]$, es decir, a las abscisas. De hecho el error se reduce a la mitad en cada paso, y tras n pasos el error es 2^{-n} el error inicial. Más tarde introduciremos el concepto de orden de convergencia y comprobaremos que este método converge de forma lineal con constante de convergencia igual a $1/2$. Además, si la función tiene varios ceros en el intervalo, algunos de éstos se pueden perder durante el proceso iterativo.

Ejemplo. Sea $f(x) = x^3 - x - 1 = 0$, $x \in [1, 2]$. Como

$$f(1) = -1, \quad f(2) = 5, \quad \Rightarrow \quad \exists \xi \in [1, 2], f(\xi) = 0.$$

Para saber cuántas raíces hay en dicho intervalo podemos calcular $f'(x) = 3x^2 - 1 = 0$,

$$f'(1) = 2, \quad f'(2) = 11,$$

y aparentemente sólo hay una (o un número impar) de soluciones en dicho intervalo. Si calculamos las raíces de $f'(x)$,

$$f'(x) = 0, \quad x = \pm \frac{1}{\sqrt{3}},$$

con lo que $f'(x) \geq 0$ en $[1, 2]$, y por tanto, existe una sola raíz de $f(x)$ en $[1, 2]$.

Aplicando el método de bisección obtenemos en cada paso

Paso	a	b	μ	$f(\mu)$
0	1	2	$3/2$	$7/8$
1	1	$3/2$	$5/4$	$-19/64$
2	$5/4$	$3/2$	$11/8$	0.2246
3	$5/4$	$11/8$	$21/16$	-0.05151
4	$21/16$	$11/8$	$43/32$	0.08261
5	$21/16$	$43/32$	$85/64$	0.01458
6	$21/16$	$85/64$	$169/128$	-0.01871
7	$169/128$	$85/64$	$339/256$	-0.002128

con lo que la raíz es $339/256 \approx 1.32422$ con un error de $1/512 \approx 0.002$, con un residuo, en este caso, del mismo orden $f(\mu) = -0.002$.

7.3.2 Método de la posición falsa o regula falsi

El método de la posición falsa también se basa en el método de Bolzano, pero para conseguir una convergencia más rápida, en lugar de dividir los intervalos por la mitad, aproxima la función en el intervalo por una línea recta y divide el intervalo en función de la raíz de dicha recta, que si estamos suficientemente cerca de la raíz será una buena aproximación a la función. La Figura 7.3.2, muestra gráficamente el proceso iterativo que presentamos en el siguiente algoritmo:

1. Encontrar un intervalo $I = [a, b]$ tal que $f(a)f(b) < 0$. Sea $x^{(0)} = a$

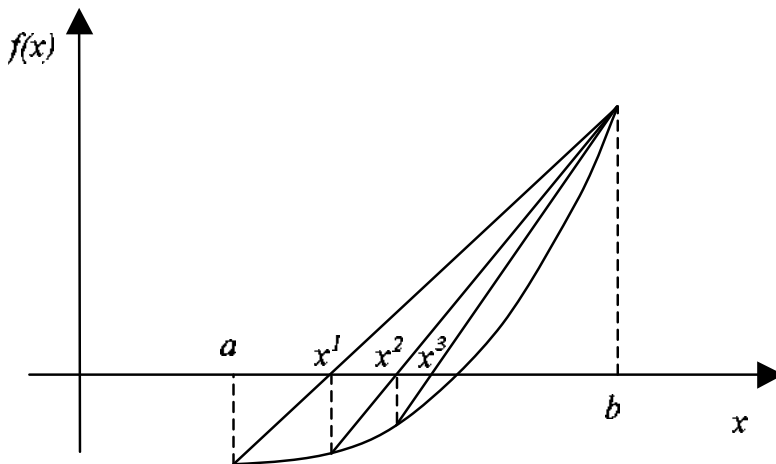


Figura 7.4. Proceso iterativo del método de la posición falsa.

2. Interpolación linealmente

$$\frac{f(b) - f(a)}{b - a} = \frac{y - f(a)}{x - a}, \quad y = f(a) + \frac{f(b) - f(a)}{b - a} (x - a),$$

y encontrar la raíz de esta recta ($y = 0$),

$$x^{(m)} = a - \frac{f(a)}{f(b) - f(a)} (b - a).$$

3. Si $f(x^{(m)}) = 0$, hemos acabado.

4. Si $f(a) f(x^{(m)}) > 0$ y $f(b) f(x^{(m)}) < 0$ elegir como intervalo $I = [x^{(m)}, b]$.

5. Si $f(b) f(x^{(m)}) > 0$ y $f(a) f(x^{(m)}) < 0$ elegir como intervalo $I = [a, x^{(m)}]$.

6. Condición de parada. Si $|x^{(m)} - x^{(m-1)}| > \epsilon_1$ entonces repetir el paso (2), en otro caso, acabar dando como resultado $x^{(m)}$ como la raíz del problema, que tendrá un error absoluto acotado por ϵ_1 .

El método de regula falsi sólo encuentra una raíz, no requiere derivadas de la función, converge linealmente para funciones convexas y requiere dos puntos en cada iteración. Este método iterativo es un caso particular de los métodos de interpolación inversa, en los que se interpola la función a resolver y se aproxima la raíz con la raíz del interpolante; en nuestro caso tenemos un método iterativo de interpolación lineal inversa.

El método de la posición falsa nos da una estimación de la raíz, pero no la acota mediante un intervalo, ya que siempre se acerca a la raíz por el mismo lado (izquierda o derecha del cero de la función en función de si ésta es cóncava o convexa, respectivamente) con una pendiente que crece. Una manera fácil de obtener un intervalo que acote la raíz es aplicar un método mixto bisección-posición falsa, en el que tras cada paso del método de posición falsa aplicamos un paso del método de bisección.

7.3.3 Método de Newton-Raphson

Sea $f(x) \in \mathbf{C}^1[a, b]$, entonces podemos desarrollar en serie de Taylor hasta primer orden alrededor de un punto $x^{(0)}$ cercano al cero x de $f(x)$ de la forma (exacta)

$$f(x) = f(x^{(0)}) + f'(\xi)(x - x^{(0)}) = 0, \quad \xi \in (x^{(0)}, x),$$

que despejando la raíz nos da

$$x = x^{(0)} - \frac{f(x^{(0)})}{f'(\xi)}, \quad x^{(0)} \leq \xi \leq x.$$

Pero como no conocemos ξ , nos tenemos que limitar a utilizar la expresión aproximada

$$x^{(1)} = x^{(0)} - \frac{f(x^{(0)})}{f'(x^{(0)})}.$$

Parece razonable construir el método iterativo

$$x^{(k+1)} = x^{(k)} - \frac{f(x^{(k)})}{f'(x^{(k)})} = g(x^{(k)}),$$

que se denomina método de Newton-Raphson.

El método de Newton se puede interpretar gráficamente como aparece en la figura 7.3.3. Por el punto $(x^{(k)}, f(x^{(k)}))$ se traza una recta cuya pendiente sea la derivada $f'(x^{(k)})$, el punto de corte de esta recta con el eje de las abscisas es la nueva aproximación $x^{(k+1)}$ a la raíz.

Este método, al contrario que los métodos anteriores, utiliza un sólo punto por interacción y no chequea el teorema de Bolzano, por lo que no devuelve un intervalo donde se incluya la raíz. Además, parece que puede converger más rápidamente que los métodos anteriores, ya que de hecho utiliza más información que ellos al hacer uso de la derivada de la función. ¿Es esto realmente cierto? El objeto de la próxima sección será responder a estas cuestiones.

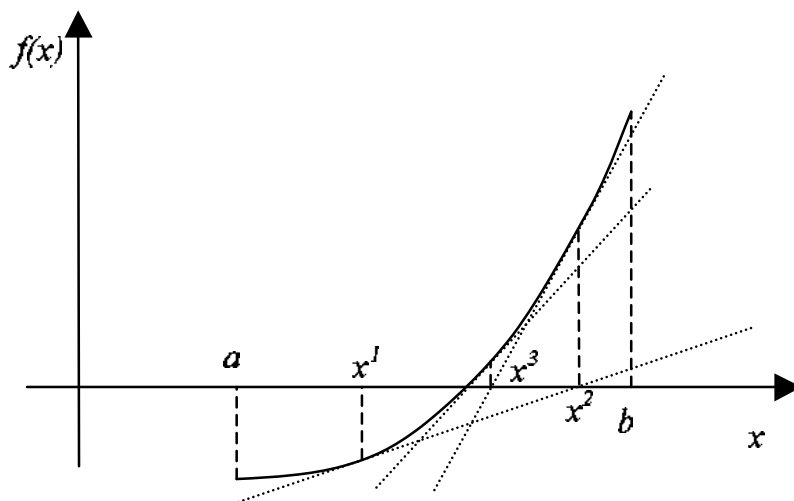


Figura 7.5. Proceso iterativo del método de Newton-Raphson.

7.3.4 Convergencia y fractales ¡NO ENTRA EXAMEN!

Sir Arthur Cayley, in 1879, considered Newton's method and asked, which zero of the equation $z^3 - 1 = 0$ in the complex plane would the method converge towards if one starts with an arbitrary initial guess. Much of Julia's work was motivated by Cayley's paper. Sixty years later, Mandelbrot aware of the work done by Pierre Fatou and Gaston Julia on iterative processes in complex plane laid the foundation of Fractal Geometry.

So much so for 4000 years of history! So what is the deal with determining to which root the Newton's method would converge if one started with an arbitrary initial guess. Sure enough Newton's method works and most assuredly' the process leads to one of the three solutions (incidentally the roots of the equation $z^3 - 1 = 0$ are 1 , $\exp(2\pi i/3)$ and $\exp(4\pi i/3)$). So, the question is: if one were to start with a large number of points in the complex plane as an initial guess to the Newton's method, how are these initial guess points distributed in the complex plane. Here is a color coded distribution of initial guesses that lead one of the three solutions from a computer experiment of the 500×500 grid in the complex plane, between -2 and 2 both along the real (X-) and imaginary (Y-) axis. The initial guesses are colored red, yellow or blue depending upon whether the initial points approached the solution 1 , $\exp(2\pi i/3)$ or $\exp(4\pi i/3)$, respectively. The coordinates of the three roots are marked by the $+$ symbol.

language of dynamical systems theory the roots can be considered as point attractors and the regions of the complex plain that asymptote to a particular root is called the basin of attraction

of that root. It is clear from the plot above that there is a very complex boundary separating the basins of attraction of the three roots. This complexity continues to exist even when looking at very small differences between initial values of the guess; this boundary never becomes smooth even at very small differences. Thus, if we were to zoom in a close-up the boundaries have a very complicated structure. which has similarity to a Cantor set. In other words, wherever two basins seem to meet, we discover upon closer examination that the third basin of attraction is there in between them, and so on ad infinitum. This complex boundary is a fractal boundary. The competition of the roots over all the points in the plane, is far from simple!

7.3.5 Métodos iterativos de punto fijo

El método de Newton-Raphson pertenece a una familia de métodos para el cálculo de ceros de funciones que se denominan métodos iterativos de punto fijo o también métodos de iteración funcional. Estudiaremos en esta sección dichos métodos de forma general y relegaremos a las próximas la aplicación de estos resultados generales al método de Newton.

Un método iterativo para el cálculo de raíces de la forma

$$x^{(n+1)} = g(x^{(n)}),$$

se denomina método de iteración de punto fijo, y se utiliza para calcular el punto fijo $\alpha = g(\alpha)$; si dicha ecuación implica que $f(\alpha) = 0$, lo que normalmente garantiza una adecuada elección de g , entonces dicho método, sin converge, permite calcular las raíces de dicha función.

Un iterativo de punto fijo se puede interpretar geoméricamente como aparece en la figura 7.3.5. El punto fijo corresponde a la intersección de la recta $y = x$ con la curva $y = g(x)$. Si se traza una línea horizontal que pasa por el punto $(x^{(n)}, g(x^{(n)}))$, y en el punto de corte de ésta con la línea $y = x$ se traza una línea vertical hacia abajo, se determina gráficamente el valor de $x^{(n+1)}$ en el eje de las abscisas. La figura 7.3.5 ilustra este hecho mediante flechas, que indican el orden de ejecución de este proceso.

