Métodos cerrados para la solución numérica de ecuaciones algebraicas y trascendentes

Cortés Rosas Jesús Javier, González Cárdenas Miguel Eduardo
Pinilla Morán Víctor Damián, Salazar Moreno Alfonso
Tovar Pérez Víctor Hugo *

2019

Resumen

Esta publicación pertenece al proyecto *Plataforma educativa para Análisis Numérico*, realizado con al apoyo del Programa UNAM-DGAPA-PAPIME PE105717.

Los métodos numéricos se encargan de obtener respuestas a problemas en donde la solución analítica es complicada. En este caso, se obtendrán raíces de ecuaciones algebraicas o trascendentes a partir del intervalo de la variable independiente que contiene a dicha raíz, de ahí el nombre de *métodos cerrados*. Específicamente, los métodos a mostrar son Bisección e Interporlación lineal[1].

Se les llama Métodos cerrados a todos aquellos que requieren de un intervalo de valores de la variable independiente $[a, b]$ para una función $f(x)$ que posee raíces reales, tal que $f(a)$ y $f(b)$ son de signos contrarios, por lo que se cumple que $f(a) \cdot f(b) < 0$. En lo general, este cambio de signos en la función valuada en a y b implica que en este intervalo existe al menos una raíz.

1. Método de bisección

El método de bisección se aplica a funciones algebraicas o trascendentes y proporciona únicamente raíces reales. Tiene su origen en un popular algoritmo de búsqueda de datos en arreglos vectoriales denominado *búsqueda binaria*. Es un método cerrado, es decir, requiere de un intervalo en el cual esté atrapada una raíz. Básicamente, consiste en cortar el intervalo en dos justo por la mitad (bisectar) considerando a este punto como una aproximación de la raíz de la función. Posteriormente, debe determinarse si la raíz verdadera se encuentra a la derecha o a la izquierda de la aproximación y, según corresponda, cerrar el intervalo con la aproximación y el límite derecho o izquierdo, pero siempre manteniendo a la raíz verdadera en el intervalo. Esta operación se repite hasta que la diferencia entre las dos últimas aproximaciones sea menor que una tolerancia preestablecida.

Bisección es un método robusto, aunque resulta lento en su proceso por lo oneroso de los cálculos que deben realizarse; por otra parte, su convergencia puede en ocasiones ser inestable.

*Profesores de la División de Ciencias Básicas de la Facultad de Ingeniería de la UNAM.

1.1. Definición del método

A partir de una función algebraica o trascendente y de un intervalo $[a, b]$ que pertenece al dominio de la función y para el cual $f(a) \cdot f(b) < 0$, lo que implica que en el intervalo $[a, b]$ existe al menos una raíz. El método consiste en bisectar el intervalo $[a, b]$:

$$x_0 = \frac{a + b}{2}$$

se obtiene una aproximación a la raíz x_0; la función se valúa en este nuevo valor y de acuerdo al signo de la función valuada en este punto, deberá sustituirse uno de los extremos del intervalo de búsqueda, de tal forma que se conserve que $f(a) \cdot f(b) < 0$. De acuerdo a la geometría de la figura, la sustitución de los intervalos deberá hacerse de la siguiente forma:

Sea a tal que $f(a) < 0$ y b tal que $f(b) > 0$:

- Si $f(x_0) < 0$, entonces x_0 sustituye a a
- Si $f(x_0) > 0$, entonces x_0 sustituye a b

En cada iteración deberá sustituirse alguno de los límites del intervalo que contiene a la raíz. Repitiendo este proceso, el intervalo se reduce paulatinamente hasta que alguna de las aproximaciones coincide razonablemente con la raíz de la función.

El proceso se detiene cuando entre la aproximación x_i y la aproximación anterior x_{i-1} se satisface un nivel de error (absoluto o relativo) preestablecido (tolerancia).

1.2. Interpretación geométrica

En la figura 1 puede observarse el intervalo $[a, b]$ en el cual está contenida una raíz de la función. Para este caso, se observa también que $f(a) < 0$ y que $f(b) > 0$ como consecuencia de la raíz contenida en el intervalo; este desarrollo es válido si se desea definir una función creciente y también es válido para una función decreciente haciendo los ajustes necesarios, pero en todo caso debe conservarse que $f(a) \cdot f(b) < 0$.

Para el caso mostrado, al bisectar el intervalo se observa que la primera aproximación x_0 se ubicó a la izquierda de la raíz y por consecuencia $f(x_0) < 0$; en virtud de esto, x_0 deberá sustituir al extremo del intervalo b, de acuerdo a la figura 2.

Por otra parte, al bisectar de nuevo el intervalo nos resulta que la siguiente aproximación x_1 se ubicó a la derecha de la raíz y por consecuencia $f(x_1) > 0$; en virtud de esto, x_1 deberá sustituir al extremo del intervalo b, conforme a la figura 3.

Una vez hecha esta sustitución, deberá bisectarse el nuevo intervalo hasta que dos aproximaciones sucesivas satisfagan la tolerancia preestablecida.

1.3. Criterio de convergencia

En todo caso, el método convergerá siempre y cuando en toda iteración se conserve: $f(a) \cdot f(b) < 0$.

