El método de Newton: de Newton a Kantorovich

por

José Antonio Ezquerro, José Manuel Gutiérrez, Miguel Ángel Hernández, Natalia Romero y María Jesús Rubio

La resolución de ecuaciones no lineales es uno de los problemas matemáticos que más frecuentemente aparecen en diversas disciplinas científicas. Así, con la notación

$$f(x) = 0 \tag{1}$$

englobamos el problema de encontrar una incógnita x, que puede ser un número real o complejo, un vector, una función, etc., a partir de los datos que nos proporciona la función f, que puede ser, por ejemplo, una función escalar, un sistema de ecuaciones, una ecuación diferencial, una ecuación integral, etc.

Incluso en el caso de que f sea una función real de variable real es bien conocido que, en general, no es posible resolver de forma exacta una ecuación no lineal. Es por ello que se recurre a técnicas iterativas para obtener aproximaciones de la solución.

Sin duda, dentro de estas técnicas iterativas, el método de Newton es el procedimiento más estudiado y empleado en la práctica. Así, con el objetivo de aproximar una solución α de una ecuación no lineal (1), el método de Newton consiste en construir, a partir de una aproximación inicial x_0 de α, una sucesión de la forma

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \quad n \geq 0. \tag{2}$$

En condiciones adecuadas, la sucesión anterior converge a la solución buscada α.

Es costumbre extendida entre los investigadores el bautizar sus descubrimientos con su propio nombre o con el de un personaje relevante en la materia. En este caso, el nombre de este método está ligado al eminente científico británico Isaac Newton. Sus trabajos de finales del siglo XVII parecen ser el germen del proceso que actualmente lleva su nombre. Sin embargo, como se muestra con más detalle en la sección 2 y en las referencias que allí se citan, el método (2) es fruto de las aportaciones de un gran número de científicos, tanto anteriores como posteriores a Newton. Otra muestra de la pluralidad de procedencias en las que se basa el método de Newton la constituye las diversas construcciones que admite. Algunas de ellas se muestran en la sección 1.

A mediados del siglo XX, el matemático soviético Leonid V. Kantorovich extendió el estudio del método de Newton a ecuaciones definidas en espacios de Banach. Así, combinando técnicas de análisis funcional y de análisis numérico, la teoría de Kantorovich permite abordar numerosos problemas no lineales tales como la resolución de ecuaciones integrales, ecuaciones diferenciales ordinarias y en derivadas parciales, o problemas de cálculo variacional, tal y como se detalla en las secciones 3 y 4.

En este contexto es donde se ha desarrollado el trabajo del grupo de investigación PRIENOL (Procesos Iterativos y Ecuaciones no Lineales) de la Universidad de La Rioja, cuyos componentes somos los autores de este artículo. En la última sección de este documento, presentamos algunas de las contribuciones que el grupo PRIENOL ha aportado a la teoría de Newton-Kantorovich.

1. Distintas formas de obtener el método de Newton

En lo que sigue mostramos brevemente algunas construcciones del método de Newton; aquéllas que a nuestro juicio nos parecen más relevantes. Las hemos seleccionado por el interés que tienen por sí mismas y porque, además, estos procedimientos nos permitirán, mediante ligeras variantes, construir otros procesos iterativos. En este listado no precisamos las condiciones que han de cumplirse para que estas construcciones puedan realizarse. Así, supondremos que las funciones que manejaremos son todo lo continuas y derivables que necesiten ser.

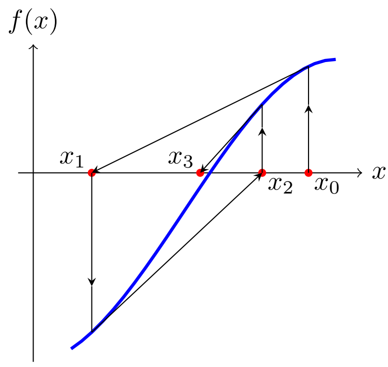

Figura 1: Interpretación geométrica del método de Newton.

Construcción geométrica. El método de Newton tiene una clara interpretación geométrica basada en la aproximación de la función $f(x)$ por una recta tangente (véase la figura 1). Por este motivo, dicho método suele denominarse también método de la tangente. Partiendo de una aproximación inicial x_0 de la solución α, la siguiente aproximación, x_1, se obtiene como la intersección de la recta tangente a la función $f(x)$ en el punto x_0 con el eje de abcisas. Así, x_1 es la intersección de las rectas $y-f(x_0) = f'(x_0)(x-x_0)$ e $y = 0$, es decir, $x_1 = x_0 - f(x_0)/f'(x_0)$. Reiterando este procedimiento se obtiene una sucesión de aproximaciones, que converge a la solución y que coincide con la obtenida mediante el método (2).

Construcción a partir del desarrollo de Taylor. A partir de x_0 buscamos una aproximación mejor de α que sea de la forma $x_0 + \varepsilon$. Lo ideal sería que $f(x_0 + \varepsilon) = 0$, en cuyo caso se obtendría la solución exacta. Usando el desarrollo de Taylor, llegamos a

$$0 = f(x_0 + \varepsilon) = f(x_0) + f'(x_0)\varepsilon + \frac{f''(x_0)}{2}\varepsilon^2 + \frac{f'''(x_0)}{6}\varepsilon^3 + \cdots \qquad (3)$$

Truncando (3) en el segundo sumando, es decir, linealizando la ecuación, se obtiene:

$$0 \simeq f(x_0) + f'(x_0)\varepsilon \;\Rightarrow\; \varepsilon \simeq -\frac{f(x_0)}{f'(x_0)}.$$

De aquí se deduce la siguiente aproximación de α: $x_0 - f(x_0)/f'(x_0)$. Reiterando el proceso obtenemos el método de Newton (2).

Construcción a partir de la interpolación racional inversa. El problema de aproximar una solución de (1) se puede transformar en el de aproximar $\phi(0)$, donde $x = \phi(y)$ denota la función inversa de $y = f(x)$. Supongamos que x_0 es una aproximación inicial de la solución α y sea $y_0 = f(x_0)$. Considerando el desarrollo de Taylor de orden uno de la función $\phi(y)$ en torno al punto y_0,

$$\phi(0) \simeq \phi(y_0) - \phi'(y_0)y_0 = x_0 - \frac{f(x_0)}{f'(x_0)},$$

se obtiene una nueva aproximación para $\alpha = \phi(0)$ que coincide, una vez más, con la expresión del método de Newton (2).

Construcción usando fórmulas de cuadratura. El método de Newton también se puede construir a partir de fórmulas de integración numérica de tipo interpolatorio. En concreto, si tenemos en cuenta la representación integral

$$f(x) = f(x_0) + \int_{x_0}^{x} f'(t)\,dt$$

y usamos la fórmula de cuadratura de Newton-Cotes de los rectángulos a izquierda,

$$\int_{x_0}^{x} f'(t)dt \simeq (x - x_0)f'(x_0),$$

llegamos a que la solución de la ecuación $f(x) = 0$ puede aproximarse por la solución de la ecuación $f(x_0) + (x - x_0)f'(x_n) = 0$, lo que conduce de nuevo al método (2).

El método de Newton continuo. El método de Newton (2) puede obtenerse también usando el método de Euler explícito para aproximar la solución de la ecuación diferencial

$$x'(t) = -\frac{f(x(t))}{f'(x(t))}, \quad t > 0. \tag{4}$$

En efecto, si se considera la condición inicial $x(t_0) = x_0$ y se toma como paso de integración $h = 1$, el método de Euler define una sucesión que coincide con la dada en (2). La ecuación (4) se conoce como método de Newton continuo y ofrece interesantes interpretaciones desde el punto de vista dinámico. En concreto, si $x(t)$ es una solución de (4) que cumple $x'(t) \to 0$ cuando $t \to \infty$ se tiene que $x(t)$ «fluye» a una solución de la ecuación (1).