La convergencia de un método de iteración de punto fijo queda garantizada para funciones contractivas gracias al teorema del punto fijo de Brower. Sea un intervalo cerrado $I = [a, b]$, una función $f(x)$ se dice contractiva si

$$\forall x \in I = [a, b], \quad g(x) \in H \subseteq I,$$

que normalmente se escribe como $g(I) \subset I$. El teorema de Brower dice que si $g(x)$ es una función contractiva en un conjunto compacto (por ejemplo, el cerrado I), entonces tiene un único punto

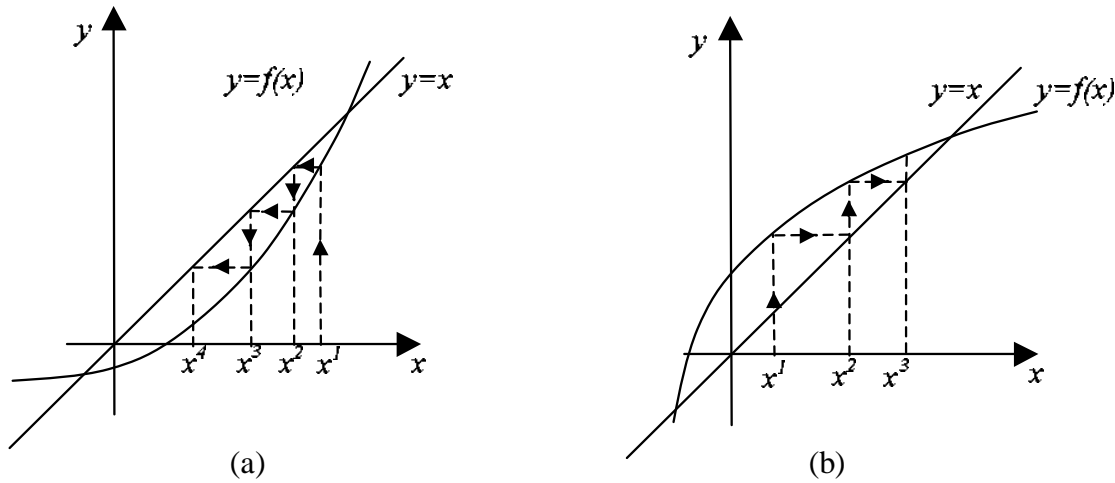


Figura 7.6. Ilustración de un método iterativo de punto fijo en el caso convergente (b) y no convergente (a).

fijo $\alpha \in I$ y además la sucesión $x^{(n+1)} = g(x^{(n)})$, para $x^{(0)} \in I$, converge a dicho punto fijo. Omitiremos en este curso la demostración general de este teorema.

En la práctica, determinar si una función es contractiva en un intervalo utilizando la definición anterior es bastante difícil. Una función $g : I \rightarrow I$ es Lipschitz-continua en I si

$$\forall x, y \in I, \quad |g(x) - g(y)| \leq \lambda |x - y|.$$

Se denomina a λ constante de Lipschitz de g . Se puede demostrar que una función Lipschitz continua con $\lambda < 1$ es contractiva en I . Además dicha función es continua. Si también es derivable, $g \in \mathbf{C}^1(I \rightarrow I)$, entonces podemos calcular fácilmente su constante de Lipschitz

$$\lambda = \|g'(x)\|_{\infty, I} = \max_{x \in I} |g'(x)|.$$

Una función derivable será contractiva si $\|g'(x)\|_{\infty, I} < 1$.

Podemos demostrar fácilmente un caso particular del teorema del punto fijo de Brower. Sea $g(x)$ Lipschitz continua con constante $\lambda < 1$ en el compacto $I \equiv [a, b]$, y además $g(I) \subset I$, entonces existe un único punto fijo $\alpha \in (a, b)$ y el método de punto fijo $x^{(n+1)} = g(x^{(n)})$, converge a dicho punto fijo para todo $x^{(0)} \in I$.

Demostración [1]. Utilizando la propiedad de contractividad

$$|x^{(n+1)} - x^{(n)}| = |g(x^{(n)}) - g(x^{(n-1)})| \leq \lambda |x^{(n)} - x^{(n-1)}| \leq \lambda^2 |x^{(n-1)} - x^{(n-2)}|,$$

y así sucesivamente obtenemos

$$|x^{(n+1)} - x^{(n)}| \leq \lambda^n |x^{(1)} - x^{(0)}|.$$

Dado que

$$\lim_{n \rightarrow \infty} x^{(n)} = x^{(0)} + \sum_{n=1}^{\infty} x^{(n)} - x^{(n-1)} = x^{(0)} + (x^{(1)} - x^{(0)}) + (x^{(2)} - x^{(1)}) + \cdots + (x^{(n)} - x^{(n-1)}),$$

la sucesión $\{x^{(n)}\}$ converge si y sólo si la serie

$$\sum_{n=1}^{\infty} (x^{(n)} - x^{(n-1)})$$

converge, para lo que basta probar que es absolutamente convergente, es decir, que converge

$$\sum_{n=1}^{\infty} |x^{(n)} - x^{(n-1)}|.$$

Esto es sencillo ya que

$$\sum_{n=1}^{\infty} |x^{(n)} - x^{(n-1)}| \leq \sum_{n=1}^{\infty} \lambda^{n-1} |x^{(1)} - x^{(0)}| = \frac{1}{1-\lambda} |x^{(1)} - x^{(0)}|.$$

Como la sucesión $\{x^{(n)}\}$ converge, llamaremos α a dicho límite,

$$\alpha = \lim_{n \rightarrow \infty} x^{(n)}.$$

Como g es contractiva, también es continua, por lo que

$$g(\alpha) = g(\lim_{n \rightarrow \infty} x^{(n)}) = \lim_{n \rightarrow \infty} g(x^{(n)}) = \lim_{n \rightarrow \infty} x^{(n+1)} = \alpha,$$

y α es un punto fijo de g . Además, como todos los $x^{(n)} \in I$, tenemos que $\alpha \in I$. Este punto fijo de g en I es único ya que si β es otro punto fijo de g en el intervalo I ,

$$|\alpha - \beta| = |g(\alpha) - g(\beta)| \leq \lambda |\alpha - \beta|,$$

y como $\lambda < 1$, tenemos $|\alpha - \beta| = 0$, **cqd**.

Sea una función $g(x) \in \mathbf{C}^1(a, b)$ que tiene una raíz $\alpha \in (a, b)$, de lo que podemos tener garantía gracias al teorema de Bolzano, por ejemplo, y el método iterativo $x^{(n+1)} = g(x^{(n)})$. Entonces si $g'(\alpha) < 1$, existe un entorno $U_\alpha(\epsilon) \equiv [\alpha - \epsilon, \alpha + \epsilon]$, para ϵ suficientemente pequeño, tal que para toda condición inicial en el mismo, $x^{(0)} \in U_\alpha$, el método numérico converge a dicha raíz.

Demostración. Dado que $g'(\alpha) < 1$, por la continuidad de la derivada, existe un entorno $U_\alpha(\epsilon)$ tal que $\|g'\|_{\infty, U_\alpha} < 1$. La ecuación que cumple el error, $e^{(k)} = \alpha - x^{(k)}$, es

$$\begin{aligned} e^{(k+1)} &= \alpha - x^{(k+1)} = g(\alpha) - g(x^{(k)}) = -g'(\alpha)(x^{(k)} - \alpha) + O\left((x^{(k)} - \alpha)^2\right) \\ &= g'(\xi_k)(\alpha - x^{(k)}) = g'(\xi_k)e^{(k)}, \quad \xi_k \in (\alpha, x^{(k)}) \end{aligned}$$

donde se ha utilizado el teorema del valor medio (o del resto de Taylor) y se ha supuesto $x^{(k)} > \alpha$, sin pérdida de generalidad. Escojamos un $x^{(0)} \in U_\alpha$, para el que se cumplirá

$$|e^{(1)}| = |g'(\xi_0)| |e^{(0)}| \leq |e^{(0)}|, \quad \xi_0 \in U_\alpha,$$

que garantiza que $x^{(1)} \in U_\alpha$. Aplicando esta relación sucesivamente,

$$|e^{(k)}| \leq |e^{(k-1)}| \leq \dots \leq |e^{(0)}|,$$

que garantiza que $x^{(k)} \in U_\alpha$ para todo k . Además, el método converge

$$\lim_{k \rightarrow \infty} |e^{(k+1)}| = 0,$$

cqd.

La condición $|g'(\alpha)| < 1$ indica que la pendiente de g en el punto fijo debe ser menor que la unidad (ángulo inferior a 45 grados). Este hecho se ilustra gráficamente en la figura 7.3.5. En la parte (a) se observa que $|g'(\alpha)| > 1$ implica la divergencia del método hacia el punto fijo de abcisa positiva, aunque el método converge hacia un segundo punto fijo de abcisa negativa. En la parte (b) se observa que $|g'(\alpha)| < 1$ garantiza la convergencia hacia el punto fijo de abcisa positiva.

Veremos a continuación algunos ejemplos de la aplicación de las condiciones anteriores. Es muy importante indicar que la condición $|g'(\alpha)| < 1$ sólo garantiza que existe un entorno (suficientemente pequeño) donde está garantizada la convergencia. En intervalos más grandes es necesario verificar la contractividad de $g(x)$.

1. Consideremos la ecuación $x + \sqrt{x} = 0$, cuya única solución en \mathbb{R} es $x = 0$. La iteración de punto fijo natural

$$x^{(n+1)} = -\sqrt{x^{(n)}} = g(x^{(n)}),$$

con $0 < x^{(0)} \in \mathbb{R}$, no converge ya que no cumple las condiciones del teorema. En concreto $g(\mathbb{R}) \not\subseteq \mathbb{R}$, ya que se conduce a valores $x^{(n)}$ complejos, por lo que lo que no puede ser contractiva en \mathbb{R} . Por ejemplo, $x^{(0)} = 1$, $x^1 = -1$, $x^2 = -i$, $x^3 = (-1)^{3/4} = (i - 1)/\sqrt{2}$,

...

2. PARA EJERCICIOS: La ecuación $3e^x = e^3 x$, que tiene dos raíces reales, se puede resolver por el método numérico $z = 3 + \ln(z/3)$ y por $z = 3e^{z-3}$.
3. Un problema general admite gran número de métodos de punto fijo, sin embargo, no todos son adecuados y habrá que estudiar la convergencia de cada uno de ellos. Por ejemplo, sea la ecuación $f(x) = x^2 - x - 2 = 0$, cuyas raíces son

$$x = \frac{1}{2} \pm \frac{1}{2} \sqrt{1+8} \in \{-1, 2\}.$$

Presentaremos varios métodos de iteración de punto fijo para ella.

- (a) Sea $x = x^2 - 2 \equiv g(x)$, cuya derivada $g'(x) = 2x$, cumple la condición

$$|g'(x)| = 2|x| < 1 \quad \Rightarrow \quad -1 < 2x < 1,$$

en el intervalo abierto $x \in (-1/2, 1/2)$, sin embargo, en dicho intervalo $g(-1/2, 1/2) \subset (-7/4, 2)$, por lo que no son aplicables las condiciones del teorema y el método no converge.

- (b) Sea $x = \sqrt{x+2}$, cuya derivada es

$$g'(x) = \frac{1}{2\sqrt{x+2}}, \quad x \in \mathbb{R} - \{-2\},$$

y se cumple la condición $|g'(x)| < 1$,

$$-1 < \frac{1}{2\sqrt{x+2}} < 1, \quad \frac{1}{4} < x+2,$$

para todo $x > -7/4 = -1.75$. Además, ya que $g(x)$ es una función monótona creciente, $g(2) = 2$ y $|g'(x)| < 1$ para $x > 2$, se comprueba fácilmente que $g(-7/4, b) \subset (-7/4, b)$, si $b \geq 2$. Por tanto, $g(x)$ es contractiva en dicho intervalo y este método numérico converge para todo $x^{(0)}$ en dicho intervalo al único punto fijo que se encuentra en él, es decir, $x = 2$. Experimentalmente se puede comprobar que este método converge a $x = 2$ para todo $x^{(0)} \in \mathbb{R}$.

- (c) Sea $x = 1 + 2/x$, entonces

$$g'(x) = -\frac{2}{x^2}, \quad \frac{2}{x^2} < 1, \quad |x| \geq \pm\sqrt{2},$$

con lo que se cumple la condición en la derivada. Además, $g(x)$ es un función decreciente que para $g(2) = 2$, con lo que $g([2, \infty)) \subset [2, \infty)$, y como $g(-1) = -1$, $g((-\infty, -1]) \subset (-\infty, -1]$, con lo que en dichos intervalos este método converge a la raíz $\alpha = 2$. para sólo puede determinar la raíz 2. De hecho, experimentalmente se puede comprobar que este método converge a $x = 2$ para todo $x^{(0)} \in \mathbb{R}$.

(d) Finalmente, tomemos un método de Picard con un parámetro de relajación (sea $1/m$), es decir,

$$x = x + \frac{x^2 - x - 2}{m} \equiv g(x),$$

por lo que

$$g'(x) = 1 + \frac{2x - 1}{m} = \frac{m - 1}{m} + \frac{2x}{m},$$

y la condición en la derivada $|g'(x)| < 1$ nos da

$$-1 < \frac{m - 1}{m} + \frac{2x}{m} < 1,$$

que conduce a las dos condiciones

$$\begin{aligned} \frac{m + 1}{m} - 1 < \frac{2x}{m}, & \quad x > \frac{1}{2}, \\ \frac{2x}{m} < \frac{m + 1}{m} + 1 = \frac{2m + 1}{m}, & \quad x < m + \frac{1}{2}, \end{aligned}$$

y eligiendo $m > 3/2$ se puede demostrar que el método converge a la raíz $\alpha = 2$.