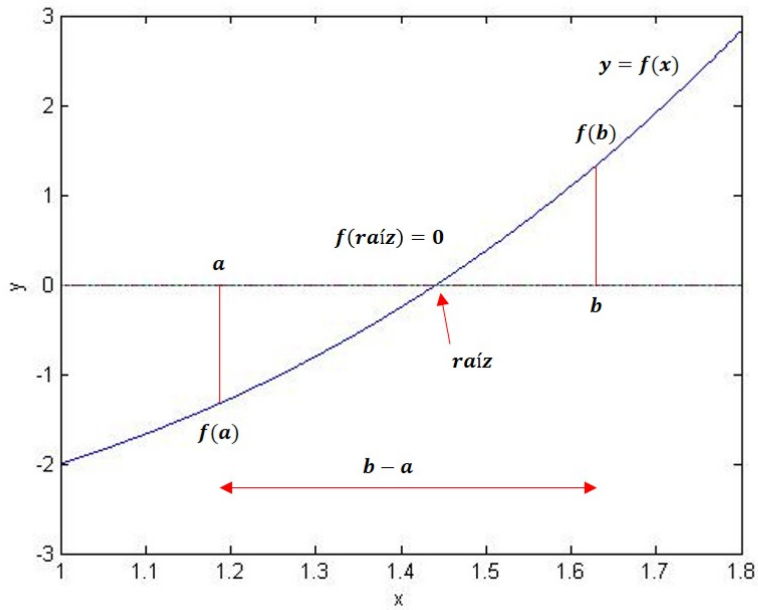

Figura 1: Intervalo $[a, b]$ que atrapa una raíz

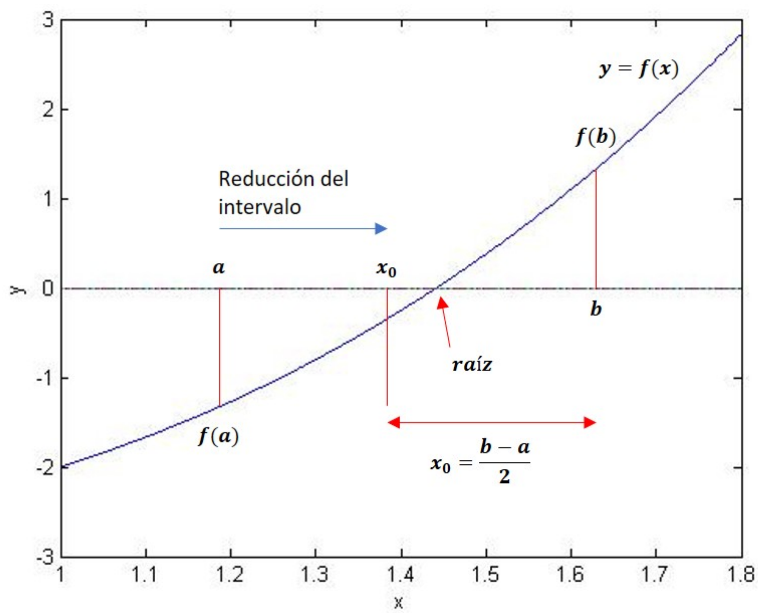

Figura 2: Intervalo reducido por la izquierda

1.4. Ejemplo de aplicación

Consideremos como ejemplo una función sencilla que nos permita verificar resultados fácilmente (Olivera Salazar, s.f.) (García B., 2017). Se propone $f(x) = x^2 - 0{,}5$. Se percibe que este polinomio de segundo grado representa a una parábola que abre hacia arriba; naturalmente, posee dos raíces

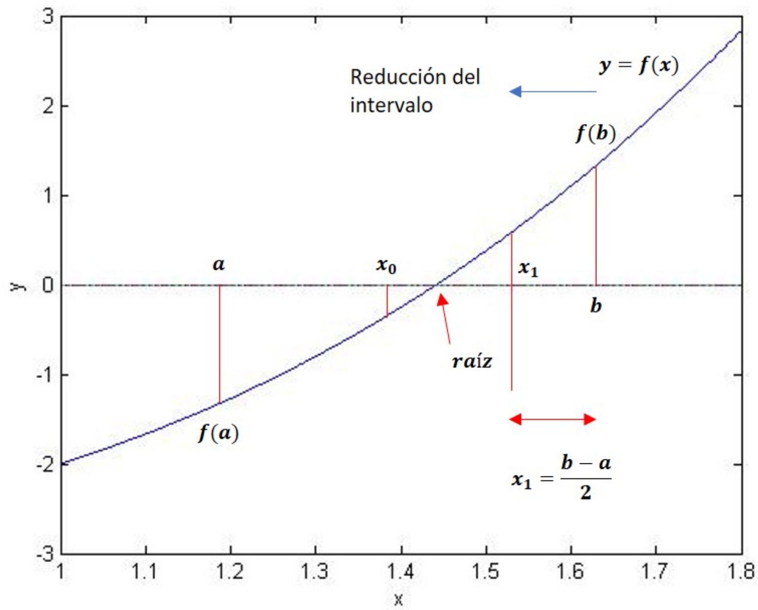

Figura 3: Intervalo reducido por la izquierda

cuyos valores son $\pm\sqrt{0{,}5}$.