2. Breve historia del método de Newton

La «paternidad» del método se atribuye a Isaac Newton, quien lo describe en varias de sus obras publicadas a finales del siglo XVII. Sin embargo, la idea de encontrar una cantidad desconocida mediante aproximaciones sucesivas se remonta a muchos siglos atrás. Así, en la Grecia clásica ya se utilizaban técnicas para aproximar números irracionales (sobre todo π) por números racionales. Pero incluso antes, dos mil años antes de Cristo, las civilizaciones mesopotámicas ya conocían técnicas para aproximar la raíz cuadrada de un número. Las referencias al respecto son abundantes. Por ejemplo, en [14, p. 42–43] se pone de manifiesto cómo en la tablilla YBC 7289 (véase la figura 2) de la *Yale Babylonian Collection* aparece un cuadrado de 30 unidades de lado en cuya diagonal están escritos los números $1;24,51,10$ y $42;25,35$[1].

La conversión al sistema decimal de la primera cantidad es 1,4142129629..., que coincide hasta la quinta cifra decimal con $\sqrt{2} = 1{,}4142135623\ldots$. Además, la segunda cantidad es el producto de 30 por la primera y es, por tanto, la longitud de la diagonal del cuadrado. Así pues, parece claro que los babilonios conocían un valor aproximado para $\sqrt{2}$ y que lo usaban para sus cálculos.

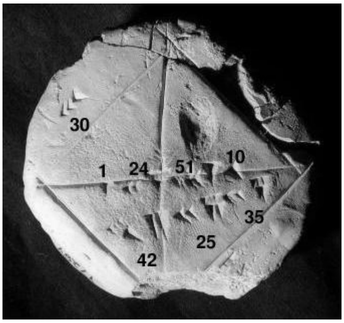

Figura 2: Tablilla YBC 7289 de la *Yale Babylonian Collection* (fotografía de Bill Casselman).

Otro indicio de que los babilonios sabían cómo aproximar cantidades irracionales aparece en la tablilla *VAT6598* que se conserva en el museo de Berlín y está fechada en 2000–1700 a.C. En ella se plantea, entre otros, el problema de encontrar la diagonal de un rectángulo de altura 40 y lado 10. Con la notación actual, el problema se traduciría en encontrar

$$\sqrt{40^2 + 10^2} = \sqrt{1700}.$$

En la misma tablilla se propone como aproximación el valor $41;15 = 41 + 15/60$. No se sabe cómo se obtuvo este valor, ni tampoco hay indicios de que se use un proceso iterativo, pero algunos autores [3] destacan el hecho de que esta cantidad coincida con la conocida aproximación para una raíz cuadrada

$$\sqrt{h^2 + l^2} \simeq h + \frac{l^2}{2h}$$

para $h = 40$ y $l = 10$.

La aproximación anterior se conoce como *fórmula de Herón* para el cálculo de raíces cuadradas, en la que, partiendo de una aproximación inicial a de $\sqrt{A}$, se propone como nueva aproximación $(a + A/a)/2$. En efecto, para $A = h^2 + l^2$ y $a = h$, la aproximación dada en la tablilla babilónica coincide con la aproximación de Herón. Aunque hay quien atribuye la fórmula de Herón al pitagórico Arquitas de Tarento

[1] Los babilonios usaban un sistema de numeración cuneiforme de base sexadecimal. En la actualidad, los expertos en el tema escriben los números babilónicos usando una mezcla de nuestra notación en base 10 y su notación en base 60. El equivalente babilónico de la coma decimal se denota con un punto y coma. El resto de los dígitos se separan por comas. Así, el número $5,51,13;2,30$ significa $5 \times 60^2 + 51 \times 60 + 13 + 2 \times 1/60 + 30 \times 1/60^2 \simeq 21073{,}0416$.


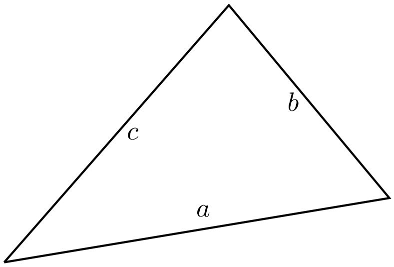

Figura 3: Herón de Alejandría (10–70 d.C. aproximadamente) y su fórmula para cálcular el área A de un triángulo, conocidos sus lados a, b y c: $A = \sqrt{s(s-a)(s-b)(s-c)}$, con $s = (a+b+c)/2$.

(428–365 a.C.) o incluso a Arquímedes (282–212 a.C.), donde el método aparece es en el primer tomo de la *Métrica* que Herón publicó en el siglo I. En este texto, descubierto por H. Schöne en 1896 (véase [3] para más detalles) se muestra cómo Herón calculaba el área de un triángulo de lados 7, 8 y 9 unidades, es decir $\sqrt{720}$. En el mismo, Herón se refiere explícitamente a que una aproximación dada puede ser tomada como punto de partida para obtener mejores aproximaciones. Parece claro, por tanto, que estamos ante la primera referencia de la utilización de un proceso iterativo para resolver un problema.

Ahora bien, ¿fue el método de Herón novedoso en su época? o ¿era una técnica ya conocida y que había sido empleada por civilizaciones anteriores? La respuesta queda en el aire, aunque la mayoría de los investigadores en esta parte de la Historia de las Matemáticas parecen inclinarse hacia la segunda opción, ya que hay constancia del uso de textos babilónicos por parte de matemáticos y astrónomos contemporáneos con Herón. Por ejemplo, Claudio Ptolomeo (100–170 d.C.) cita en su *Almagesto* datos astronómicos de la época del rey asirio Nabonassar (747 a.C.).

A partir de la fórmula de Herón, las técnicas para calcular la raíz cuadrada de un número (y, en general, las raíces n-ésimas) se fueron transmitiendo y/o redescubriendo a través de los siglos y de las civilizaciones hasta el siglo XVII. Aunque no hay muchas evidencias escritas de lo que ocurrió durante este largo periodo de tiempo, sí que, a modo de ejemplo, se pueden encontrar algunas referencias sobre métodos para el cálculo de raíces n-ésimas [3]. Podemos citar, por ejemplo, el libro chino de Matemáticas por excelencia, el *Jiuzhang suanshu*, que se traduce por *Nueve capítulos del arte matemático*. Se conoce una versión del siglo III, con comentarios de Liu Hui (220–280 d.C. aproximadamente), que contiene una colección de problemas que requieren el cálculo de raíces cuadradas y cúbicas. Posteriormente, en el siglo IV, Teón de Alejandría (335–405 d.C. aprox.), padre de Hypatia, desarrolló un método totalmente geométrico para el cálculo aproximado de raíces cuadradas. En los trabajos del matemático persa Sharaf al-Din al-Tusi (1135–1213) se encuentran las soluciones, tanto algebraicas como numéricas, de algunas ecuaciones cúbicas. Además, parece ser que al-Tusi fue el primero en calcular la derivada de un polinomio de tercer grado.

En el trabajo *Raf al-Hijab* del matemático árabe Al-Marrakushi ibn Al-Banna (1256–1321), que podríamos traducir por *Que levanta el velo*, se indica cómo calcular raíces cuadradas usando series y fracciones continuas. Parece ser que Al-Banna fue un gran recopilador de los conocimientos matemáticos de su época ,y que en sus escritos nos muestra su versión de los trabajos de matemáticos árabes anteriores.

El problema de encontrar la raíz n-ésima de un número fue evolucionando hacia el problema más general de encontrar las raíces de una ecuación polinómica e, incluso, de una ecuación trascendental (por ejemplo la ecuación de Kepler). A partir del siglo XV, el problema se fue bifurcando en varias líneas (resolución algebraica de ecuaciones polinómicas, resolución aproximada usando iteraciones de punto fijo, aproximaciones mediante fracciones continuas, etc.). Un análisis detallado del desarrollo histórico de estos problemas escapa de los objetivos de este trabajo, así que debemos remitir al lector interesado a alguno de los textos especializados, como [3] o [21].

Centrándonos en el «nacimiento» del método de Newton, podemos destacar el antecedente del matemático francés François Viète (1540–1603), quien desarrolló un ambicioso proyecto para buscar, de forma genérica, las soluciones positivas de ecuaciones polinómicas de grados 2 a 6. Viète fue el primero en representar los parámetros de una ecuación con letras, no sólo las incógnitas. El método empleado por Viète («logística especiosa» o «arte del cálculo sobre símbolos») estaba inspirado en la tradición geométrica griega. El método de Viète, escrito en un lenguaje arcaico y con notaciones engorrosas, no tuvo continuación y pronto pasó al ostracismo, apartado por la geometría cartesiana. Sin embargo, Viète fue el primero en darse cuenta de la relación existente entre las raíces y los coeficientes de un polinomio, e intentó usar el álgebra. Parece ser que el trabajo de Viète fue el que inspiró a Isaac Newton (1643–1727) para desarrollar su método para resolver ecuaciones. La primera referencia escrita sobre el método de Newton se encuentra en *De analysi per aequationes numero terminorum infinitas*, en una carta escrita a sus colegas Barrow y Collins en 1669, pero que no fue publicada hasta 1711. A los dos años de escribir la citada carta, en 1671, Newton desarrollaba su método en *De metodis fluxionum et serierum infinitarum*. De nuevo, esta obra tardó en publicarse y no es hasta 1736 cuando se publica una traducción de la misma bajo el título de *Method of Fluxions*.