4. Sea la función $f(x) = x - 2 \sin x$, y hagamos

$$x = g(x) = 2 \sin x, \quad g'(x) = 2 \cos x,$$

cuya convergencia requiere

$$|g'| = 2 |\cos x| < 1, \quad \frac{\pi}{3} < x < \frac{2\pi}{3}.$$

Además, como $g(\pi/3) = \sqrt{3}$, $g(2\pi/3) = -\sqrt{3}$ y el máximo de $g(x)$ en dicho intervalo es $g(\pi/2) = 2$, tenemos que $g([\pi/3, 2\pi/3]) \in [\sqrt{3}, 2] \subseteq [\pi/3, 2\pi/3]$, y la función es contractiva, con lo que queda garantizada la convergencia.

7.3.6 Métodos de punto fijo más generales

En general, para obtener un método de punto fijo se escribe la ecuación $f(x) = 0$ de la forma $g_1(x) = g_2(x)$, con g un función fácilmente invertible. De esta forma se monta la iteración $g_1(x^{(n+1)}) = g_2(x^{(n)})$, que debe permitir calcular $x^{(n+1)}$ en función de $x^{(n)}$ fácilmente.

Si $h'_1(x)/h'_2(x)$ es continua y $|h'_1(\alpha)/h'_2(\alpha)| < 1$, siendo α la raíz, la iteración converge para $x^{(0)}$ suficientemente cerca de α , utilizando un teorema similar al utilizado en la sección anterior.

También se pueden desarrollar métodos iterativos de punto fijo de la forma general $x^{(k+1)} = g(x^{(k)}, x^{(k-1)}, \dots)$, que permitirá calcular la raíz de $f(x)$ si $x = g(x, x, \dots)$ implica que $f(x) = 0$. El estudio de estos métodos no requiere ninguna dificultad adicional, luego no los estudiaremos específicamente.

7.3.7 Orden de convergencia de un método de punto fijo

Se denomina orden de convergencia de un método iterativo de punto fijo al mayor número real q tal que el límite

$$\lim_{n \rightarrow \infty} \frac{|x^{(n+1)} - \alpha|}{|x^{(n)} - \alpha|^q},$$

existe y es diferente de cero. Es importante notar que no siempre existe dicho número, y cuando existe, no necesariamente es entero [1].

Un método de punto fijo $x^{(n+1)} = g(x^{(n)})$, con punto fijo $\alpha = g(\alpha)$, es de orden q si cumple que

$$\frac{d^j g}{dx^j}(\alpha) = 0, \quad j = 1, 2, \dots, q-1, \quad \frac{d^q g}{dx^q}(\alpha) \neq 0,$$

Demostración. Desarrollando en serie de Taylor,

$$x^{(n+1)} = g(x^{(n)}) = \alpha + g'(\alpha)(x^{(n)} - \alpha) + \dots + \frac{1}{q!} \frac{d^q g}{dx^q}(\xi)(x^{(n)} - \alpha)^q, \quad \xi \in (x^{(n)}, \alpha),$$

y aplicando la condición del enunciado,

$$x^{(n+1)} - \alpha = \frac{1}{q!} \frac{d^q g}{dx^q}(\xi)(x^{(n)} - \alpha)^q,$$

con lo que

$$\lim_{n \rightarrow \infty} \frac{|x^{(n+1)} - \alpha|}{|x^{(n)} - \alpha|^q} = \frac{1}{q!} \frac{d^q g}{dx^q}(\alpha) \neq 0,$$

ya que $x_i \rightarrow \alpha$ conforme n crece. Es decir, el método tiene orden de convergencia q , **cgqd**.

En la práctica, conviene utilizar métodos de convergencia supralineal, como el método de Newton que converge cuadráticamente. Los métodos numéricos de mayor orden se utilizan poco, fundamentalmente por su coste más alto, y sólo son más eficientes cuando se requieren soluciones de muy alta precisión.

7.3.8 Convergencia del método de Newton-Raphson

El método de Newton-Raphson es un método de punto fijo como los estudiados en la sección anterior,

$$x^{(k+1)} = x^{(k)} - \frac{f(x^{(k)})}{f'(x^{(k)})} \equiv g(x^{(k)}).$$

El punto fijo para este método es

$$\alpha = \alpha - \frac{f(\alpha)}{f'(\alpha)} \Rightarrow f(\alpha) = 0,$$

es decir, una raíz de $f(x)$. Además, en un entorno suficientemente pequeño del punto fijo α , se cumple que el método converge, ya que

$$g'(x) = 1 - \frac{f'}{f'} + \frac{f f''}{f'^2} = \frac{f f''}{f'^2},$$

implica que $g'(\alpha) = 0$, por lo que $|g'(x)| \ll 1$ en dicho entorno.

Además, el método de Newton tiene convergencia cuadrática, ya que

$$\begin{aligned} e^{(n+1)} &= \alpha - x^{(n+1)} = \alpha - g(x^{(n)}) \\ &= \alpha - \underbrace{(g(\alpha) + g'(\alpha)(x^{(n)} - \alpha))}_{=0} + \frac{g''(\alpha)}{2} (x^{(n)} - \alpha)^2 + O((x^{(n)} - \alpha)^3) \\ &= -\frac{g''(\xi)}{2} (e^{(n)})^2 = C (e^{(n)})^2, \end{aligned}$$

donde $x^{(n)} \leq \xi \leq \alpha$ y lo que indica que el método converge cuadráticamente con C como constante asintótica de error (es decir, su orden de convergencia es dos).

El método de Newton tiene una convergencia cuadrática siempre que $x^{(0)}$ esté suficientemente cerca de α . Si no es así, puede llegar a diverger. Normalmente se usan los algoritmos de bisección o regula falsi para acercarnos a la raíz lo suficiente y luego aplicamos Newton.

El método de Newton puede tener un problema cuando la función $f(x)$ tiene raíces múltiples. En dicho caso $f(\alpha) = f'(\alpha) = 0$ y la evaluación de la función $g(x)$ puede tener problemas para $x^{(n)}$ cerca de α . Sea una función con una raíz de multiplicidad p , entonces se puede escribir como

$$f(x) = (x - \alpha)^p F(x), \quad \mathbb{Z} \ni p > 1,$$

donde $F(\alpha) \neq 0$, y su derivada será

$$f'(x) = \frac{df}{dx} = (x - \alpha)^{p-1} (p F(x) + (x - \alpha) F'(x)) = (x - \alpha)^{p-1} G(x),$$

donde $G(\alpha) \neq 0$. Por tanto, al aplicar el método de Newton en aritmética exacta no hay problemas en la evaluación de g ya que

$$\lim_{x \rightarrow \alpha} \frac{f(x)}{f'(x)} = \lim_{x \rightarrow \alpha} (x - \alpha) \frac{F(x)}{G(x)} = \frac{F(\alpha)}{G(\alpha)} \lim_{x \rightarrow \alpha} (x - \alpha) = 0.$$

Sin embargo, en un ordenador con aritmética finita, la aparición en el denominador de un término pequeño puede llevar a una acumulación pernicioso de los errores de redondeo. Más aún, el resultado puede depender del orden con el que se realicen las operaciones.

7.3.9 Método de la secante

En el método de Newton es necesario evaluar la derivada $f'(x^{(n)})$. Para ello se puede utilizar la forma analítica (si se conoce) o se puede evaluar numéricamente

$$f'(x) = \frac{f(x + \epsilon) - f(x)}{\epsilon},$$

donde se puede tomar un número suficientemente pequeño, como un múltiplo del épsilon de la máquina. Sin embargo, en este caso, al dividir por un número pequeño un número que también es pequeño se pueden producir errores de redondeo bastante significativos, como estudiaremos en el tema 8 de derivación numérica.

Una posibilidad utilizada bastante es aproximar

$$f'(x^{(n)}) \approx \frac{f(x^{(n)}) - f(x^{(n-1)})}{x^{(n)} - x^{(n-1)}},$$

con lo que

$$x^{(n+1)} = x^{(n)} - \frac{f(x^{(n)})}{\frac{f(x^{(n)}) - f(x^{(n-1)})}{x^{(n)} - x^{(n-1)}}} = x^{(n)} - \frac{f(x^{(n)})}{f(x^{(n)}) - f(x^{(n-1)})} (x^{(n)} - x^{(n-1)}),$$

y hemos obtenido el método de la secante. Es importante notar que cuando $x^{(n)}$ se aproxime a α el método de la secante puede sufrir pérdidas de dígitos significativos debido a diferencias cancelativas en su evaluación.

El método de la secante se puede interpretar gráficamente como aparece en la figura 7.3.9. Se traza una recta secante a los puntos $(x^{(k)}, f(x^{(k)}))$ y $(x^{(k-1)}, f(x^{(k-1)}))$ y se toma como nueva iteración el punto de corte de esta recta con el eje de las abscisas.

Este método está relacionado con el método regula falsi, ya que se puede escribir

$$\frac{f(x^{(n)}) - f(x^{(n-1)})}{x^{(n)} - x^{(n-1)}} = \frac{-f(x^{(n)})}{x^{(n+1)} - x^{(n)}} = \frac{f(x^{(n+1)}) - f(x^{(n)})}{x^{(n+1)} - x^{(n)}},$$

que suponiendo $f(x^{(n+1)}) = 0$ y despejando $x^{(n+1)}$ en función de $x^{(n-1)}$ y $x^{(n)}$ nos da dicho método.

Para estudiar la convergencia del método de la secante [1] estudiaremos el comportamiento del error $e^{(n)} = x^{(n)} - \alpha$, donde α es el punto fijo. Este error toma la forma

$$e^{(n+1)} = x^{(n+1)} - \alpha = \frac{f(x^{(n)}) x^{(n-1)} - f(x^{(n-1)}) x^{(n)}}{f(x^{(n)}) - f(x^{(n-1)})} = \frac{f(x^{(n)}) e^{(n-1)} - f(x^{(n-1)}) e^{(n)}}{f(x^{(n)}) - f(x^{(n-1)})},$$

de donde podemos factorizar el término $e^{(n-1)} e^{(n)}$,

$$e^{(n+1)} = \frac{x^{(n)} - x^{(n-1)}}{f(x^{(n)}) - f(x^{(n-1)})} \frac{f(x^{(n)})/e^{(n)} - f(x^{(n-1)})/e^{(n-1)}}{x^{(n)} - x^{(n-1)}} e^{(n-1)} e^{(n)},$$

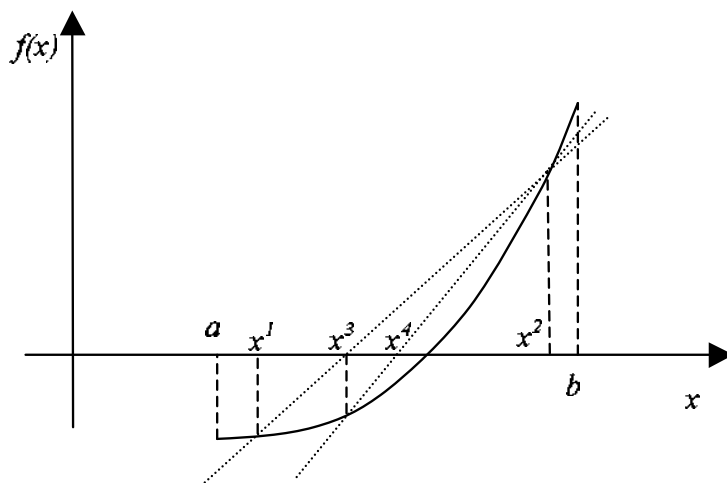


Figura 7.7. Proceso iterativo del método de la secante.

donde desarrollando por Taylor $f(x^{(n)}) = f(\alpha + e^{(n)})$, utilizando $f(\alpha) = 0$ y despreciando términos de segundo orden (o cuadráticos), es fácil obtener

$$\frac{x^{(n)} - x^{(n-1)}}{f(x^{(n)}) - f(x^{(n-1)})} \approx \frac{1}{f'(\alpha)},$$

y, con más cuidado,

$$\frac{f(x^{(n)})/e^{(n)} - f(x^{(n-1)})/e^{(n-1)}}{x^{(n)} - x^{(n-1)}} \approx \frac{1}{2} \frac{e^{(n)} - e^{(n-1)}}{x^{(n)} - x^{(n-1)}} f''(\alpha) = \frac{1}{2} f''(\alpha),$$

con lo que el error sigue la ecuación

$$e^{(n+1)} \approx \frac{f''(\alpha)}{2 f'(\alpha)} e^{(n-1)} e^{(n)} = C e^{(n-1)} e^{(n)},$$

donde la aproximación mejora conforme n crece.

Suponiendo ahora que p es el orden de convergencia del método de la secante, es decir,

$$|e^{(n+1)}| \approx A |e^{(n)}|^p, \quad |e^{(n-1)}| \approx \left(\frac{|e^{(n)}|}{A} \right)^{1/p},$$

y sustituyendo en la ecuación del error

$$A |e^{(n)}|^p \approx C A^{-1/p} |e^{(n)}|^{1+1/p},$$

que nos indica que

$$|e^{(n)}|^{1-p+1/p} \approx \text{const.},$$

con lo que $1 - p + 1/p = 0$, que nos da $p = (1 + \sqrt{5})/2 \approx 1.618$, que se denomina número de oro.