Ahora bien, suponiendo desconocida esta información, se realizará la exploración de la función para encontrar sus raíces. El paso más recomendado es graficar la función.

A partir de la figura 4 se perciben los intervalos que atrapan a cada una de las raíces: $[-1, 0]$ y $[0, 1]$. Se propone obtener la raíz negativa y al mismo tiempo la obtención de la raíz positiva queda como ejercicio para practicar el método.

Siendo nuestro intervalo de búsqueda $[-1, 0]$ y la función valuada en sus extremos como $f(-1) = 0{,}5$ y $f(0) = -0{,}5$, se observa que se comprueba el criterio de convergencia $(f(a)\cdot f(b) < 0)$ y que $a = -1$ y $b = 0$, por lo tanto la primera aproximación a la raíz es:

$$x_0 = \frac{a+b}{2} = \frac{-1+0}{2} = -0{,}5$$

Valuando la función en esta nueva aproximación se obtiene:

$$f(x_0) = -0{,}25$$

Ya que la función valuada en la primera aproximación es de signo negativo, el intervalo se cerró por la derecha y corresponde sustituir el límite a por x_0, es decir: $b \leftarrow x_0$. El nuevo intervalo que atrapa la raíz y que cumple con el criterio de convergencia es $[-1, -0{,}5]$.

En una nueva iteración, el proceso se realiza de nuevo:

$$x_1 = \frac{(-1) + (-0{,}5)}{2} = -0{,}75$$

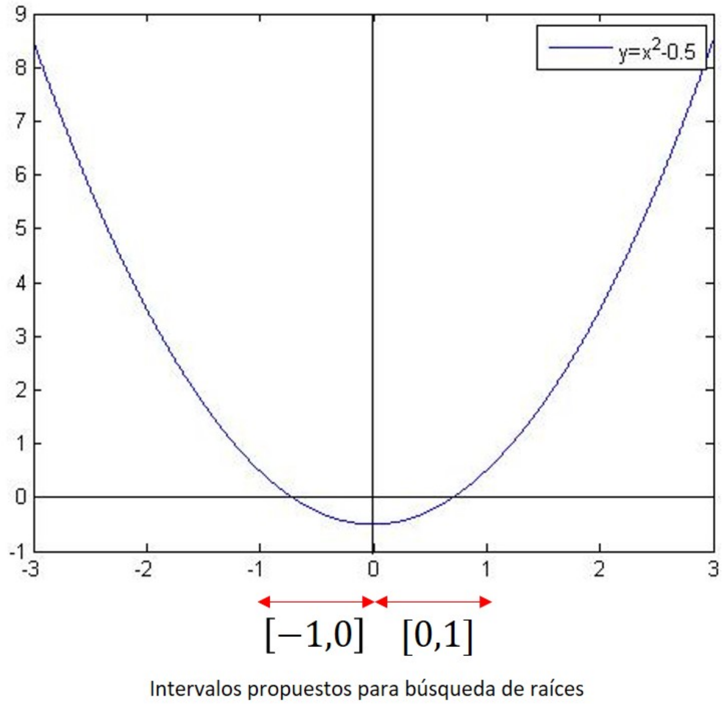

Figura 4: Intervalos iniciales de solución

$$f(x_1) = 0{,}0625$$

La función valuada resulta ahora de signo positivo, lo que obliga a que $a \leftarrow x_1$ y el nuevo intervalo que atrapa la raíz y que cumple con el criterio de convergencia es $[-0{,}75, -0{,}5]$. Asimismo, en esta nueva iteración se observa que la función valuada en ella tiende a cero, señal inequívoca que x_1 es una aproximación a la raíz.

Por otra parte, el error absoluto E entre las dos aproximaciones es:

$$E = \mid x_1 - x_0 \mid = \mid (-0{,}75) - (-0{,}5) \mid = 0{,}25$$

Cuando este error E cumpla con una tolerancia preestablecida, el método se detiene y la última x_i obtenida será considerada como la raíz de $f(x)$. En el cuadro 1 se muestra la evolución del método.

Podemos afirmar que con una tolerancia absoluta de $E = 0{,}00781$ la raíz de $f(x) = x^2 - 0{,}5$ es $x = -0{,}70703$ obtenida en ocho iteraciones.

Por motivos de espacio no se presenta una cantidad mayor de iteraciones; no obstante, repitiendo el método para una tolerancia absoluta de $E = 0{,}0001$ se obtiene que la raíz es $x = -0{,}70712$ en quince iteraciones.