Figura 4: Isaac Newton (1643–1727), el «padre» del conocido método para resolver ecuaciones no lineales que lleva su nombre.

Para hacernos una idea de cómo trabajaba Newton, podemos ilustrar su método con el mismo ejemplo que él consideró, la ecuación $x^3 - 2x - 5 = 0$. Newton argumentaba de la siguiente manera:

> Por tanteo, se ve que la solución está cerca de 2. Haciendo $x = 2 + \varepsilon$ y sustituyendo en la ecuación se obtiene:
>
> $$\varepsilon^3 + 6\varepsilon^2 + 10\varepsilon - 1 = 0. \tag{5}$$
>
> Ignorando los términos $\varepsilon^3 + 6\varepsilon^2$ con el pretexto de que ε es pequeño, se llega a que $10\varepsilon - 1 \simeq 0$ ó $\varepsilon = 0{,}1$. Entonces, $x = 2{,}1$ es una aproximación de la solución mejor que la inicial.
>
> Haciendo ahora $\varepsilon = 0{,}1 + \nu$ y sustituyendo en (5) se sigue que
>
> $$\nu^3 + 6{,}3\nu^2 + 11{,}23\nu + 0{,}061 = 0.$$
>
> Ignorando de nuevo los términos en ν de grado mayor o igual que dos, se llega a que $\nu \simeq -0{,}054$ y, por tanto, $x = 2{,}046$ es una aproximación que mejora las anteriores. Newton indicaba que el proceso se puede repetir las veces que sean necesarias.

Como vemos, la idea de Newton consiste en añadir un término corrector a una aproximación inicial dada. Para obtener esta aproximación, lo que hace es truncar el binomio de Newton en el segundo término, en expresiones del tipo

$$(a + \varepsilon)^n \simeq a^n + na^{n-1}\varepsilon.$$

De esta manera, para obtener el valor aproximado de ε, simplemente hay que resolver una ecuación lineal.

Escribiendo el problema con la notación actual y llamando $p(x) = x^3 - 2x - 5$, tenemos que la nueva aproximación es

$$2 - \frac{p(2)}{p'(2)} = 2 + \frac{1}{10} = 2{,}1,$$

que se corresponde con la conocida formulación del método de Newton (2) cuando $f(x)$ es el polinomio $p(x)$ anterior. Ahora bien, no se tiene constancia de que Newton usara el cálculo diferencial ni de que expresara el proceso como un método iterativo en el sentido de que una aproximación pueda ser considerada como punto de partida de la siguiente. Además, Newton usaba «su método» sólo para resolver ecuaciones polinómicas. Por lo tanto, la idea que Newton tenía de su método dista bastante de la que tenemos hoy en día.

La idea de iteración se atribuye (véanse, por ejemplo, [3] y [21]) a Joseph Raphson (1648–1715), quien además simplifica el aspecto operacional de la técnica de Newton. En 1690 publica un tratado, *Analysis Aequationum Universalis*, en el que se dan fórmulas explícitas para el término corrector para algunos casos particulares de ecuaciones. En concreto, calcula los términos correctores para las ecuaciones $x^3 - r = 0$ y $x^3 - px - q = 0$ que son, respectivamente,

$$\frac{r - x_0^3}{3x_0^2} \quad \text{y} \quad \frac{q + px_0 - x_0^3}{3x_0^2 - p},$$

siendo x_0 la aproximación inicial. Notemos que Raphson publicó su obra 46 años antes que el «Método de las fluxiones» de Newton. No obstante, el propio Raphson es el primero en reconocer que el método de Newton ya era conocido en los ambientes científicos de la época y que su método era una versión, mejorada, eso sí.

La contribución de Raphson ha sido tenida en cuenta históricamente, no en vano muchos autores denominan el proceso como método de Newton-Raphson. Sin embargo, en los trabajos de Raphson no se aprecia la conexión existente entre el término corrector, la función que define la ecuación, y su derivada.

La incorporación del cálculo diferencial se debe a Thomas Simpson (1710–1761). Como se puede ver en [21], Simpson, en su obra *Essays on Mathematics*, publicada en 1740, fue quien estableció el método tal y como lo conocemos actualmente, salvo aspectos notacionales (Simpson explicaba de forma retórica cómo obtener las aproximaciones sucesivas). Además, Simpson extendió el proceso a funciones cualesquiera, no solamente polinomios.

Con motivo de ciertas observaciones a propósito de la utilización de series infinitas, Newton parece estar preocupado por el concepto de convergencia, pero no aporta ninguna solución a este problema. La primera vez que aparece la discusión de la convergencia del método de Newton es en 1768, en el *Traité de la résolution des équations en general* de Jean Raymond Mourraille (1720–1808). A pesar de contener ideas novedosas, la mayor parte del trabajo de Mourraille pasó inadvertido.

Figura 5: Thomas Simpson (1710–1761), el «gran olvidado» en cuanto a sus aportaciones a lo que hoy conocemos como método de Newton.

Contrariamente a Newton y Raphson, Mourraille hace hincapié en el aspecto geométrico del método de Newton, justificando que éste sea también conocido como método de la tangente. Mourraille utiliza la representación geométrica para explicar el comportamiento de la sucesión iterativa producida por el algoritmo de Newton. Además, Mourraille observa por primera vez que, dependiendo del punto de salida elegido, la sucesión generada por el método puede converger a alguna de las raíces de la ecuación, oscilar, tender a infinito, o acercarse a un límite que no es solución de la ecuación. Finalmente, Mourraille también muestra que la convergencia puede ser más o menos rápida, pero solamente lo indica en forma cuantitativa.

Posteriormente Joseph-Louis Lagrange (1736–1813), en su *Traité de la résolution des équations numériques de tous les degrés* publicado en 1808 [8], afirma que el método atribuido a Newton es el que se emplea habitualmente para resolver ecuaciones numéricas. Ahora bien, advierte que este método sólo se puede usar para ecuaciones que están ya «casi resueltas» en el sentido de que para aplicarlo se necesita una buena aproximación de la solución. Además, plantea dudas sobre la exactitud de cada nueva iteración y observa que el método puede tener problemas para el caso de raíces múltiples o muy próximas entre sí.

Jean Baptiste Joseph Fourier (1768–1830) fue el primero en analizar la velocidad de convergencia del método de Newton en una nota titulada *Question d'analyse algébraique* (1818) [8]. En este trabajo, Fourier expresa el método con la notación actual (2) y lo bautiza como *la méthode newtonienne*, haciendo referencia explícita a las obras de Newton, Raphson y Lagrange. Quizás, Fourier es el «causante» de la falta de reconocimiento para el trabajo de Simpson.

El siguiente matemático importante en estudiar el método de Newton fue Augustin Louis Cauchy (1789–1857), quien estudió este tema desde 1821, pero no dio una formulación satisfactoria hasta la publicación de las *Leçons sur le Calcul différentiel* en 1829 [8]. Cauchy da condiciones, en términos de las derivadas f' y f'', para asegurar que el método de Newton es convergente a una solución α de la ecuación (1) para todo punto de partida x_0 perteneciente a un intervalo determinado. Lo que Cauchy estaba buscando son, por tanto, resultados de convergencia global para el método de Newton; es decir, caracterizar los intervalos $I_\alpha \subseteq \mathbb{R}$ para los que

$$\lim_{n\to\infty} x_n = \alpha, \text{ con } x_0 \in I_\alpha.$$

Aunque la mayoría del trabajo de Cauchy se centra en el campo real, al final del mismo dedica un apartado al estudio de raíces complejas. Pero el estudio del método de Newton para aproximar las soluciones complejas de una ecuación encerraba ciertas sorpresas. Quizás, el que mejor pueda dar fe de ello es Arthur Cayley, quien en 1879 planteó el problema de caracterizar las regiones S_α del plano complejo para las cuales el método de Newton converge a la raíz α si $x_0 \in S_\alpha$. A S_α se le conoce como *cuenca de atracción* de la raíz α. En concreto, Cayley comenzó estudiando el problema de caracterizar las cuencas de atracción para el caso de un polinomio de segundo grado con dos raíces distintas:

$$p(z) = (z-\alpha)(z-\beta),\ \alpha \neq \beta,\ \alpha, \beta \in \mathbb{C}.$$

Figura 6: Jean Baptiste Joseph Fourier (1768–1830) «bautiza» a (2) como método de Newton.