De esta forma, hemos demostrado que el método de la secante tiene un orden de convergencia no entero, $p \approx 1.62$, denominado superlineal (mejor que lineal), y que converge más rápidamente que los métodos de bisección y de regula falsi, de convergencia lineal, pero más lentamente que el método de Newton, de convergencia cuadrática.

7.3.10 Aproximación cuadrática, métodos de Newton y de la secante

Tanto el método de Newton como el método de la secante están relacionados con las técnicas de interpolación cuadrática (que estudiaremos en el siguiente tema). Para definir unívocamente una parábola es necesario especificar tres puntos (para una recta son suficientes dos). Sean los tres puntos $(\alpha, f(\alpha))$, $(\beta, f(\beta))$ y $(x^{(s)}, f(x^{(s)}))$. Sea $p(x)$ un polinomio de segundo orden (una parábola) que interpola dichos puntos, es decir,

$$p(\alpha) = f(\alpha), \quad p(\beta) = f(\beta), \quad p(x^{(s)}) = f(x^{(s)}).$$

Este polinomio se puede escribir

$$p(x) = f[\alpha] + f[\alpha, \beta](x - \alpha) + f[\alpha, \beta, x^{(s)}](x - \alpha)(x - \beta).$$

Esta manera de escribir un polinomio se denomina forma de Newton. Los coeficientes de la forma de Newton se denominan diferencias divididas, y los estudiaremos en más detalle en el próximo tema. Aplicando las condiciones requeridas (de interpolación) podemos obtener fácilmente dichos coeficientes. Para $x = \alpha$,

$$p(\alpha) = f[\alpha] = f(\alpha).$$

Para $x = \beta$,

$$p(\beta) = f(\alpha) + f[\alpha, \beta](\beta - \alpha) = f(\beta),$$

de donde

$$f[\alpha, \beta] = \frac{f(\beta) - f(\alpha)}{\beta - \alpha}.$$

Es obvio que $f[\beta, \alpha] = f[\alpha, \beta]$. Para $x = x^{(s)}$,

$$p(x^{(s)}) = f(x^{(s)}) = f[\alpha] + f[\alpha, \beta](x^{(s)} - \alpha) + f[\alpha, \beta, x^{(s)}](x^{(s)} - \alpha)(x^{(s)} - \beta),$$

de donde podemos escribir

$$f[\alpha, \beta, x^{(s)}](x^{(s)} - \alpha)(x^{(s)} - \beta) = f(x^{(s)}) - f(\alpha) - f[\alpha, \beta](x^{(s)} - \alpha),$$

de donde despejando el coeficiente $f[\alpha, \beta, x^{(s)}]$ obtenemos

$$\begin{aligned} f[\alpha, \beta, x^{(s)}] &= \frac{(f(x^{(s)}) - f(\alpha))(\beta - \alpha) - (f(\beta) - f(\alpha))(x^{(s)} - \alpha)}{(\beta - \alpha)(x^{(s)} - \alpha)(x^{(s)} - \beta)}. \\ &= \frac{\frac{f(x^{(s)}) - f(\alpha)}{x^{(s)} - \alpha} - \frac{f(\beta) - f(\alpha)}{\beta - \alpha}}{x^{(s)} - \beta} = \frac{f[\alpha, x^{(s)}] - f[\alpha, \beta]}{x^{(s)} - \beta}. \end{aligned}$$

En el próximo tema llamaremos a estos coeficientes diferencias divididas y estudiaremos sus propiedades con detalle.

Como hemos aproximado $f(x)$ por su interpolante cuadrático $p(x)$, es razonable pensar que un cero de $f(x)$ se puede aproximar por uno de $p(x)$.

Escojamos como $x^{(s)}$ un cero de $f(x)$, de tal forma que se cumple de forma exacta la expresión

$$f(x^{(s)}) = 0 = p(x^{(s)}) = f(\alpha) + f[\alpha, \beta](x^{(s)} - \alpha) + f[\alpha, \beta, x^{(s)}](x^{(s)} - \alpha)(x^{(s)} - \beta),$$

que podemos escribir como

$$x^{(s)} = \alpha - \frac{f(\alpha)}{f[\alpha, \beta]} - \frac{f[\alpha, \beta, x^{(s)}]}{f[\alpha, \beta]}(x^{(s)} - \alpha)(x^{(s)} - \beta),$$

A partir de dicha expresión podemos obtener el método de Newton si observamos que

$$f[\alpha, \alpha] = \lim_{\beta \rightarrow \alpha} f[\alpha, \beta] = \lim_{\beta \rightarrow \alpha} \frac{f(\beta) - f(\alpha)}{\beta - \alpha} = f'(\alpha),$$

con lo que tomando $\alpha = \beta = x^{(n)}$ en la expresión anterior, obtenemos

$$x^{(s)} = \underbrace{x^{(n)} - \frac{f(x^{(n)})}{f[x^{(n)}, x^{(n)}]}}_{x^{(n+1)} \text{ del método de Newton}} - \frac{f[x^{(n)}, x^{(n)}, x^{(s)}]}{f[x^{(n)}, x^{(n)}]}(x^{(s)} - x^{(n)})^2.$$

Esta expresión permite determinar de manera exacta el error cometido por el método de Newton al aproximar el cero por $x^{(n+1)}$ en lugar de $x^{(s)}$. Llamando a este error $e^{(n)} = x^{(s)} - x^{(n)}$, obtenemos la siguiente ecuación

$$e^{(n+1)} = -\frac{f[x^{(n)}, x^{(n)}, x^{(s)}]}{f[x^{(n)}, x^{(n)}]}(e^{(n)})^2 = C^{(n)}(e^{(n)})^2,$$

que nos indica que el método de Newton es de segundo orden de convergencia, como ya observamos en la sección anterior.

Para obtener el método de la secante, llamaremos $\alpha = x^{(n)}$ y $\beta = x^{(n-1)}$, con lo que obtenemos

$$x^{(s)} = \underbrace{x^{(n)} - \frac{f(x^{(n)})}{f[x^{(n)}, x^{(n-1)}]}}_{x^{(n+1)} \text{ del método de la secante}} - \frac{f[x^{(n)}, x^{(n-1)}, x^{(s)}]}{f[x^{(n)}, x^{(n-1)}]}(x^{(s)} - x^{(n)})(x^{(s)} - x^{(n-1)}),$$

que es una expresión exacta para el error de dicho método. Escribiendo como antes $e^{(n)} = x^{(s)} - x^{(n)}$, se tiene

$$e^{(n+1)} = -\frac{f[x^{(n-1)}, x^{(n)}, x^{(s)}]}{f[x^{(n-1)}, x^{(n)}]} e^{(n)} e^{(n-1)} = C_n e^{(n)} e^{(n-1)}.$$

Podemos utilizar esta expresión para determinar el orden de convergencia del método de la secante que llamaremos p . Como veremos dicho orden no será entero. Aplicando valores absolutos a la expresión anterior

$$\frac{|e^{(n+1)}|}{|e^{(n)}|^p} = C_n |e^{(n)}|^{1-p} |e^{(n-1)}| = C_n \left(\frac{|e^{(n)}|}{|e^{(n-1)}|^p} \right)^\gamma,$$

donde hemos hecho $\gamma = 1 - p$, es decir, $\gamma p = -1$. Podemos calcular p fácilmente como

$$(1 - p)p = -1, \quad p^2 - p - 1 = 0, \quad p = \frac{1 + \sqrt{5}}{2} \approx 1.618.$$

Note que la raíz negativa no tiene sentido en nuestro problema. La definición de C_n cuando n crece nos permite escribir, si $x^{(n)} \rightarrow x^{(s)}$,

$$\lim_{n \rightarrow \infty} C_n = -\lim_{n \rightarrow \infty} \frac{f[x^{(n-1)}, x^{(n)}, x^{(s)}]}{f[x^{(n-1)}, x^{(n)}]} = -\frac{f''(x^{(s)})}{2f'(x^{(s)})} \equiv C.$$

Para n suficientemente grande, $C_n \approx C$, y podemos estudiar la siguiente iteración (aproximada)

$$y^{(n+1)} = C (y^{(n)})^\gamma, \quad y^{(n+1)} \equiv \frac{|e^{(n+1)}|}{|e^{(n)}|^p},$$

cuyo punto fijo, sea α , cumple

$$\alpha = C \alpha^\gamma, \quad C = \alpha^{1-\gamma} = \alpha^p,$$

con lo que $y^{(n)} \approx C^{1/p}$ para n suficientemente grande, y por tanto, en dicho caso,

$$y^{(n+1)} = \frac{|e_{n+1}|}{|e_n|^p} = C y^{(n)} = C C^{\gamma/p} = C C^{(p+\gamma)/p} = C^{1/p} = \left| \frac{f''(x^{(s)})}{2f'(x^{(s)})} \right|^{1/p},$$

que es la constante de convergencia del método de la secante.

7.3.11 Método de Steffensen

Con el método de la secante hemos visto un cálculo aproximado de la derivada que aparece en el método de Newton reduce su orden de convergencia. ¿Se puede obtener un aproximación

de la derivada que preserve la convergencia cuadrática del método de Newton? El método de Steffensen nos da la respuesta y aproxima

$$f'(x) \approx \frac{f(x + f(x)) - f(x)}{f(x)} = \hat{f}(x),$$

conduciendo al método

$$x^{(n+1)} = x^{(n)} - \frac{f(x^{(n)})}{g(x^{(n)})},$$

que se puede probar fácilmente que es de convergencia cuadrática (como puede comprobar el lector). La idea de aproximar derivadas con una evaluación reiterada de $f(x)$ la volveremos a ver cuando estudiemos métodos de Runge-Kutta en el tema 9.

7.3.12 Iteración inversa y método de Newton

A la hora de calcular las raíces de $f(x) = 0$, podemos determinar su inversa, y evaluarla en el origen. Llamemos g a la inversa de f , es decir, si $y = f(x)$ entonces $x = g(y)$. Las raíces de $f(x) = 0$ son los puntos $\alpha = g(0)$.

Podemos definir una iteración basada en la inversa de la forma

$$x^{(n+1)} = g(y^{(n+1)}) = g(0) + \sum_{j=1}^{\infty} \frac{1}{j!} \frac{d^j g(0)}{dx^j} (y^{(n)})^j,$$

donde las derivadas se determinan fácilmente, por ejemplo, para la primera

$$g'(y) = \frac{dg}{dy} = \frac{dx}{dy} = \frac{1}{f'(x)},$$

y para la segunda,

$$g''(x) = \frac{dg'(y)}{dy} = \frac{d}{dy} \left(\frac{1}{f'(x)} \right) = \frac{dx}{dy} \frac{d}{dx} \left(\frac{1}{f'(x)} \right) = -\frac{f''(x)}{(f'(x))^3}.$$

Observando que $y^{(n)} = f(x^{(n)})$, obtenemos

$$x^{(n+1)} - \alpha = \sum_{j=1}^{\infty} \frac{1}{j!} \frac{d^j g(0)}{dx^j} (f(x^{(n)}))^j,$$

y como $f(x^{(n)}) - f(\alpha) = (x^{(n)} - \alpha) f'(\xi)$,

$$x^{(n+1)} - \alpha = \sum_{j=1}^{\infty} \frac{1}{j!} \frac{d^j g(0)}{dx^j} (x^{(n)} - \alpha)^j (f'(\xi))^j.$$

Para un método de orden de convergencia q , un número natural,

$$\frac{d^j g(0)}{dx^j} = 0, \quad j = 1, 2, \dots, q-1,$$

por lo que, aplicando el teorema del valor medio,

$$x^{(n+1)} - \alpha = \frac{1}{q!} \frac{d^q g(0)}{dx^q} (x^{(n)} - \alpha)^q (f'(\eta))^q,$$

es decir, podemos calcular fácilmente la constante de convergencia del método

$$\frac{|x^{(n+1)} - \alpha|}{|x^{(n)} - \alpha|} = \frac{1}{q!} \frac{d^q g(0)}{dx^q} (f'(\eta))^q = C.$$

Por ejemplo, para un método de orden $q = 2$,

$$\frac{|x^{(n+1)} - \alpha|}{|x^{(n)} - \alpha|} = C = -\frac{1}{2!} \frac{f''(x)}{(f'(x))^3} (f'(x))^2 = -\frac{f''(x)}{2 f'(x)},$$

que es la constante de convergencia del método de Newton.