1.5. Conclusiones

Como se mencionó en su oportunidad, bisección es un método robusto y muy fácil de programar como podrá constatarse en el algoritmo que acompaña a este texto. Su debilidad radica en la necesidad de

Cuadro 1: Solución al ejemplo

Iteraciones	0	1	2	3	4	5	6	7
a	-1	-1	$-0{,}75$	$-0{,}75$	$-0{,}75$	$-0{,}71875$	$-0{,}71875$	$-0{,}71094$
b	0	$-0{,}5$	$-0{,}5$	$-0{,}625$	$-0{,}6875$	$-0{,}6875$	$-0{,}70313$	$-0{,}70313$
$f(a)$	$0{,}5$	$0{,}5$	$0{,}625$	$0{,}625$	$0{,}625$	$0{,}01660$	$0{,}01660$	$0{,}00543$
$f(b)$	$-0{,}5$	$-0{,}25$	$-0{,}25$	$-0{,}1938$	$-0{,}02734$	$-0{,}02734$	$-0{,}00562$	$-0{,}00562$
Convergencia	Cumple	Cumple	Cumple	Cumple	Cumple	Cumple	Cumple	Cumple
x_i	$-0{,}5$	$-0{,}75$	$-0{,}625$	$-0{,}6875$	$-0{,}71875$	$-0{,}70313$	$-0{,}71094$	$-0{,}70703$
$f(x_i)$	$-0{,}25$	$0{,}625$	$-0{,}10938$	$-0{,}02734$	$0{,}01660$	$-0{,}00562$	$0{,}00543$	$-0{,}00011$
Error	-	$-0{,}5$	$-0{,}25$	$-0{,}125$	$0{,}0625$	$0{,}03125$	$0{,}01563$	$0{,}00781$

calcular continuamente los valores de la función para los diferentes valores de las aproximaciones.

Por otra parte, como puede verificarse en el cuadro 1, los valores de la función valuada en las diferentes $f(x_i)$ tiende a cero, pero no en forma directa, sino oscilando entre valores positivos y negativos. Este comportamiento indica que el método puede ser inestable pero convergente si se respeta el criterio que corresponde.

2. Método de la Interpolación lineal

Por su construcción geométrica, a este método también se le conoce como de las *cuerdas*. Una cuerda, geométricamente hablando, es el segmento de una recta que une dos puntos de un arco (internamente). Como se verá más adelante, la geometría que se forma en torno a esta cuerda, asociada a la búsqueda de la raíz de la función, permite establecer un método numérico considerado más eficiente que Bisección y que, con las debidas precauciones, suele ser más rápido. Se aplica a funciones algebraicas y trascendentes y proporciona sólo raíces reales.

Este método es de tipo cerrado, es decir, requiere de un intervalo $[a, b]$ que pertenece al dominio de la función y para el cual $f(a) \cdot f(b) < 0$, lo que implica que en el intervalo $[a, b]$ existe al menos una raíz.

2.1. Definición del método e interpretación geométrica

A partir de una $f(X)$ algebraica o trascendente, se requiere determinar el intervalo $[a, b]$ que contiene al menos una raíz, como lo muestra la figura 5.

El método propone unir los puntos $f(a)$ y $f(b)$ con una línea recta de tal forma que se construya la cuerda de la función 6. Esta línea recta deberá cortar al eje de las abscisas en un punto al que llamaremos x_0 porque será la primera aproximación a la raíz buscada.

El paso siguiente será determinar cuál de los extremos del intervalo $[a, b]$ es sustituido por x_0, de tal forma que se cumpla con el criterio de convergencia $f(a) \cdot f(b) < 0$. Para el caso de la geometría que se muestra en la figura 7, el extremo b es sustituido por x_0.