Cayley comprobó que las cuencas de atracción de las dos raíces estaban formadas por los semiplanos en los que queda dividido el plano complejo por la recta equidistante de las dos raíces (la mediatriz del segmento de extremos α y β). Si tomamos un punto de partida en el mismo semiplano que una raíz, el método de Newton converge a dicha raíz, es decir, en este caso el método de Newton converge a la raíz más cercana al punto de partida. Si tomamos un punto de partida en la mediatriz, el método de Newton proporciona una sucesión de puntos en la propia mediatriz sin ningún orden aparente, apareciendo así un comportamiento caótico. Pero el problema se complica sobremanera cuando se pasa de un polinomio de segundo grado a uno de tercer grado. En palabras del propio Cayley: «el caso de las ecuaciones cúbicas parece que presenta considerables dificultades». Parece ser que Cayley continuó trabajando, sin éxito, en este problema. Once años después, en 1890, vuelve a escribir: «Espero poder aplicar esta teoría al caso de una ecuación cúbica, pero los cálculos son mucho más difíciles».

Cuarenta años más tarde, los trabajos de Gaston M. Julia (1918) y Pierre J. L. Fatou (1920) revelaron que el problema al que se enfrentaba Cayley no era en absoluto trivial, sino prácticamente inabordable con los conocimientos y técnicas de su época. La figura 7 muestra las cuencas de atracción para el método de Newton aplicado al polinomio cúbico $p(z) = z^3 - 1$. En ella se aprecia la estructura fractal de las cuencas de atracción y al verla se puede entender el fracaso de Cayley al intentar caracterizar dichas cuencas. En la actualidad no es difícil la representación gráfica de las cuencas de atracción, usando diversos programas informáticos. En concreto, la figura 7 se ha obtenido usando *Mathematica* y siguiendo las instrucciones dadas en [18].

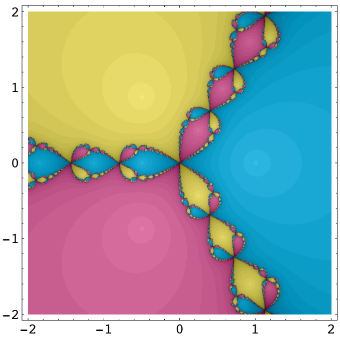

Figura 7: Cuencas de atracción de las tres raíces del polinomio $p(z) = z^3 - 1$.

El hecho de pensar en el método de Newton para encontrar las raíces de una función de variable compleja nos lleva, de forma natural, a pensar también en el método de Newton para encontrar las raíces de una función vectorial de dos variables $f : \mathbb{R}^2 \to \mathbb{R}^2$. En este caso, el método de Newton admite unas interesantes interpretaciones geométricas, como puede verse en [1] y [20]. Volviendo al punto de vista histórico, fue el propio Simpson quien estudió por primera vez un sistema de dos ecuaciones transcendentales. En el sexto de sus ensayos, escritos en 1740, Simpson describe la técnica para resolver sistemas no lineales de dos ecuaciones con dos incógnitas usando lo que hoy en día llamamos el método de Newton para sistemas

$$F(x) = 0 \quad \text{con} \quad F : \mathbb{R}^d \to \mathbb{R}^d.$$

De hecho (véase [21] para más detalles), Simpson indica cómo calcular los elementos de la matriz jacobiana $F'(x)$ para el caso $n = 2$ y la solución del correspondiente sistema de ecuaciones $F'(x)(x_{n+1} - x_n) = -F(x_n)$ usando la regla de Cramer. Simpson ilustra su técnica con tres ejemplos. El primero de ellos, con la notación actual, es:

$$\begin{cases} y + \sqrt{y^2 - x^2} = 10, \\ x + \sqrt{y^2 + x} = 12. \end{cases}$$

A pesar del carácter innovador del trabajo de Simpson, éste no tuvo mucha repercusión en su época.

El problema de la extensión del método de Newton al caso multidimensional cayó en el olvido hasta comienzos del siglo XX. En 1916, Henry B. Fine y Albert A. Bennet, ambos profesores de la Universidad de Princeton, publicaron sendos artículos *On Newton's Method of Approximation* y *Newton's method in general analysis* para sistemas de ecuaciones y para ecuaciones funcionales respectivamente. En el trabajo de Fine se da un resultado de convergencia para el método de Newton aplicado a un sistema de ecuaciones en el cual no se asume la existencia de solución, sino que sólo se exigen condiciones sobre el punto de partida. En ese mismo trabajo, Fine afirma que no había encontrado anteriormente resultados de este tipo para funciones vectoriales. En el trabajo de Bennet se justifica el empleo del método de Newton también en el caso de ecuaciones funcionales, extendiendo el trabajo de Fine. Ambos trabajos resultaron muy innovadores en una época en la que el análisis funcional era una rama de las Matemáticas que estaba dando sus primeros pasos. De hecho, sólo unos años más tarde, en 1932, Stefan Banach introdujo la noción de espacios de Banach en su famoso libro *Theorie des opérations linéaires* [2].

A partir de ese momento, varios investigadores muestran interés en la extensión del método de Newton para sistemas de ecuaciones y las publicaciones sobre el tema comienzan a ser numerosas. Pueden consultarse algunas bibliografías específicas sobre el método de Newton ([12], [13]) o buscar en las bases de datos especializadas.

Durante este periodo, destacamos las publicaciones de Alexander Ostrowski [15] y [16], en las que se estudian y comparan distintas condiciones de convergencia dadas previamente por otros autores. Además, se dan estimaciones del error cometido al aplicar el método de Newton para resolver un sistema de ecuaciones no lineales.

Es en este contexto cuando Leonid V. Kantorovich se plantea en [10] la extensión del método de Newton para resolver ecuaciones funcionales definidas entre dos espacios de Banach. Con Kantorovich se inicia el estudio «moderno» del método de Newton con numerosas publicaciones interesantes, tanto desde el punto de vista teórico como práctico. Desde el punto de vista teórico, existen numerosas variantes del teorema de Kantorovich, que modifican sus condiciones y resultados (en los trabajos recopilatorios [17] o [19] pueden verse algunas de estas contribuciones). Desde el punto de vista práctico, la teoría de Newton-Kantorovich tiene aplicaciones, no sólo para resolver ecuaciones no lineales, sino en otros contextos tales como la optimización con o sin restricciones.

3. El teorema de Newton-Kantorovich

La generalización del método de Newton (2) a espacios de Banach se debe al matemático soviético Leonid V. Kantorovich[2], quien a mediados del siglo XX publica varios trabajos, de entre los que destacamos [10] y [11].

La generalización del método de Newton para resolver ecuaciones

$$F(x) = 0, \tag{6}$$

[2]Además de sus aportaciones al estudio del método de Newton, Kantorovich es conocido por sus trabajos sobre asignación óptima de recursos escasos, por los cuales fue galardonado con el Premio Nobel de Economía en 1975, compartido con Tjalling Koopmans.

donde F es un operador no lineal diferenciable definido entre dos espacios de Banach X e Y, se escribe de la forma

$$x_{n+1} = x_n - [F'(x_n)]^{-1}F(x_n), \; n \geq 0, \tag{7}$$

siendo $F'(x_n)$ la derivada de Fréchet del operador $F(x)$ en el punto x_n y $[F'(x_n)]^{-1}$ su operador inverso.

La contribución más destacada de Kantorovich no es la demostración en sí de sus resultados, sino el hecho de apoyarse en técnicas de análisis funcional para demostrar resultados de análisis numérico. Este enfoque permite que numerosos problemas no lineales, tales como la resolución de ecuaciones integrales, ecuaciones diferenciales ordinarias y en derivadas parciales, o problemas de cálculo variacional, puedan escribirse de la forma (6) y utilizar entonces el método de Newton (7) para su resolución.