7.3.13 Aceleración de Aitken

El procedimiento de la δ^2 de Aitken se puede utilizar para acelerar la convergencia de cualquier método iterativo, sea

$$x^{(n+1)} = g(x^{(n)}), \quad \alpha = g(\alpha).$$

Escribiendo el error de la forma

$$e^{(n+1)} = \alpha - x^{(n+1)} = \alpha - (g(\alpha) + g'(\xi)(x^{(n)} - \alpha)) = g'(\xi)(\alpha - x^{(n)}),$$

para $x^{(n)} < \xi < \alpha$, sin pérdida de generalidad, y despejando α de la forma

$$\alpha(1 - g'(\xi)) = x^{(n+1)} - g'(\xi)x^{(n)} = x^{(n+1)} - g'(\xi)x^{(n+1)} + g'(\xi)x^{(n+1)} - g'(\xi)x^{(n)},$$

$$\alpha(1 - g'(\xi)) = (1 - g'(\xi))x^{(n+1)} + g'(\xi)(x^{(n+1)} - x^{(n)}),$$

obtenemos una expresión exacta para el punto fijo

$$\alpha = x^{(n+1)} + \frac{x^{(n+1)} - x^{(n)}}{\frac{1}{g'(\xi)} - 1},$$

aunque en la práctica ignoramos ξ . Sin embargo, podemos hacer

$$\begin{aligned} r_n &\equiv \frac{x^{(n)} - x^{(n-1)}}{x^{(n+1)} - x^{(n)}} = \frac{x^{(n)} - x^{(n-1)}}{g(x^{(n)}) - g(x^{(n-1)})} \\ &= \frac{x^{(n)} - x^{(n-1)}}{g'(x^{(n-1)})(x^{(n)} - x^{(n-1)}) + O((x^{(n)} - x^{(n-1)})^2)} \\ &= \frac{x^{(n)} - x^{(n-1)}}{g'(\mu)(x^{(n)} - x^{(n-1)})}, \end{aligned}$$

donde $x^{(n-1)} < \mu < x^{(n)}$ y podemos aproximar

$$g'(\mu) = \frac{1}{r_n} = \frac{x^{(n+1)} - x^{(n)}}{x^{(n)} - x^{(n-1)}}.$$

Si ahora aproximamos ξ por μ , algo razonable para n suficientemente grande, podemos hacer

$$\begin{aligned} \alpha \approx \hat{x}^{(n+1)} &= x^{(n+1)} + \frac{x^{(n+1)} - x^{(n)}}{\frac{1}{g'(\mu)} - 1} = x^{(n+1)} + \frac{x^{(n+1)} - x^{(n)}}{\frac{x^{(n)} - x^{(n-1)}}{x^{(n+1)} - x^{(n)}} - 1} \\ &= x^{(n+1)} + \frac{(x^{(n+1)} - x^{(n)})^2}{2x^{(n)} - x^{(n+1)} - x^{(n-1)}}. \end{aligned}$$

Es habitual escribir la expresión anterior utilizando operadores en diferencias finitas, que estudiaremos con más detenimiento en el tema 8 sobre diferenciación numérica,

$$\Delta^+ x^{(n)} = x^{(n+1)} - x^{(n)}, \quad \Delta^- x^{(n)} = x^{(n)} - x^{(n-1)},$$

$$\Delta^2 x^{(n)} = \Delta^+ \Delta^- x^{(n)} = \Delta^+ (x^{(n)} - x^{(n-1)}) = x^{(n+1)} - x^{(n)} - (x^{(n)} - x^{(n-1)}),$$

con lo que podemos escribir finalmente

$$\hat{x}^{(n+1)} = x^{(n+1)} - \frac{(\Delta^+ x^{(n)})^2}{\Delta^2 x^{(n)}}.$$

La sucesión $\{\hat{x}_n\}$ converge más rápido que $\{x_n\}$.

7.4 Recomendaciones generales

Hemos presentado diferentes técnicas para la resolución de ecuaciones no lineales. El problema ya hemos indicado que es de solución general imposible, pero en la práctica, un conocimiento detallado de la función a resolver permite simplificar dicho problema y atacarlo con bastante seguridad de éxito.

Hay que tener en cuenta las siguientes consideraciones [2]:

Reescribir la ecuación. En muchas ocasiones se puede reescribir la ecuación de forma que sea más fácil de resolver. Veamos un ejemplo: la ecuación $f(x) = 0$ y $\exp(f(x)) - 1 = 0$ tienen exactamente las mismas soluciones, pero a veces, una de ellas es más fácil de resolver que la otra; para $f(x) = 6x - x^4 - 1$, la primera es más fácil de resolver, pero para $f(x) = \ln(6x - x^4)$ lo es la segunda. En general cada ecuación representa un miembro de una familia infinita de posibilidades. Si un método falla para una ecuación puede que con una reescritura adecuada,

funcione. Se requiere experiencia para conocer cuando un método un funcionará adecuadamente para una ecuación concreta.

Evitar errores de redondeo en la función. Se debe intentar reescribir la función de forma que su evaluación esté bien condicionada dentro del intervalo de interés donde vamos a buscar sus raíces. En caso contrario podemos encontrarnos con que los errores de redondeo no le permitirán converger al método numérico. Pongamos un ejemplo sencillo que aparece en el problema financiero de calcular las tasas de interés en inversiones. Consideremos el parámetro $p \geq 0$ y la ecuación

$$x - h(px) = 0,$$

donde $h(x) = (1 - \exp(-x))/x$, si $x \neq 0$, y $h(0) = 1$. La función $h(x)$ sólo tiene una raíz en el intervalo $0 < x \leq 1$.

7.5 Sistemas de ecuaciones no lineales

7.5.1 Método de Picard á la Gauss-Jacobi

En este tema consideraremos sólo el caso de que el número de ecuaciones no lineales y el número de incógnitas son iguales. Sea por ejemplo el sistema no lineal de dos ecuaciones con dos incógnitas

$$f(x, y) = 0, \quad g(x, y) = 0,$$

que podemos escribir

$$x = x + \mu f(x, y) = F(x, y; \mu),$$

$$y = y + \nu g(x, y) = G(x, y; \nu),$$

donde $\mu \neq 0$ y $\nu \neq 0$.

El método de iteración de Picard á la Gauss-Jacobi toma la siguiente forma

$$x^{(n+1)} = F(x^{(n)}, y^{(n)}; \mu), \quad y^{(n+1)} = G(x^{(n)}, y^{(n)}; \nu),$$

cuyos puntos fijos

$$\alpha = F(\alpha, \beta; \mu), \quad \beta = G(\alpha, \beta; \nu),$$

son soluciones del sistema original

$$f(\alpha, \beta) = 0 = g(\alpha, \beta).$$

El error en x , $e_x^{(n)} = x^{(n)} - \alpha$ sigue la ecuación

$$\begin{aligned} e_x^{(n+1)} &= x^{(n+1)} - \alpha = F(x^{(n)}, y^{(n)}; \mu) - \alpha \\ &= F(\alpha, \beta; \mu) + \frac{\partial F(\xi_1, y^{(n)})}{\partial x} (x^{(n)} - \alpha) + \frac{\partial F(\xi_2, x^{(n)})}{\partial y} (y^{(n)} - \beta) \\ &= \frac{\partial F(\xi_1, y^{(n)})}{\partial x} e_x^{(n)} + \frac{\partial F(\xi_2, x^{(n)})}{\partial y} e_y^{(n)}, \end{aligned}$$

donde $e_y^{(n)} = y^{(n)} - \beta$, $x^{(n)} \leq \xi_1 \leq \alpha$ y $y^{(n)} \leq \xi_2 \leq \beta$. Para el error en y se sigue una ecuación del todo similar y por tanto

$$e^{(n+1)} = \begin{pmatrix} e_x \\ e_y \end{pmatrix}^{(n+1)} = \begin{pmatrix} \frac{\partial F}{\partial x} & \frac{\partial F}{\partial y} \\ \frac{\partial G}{\partial x} & \frac{\partial G}{\partial y} \end{pmatrix}^{(n)} \begin{pmatrix} e_x \\ e_y \end{pmatrix}^{(n)} = J_n e^{(n)},$$

con lo que

$$e^{(n+1)} = J_n J_{n-1} \cdots J_0 e^{(0)},$$

donde J_i es el Jacobiano evaluado en $x^{(i)} \leq \xi_x^{(i)} \leq \alpha$ y $y^{(i)} \leq \xi_y^{(i)} \leq \beta$. Tomando $\|J\| = \max_i \|J_i\|$, y aplicando normas a la expresión anterior,

$$\|e^{(n)}\| \leq \|J\|^n \|e^{(0)}\|,$$

y una condición necesaria para la convergencia del método iterativo es que $\|J\| < 1$, que podemos garantizar que se cumple si $\rho(J) < 1$. Pero, como tenemos dos parámetros libres μ y ν , podemos elegirlos para que $\|J\|$ sea mínima y menor que la unidad; recuerde que

$$J = \begin{pmatrix} 1 + \mu \frac{\partial f}{\partial x} & \mu \frac{\partial f}{\partial y} \\ \nu \frac{\partial g}{\partial x} & 1 + \nu \frac{\partial g}{\partial y} \end{pmatrix}.$$

7.5.2 Método de Picard á la Gauss-Seidel

El método de iteración de Picard á la Gauss-Seidel para un sistema de dos ecuaciones con dos incógnitas toma la siguiente forma

$$x^{(n+1)} = F(x^{(n)}, y^{(n)}; \mu), \quad y^{(n+1)} = G(x^{(n)} + 1, y^{(n)}; \nu).$$

El error se comporta como

$$e_x^{(n+1)} = x^{(n+1)} - \alpha = F(x^{(n)}, y^{(n)}; \mu) - \alpha = \frac{\partial F}{\partial x} e_x^{(n)} + \frac{\partial F}{\partial y} e_y^{(n)},$$

$$e_y^{n+1} = y^{(n+1)} - \beta = G(x^{(n)}, y^{(n)}; \mu) - \beta = \frac{\partial G}{\partial x} e_x^{(n+1)} + \frac{\partial G}{\partial y} e_y^{(n)},$$

donde las derivadas parciales del Jacobiano se evalúan en puntos intermedios (teorema del resto de Taylor), por tanto, podemos escribir

$$\begin{aligned} \begin{pmatrix} 1 & 0 \\ -\frac{\partial G}{\partial x} & 1 \end{pmatrix}^{(n)} \begin{pmatrix} e_x \\ e_y \end{pmatrix}^{(n+1)} &= \begin{pmatrix} \frac{\partial F}{\partial x} & \frac{\partial F}{\partial y} \\ 0 & \frac{\partial G}{\partial y} \end{pmatrix}^{(n)} \begin{pmatrix} e_x \\ e_y \end{pmatrix}^{(n)} \\ \begin{pmatrix} e_x \\ e_y \end{pmatrix}^{(n+1)} &= \begin{pmatrix} 1 & 0 \\ \frac{\partial G}{\partial x} & 1 \end{pmatrix}^{(n)} \begin{pmatrix} \frac{\partial F}{\partial x} & \frac{\partial F}{\partial y} \\ 0 & \frac{\partial G}{\partial y} \end{pmatrix}^{(n)} \begin{pmatrix} e_x \\ e_y \end{pmatrix}^{(n)} \\ &= \begin{pmatrix} \frac{\partial F}{\partial x} & \frac{\partial F}{\partial y} \\ \frac{\partial F}{\partial x} \frac{\partial G}{\partial x} & \frac{\partial F}{\partial y} \frac{\partial G}{\partial x} + \frac{\partial G}{\partial y} \end{pmatrix}^{(n)} \begin{pmatrix} e_x \\ e_y \end{pmatrix}^{(n)} = S e^{(n)}, \end{aligned}$$

con lo que

$$\|e_n\| \leq \|S\|^n \|e_0\|,$$

y una condición necesaria para la convergencia es $\|S\| < 1$ (garantizado por $\rho(S) < 1$). Pero como tenemos dos parámetros libres μ y ν , debemos elegirlos para que $\|S\|$ sea mínima y menor que la unidad.

7.5.3 Método de Newton-Raphson y sus modificaciones

El método de Newton-Raphson también se puede generalizar fácilmente a sistemas de ecuaciones, basta tomar el desarrollo en serie de Taylor

$$f(x^{(k+1)}, y^{(k+1)}) = f(x^{(k)}, y^{(k)}) + \frac{\partial f}{\partial x} \Delta x^{(k)} + \frac{\partial f}{\partial y} \Delta y^{(k)},$$

$$g(x^{(k+1)}, y^{(k+1)}) = g(x^{(k)}, y^{(k)}) + \frac{\partial g}{\partial x} \Delta x^{(k)} + \frac{\partial g}{\partial y} \Delta y^{(k)},$$

donde $\Delta x^{(k)} = x^{(k+1)} - x^{(k)}$, $\Delta y^{(k)} = y^{(k+1)} - y^{(k)}$, y las derivadas parciales están evaluadas en puntos intermedios adecuadamente elegidos. Igualando estas expresiones a cero y escribiendo el sistema lineal resultante en forma matricial,

$$J^{(k)} \Delta x^{(k)} \equiv \begin{pmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \\ \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \end{pmatrix}^{(k)} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix}^{(k)} = - \begin{pmatrix} f \\ g \end{pmatrix}^{(k)} \equiv -F^{(k)},$$

con lo que

$$\Delta x^{(k)} = - \left(J^{(k)} \right)^{-1} F^{(k)}, \quad x^{(k+1)} = x^{(k)} + \Delta x^{(k)}.$$

Normalmente no se determina la inversa del Jacobiano sino que se resuelve el sistema lineal $J^{(k)} \Delta x^{(k)} = -F^{(k)}$ utilizando un método directo o iterativo.