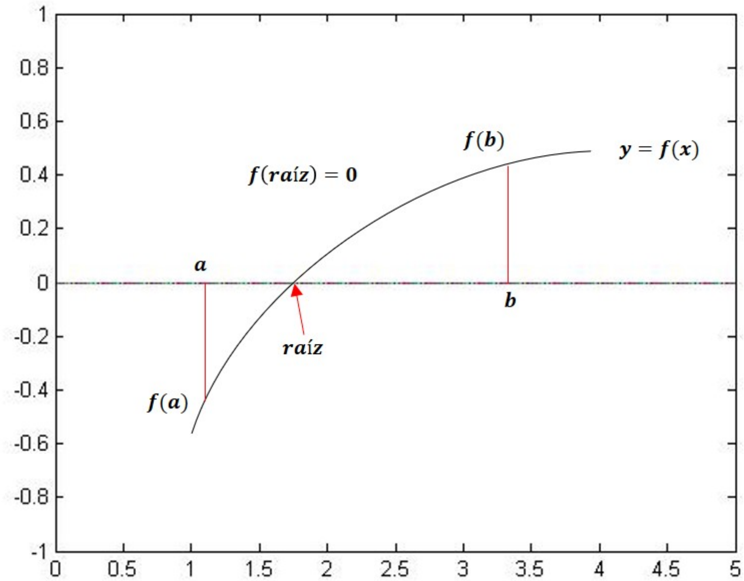

Figura 5: Intervalos [a,b] que atrapa una raíz

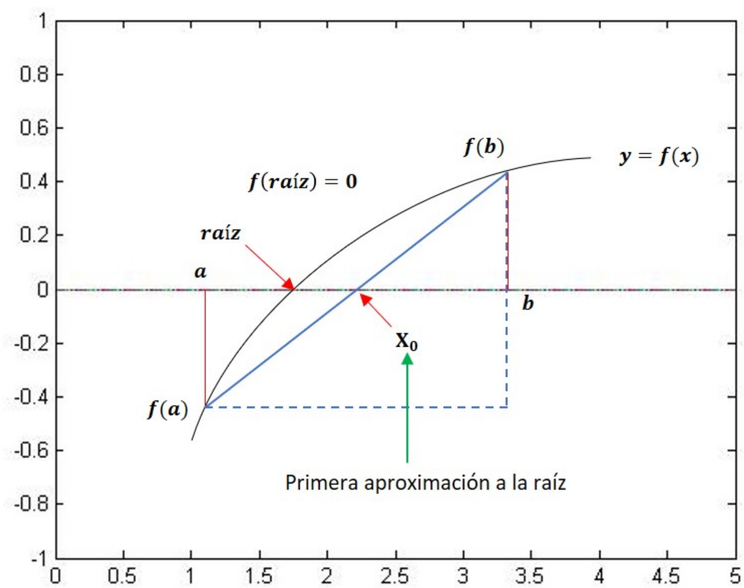

Figura 6: Establecimiento de la cuerda y obtención de la primera aproximación

Se traza otra vez una nueva cuerda, ahora entre $f(a)$ y $f(x_0)$ la cual nuevamente deberá cortar al eje horizontal más cerca de la raíz de la función. Vigilando el criterio de convergencia, este proceso se repetirá el número de veces que sea necesario hasta que la diferencia entre las aproximaciones $f(x_i)$ y $f(x_{i-1})$ sea menor que una tolerancia preestablecida (figura 8).

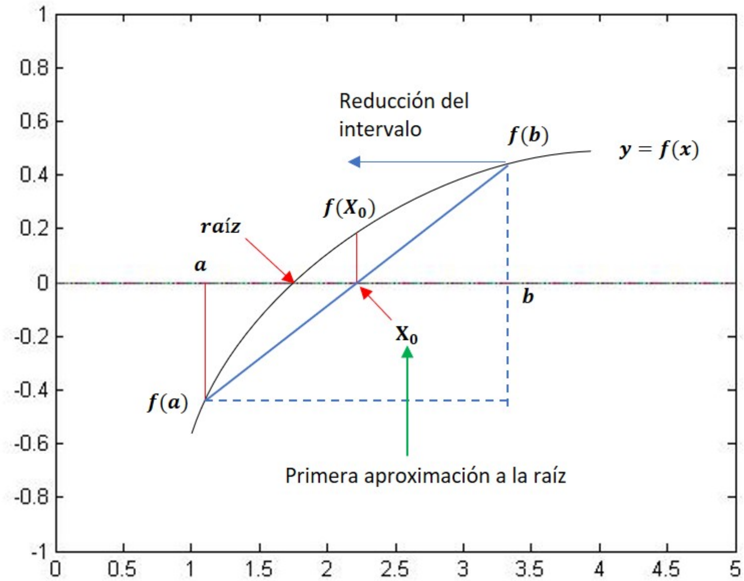

Figura 7: Reducción del intervalo que atrapa a la raíz

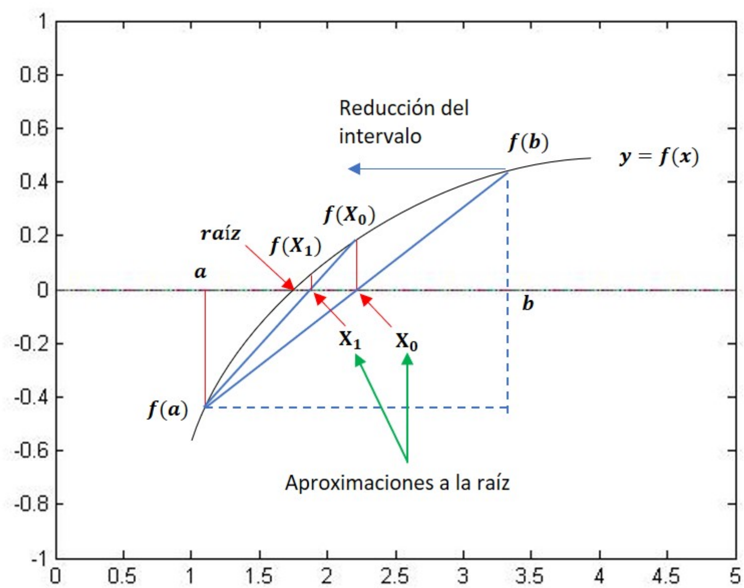

Figura 8: Obtención de las siguientes aproximaciones a la raíz

La fórmula de recurrencia [1] se obtiene de la geometría inicialmente planteada en la figura 6. Puede observarse en la figura 9 que con la recta de la primera iteración se forman dos triángulos semejantes que pueden modelarse con la siguiente expresión:

[1]Es la expresión mediante la cual se calculan las aproximaciones a la raíz

$$\frac{f(b)-f(a)}{b-a}=\frac{0-f(a)}{x_0-a}$$

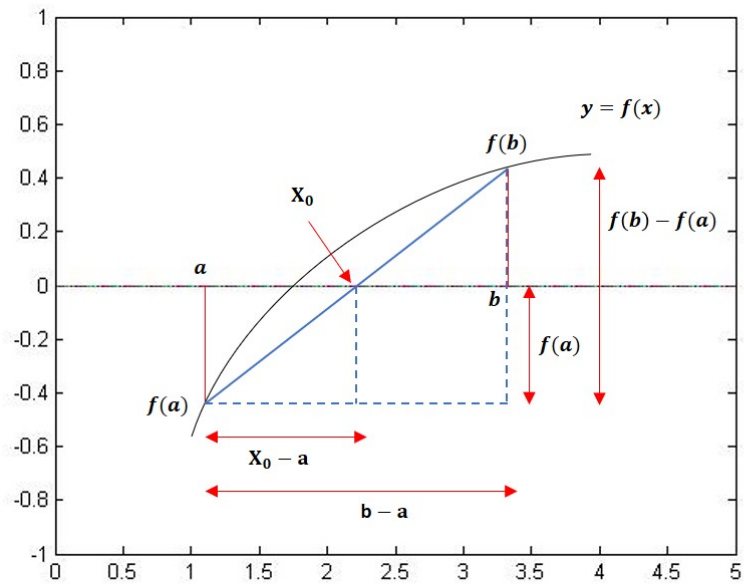

Figura 9: Triángulos semejantes a partir de la primera aproximación a la raíz

A partir de esta expresión, el método toma sus diversos nombres: Por una parte, se le denomina *Interpolación lineal* ya que por esta técnica (triángulos semejantes) se busca la raíz de la función; por otra parte, se observa claramente que el punto $f(x_0)$, que es el valor de la función en x_0 en su primera iteración no corresponde a verdadera raíz de $f(x)$ sino al cruce de la primera cuerda por el eje horizontal. Al hacer el supuesto de que en cada iteración $f(x) = f(x_0) = 0$ permite en forma iterativa aproximarse a la función. Este supuesto se denomina comúnmente *Regla falsa*[2]. Si aceptamos el concepto de la regla falsa, la incógnita de la ecuación es x_0. Despejándola se obtiene:

$$x_0 = a + \frac{(a-b)f(a)}{f(b)-f(a)}$$

Una vez obtenida la aproximación x_0 deberá evaluarse $f(x_0)$ y de acuerdo a la geometría utilizada en el desarrollo de la ecuación de recurrencia, sustituirse alguno de los extremos del intervalo $[a, b]$:

- Si $f(x_0) < 0$, entonces x_0 sustituye a a
- Si $f(x_0) > 0$, entonces x_0 sustituye a b

El proceso se detiene cuando entre la aproximación x_i y la aproximación anterior x_{i-1} se satisface un nivel de error (absoluto o relativo) preestablecido (tolerancia).

[2] Algunos autores utilizan el término original *Regula falsi*

2.2. Criterio de convergencia

Como se ha mencionado, en cada una de las iteraciones debe vigilarse el cumplimiento de la expresión $f(a) \cdot f(b) < 0$ lo que garantiza que en el intervalo $[a, b]$ siempre esté contenida una raíz.

2.3. Ejercicio de aplicación

Con el fin de comparar este método con el de Bisección, retomemos la sencilla función $f(x) = x^2 - 0{,}5$ y su gráfica en la figura 4. Calculemos ahora la raíz positiva que se encuentra en el intervalo $[0, 1]$, tal que $a = 0$ ya que $f(0) = -0{,}5$ y $b = 1$ ya que $f(1) = 0{,}5$, garantizando el cumplimiento del criterio de convergencia.

La primera aproximación es:

$$x_0 = 0 + \frac{(0-1)\cdot(-0{,}5)}{(0{,}5)-(-0{,}5)} = 0{,}5$$
$$f(x_0) = -0{,}25$$

Ya que $f(0{,}5)$ es negativo, se sustituye a a y el nuevo intervalo que atrapa a la raíz es $[0{,}5, 1]$, conservándose el criterio de equivalencia. La siguiente iteración sera:

$$x_1 = 0{,}5 + \frac{(0{,}5-1)\cdot(-0{,}25)}{(0{,}5)-(-0{,}25)} = 0{,}66667$$
$$f(x_1) = -0{,}05556$$

Por otra parte, el error absoluto E entre las dos aproximaciones es:

$$E = \mid x_1 - x_0 \mid = \mid (0{,}6667) - (0{,}5) \mid = 0{,}1667$$

Cuando este error E cumpla con una tolerancia preestablecida, el método se detiene y la última x_i obtenida será considerada como la raíz de $f(x)$. En el cuadro 2 se muestra la evolución del método.

Como conclusión podemos afirmar que con una tolerancia absoluta de $E = 0{,}00003$ la raíz de $f(x) = x^2 - 0{,}5$ es $x = 0{,}70710$ obtenida en 7 iteraciones.

De este ejercicio pueden obtenerse varias conclusiones previas:

- El método de Interpolación lineal alcanzó una tolerancia predefinida de una milésima en la mitad de iteraciones que bisección.
- En el desarrollo del método se pudo observar que un punto del intervalo $[a, b]$, en este caso $b = 1$ y en consecuencia $f(b) = 0{,}5$, se mantuvo fijo.