Figura 8: El matemático soviético Leonid V. Kantorovich (1912–1986).

Una característica destacable de los resultados de Kantorovich es que no asumen la existencia de soluciones de la ecuación (6). Este tipo de estudio de la convergencia, donde se exigen condiciones al punto de partida, pero no a la solución, se denomina convergencia semilocal. Así, el teorema clásico de Newton-Kantorovich no es sólo un resultado sobre convergencia del método de Newton, sino también un resultado sobre existencia de soluciones de la ecuación (6).

En esta sección presentamos el teorema clásico de Newton-Kantorovich y demostramos una variante de éste que garantiza una mayor aplicación, ya que se suaviza la condición más exigente del teorema clásico. Para demostrarlo, utilizamos una técnica más sencilla que la dada originariamente por Kantorovich, y que permite también una mayor generalización de la aplicación del teorema cuando se suavizan las condiciones de convergencia para el método de Newton. Comenzamos enunciando el teorema que dio inicio a esta teoría y que fue publicado en 1948 [10].

Teorema 1 (de Kantorovich). *Sean X e Y dos espacios de Banach y $F : \Omega \subseteq X \to Y$ un operador dos veces diferenciable Fréchet en un dominio abierto $\Omega \subset X$, y suponemos que para algún $x_0 \in \Omega$:*

(i) existe $[F'(x_0)]^{-1} \in \mathcal{L}(Y, X)$, donde $\mathcal{L}(Y, X)$ es el conjunto de los operadores lineales de Y en X, y es tal que $\|[F'(x_0)]^{-1}\| \leq \beta$,

(ii) $\|[F'(x_0)]^{-1}F(x_0)\| \leq \eta$,

(iii) $\|F''(x)\| \leq M$ en Ω.

Si $h = M\beta\eta \leq 1/2$ y $\overline{B(x_0, t^*)} \subset \Omega$, donde $t^* = \frac{1-\sqrt{1-2h}}{M\beta}$, entonces la sucesión $\{x_n\}$, dada por (7), está bien definida, $x_n \in B(x_0, t^*)$ para todo $n \geq 0$, y $\{x_n\}$ converge *a una solución* $x^* \in \overline{B(x_0, t^*)}$ *de (6). La solución* x^* *es única en* $B(x_0, t^{**}) \cap \Omega$ *si* $h < 1/2$ *o en* $\overline{B(x_0, t^{**}) \cap \Omega}$ *si* $h = 1/2$*, donde* $t^{**} = \frac{1+\sqrt{1-2h}}{M\beta}$*. Además, se verifica*

$$\|x^* - x_n\| \leq \frac{2\eta_n}{1 + \sqrt{1 - 2h_n}} \leq 2^{1-n}\,(2h)^{2^n - 1}\,\eta, \qquad n \geq 0, \tag{8}$$

donde $\{\eta_n\}$ *y* $\{h_n\}$ *están definidas por las siguientes relaciones de recurrencia:*

$$\begin{aligned} &\beta_0 = \beta, && \eta_0 = \eta, && h_0 = h, \\ &\beta_n = \frac{\beta_{n-1}}{1 - h_{n-1}}, && \eta_n = \frac{h_{n-1}\eta_{n-1}}{2(1 - h_{n-1})}, && h_n = M\beta_n\eta_n, \qquad n \geq 1. \end{aligned}$$

La demostración del resultado anterior se basa en el estudio de las sucesiones recurrentes $\{\beta_n\}$, $\{\eta_n\}$ y $\{h_n\}$. Tres años más tarde, Kantorovich introduce en [11] el conocido «principio de la mayorante» para demostrar nuevamente el teorema anterior, que se basa en el concepto de «sucesión mayorizante». Se dice que una sucesión de números reales $\{t_n\}$ mayoriza a una sucesión $\{x_n\}$, definida en un espacio de Banach X, si y sólo si

$$\|x_{n+1} - x_n\| \leq t_{n+1} - t_n, \qquad n \geq 0.$$

El interés de las sucesiones mayorizantes reside en el hecho de que su convergencia asegura la convergencia de la sucesión en el espacio de Banach. En efecto, si $\{t_n\}$ converge a t^*, entonces existe $x^* \in X$ de manera que la sucesión $\{x_n\}$ converge a x^* y se cumple

$$\|x^* - x_n\| \leq t^* - t_n, \qquad n \geq 0.$$

Kantorovich prueba también que la sucesión mayorizante $\{t_n\}$ puede obtenerse aplicando el método de Newton a la función escalar

$$f(t) = \frac{M}{2}t^2 - \frac{1}{\beta}t + \frac{\eta}{\beta}, \tag{9}$$

es decir,

$$\begin{cases} t_0 = 0, \\ t_{n+1} = t_n - f(t_n)/f'(t_n), \quad n \geq 0. \end{cases}$$

Por otra parte, es conocido que las condiciones (iii) y $\overline{B(x_0, t^*)} \subset \Omega$ del teorema 1 pueden suavizarse, respectivamente, por (véase [19] y las referencias que allí aparecen)

$$\|F'(x) - F'(y)\| \leq L\|x - y\|, \quad x, y \in \Omega, \tag{10}$$

$$\overline{B(x_1, t^* - \eta)} \subset \Omega.$$

Notemos que la función «mayorizante» f definida en (9) se consigue fácilmente como función interpoladora de las condiciones (i)–(iii) del teorema 1, incluso con la modificación (10) de (iii). Sin embargo, la evolución histórica de la condición (iii) a condiciones más suaves (por ejemplo, operadores Hölder-continuos [9],

ω-condicionados [5], [6], etc.) hace que la construción de f no sea evidente, incluso excesivamente complicada. Así, el grupo de investigación PRIENOL ha desarrollado una nueva técnica de demostración de la convergencia del método de Newton que se apoya en la construcción de una sucesión real, no mayorizante, que verifica un determinado conjunto de relaciones de recurrencia, a partir de la cual podemos asegurar la convergencia de la sucesión de Newton en el espacio de Banach. Su aplicación es simple y tiene ciertas ventajas sobre la técnica habitual de las sucesiones mayorizantes. Por un lado, se pueden generalizar los resultados obtenidos bajo las condiciones de tipo Newton-Kantorovich y, por otro lado, se pueden mejorar los resultados obtenidos mediante sucesiones mayorizantes cuando F' satisface (10). Además, también se pueden obtener estimaciones a priori del error ([7]). El comportamiento de esta nueva técnica es óptimo cuando se suavizan las condiciones impuestas al operador F (en particular, cuando se suaviza la condición (*iii*) del teorema 1). En el siguiente teorema presentamos la nueva técnica.

Teorema 2. *Sean X e Y dos espacios de Banach y $F : \Omega \subseteq X \to Y$ un operador diferenciable Fréchet en un dominio abierto convexo no vacío Ω. Supongamos que $[F'(x_0)]^{-1} \in \mathcal{L}(Y, X)$ existe en algún $x_0 \in \Omega$ y que se verifican las condiciones (i), (ii) y (10). Si $a_0 = L\beta\eta \leq 1/2$ y $\overline{B(x_0, R)} \subset \Omega$, donde $R = \frac{2(1-a_0)}{2-3a_0}\eta$, la sucesión $\{x_n\}$ dada por (7), empezando en x_0, converge a una solución x^* de la ecuación (6), la solución x^* y las iteraciones x_n pertenecen a $\overline{B(x_0, R)}$, y la solución x^* es única en la región $B(x_0, \eta/a_0) \cap \Omega$. Además, la sucesión (7) tiene al menos R-orden de convergencia dos si $a_0 \in (0, 1/2)$ o al menos uno si $a_0 = 1/2$, y se obtienen las siguientes estimaciones del error:*

$$\|x^* - x_n\| \leq \left(\gamma^{2^n - 1}\right) \frac{\Delta^n}{1 - \gamma^{2^n}\Delta}\eta, \quad n \geq 0, \tag{11}$$

donde $\gamma = \frac{a_0}{2(1-a_0)^2}$ y $\Delta = 1 - a_0$.

Para estudiar la convergencia semilocal del método de Newton a una única solución de (6), construimos, a partir de unos parámetros reales, un sistema de dos relaciones de recurrencia en las que está implicada una sucesión de números reales positivos, cuyo análisis garantiza la convergencia de la sucesión de Newton $\{x_n\}$ en el espacio de Banach X.