Para cortar la iteración de Newton, es decir, para chequear la convergencia, se pueden utilizar cualquiera de las siguientes condiciones

1. $\|x^{(k+1)} - x^{(k)}\| \leq \epsilon_1$,
2. $\frac{\|x^{(k+1)} - x^{(k)}\|}{\|x^{(k+1)}\|} \leq \epsilon_2$,
3. $\|F(x^{(k)})\| \leq \epsilon_3$, ya que

$$\frac{\|F\|}{\|J\|} \leq \|\Delta x\| \leq \|J^{-1}\| \|F\|.$$

Existen muchas variantes del método de Newton que se basan en evaluar J sólo cuando sea estrictamente necesario y mantenerla constante si sus elementos no cambian mucho. Veamos algunas variantes

1. Método de Newton modificado

$$J \Delta x^{(k+1)} = -F(x^{(k)}),$$

donde J se mantiene fijo durante las iteraciones.

2. Método de Newton con relajación (*damped*)

$$J^{(k)} \Delta x^{(k+1)} = -\alpha F(x^{(k)}), \quad 0 < \alpha < 1,$$

donde α se selecciona para que

$$\|F(x^{(k+1)})\| \leq \|F(x^{(k)})\|,$$

es decir, para aumentar la velocidad de convergencia.

3. Método de Newton modificado con relajación (*damped*)

$$J \Delta x^{(k+1)} = -\alpha F(x^{(k)}).$$

Existen más variantes que no comentaremos, como cambiar en cada iteración sólo la parte diagonal, o tridiagonal, del jacobiano, utilizar su parte triangular inferior, o superior, etc.

El método de Newton-Raphson tiene como gran ventaja su convergencia cuadrática si $x^{(0)}$ está suficientemente próximo al punto fijo, aunque esta velocidad de convergencia se reduce cuando se utilizan sus modificaciones. Además, en muchos casos el jacobiano se puede evaluar analíticamente. Su desventaja es su mayor coste y que en una evaluación aproximada del jacobiano se pueden producir diferencias cancelativas y una pérdida significativa de precisión.

7.5.4 Métodos de continuación

Una manera de mejorar la velocidad de convergencia del método de Newton es utilizar técnicas de continuación o de inclusión (*imbedding*). Para resolver la ecuación $f(x) = 0$ la podemos reemplazar por la ecuación

$$g(x; t) = 0, \quad g(x; 1) = f(x), \quad g(x; 0) = 0,$$

elegida de tal forma que $g(x; 0) = 0$ sea más fácil de resolver que la ecuación original $f(x) = 0$. De esta forma calculamos la estimación inicial como la solución de

$$g(x^{(0)}; 0) = 0.$$

Seguidamente dividimos el intervalo $0 \leq t \leq 1$ en una partición $0 = t_0 < t_1 < \dots < t_N = 1$ y resolveremos las ecuaciones

$$g(x^{(i)}; t_i) = 0,$$

utilizando el método de Newton

$$g(x^{(i)}; t_i) = g(x^{(i-1)}; t_{i-1}) + \frac{\partial g(x^{(i-1)}; t_{i-1})}{\partial x} (x^{(i)} - x^{(i-1)}) + \frac{\partial g(x^{(i-1)}; t_{i-1})}{\partial t} (t_{i+1} - t_i) = 0,$$

obtenemos una ecuación para $x^{(i)}$ fácilmente resoluble.

Ejemplos de funciones simples de continuación son

$$g(x; t) = t f(x) + (1 - t) (f(x) - f(x^{(0)})),$$

$$g(x; t) = t f(x) + (1 - t) (x - x^{(0)}).$$

7.5.5 Optimización y descenso más rápido

El problema de determinar máximos (o mínimos) locales de una función $F(x)$, si esta es diferenciable, se reduce a calcular los ceros de $\nabla F = 0$. En este sentido este tipo de problemas están muy relacionados con la resolución de ecuaciones no lineales. Trataremos este tipo de problemas más en detalle en la segunda parte de la asignatura Métodos Matemáticos y Técnicas Computacionales de cuarto curso. Aquí nos limitaremos a una introducción breve al método de descenso más rápido.

Definamos

$$g(t) = F(x + tu), \quad g'(t) = (\nabla F)^\top u = \|\nabla F\|_2 \|u\|_2 \cos \theta,$$

donde el ángulo entre los vectores es

$$\theta = \arccos \frac{(\nabla F)^\top u}{\|\nabla F\|_2 \|u\|_2}.$$

Cuando $(\nabla F)^\top u > 0$, F es creciente y su crecimiento más rápido ocurre cuando u es paralelo a ∇F , el gradiente de F .

Con objeto de obtener el mínimo x^* , tal que $(\nabla F(x^*)) = 0$, podemos acercarnos al mínimo a partir de un valor x a lo largo de una línea recta en la dirección de $u = -\nabla F$. Definiendo

$$g(t) = F(x - t \nabla F),$$

para el mínimo tenemos $g'(t) = 0$.

El método del descenso más rápido parte de una aproximación inicial $x^{(0)}$ y define la siguiente iteración

$$x^{(k+1)} = x^{(k)} - t \nabla F(x^{(k)}).$$

Ejemplo: $F(x, y) = x^3 + y^3 - 2x^2 + 3y^2 - 8$, con lo que el punto estacionario es cero del Jacobiano,

$$\frac{\partial F}{\partial x} = 3x^2 - 4x = 0, \quad x = 0, \quad x = 4/3,$$

$$\frac{\partial F}{\partial y} = 3y^2 + 6y = 0, \quad y = 0, \quad y = -2.$$

Para saber si este punto estacionario es mínimo, máximo o punto de silla estudiamos el Hessiano,

$$\frac{\partial^2 F}{\partial x^2} = 6x - 4 = 0, \quad F_{xx}(0) = -4, \quad F_{xx}(4/3) = 4,$$

$$\frac{\partial^2 F}{\partial y^2} = 6y + 6 = 0, \quad F_{yy}(0) = 6, \quad F_{yy}(-2) = -6,$$

$$\frac{\partial^2 F}{\partial x \partial y} = 0.$$

Aplicando Taylor alrededor del punto estacionario

$$\begin{aligned} F(x, y) &= F(x^*, y^*) + \frac{\partial F}{\partial x} \Delta x + \frac{\partial F}{\partial y} \Delta y \\ &\quad + \frac{1}{2} \frac{\partial^2 F}{\partial x^2} \Delta x^2 + \frac{\partial^2 F}{\partial x \partial y} \Delta x \Delta y + \frac{1}{2} \frac{\partial^2 F}{\partial y^2} \Delta y^2 \\ &= F(x^*, y^*) + \frac{1}{2} \frac{\partial^2 F}{\partial x^2} \Delta x^2 + \frac{1}{2} \frac{\partial^2 F}{\partial y^2} \Delta y^2, \end{aligned}$$

tenemos un máximo en $(0, -2)$ y un mínimo en $(4/3, 0)$, y puntos de silla para $(0, 0)$ y $(4/3, -2)$.

Para determinar numéricamente el mínimo $(4/3, 0)$, partiríamos de una estimación inicial, p.ej. $x^{(0)} = (1, -1)$, y calculamos el gradiente $\nabla F(x^{(0)}) = (-1, -3)$, y estudiamos el mínimo de la función $g(t) = F(1+t, -1+3t)$, es decir, la raíz de $g'(t) = 0$.

De esta forma tomando la dirección de máximo descenso $u = -\nabla F$,

$$\frac{d}{dt} F(x + tu) = (\nabla F)^\top u = -(\nabla F)^\top \nabla F = 0$$

y podemos utilizar una iteración $x^{(n+1)} = x^{(n)} + t_n u_n$ haciendo

$$(\nabla F)_{n+1}^\top (\nabla F)_n = 0, \quad (\nabla F)_{n+1}^\top u_n < 0.$$

Para la función $F = \|f\|_2^2$, tenemos $\nabla F = 2(\nabla f)^\top f$, tomamos la iteración $x^{(n)} + 1 = x^{(n)} + h$, donde h es la dirección de máximo descenso,

$$(\nabla F)_{n+1}^\top h = 0, \quad 2(\nabla f)^\top f h < 0,$$

con lo que elegimos $h = -(\nabla f)^{-1} f$, con lo que

$$(\nabla F)_{n+1}^\top h = -2((\nabla f)^\top f)^\top (\nabla f)^{-1} f = -2f^\top f = -2\|f\|_2^2 < 0.$$

El procedimiento del descenso más rápido se consigue que $f(x^{(n)}) > f(x^{(n+1)})$. La iteración $x^{(n+1)} = x^{(n)} + h$ no es aceptable si $\|f_{n+1}\|_2 > \|f_n\|_2$, y en ese caso se suele hacer $x^{(n+1)} = x^{(n)} + h/2$.

7.6 Raíces de polinomios

Las funciones no lineales más simples son los polinomios

$$p(x) = a_0 + a_1 x + \cdots + a_n x^n.$$

El teorema fundamental del álgebra nos dice que un polinomio tiene n raíces complejas. Si los coeficientes a_i son reales, las raíces complejas aparecen en pares de complejos conjugados entre sí, sin embargo, ello no es cierto si los coeficientes son complejos.

7.6.1 Acotación de raíces. Reglas elementales y sucesiones de Sturm

Acotaciones de raíces de polinomios.

1. Regla de los signos de Descartes. Sea σ el número de variaciones de signos en la secuencia de coeficientes a_i y n_p el número de raíces positivas, entonces

$$n_p \leq \sigma,$$

y $\sigma - n_p$ siempre es un número entero par. La misma regla aplicada a $p(-x)$ permite determinar el número de raíces negativas.

Ejemplo: $p(x) = x^3 + 2x^2 - x - 1$, que tiene $\sigma = 1$ y $\sigma - n_p = 1 - n_p \in \{0, 2\}$, luego $n_p = 1$, es decir, al menos tiene una raíz positiva. Estudiando $p(-x) = -x^3 + 2x^2 + x - 1$, tenemos $\sigma = 2$, $\sigma - n_p = 2 - n_p \in \{0, 2\}$, y tenemos o ninguna raíz negativa o dos raíces negativas (de hecho, tiene dos raíces negativas pero el teorema no nos lo asegura).

2. Teorema Sean los polinomios

$$P(x) = |a_n| x^{(n)} - |a_{n-1}| x^{(n-1)} - \cdots - |a_0|,$$

$$Q(x) = |a_n| x^{(n)} + |a_{n-1}| x^{(n-1)} + \cdots + |a_1| x - |a_0|,$$

a los que podemos aplicar el teorema de Descartes, para obtener que P tiene una única raíz real positiva R y Q tiene una única raíz real positiva r . Entonces se verifica que todos los ceros de $p(x) = a_n x^{(n)} + a_{n-1} x^{(n-1)} + \cdots + a_0$ se encuentran en el intervalo

$$r \leq |x| \leq R.$$

3. Teorema Sea

$$r = \max_{0 \leq k \leq n-1} \left| \frac{a_k}{a_n} \right| + 1,$$

entonces todos los ceros de $p(x)$ se encuentran en $|x| \leq r$.

4. Teorema de Gershgorin Al menos una raíz del polinomio $p(x) = a_n x^{(n)} + a_{n-1} x^{(n-1)} + \dots + a_0$ se encuentra en el círculo centrado en el origen de radio $\min\{\rho_1, \rho_n\}$, donde

$$\rho_1 = n \frac{|a_0|}{|a_1|}, \quad \rho_n = \left(\frac{|a_0|}{|a_n|} \right)^{1/n}.$$

5. Sucesiones de Sturm El siguiente teorema debido a Sturm (1835) nos permite conocer el número de raíces positivas distintas que tiene un polinomio en un intervalo $[a, b]$ dado.

Teorema de Sturm. Existe un conjunto de polinomios con coeficientes reales $f_1(x), f_2(x), \dots, f_m(x)$ (una sucesión de Sturm), cuyos grados son de orden decreciente, tales que, si $b > a$, el número de raíces reales distintas de la ecuación

$$f(x) \equiv a_n x^{(n)} + \dots + a_1 x + a_0 = 0,$$

en el intervalo $[a, b]$ está dado por la diferencia $N = s(a) - s(b)$, donde $s(a)$ y $s(b)$ son el número de cambios de signo en la secuencia f, f_1, f_2, \dots, f_m en $x = a$ y en $x = b$, respectivamente.

Aplicando este teorema a $(-\infty, \infty)$ podemos determinar el número total de raíces reales de $f(x)$.

Existen varios procedimientos para construir una sucesión de Sturm. El más usado es el siguiente. Sea $f_1(x) = f(x)$ y $f_2(x) = f'(x)$. Hágase $f_3(x)$ igual al resto, con su signo cambiado, de la división $f_1(x)/f_2(x)$. Los términos siguientes de la sucesión se obtienen iterando este mismo procedimiento, es decir, f_{k+1} es el resto, cambiado de signo, de dividir f_{k-1} entre f_k .

Ejemplo. Sea el polinomio

$$f(x) = x^3 + 3x^2 - 1,$$

cuya sucesión de Sturm es

$$f_1 = f(x), \quad f_2 = 3x^2 + 6x, \quad f_3 = 2x + 1, \quad f_4 = 9/4.$$

La siguiente tabla muestra los signos de la sucesión de Sturm para diferentes valores de x .