Esta última observación permite hacer una mejora al método con el fin de aumentar su velocidad. Cuando se detecta que uno de los extremos del intervalo permanece fijo (en forma empírica, después de al menos dos iteraciones en las cuales permanece inalterable), su correspondiente valor $f(x)$ se sustituye por $\frac{f(x)}{2}$. Geométricamente ocurre que la recta que se forma entre los extremos del intervalo se acerca más a la raíz disminuyendo el número de iteraciones necesarias para satisfacer la tolerancia preestablecida.

Cuadro 2: Solución al ejemplo

Iteración	0	1	2	3	4	5	6
a	0	0.5	0.6667	0.70000	0.70588	0.70690	0.70707
b	1	1	1	1	1	1	1
$f(a)$	-0.5	-0.25	-0.05556	-0.01000	-0.00173	-0.00030	-0.00005
$f(b)$	0.5	0.5	0.5	0.5	0.5	0.5	0.5
Criterio de convergencia	Cumple	Cumple	Cumple	Cumple	Cumple	Cumple	Cumple
Aproximación	0.5	0.66667	0.70000	0.70588	0.70690	0.70707	0.70710
$f(x_i)$	-0.25	-0.05556	-0.01000	-0.00173	-0.00030	-0.00005	-0.00001
Tolerancia	-	0.16667	0.03333	0.00588	0.00101	0.00017	0.00003

3. Pistas para detectar raíces

La inspección de las gráficas de las funciones es la mejor de las maneras de encontrar los intervalos que contienen raíces de una función (Chapra y Canale, 2015). No obstante, cuando se realizan subrutinas de cómputo con un método cerrado como parte de un sistema mayor, una gráfica no necesariamente es lo adecuado. El algoritmo deberá localizar en forma automática los extremos del intervalo $[a, b]$ que contiene a la raíz. Si el caso deberemos tomar en cuenta lo siguiente:

Obsérvese al figura 10:

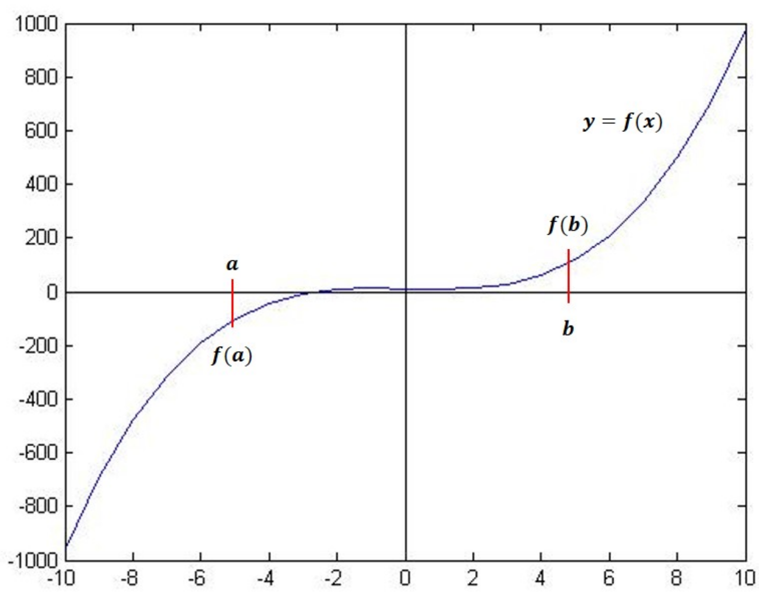

Figura 10: Intervalo que atrapa una raíz

De acuerdo a la metodología propuesta, existe una raíz en el intervalo señalado. No obstante, para la misma función se hace un acercamiento a la zona cercana a la intersección de los ejes (figura 11):

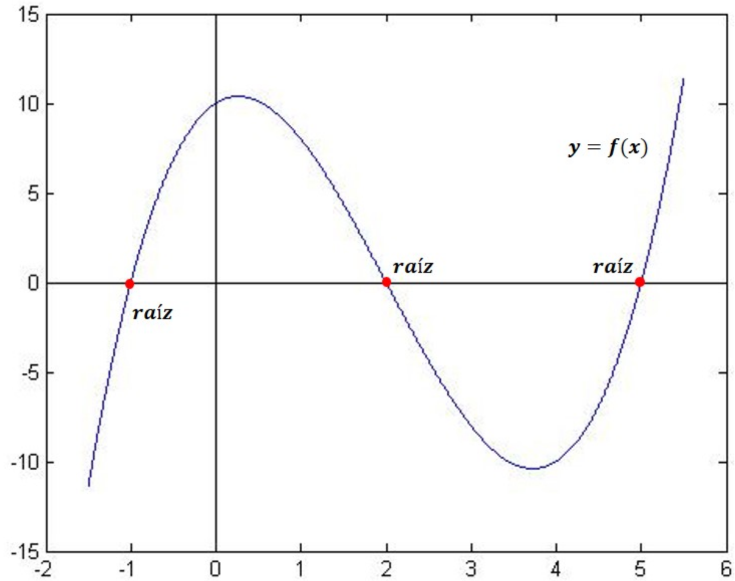

Figura 11: Total de raíces de la función

La función $f(x)$ tiene en realidad tres raíces y no una como originalmente se pensó. Puede hacerse una regla general para esta situación: en el intervalo $[a, b]$, cuando se detecta una raíz (el cambio de signo entre $f(a)$ y $f(b)$) puede existir un número *non* de raíces.