Relaciones de recurrencia

A partir de $a_0 = L\beta\eta$, definimos la siguiente sucesión escalar:

$$a_{n+1} = a_n f(a_n)^2 g(a_n), \quad n \geq 0, \tag{12}$$

donde

$$f(x) = \frac{1}{1-x} \qquad \text{y} \qquad g(x) = x/2. \tag{13}$$

Notemos que consideraremos $a_0 > 0$ pues con $a_0 = 0$ queda un problema trivial, ya que la solución de (6) sería x_0.

A continuación se prueba, mediante inducción sobre n, que se cumplen, para todo $n \geq 1$, las siguientes dos relaciones de recurrencia entre las sucesiones $\{a_n\}$ y $\{x_n\}$:

[I] $\|\Gamma_n\| = \|[F'(x_n)]^{-1}\| \leq f(a_{n-1})\|\Gamma_{n-1}\|$,

[II] $\|x_{n+1} - x_n\| \leq f(a_{n-1})g(a_{n-1})\|x_n - x_{n-1}\|$.

Para ello, supondremos que

$$x_n \in \Omega \qquad \text{y} \qquad a_n < 1, \qquad \text{para todo } n \geq 0. \tag{14}$$

Notemos que ambas suposiciones se demostrarán posteriormente. Entonces,

$$\|I - \Gamma_0 F'(x_1)\| \leq \|\Gamma_0\|\|F'(x_0) - F'(x_1)\| \leq L\beta\|x_1 - x_0\| \leq L\beta\eta = a_0 < 1.$$

Luego, por el teorema de Banach sobre inversión de operadores, Γ_1 está definido y

$$\|\Gamma_1\| \leq \frac{\|\Gamma_0\|}{1 - \|I - \Gamma_0 F'(x_1)\|} \leq f(a_0)\|\Gamma_0\|.$$

A partir de la fórmula de Taylor y (7), se obtiene

$$F(x_1) = F(x_0) + F'(x_0)(x_1 - x_0) + \int_{x_0}^{x_1} [F'(x) - F'(x_0)]\, dx$$
$$= \int_{x_0}^{x_1} [F'(x) - F'(x_0)]\, dx = \int_0^1 [F'(x_0 + t(x_1 - x_0)) - F'(x_0)](x_1 - x_0)\, dt,$$

y en consecuencia

$$\|F(x_1)\| \leq \frac{L}{2}\|x_1 - x_0\|^2 \leq \frac{L\eta}{2}\|x_1 - x_0\|.$$

Luego

$$\|x_2 - x_1\| = \|\Gamma_1 F(x_1)\| \leq \|\Gamma_1\|\|F(x_1)\| \leq f(a_0)g(a_0)\|x_1 - x_0\|.$$

Si suponemos ahora que las relaciones **[I]** y **[II]** se verifican para un cierto $n \geq 1$, se sigue de forma análoga al caso $n = 1$ que **[I]** y **[II]** también se cumplen para $n+1$. En consecuencia, queda probado por inducción que **[I]** y **[II]** se verifican para cualquier $n \in \mathbb{N}$.

Análisis de la sucesión escalar $\{a_n\}$

Nuestro próximo objetivo es analizar la sucesión real (12) de manera que podamos asegurar la convergencia de la sucesión (7) definida en el espacio de Banach X. Para hacer esto, basta con probar (14) y ver que (7) es una sucesión de Cauchy. En primer lugar, damos un lema técnico cuya demostración se omite por su sencillez.

Lema 1. *Sean f y g las funciones reales dadas en (13). Entonces*

(a) f es creciente y $f(x) > 1$ en $(0, 1)$,

(b) g es creciente,

(c) para $\gamma \in (0, 1)$, tenemos $f(\gamma x) < f(x)$ si $x > 0$ y $g(\gamma x) = \gamma g(x)$.

Ahora, damos algunas propiedades de la sucesión escalar $\{a_n\}$ definida en (12).

Lema 2. *Sean f y g las funciones reales dadas en (13). Si $a_0 \in (0, 1/2)$, entonces*

(a) $f(a_0)^2 g(a_0) < 1$,

(b) la sucesión $\{a_n\}$ es estrictamente decreciente y $a_n < 1$ para todo $n \geq 1$.

Si $a_0 = 1/2$, entonces $a_n = a_0 < 1$ para todo $n \geq 1$.

Demostración. En primer lugar, consideramos el caso $a_0 \in (0, 1/2)$. El apartado (a) se sigue de forma inmediata. En cuanto a (b), lo demostraremos por inducción sobre n. Como $f(a_0)^2 g(a_0) < 1$, es claro que $a_1 < a_0$. Suponemos ahora que $a_j < a_{j-1}$ para $j = 1, 2, \ldots, n$. Entonces,

$$a_{n+1} = a_n f(a_n)^2 g(a_n) < a_n f(a_0)^2 g(a_0) < a_n,$$

puesto que f y g son crecientes. Por tanto, la sucesión $\{a_n\}$ es estrictamente decreciente. Y consecuentemente $a_n < 1$ para todo $n \geq 1$.

Si, por otra parte, $a_0 = 1/2$, entonces $f(a_0)^2 g(a_0) = 1$, y obviamente $a_n = a_0 = 1/2 < 1$ para todo $n \geq 0$. □

Lema 3. *Sean f y g las funciones reales dadas en (13). Si $a_0 \in (0, 1/2)$, definimos $\gamma = a_1/a_0$, y entonces*

(a) $a_n < \gamma^{2^{n-1}} a_{n-1}$ *y* $a_n < \gamma^{2^n - 1} a_0$, *para todo $n \geq 2$,*

(b) $f(a_n) g(a_n) < \gamma^{2^n - 1} f(a_0) g(a_0) = \gamma^{2^n} / f(a_0)$, *para todo $n \geq 1$.*

Si $a_0 = 1/2$, entonces $f(a_n) g(a_n) = f(a_0) g(a_0) = 1/f(a_0)$, para todo $n \geq 1$.

Demostración. Comenzamos con el caso en que $a_0 \in (0, 1/2)$. Probaremos (a) siguiendo un procedimiento inductivo. Si $n = 2$, aplicando el apartado (b) del lema 1, obtenemos

$$a_2 = a_1 f(a_1)^2 g(a_1) = \gamma a_0 f(\gamma a_0)^2 g(\gamma a_0) < \gamma^2 a_1 = \gamma^3 a_0.$$

Supongamos ahora que

$$a_{n-1} < \gamma^{2^{n-2}} a_{n-2} < \gamma^{2^{n-1}-1} a_0.$$

Entonces, por el mismo razonamiento,

$$\begin{aligned} a_n = a_{n-1} f(a_{n-1})^2 g(a_{n-1}) &< \gamma^{2^{n-2}} a_{n-2} f\left(\gamma^{2^{n-2}} a_{n-2}\right)^2 g\left(\gamma^{2^{n-2}} a_{n-2}\right) \\ &< \gamma^{2^{n-1}} a_{n-1} < \gamma^{2^{n-1}} \gamma^{2^{n-2}} a_{n-2} < \cdots < \gamma^{2^n - 1} a_0. \end{aligned}$$

Para probar (b), observamos que

$$f(a_n) g(a_n) < f\left(\gamma^{2^n-1} a_0\right) g\left(\gamma^{2^n-1} a_0\right) < \gamma^{2^n-1} f(a_0) g(a_0) = \gamma^{2^n} / f(a_0), \quad n \geq 1,$$

con lo que se completa la demostración.

El caso en que $a_0 = 1/2$ se sigue de forma análoga. □

Demostración del teorema 2

Ya estamos en condiciones de demostrar el teorema 2 de convergencia semilocal para el método de Newton cuando se aplica a operadores cuya primera derivada es Lipschitz continua (es decir, F' satisface (10)).