	$f_1(x)$	$f_2(x)$	$f_3(x)$	$f_4(x)$	cambios signo
$x = -\infty$	-	+	-	+	3
$x = -3$	-	+	-	+	3
$x = -2$	+	0	-	+	2
$x = -1$	+	-	-	+	2
$x = 0$	-	0	+	+	1
$x = 1$	+	+	+	+	0
$x = \infty$	+	+	+	+	0

que indica que el polinomio tiene tres raíces reales positivas cada una en los siguientes intervalos

$$(-3, -2), \quad (-1, 0), \quad (0, 1).$$

7.6.2 Método de Horner para la evaluación de polinomios

Un polinomio se puede escribir de múltiples formas, de forma estándar o mediante raíces,

$$p_n(x) = a_0 + a_1 x + \cdots + a_n x^{(n)} = a_n (x - x^1) (x - x^2) \cdots (x - x^{(n)}),$$

o en forma de Taylor

$$p_n(x) = p_n(c) + p'_n(c) (x - c) + \cdots + \frac{p^{(n)}(c)}{n!} (x - c)^n,$$

o en forma de Newton

$$p_n(x) = a_0 + a_1 (x - c_1) + \cdots + a_n (x - c_1) (x - c_2) \cdots (x - c_n)^n.$$

Si se tiene un polinomio $p(x)$ y un número complejo z_0 , el algoritmo de Horner proporciona el número $p(z_0)$ y el polinomio $q(x) = \frac{p(x) - p(z_0)}{x - z_0}$. El grado del polinomio q es uno menos que el grado de p . Representándolo por $q(x) = b_{n-1}x^{(n-1)} + \cdots + b_1x + b_0$ y sustituyendo en la ecuación

$p(x) = (x - z_0)q(x) + p(z_0)$ se obtiene

$$\begin{aligned} b_{n-1} &= a_n \\ b_{n-2} &= a_{n-1} + z_0 b_{n-1} \\ &\vdots \\ b_0 &= a_1 + z_0 b_1 \\ p(z_0) &= a_0 + b_0 \end{aligned}$$

Si se observa además que $p'(z_0) = q(z_0)$ se puede implementar el algoritmo de Horner de forma que nos permita calcular simultáneamente $p(z_0)$ y $p'(z_0)$ lo que resulta muy apropiado cuando se utiliza el método de Newton para el cálculo de las raíces de un polinomio.

El algoritmo sería el siguiente:

$$\begin{aligned} b_n &= a_n \\ b_{n-1} &= b_n z_0 + a_{n-1} & c_{n-1} &= b_n \\ b_{n-2} &= b_{n-1} z_0 + a_{n-2} & c_{n-2} &= c_{n-1} z_0 + b_{n-1} \\ &\vdots \\ b_0 &= a_1 + z_0 b_1 & c_0 &= c_1 z_0 + b_1 \\ p(z_0) &= b_0 & p'(z_0) &= c_0 \end{aligned}$$

7.6.3 Método de Newton-Raphson para raíces reales

Podemos calcular las raíces de un polinomio $p(x) = 0$ mediante el método de Newton-Raphson,

$$p(x) = p(z) + p'(\xi)(x - z) = 0$$

aproximando $p'(\xi)$ por $p'(z)$, y despejando

$$x = z - \frac{p(z)}{p'(z)},$$

donde se aplicará la regla de Horner para evaluar el polinomio en el numerador y su derivada en el denominador, con objeto de reducir la propagación de errores.

Los problemas fundamentales de este método son

1. Es necesario calcular $p'(x)$ y esta derivada es susceptible de propagar errores de redondeo,
2. si $p(x)$ tiene alguna raíz doble, será $p'(x) = 0$ y el método no está bien definido. Una manera de evitar el problema de las raíces múltiples es aplicar el método de Newton a la función

$$u(x) = \frac{p(x)}{p'(x)}, \quad x^{(n+1)} = x^{(n)} - \frac{u(x^{(n)})}{u'(x^{(n)})},$$

donde

$$u' = \frac{p'^2 - p p''}{p'^2},$$

y por tanto

$$x^{(n+1)} = x^{(n)} - \frac{p p'}{p'^2 - p p''},$$

pero que tiene un orden de convergencia lineal solamente.

3. El cálculo de la raíces de un polinomio, por ejemplo, por el método de Newton, podría estar mal condicionado ante perturbaciones en los coeficientes. Por ejemplo,

$$x^2 - 1 = 0, \quad x = \pm 1,$$

que para una pequeña perturbación

$$x^2 + \epsilon x - 1 = 0, \quad x = \frac{-\epsilon}{2} \pm \sqrt{1 + \left(\frac{\epsilon}{2}\right)^2},$$

y desarrollando en serie de Taylor

$$x = \frac{-\epsilon}{2} \pm \left(1 + \frac{1}{2} \frac{\epsilon^2}{4} + O(\epsilon^4)\right) = \pm 1 - \frac{\epsilon}{2} + O(\epsilon^2)$$

En este caso el polinomio está bien condicionado en el sentido de Hadamard, ya que pequeños cambios en los coeficientes conducen a pequeños cambios en la solución. Pero en otros casos esto no es cierto.

4. ¿Cómo se determinan las demás raíces del polinomio? Se puede utilizar deflación, que consiste en calcular las raíces del polinomio de grado $n - 1$

$$p_{n-1}(x) = \frac{p_n(x)}{x - x^1},$$

donde x^1 es la raíz previamente calculada. El problema fundamental de la deflación es que la división puede causar graves errores de redondeo. Para minimizar estos errores es necesario calcular las raíces en orden creciente de magnitud, y deberíamos intentar calcular las raíces iterando con el polinomio original.

5. Encontrar una semilla o valor inicial en el dominio de atracción de cada raíz del polinomio no suele ser fácil.

7.6.4 Deflación (reducción) de Maehly

Sean z_1, z_2, \dots, z_k , k raíces del polinomio $p_n(x) = 0$, definimos el polinomio

$$\bar{p}(x) = \frac{p_n(x)}{(x - z_1) \cdots (x - z_k)},$$

al que podemos aplicar el método de Newton

$$x^{(n+1)} = x^{(n)} - \frac{\bar{p}(x^{(n)})}{\bar{p}'(x^{(n)})}.$$

Es usual reescribir esta expresión un poco. Tomemos logaritmos

$$\ln \bar{p} = \ln p_n - \sum_{i=1}^k \ln(x - z_i),$$

cuya derivada da

$$\frac{\bar{p}'}{\bar{p}} = \frac{p_n'}{p_n} - \sum_{i=1}^k \frac{1}{x - z_i},$$

que podemos escribir

$$\begin{aligned} \bar{p}' &= \frac{p_n' \bar{p}}{p_n} - \sum_{i=1}^k \frac{\bar{p}}{x - z_i} \\ &= \frac{1}{\prod_{i=1}^k (x - z_i)} \left(p_n' - \sum_{i=1}^k \frac{p_n}{x - z_i} \right), \end{aligned}$$

y la iteración de Newton es

$$\begin{aligned} x^{(n+1)} &= x^{(n)} - \frac{p_n}{\prod_{i=1}^k (x - z_i)} \frac{\prod_{i=1}^k (x - z_i)}{p_n' - \sum_{i=1}^k \frac{p_n}{x - z_i}} \\ &= x^{(n)} - \frac{p_n(x^{(n)})}{p_n'(x^{(n)}) - \sum_{i=1}^k \frac{p_n(x^{(n)})}{x^{(n)} - z_i}}. \end{aligned}$$

7.6.5 Método de Müller para raíces complejas

El método de Newton calcula las raíces de una en una y tiene problemas cuando se quieren determinar todas las raíces del polinomio. Sin embargo, existen métodos que permiten determinar todas las raíces y entre ellos se encuentran los métodos de los cocientes de diferencias, el método de Graeffe y el método de Müller.

El método de Müller permite encontrar tanto raíces reales como complejas, es iterativo, no requiere el cálculo de derivadas, permite encontrar tanto raíces simples como múltiples, tiene

convergencia casi cuadrática, no requiere una aproximación inicial a la raíz y está basado en el método de la secante.

Dados tres puntos, x^{i-2} , $x^{(i-1)}$ y $x^{(i)}$, primero se determina una parábola de interpolación que pase por los valores del polinomio en dichos puntos, es decir,

$$f(x^{i-2}) = p(x^{i-2}), \quad f(x^{(i-1)}) = p(x^{(i-1)}), \quad f(x^{(i)}) = p(x^{(i)}).$$

Para ello escribamos esta parábola como

$$p(x) = a_i + b_i x + c_i x^2 = f_i + f[x^{(i-1)}, x^{(i)}] (x - x^{(i)}) + f[x^{i-2}, x^{(i-1)}, x^{(i)}] (x - x^{(i)}) (x - x^{(i-1)}),$$

y tenemos fácilmente que

$$f_{i-1} = f_i + f[x^{(i-1)}, x^{(i)}] (x^{(i-1)} - x^{(i)}), \quad f[x^{(i-1)}, x^{(i)}] = \frac{f_{i-1} - f_i}{x^{(i-1)} - x^{(i)}},$$

y restando las ecuaciones

$$f_{i-1} = f_i + f[x^{(i-1)}, x^{(i)}] (x^{(i-1)} - x^{(i)}),$$

$$f_{i-2} = f_i + f[x^{(i-1)}, x^{(i)}] (x^{i-2} - x^{(i)}) + f[x^{i-2}, x^{(i-1)}, x^{(i)}] (x^{i-2} - x^{(i)}) (x^{i-2} - x^{(i-1)}),$$

obtenemos

$$\frac{f_{i-1} - f_{i-2}}{x^{(i-1)} - x^{i-2}} = f[x^{(i-1)}, x^{(i)}] - f[x^{i-2}, x^{(i-1)}, x^{(i)}] (x^{i-2} - x^{(i)}),$$

con lo que

$$f[x^{i-2}, x^{(i-1)}, x^{(i)}] = \frac{f[x^{(i-1)}, x^{(i)}] - f[x^{i-2}, x^{(i-1)}]}{x^{i-2} - x^{(i)}}.$$

Ahora podemos aproximar las raíces de $f(x) = 0$ por las de

$$\begin{aligned} p(x) = 0 &= f_i + f[x^{(i-1)}, x^{(i)}] (x - x^{(i)}) \\ &\quad + f[x^{i-2}, x^{(i-1)}, x^{(i)}] (x - x^{(i)}) (x - x^{(i)} + x^{(i)} - x^{(i-1)}) \\ &= f_i + \left(f[x^{i-2}, x^{(i-1)}, x^{(i)}] (x^{(i)} - x^{(i-1)}) + f[x^{(i-1)}, x^{(i)}] \right) (x - x^{(i)}) \\ &\quad + f[x^{i-2}, x^{(i-1)}, x^{(i)}] (x - x^{(i)})^2 \\ &= \alpha_i + \beta_i (x - x^{(i)}) + \gamma_i (x - x^{(i)})^2, \end{aligned}$$

cuyas raíces son

$$x - x^{(i)} = \frac{-\beta_i \pm \sqrt{\beta_i^2 - 4\alpha_i \gamma_i}}{2\gamma_i},$$

aunque para evaluarlas numéricamente evitando posibles diferencias cancelativas, se deberá utilizar el algoritmo desarrollo en el tema 2.

Como criterio de parada para este método iterativo se puede utilizar cualquiera de los siguientes criterios

$$|x^{(i)} - x^{(i-1)}| \leq \epsilon_1, \quad \frac{|x^{(i)} - x^{(i-1)}|}{|x^{(i)}|} \leq \epsilon_2, \quad |f(x^{(i)})| \leq \epsilon_3, \dots$$

Una vez que hayamos calculado dos raíces, aplicaremos el procedimiento de deflación y usaremos iteraremos sucesivamente el método de Müller. Normalmente se recomienda calcular las raíces en orden creciente de magnitud e iterar cada raíz obtenida mediante el método de Newton con el polinomio original para mejorar su precisión.

7.6.6 Método de Bairstow

Este método sirve para calcular las raíces complejas de un polinomio de coeficientes reales. Utiliza el hecho de que si $w = \alpha + i\beta$ es una raíz del polinomio $p(x) = a_n x^n + \dots + a_0$ de coeficientes reales entonces $\bar{w} = \alpha - i\beta$ es también raíz de p y además $(x - w)(x - \bar{w})$ es un factor cuadrático real de p .