Por otra parte, en la figura 12:

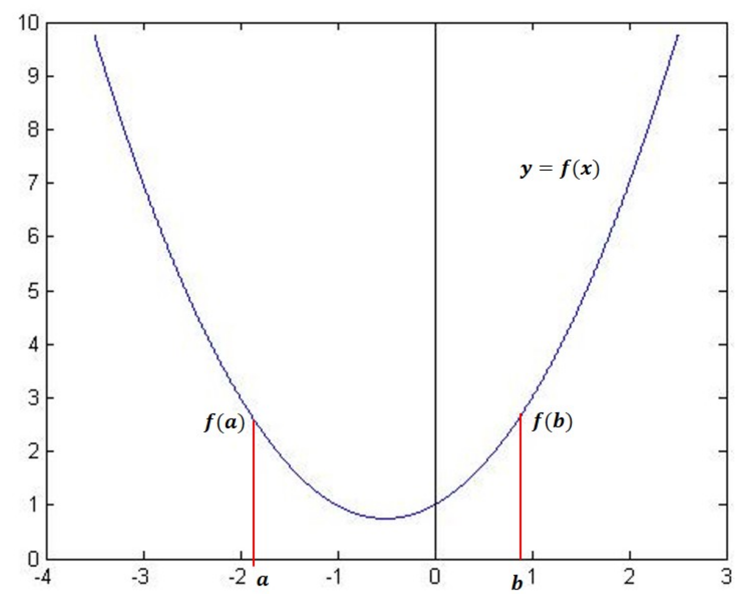

Figura 12: Función sin raíces reales

La función $f(x)$ no cruza el eje de las abscisas, por lo que no posee raíces reales; esto verifica el criterio de convergencia en el sentido que no hay cambio de signo entre $f(a)$ y $f(b)$. En la siguiente figura (13):

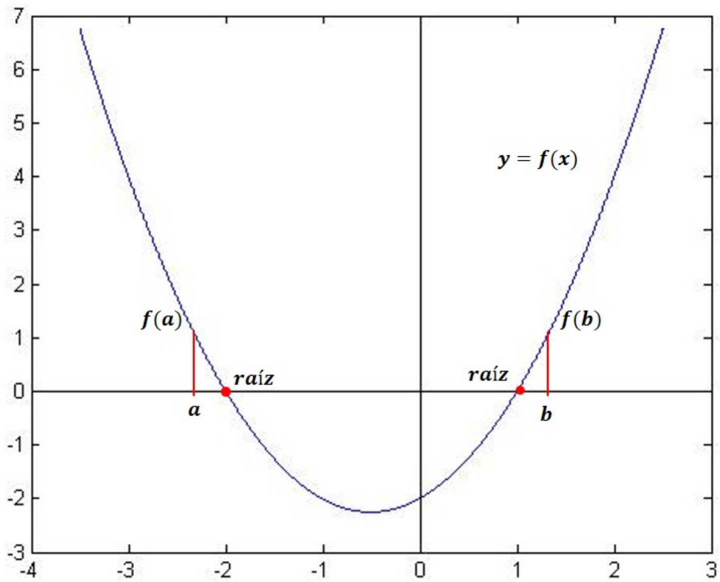

Figura 13: Rices pares en la función

No existe cambio de signo entre $f(a)$ y $f(b)$, por lo que en primera instancia no se cumple con el criterio de equivalencia. No obstante, la función tiene dos raíces. Puede hacerse una regla general para esta situación: en el intervalo $[a, b]$, cuando no existe un cambio de signo entre $f(a)$ y $f(b)$ puede no existir alguna raíz real o bien, un número *par* de raíces.

Por último, deberán contemplarse situaciones más específicas como raíces múltiples o funciones no continuas.

3.1. Conclusiones

El método de la Interpolación lineal permite alcanzar una aproximación a la raíz de una forma más rápida que Bisección. Sin embargo, presenta fallas cuando en el intervalo $[a, b]$ hay más de una raíz. Es quizá por este motivo por el cual se prefiere a Bisección pese a su lentitud (por robusto) que a este método más rápido, pero potencialmente inestable.

Notas

[1]Las figuras y gráficas incluidas en este trabajo fueron elaboradas por los autores

Referencias

Borras, H., Duran, R., y Iriarte, R. (1984). *Apuntes de métodos numéricos* (F. de Ingeniería UNAM, Ed.).

Burden, R., y Faires, D. (2011). *Análisis numérico* (C. Learning, Ed.).

Chapra, S., y Canale, R. (2015). *Métodos numéricos para ingenieros* (M. Hill, Ed.).

García B., S. (2017). *Métodos numéricos.*

Luthe, R., Olivera, A., y Schutz, F. (1985). *Métodos numéricos.*

Olivera Salazar, A. (s.f.). *Métodos numéricos* (Limusa, Ed.).

Sandoval, H. (2017). *Métodos numéricos.*