Comenzamos considerando el caso $a_0 \in (0, 1/2)$. Probamos en primer lugar que $x_n \in B(x_0, R)$. Observamos que

$$\begin{aligned}
\|x_n - x_0\| &\leq \sum_{i=0}^{n-1} \|x_{i+1} - x_i\| \overset{[\mathrm{II}]}{\leq} \left[1 + \sum_{i=0}^{n-2}\left(\prod_{j=0}^{i} f(a_j)g(a_j)\right)\right] \|x_1 - x_0\| \\
&\overset{\text{lema 3 }(b)}{<} \left[1 + \sum_{i=0}^{n-2}\left(\prod_{j=0}^{i} f(a_0)g(a_0)\gamma^{2^j-1}\right)\right] \|x_1 - x_0\| \\
&= \left[1 + \sum_{i=0}^{n-2}\left(\prod_{j=0}^{i}\left(\gamma^{2^j}\Delta\right)\right)\right] \|x_1 - x_0\| = \left[1 + \sum_{i=0}^{n-2}\left(\gamma^{2^{1+i}-1}\Delta^{1+i}\right)\right] \|x_1 - x_0\|,
\end{aligned}$$

donde $\gamma = a_1/a_0 < 1$ y $\Delta = f(a_0)g(a_0)/\gamma = 1/f(a_0) = 1 - a_0 < 1$. Ahora, por la desigualdad de Bernoulli ($(1+z)^n - 1 \geq nz$ si $z > -1$), se sigue $\gamma^{2^{1+i}-1} = \gamma^{2(2^i-1)+1} \leq \gamma^{2i+1}$, y en consecuencia

$$\begin{aligned}
\|x_n - x_0\| &< \left[1 + \gamma\Delta \sum_{i=0}^{n-2} \gamma^{2i}\Delta^i\right] \|x_1 - x_0\| < \left[1 + \gamma\Delta \frac{1 - \left(\gamma^2\Delta\right)^{n-1}}{1 - \gamma^2\Delta}\right] \eta \\
&< \frac{\eta}{1 - \gamma\Delta} = \frac{2(1 - a_0)}{2 - 3a_0}\eta = R.
\end{aligned}$$

Por tanto, $x_n \in B(x_0, R)$ y, como $\overline{B(x_0, R)} \subseteq \Omega$, entonces $x_n \in \Omega$ para todo $n \geq 0$. Notamos que ahora se satisfacen las condiciones exigidas en (14).

En segundo lugar, veamos que $\{x_n\}$ es una sucesión de Cauchy. Para ello, seguimos un procedimiento totalmente análogo al del punto anterior. Así, para $m \geq 1$ y $n \geq 1$,

$$\begin{aligned}
\|x_{n+m} - x_n\| &\leq \sum_{i=n}^{n+m-1} \|x_{i+1} - x_i\| \overset{[\mathrm{II}]}{\leq} \left(1 + \sum_{i=n}^{n+m-2}\left(\prod_{j=n}^{i} f(a_j)g(a_j)\right)\right) \|x_{n+1} - x_n\| \\
&\overset{[\mathrm{II}]}{\leq} \sum_{i=n-1}^{n+m-2}\left(\prod_{j=0}^{i} f(a_j)g(a_j)\right) \|x_1 - x_0\| \\
&\overset{\text{lema 3 }(b)}{<} \sum_{i=n-1}^{n+m-2}\left(\prod_{j=0}^{i} \gamma^{2^j-1} f(a_0)g(a_0)\right) \|x_1 - x_0\|
\end{aligned}$$

$$= \sum_{i=n-1}^{n+m-2} \left(\prod_{j=0}^{i} \left(\gamma^{2^j} \Delta \right) \right) \|x_1 - x_0\| = \sum_{i=n-1}^{n+m-2} \left(\gamma^{2^{1+i}-1} \Delta^{1+i} \right) \|x_1 - x_0\|$$
$$= \sum_{i=0}^{m-1} \left(\gamma^{2^{n+i}-1} \Delta^{n+i} \right) \|x_1 - x_0\|.$$

Ahora, de nuevo por la desigualdad de Bernoulli, $\gamma^{2^{n+i}-1} = \gamma^{2^n-1}\gamma^{2^n(2^i-1)} \leq \gamma^{2^n-1}\gamma^{2^n i}$, y por tanto

$$\|x_{n+m} - x_n\| < \left(\sum_{i=0}^{m-1} \left(\gamma^{2^n i} \Delta^i \right) \right) \gamma^{2^n-1} \Delta^n \eta < \frac{1 - \left(\gamma^{2^n} \Delta\right)^m}{1 - \gamma^{2^n} \Delta} \gamma^{2^n-1} \Delta^n \eta. \tag{15}$$

Luego la sucesión $\{x_n\}$ es de Cauchy.

Para ver que x^* es una solución de (6) seguimos el procedimiento habitual. Hemos probado que $\|\Gamma_n F(x_n)\| \to 0$ cuando $n \to \infty$. Teniendo en cuenta $\|F(x_n)\| \leq \|F'(x_n)\|\|\Gamma_n F(x_n)\|$ y que la sucesión $\{\|F'(x_n)\|\}$ está acotada, puesto que

$$\|F'(x_n)\| \overset{(10)}{\leq} \|F'(x_0)\| + L\|x_n - x_0\| < \|F'(x_0)\| + LR,$$

se cumple que $\|F(x_n)\| \to 0$ cuando $n \to \infty$. En consecuencia, $F(x^*) = 0$ por la continuidad de F en $B(x_0, R)$.

Para probar la unicidad, suponemos que z^* es otra solución de (6) en la región $B\left(x_0, \eta/a_0\right) \cap \Omega$; a partir de la aproximación

$$0 = \Gamma_0[F(z^*) - F(x^*)] = \left[\int_0^1 \Gamma_0 F'(x^* + t(z^* - x^*))\, dt \right] (z^* - x^*) = P(z^* - x^*),$$

se prueba que el operador $P = \int_0^1 \Gamma_0 F'(x^* + t(z^* - x^*))\, dt$ es invertible y, por lo tanto, $z^* = x^*$. Para ello, basta con aplicar el teorema de Banach sobre inversión de operadores y ver fácilmente que $\|I - P\| < 1$.

Finalmente, si tomamos $m \to \infty$ en (15), obtenemos (11) para todo $n \geq 0$. Además, de (11), es sencillo ver que el R-orden de convergencia de la sucesión (7) es al menos dos (véase [9]).

Por otra parte, si $a_0 = 1/2$, entonces $a_n = a_0 = 1/2$ para todo $n \geq 0$. Siguiendo un procedimiento totalmente análogo al del caso anterior en el que $a_0 \in (0, 1/2)$, obtenemos los mismos resultados, teniendo en cuenta ahora que $\gamma = 1$ y $\Delta = f(a_0)g(a_0) < 1$, a excepción del R-orden de convergencia, que en este caso es al menos uno.

4. Aplicación del teorema de Newton-Kantorovich

A continuación, vemos cómo se puede utilizar el teorema de Newton-Kantorovich (versión del teorema 2) en un ejemplo concreto: la ecuación de Chandrasekhar [4]. El análisis se realizará en dos direcciones: por una parte, utilizamos el teorema 2 de Newton-Kantorovich para probar la existencia y unicidad de una solución de la ecuación y, por otra, para garantizar la convergencia del método de Newton a dicha solución.

La ecuación de Chandrasekhar es una ecuación integral que aparece en diversos problemas físicos (transferencia radiactiva, cinética de gases, etc.). Consideramos el espacio $X = C[0,1]$ de las funciones continuas en $[0,1]$ y dotado de la norma del máximo $\|x\| = \text{máx}_{t\in[0,1]} |x(t)|$, $x \in X$. Se trata de encontrar una función $x \in X$ que cumpla

$$x(s) = 1 + \lambda x(s) \int_0^1 \frac{s}{s+t} x(t)\, dt, \quad s \in [0,1]. \tag{16}$$

Con notación de operadores, la resolución de (16) es equivalente a resolver (6), siendo F el operador definido de X en X y tal que

$$[F(x)](s) = x(s) - 1 - \lambda x(s) \int_0^1 \frac{s}{s+t} x(t)\, dt, \quad s \in [0,1].$$

Consideramos, por ejemplo, el caso particular de $\lambda = 1/4$.