La idea del método es utilizar un método de resolución de ecuaciones no lineales (por ejemplo el método de Newton) para encontrar los factores cuadráticos de p . Para ello se utiliza el siguiente teorema:

Si el polinomio $p(x) = a_n x^n + \dots + a_0$ se divide entre el polinomio cuadrático $x^2 - ux - v$ entonces el cociente y el residuo son:

$$q(x) = b_n x^{(n-2)} + b_{n-1} x^{(n-3)} + \dots + b_3 x + b_2$$

$$r(x) = b_1(x - u) + b_0$$

y pueden calcularse de forma recursiva haciendo $b_{n+1} = b_{n+2} = 0$ y utilizando la recursión

$$b_k = a_k + ub_{k+1} + vb_{k+2}, \quad k = n, n-1, \dots, 0$$

La demostración de este resultado es sencilla. Basta tener en cuenta la relación

$$p(x) = q(x)(x^2 - ux - v) + r(x)$$

En el caso que nos ocupa todos los coeficientes a_k de p son reales y también los coeficientes del factor cuadrático u y v . En el proceso de división b_0 y b_1 son funciones de u y v y para que

el factor cuadrático sea divisor de p se debe verificar $r(x) = 0$. Nos planteamos el problema: encontrar valores u y v tales que

$$b_0(u, v) = 0, \quad b_1(u, v) = 0$$

Si encontramos una raíz de esta ecuación se tendrá que $w = \alpha + i\beta$ y \bar{w} son raíces complejas conjugadas de p siendo

$$w + \bar{w} = u, \quad w\bar{w} = v$$

El método de Bairstow utiliza el método de Newton para resolver el sistema no lineal, aunque podría utilizar cualquier otro método de resolución de sistemas no lineales como Picard á lo Gauss-Seidel' o Picard á lo Jacobi'.

Si se utiliza el método de Newton es necesario utilizar las derivadas parciales:

$$c_k = \frac{\partial b_k}{\partial u}, \quad d_k = \frac{\partial b_{k-1}}{\partial v}, \quad k = 0, 1, \dots, n$$

Estas derivadas se pueden obtener derivando la relación de recurrencia que tenemos para b_k :

$$c_k = b_{k+1} + uc_{k+1} + vc_{k+2}, \quad (c_{n+1} = c_n = 0)$$

$$d_k = b_{k+1} + ud_{k+1} + vd_{k+2}, \quad (d_{n+1} = d_n = 0)$$

Como se obtiene la misma recurrencia para c_k y d_k nos quedamos con la primera. El proceso consistirá en asignar unos valores iniciales a u y v y buscar correcciones δu y δv de forma que se verifique:

$$b_0(u + \delta u, v + \delta v) = b_1(u + \delta u, v + \delta v) = 0$$

Se linealizan estas ecuaciones escribiendo

$$b_0(u, v) + \frac{\partial b_0}{\partial u} \delta u + \frac{\partial b_0}{\partial v} \delta v = 0$$

$$b_1(u, v) + \frac{\partial b_1}{\partial u} \delta u + \frac{\partial b_1}{\partial v} \delta v = 0$$

Teniendo en cuenta las relaciones previamente establecidas, el sistema se puede reescribir como:

$$\begin{bmatrix} c_0 & c_1 \\ c_1 & c_2 \end{bmatrix} \begin{bmatrix} \delta u \\ \delta v \end{bmatrix} = - \begin{bmatrix} b_0(u, v) \\ b_1(u, v) \end{bmatrix}$$

La solución de este sistema queda de la forma

$$J = c_0 c_2 - c_1^2$$

$$\delta u = (c_1 b_1 - c_2 b_0)/J$$

$$\delta v = (c_1 b_0 - c_0 b_1)/J$$

donde J es el determinante jacobiano del par de funciones no lineales $b_0(u, v)$ y $b_1(u, v)$. Además se puede demostrar (ver Kincaid) que J es distinto de 0 si u y v son tales que las raíces de $x^2 - ux - v$ son raíces simples de p .

7.6.7 Método de Graeffe

La idea del método de Graeffe es reemplazar la ecuación original por otra cuyas raíces sean los cuadrados de las del original, de tal forma que estas nuevas raíces estén bien separadas entre ellas.

Sea el polinomio

$$f_0(z) = a_n (z - \alpha_1) \cdots (z - \alpha_n),$$

donde supondremos que $a_n = 1$, entonces definimos el polinomio

$$\begin{aligned} f_1(z) &= (-1)^n f_0(z) f_0(-z) = (-1)^n (z - \alpha_1) \cdots (z - \alpha_n) (-z - \alpha_1) \cdots (-z - \alpha_n) \\ &= (z^2 - \alpha_1^2) \cdots (z^2 - \alpha_n^2) \\ &= (w - \alpha_1^2) \cdots (w - \alpha_n^2) = f_1(w), \end{aligned}$$

y repitiendo este procedimiento podemos llegar a obtener

$$f_{r+1}(w) = (-1)^n f_r(z) f_r(-z),$$

Los coeficientes del polinomio f_{r+1} se calculan

$$a_j^{(r+1)} = (-1)^{n-j} \left((a_j^{(r)})^2 + 2 \sum_{k=1}^{\min\{j, n-j\}} (-1)^k a_{j-k}^{(r)} a_{j+k}^{(r)} \right)$$

con lo que se obtiene

$$a_j^{(r)} = (-1)^{n-j} S_{n-j}(\alpha_1^{2r}, \dots, \alpha_n^{2r}),$$

y finalmente

$$a_{n-1}^{(r)} = -1 S_1(\alpha_1^{2r}, \dots, \alpha_n^{2r}) = - \sum_{k=1}^n \alpha_k^{2r},$$

y haciendo las raíces

$$\alpha_k = \rho_k e^{i\phi_k}, \quad \rho_1 > \rho_2 > \cdots > \rho_n,$$

reescribimos

$$a_{n-1}^{(r)} = -\alpha_1^{2^r} \left(1 + \sum_{k=2}^n \left(\frac{\alpha_k}{\alpha_1} \right)^{2^r} \right),$$

con lo que

$$\lim_{r \rightarrow \infty} |a_{n-1}^{(r)}|^{2^{-r}} = |\alpha_1|$$

luego la primera raíz se determina haciendo

$$\rho_1 \approx |a_{n-1}^{(r)}|^{2^{-r}},$$

y en general

$$\rho_k \approx \left| \frac{a_{n-k}^{(r)}}{a_{n-k+1}^{(r)}} \right|^{2^{-r}}.$$

Las desventajas de este método se dan cuando las raíces del polinomio original tienen magnitud parecida y porque hay que tener cuidado a la hora de determinar los ángulos ϕ_k .

Ejemplo. Sea el polinomio cuadrático

$$p(z) = (z - c_1)(z - c_2) = f_0(z),$$

definamos los polinomios

$$\begin{aligned} f_1(z) &= (-1)^2 (z^2 - c_1^2)(z^2 - c_2^2) = f_1(w) = (-1)^2 (w - c_1^2)(w - c_2^2) \\ &= w^2 - (c_1^2 + c_2^2)w + c_1^2 c_2^2, \end{aligned}$$

$$f_2(z) = f_2(q) = (-1)^2 (q - c_1^4)(q - c_2^4) = q^2 - (c_1^4 + c_2^4)q + c_1^4 c_2^4,$$

con lo que

$$\begin{aligned} a_1^{(r+1)} &= -(c_1^4 + c_2^4) = -c_1^4 - c_2^4 + 2c_1^2 c_2^2 - 2c_1^2 c_2^2 \\ &= -(c_1^2 + c_2^2)^2 + 2c_1^2 c_2^2 = a_1^{(r)2} + 2c_1^2 c_2^2 = -a_1^{(r)2} + 2a_0^{(r)}, \end{aligned}$$

con lo que obtenemos finalmente

$$a_1^{(r+1)} = -a_1^{(r)2} \left(1 - 2 \frac{a_0^{(r)}}{a_1^{(r)2}} \right).$$

7.6.8 Método de Bernoulli

El método de Bernoulli es bastante útil para determinar la raíz de mayor magnitud. Sea el polinomio

$$f(x) = a_0 + a_1 x + \cdots + a_n x^{(n)},$$

y escribamos el polinomio

$$a_n u_k + a_{n-1} u_{k-1} + \cdots + a_0 u_{k-n} = 0,$$

si ahora hacemos $u_k = \beta x^{(k)}$, tenemos

$$a_n x^{(k)} + a_{n-1} x^{k-1} + \cdots + a_0 x^{k-n} = 0,$$

y en el caso $k = n$ obtenemos el polinomio original.

Ahora asumamos que todas las raíces son diferentes

$$u_k = \sum_{i=1}^n c_i \alpha_i^k = c_1 \alpha_1^k \left(1 + \sum_{i=2}^n \frac{c_i}{c_1} \left(\frac{\alpha_i}{\alpha_1} \right)^k \right),$$

y si $c_1 \neq 0$,

$$\lim_{k \rightarrow \infty} \frac{u_k}{u_{k-1}} = \alpha_1 \in \mathbb{R},$$

donde c_i dependen de las condiciones iniciales

$$a_n u_m + a_{n-1} u_{m-1} + \cdots + m u_{m-n} = 0, \quad m = 1, 2, \dots, n.$$

Si las raíces son reales $\alpha_i \in \mathbb{R}$ y diferentes en magnitud el método es útil para calcular la raíz real más grande, o la raíz real más pequeña si se aplica a $f(1/z)$. Sin embargo, el método es de convergencia lenta si $\alpha_2 \approx \alpha_1$.

Ejemplo. Sea el polinomio cúbico

$$z^3 - 5z^2 - 17z + 21 = 0,$$

tomando

$$u_k - 5u_{k-1} - 17u_{k-2} + 21u_{k-3} = 0,$$

y el general

$$a_n u_m + a_{n-1} u_{m-1} + \cdots + m u_{m-n} = 0, \quad m = 1, 2, 3.$$

Para $m = 1$,

$$a_n u_1 + a_{n-1} m = 0, \quad u_1 - 5 = 0, \quad u_1 = 5.$$

Para $m = 2$,

$$a_n u_2 + a_{n-1} u_1 + a_{n-2} m = 0, \quad u_2 - 5 u_1 - 17 \cdot 2 = 0, \quad u_2 = 59.$$

Para $m = 3$,

$$a_n u_3 + a_{n-1} u_2 + a_{n-2} u_1 + a_{n-3} m = 0,$$

$$u_3 - 5 u_2 - 17 u_1 + 3 \cdot 21 = 0,$$

$$u_3 = 295 + 85 - 63 = 317.$$

Y para $m > 3$

$$u_k - 5 u_{k-1} - 17 u_{k-2} + 21 u_{k-3} = 0$$

nos permite obtener

$$u_k = 5 u_{k-1} + 17 u_{k-2} - 21 u_{k-3}, \quad k = 4, 5, \dots$$

Este método es de convergencia más lenta que el método de Graeffe, tiene convergencia lineal (que se puede acelerar mediante el proceso δ^2 de Aitken) y, finalmente, puede tener problemas cuando las raíces son mayores que la unidad, en cuyo caso los cocientes para calcular las raíces pueden ser muy grandes, aunque esto se puede evitar trabajando con u_{k-1}/u_k .

7.7 Aplicaciones a ecuaciones diferenciales ordinarias

Si queremos resolver una ecuación diferencial ordinaria

$$\frac{dy}{dt} = f(t, y),$$

mediante un método implícito, por ejemplo, el método de Euler hacia atrás

$$\frac{y^{n+1} - y^{(n)}}{\Delta t} = f_{n+1}(t, y) = f(t_{n+1}, y^{n+1}),$$

deberemos resolver una ecuación algebraica no lineal. Definiendo $z = y^{n+1}$, tenemos

$$F(z) = z - f(t_{n+1}, z) \Delta t - y^{(n)} = 0,$$

que, en general, es una ecuación no lineal que tiene soluciones analíticas sólo para algunas pocas ecuaciones diferenciales, por lo que habrá que utilizar un método numérico.

El problema fundamental de este método es que la ecuación diferencial puede tener soluciones que dejen de existir en un tiempo finito, es decir, que se vuelvan no acotadas en tiempo finito (*blow-up*). Veamos un ejemplo,

$$\frac{dy}{dt} = y^2, \quad y(0) = 1,$$

que se resuelve fácilmente integrando

$$-\frac{1}{y} = t + C, \quad y = -\frac{1}{C+t},$$

y aplicando la condición inicial

$$y(0) = 1 = -\frac{1}{C}, \quad C = -1,$$

con lo que obtenemos que la solución

$$y(t) = \frac{1}{1-t},$$

existe sólo para $t < 1$.

BIBLIOGRAFÍA

- [1] David Kincaid, Ward Cheney, “*Análisis Numérico. Las matemáticas del cálculo científico*,” Addison-Wesley Iberoamericana, 1994.
- [2] William M. Kahan, “Personal Calculator Has Key to Solve Any Equation $f(x) = 0$,” *Hewlett-Packard Journal*, December 1979, pp. 20-26.
- [3] William M. Kahan, “*Lecture Notes on Real Root-Finding*,” Math. 128A/B, 2002, available at <http://>.
- [4] J.M.A. Danby, “The Solution of Kepler’s Equation,” *Celestial Mechanics*, vol. 31 (1983), pp. 95-107, 317-328, y vol. 40 (1987), pp. 303-312.
- [5] Ramon E. Moore, “*Interval Analysis*,” Prentice-Hall, Englewood Cliffs, NJ, 1966.
- [6] Vladik Kreinovich, Daniel J. Berleant, Misha Koshelev, “*Interval Computation*,” web home-page: <http://www.cs.utep.edu/interval-comp/main.html>.