Para aplicar el teorema 2 necesitamos partir de una función inicial adecuada. Como de (16) se deduce que $x(0) = 1$, una elección razonable parece ser $x_0(s) = 1$, para todo $s \in [0,1]$. Además, hay que calcular el operador derivada primera de Fréchet F'. Para ello, hay que tener en cuenta que F' actúa de X en $\mathcal{L}(X)$, el conjunto de los operadores lineales de X en X. Así, para cada $x \in X$, $F'(x) \in \mathcal{L}(X)$ y $F'(x)y$ es una función continua, para todo $x, y \in X$, definida de la siguiente manera:

$$[F'(x)y](s) = y(s) - \frac{1}{4} x(s) \int_0^1 \frac{s}{s+t} y(t)\, dt - \frac{1}{4} y(s) \int_0^1 \frac{s}{s+t} x(t)\, dt, \quad s \in [0,1].$$

Podemos considerar el subconjunto Ω que aparece en el enunciado del teorema 2 como el propio espacio X, que es un conjunto convexo. A continuación, calculamos las constantes β, η y L.

En primer lugar, vemos que $\|\Gamma_0 F(x_0)\| \le \|\Gamma_0\| \|F(x_0)\|$. Por un lado, como

$$[F(x_0)](s) = -\frac{1}{4} \int_0^1 \frac{s}{s+t}\, dt = -\frac{s}{4} \log \frac{1+s}{s},$$

tenemos $\|F(x_0)\| = \log 2/4$. Por otro lado, por el lema de Banach sobre inversión de operadores, podemos encontrar una cota para Γ_0. Dada una función cualquiera $y \in X$, se tiene

$$\|[I - F'(x_0)]y\| = \frac{1}{4} \max_{s\in[0,1]} \left| \int_0^1 \frac{s}{s+t} y(t)\, dt + y(s) \int_0^1 \frac{s}{s+t}\, dt \right| \le \frac{\log 2}{2} \|y\|,$$

luego

$$\|I - F'(x_0)\| \le \log 2/2 = 0{,}3465\ldots < 1,$$

con lo que se garantiza que Γ_0 existe y es tal que

$$\|\Gamma_0\| \leq \frac{1}{1 - \|I - F'(x_0)\|} \leq \frac{1}{1 - (\log 2/2)} = 1{,}5303\ldots = \beta.$$

En consecuencia,

$$\|\Gamma_0 F(x_0)\| \leq \frac{(\log 2/4)}{1 - (\log 2/2)} = 0{,}2651\ldots = \eta.$$

De forma parecida deducimos que

$$\|F'(x) - F'(y)\| \leq \|F'(x) - F'(y)\| \leq (\log 2/2)\|x - y\|,$$

y, por tanto, $L = \log 2/2$.

Así, tenemos que $a_0 = L\beta\eta = 0{,}1406\ldots < 1/2$, con lo que se cumplen las condiciones del teorema 2 y se prueba la existencia de solución de la ecuación (16) con $\lambda = 1/4$, que está en la bola $\overline{B(x_0, 0{,}2888\ldots)}$ y es única en $B(x_0, 1{,}8853\ldots)$.

Una vez probada la existencia de una solución de la ecuación (16) con $\lambda = 1/4$, vamos a encontrar una aproximación numérica de ella mediante el método de Newton. Para ello, transformamos el problema continuo en un problema discreto utilizando una fórmula de integración numérica (en concreto. la de Gauss-Legendre) para aproximar la integral que aparece en (16). Obtenemos así un sistema de ecuaciones no lineales que resolvemos por el método de Newton. Las incógnitas son los valores aproximados de la solución en una serie de puntos del intervalo $[0, 1]$. Finalmente, a partir de estos valores y mediante un proceso de interpolación, obtenemos una aproximación de la solución.

Pasamos ahora a encontrar la aproximación numérica de la solución de (16) con $\lambda = 1/4$. Utilizando la fórmula de Gauss-Legendre con m nodos, podemos hacer

$$\int_0^1 f(t)\,dt \simeq \sum_{j=1}^{m} p_j f(t_j),$$

donde t_j y p_j son respectivamente los nodos y los pesos de la fórmula de cuadratura en $[0, 1]$, que están tabulados para distintos valores de m. Denotando x_i a las aproximaciones de $x(t_i)$, para $i = 1, 2, \ldots, m$, se llega al siguiente sistema de ecuaciones no lineales

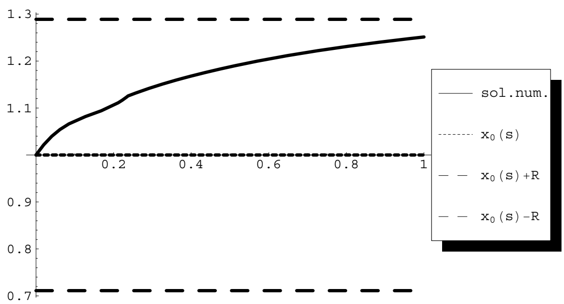

Figura 9: Gráfica de la solución aproximada de (16) con $\lambda = 1/4$.

$$x_i = 1 + \frac{1}{4} x_i t_i \sum_{j=1}^{m} p_j \frac{x_j}{t_i + t_j}, \quad i = 1, 2, \ldots, m,$$

que resolvemos aplicando el método de Newton. Para $m = 8$ y utilizando 6 dígitos significativos, la solución numérica que se obtiene es la dada en el cuadro 1.

i	x_i	i	x_i	i	x_i	i	x_i
1	1,02176...	3	1,12599...	5	1,20351...	7	1,24205...
2	1,07334...	4	1,17011...	6	1,22698...	8	1,24999...

Cuadro 1: Solución numérica de (16) con $\lambda = 1/4$.

Interpolando la función que pasa por los puntos (t_i, x_i), para $i = 1, 2, \ldots, 8$, y sabiendo que $x(0) = 1$, obtenemos la gráfica de la solución aproximada en la figura 9.

5. Alternativas a las condiciones de Kantorovich

Como ya se ha indicado anteriormente, la utilización de sucesiones mayorizantes es la técnica habitual para probar la convergencia semilocal del método de Newton. El grupo de investigación PRIENOL ha desarrollado a lo largo de los últimos años la técnica alternativa dada en la sección anterior para demostrar el teorema 2. Esta técnica simplifica el análisis de la convergencia semilocal de cualquier proceso iterativo bajo condiciones más suaves que las tradicionales de Newton-Kantorovich.

Como puede verse en el teorema 1, la condición (iii),

$$\|F''(x)\| \leq M \quad \text{en} \quad \Omega, \tag{17}$$

es la más exigente. Sin embargo, hay situaciones en las que esta condición no se satisface. Por ejemplo, en algunas ecuaciones integrales no lineales de tipo Hammerstein mixto:

$$x(s) = u(s) + \sum_{i=1}^{m} \int_a^b G_i(s,t) H_i(x(t))\, dt, \quad s \in [a,b], \tag{18}$$

donde $-\infty < a < b < \infty$, G_i, H_i $(i = 1, 2, \ldots, m)$ y u son funciones conocidas y x es una función solución a determinar. (Ecuaciones integrales de este tipo pueden encontrarse, por ejemplo, en modelos dinámicos de reactores químicos.) En particular, para ecuaciones de la forma

$$x(s) = u(s) + \int_a^b G(s,t)\big(\lambda_1 x(t)^{2+q} + \lambda_2 x(t)^n\big)\, dt, \tag{19}$$

con $s \in [a,b]$, $q \in [0,1]$, $n \geq 2$ $(n \in \mathbb{N})$ y $\lambda_1, \lambda_2 \in \mathbb{R}$, donde u es una función continua y el núcleo G es continuo y no negativo en $[a,b] \times [a,b]$.

La resolución de la ecuación integral (19) es equivalente a resolver la ecuación (6), donde $F : C[a,b] \to C[a,b]$ y

$$[F(x)](s) = x(s) - u(s) - \int_a^b G(s,t)\big(\lambda_1 x(t)^{2+q} + \lambda_2 x(t)^n\big)\, dt.$$

En este caso,

$$[F'(x)y](s) = y(s) - \int_a^b G(s,t)\big((2+q)\lambda_1 x(t)^{1+q} + n\lambda_2 x(t)^{n-1}\big)y(t)\,dt,$$

$$[F''(x)(yz)](s) = -\int_a^b G(s,t)\big((2+q)(1+q)\lambda_1 x(t)^{q} + n(n-1)\lambda_2 x(t)^{n-2}\big)z(t)y(t)\,dt.$$

Luego la condición (17) no se cumple porque F'' no está acotada en $C[a,b]$.

Para solventar el problema anterior, podemos prelocalizar la raíz en algún dominio $\Omega \subset C[a,b]$ en el que se pueda encontrar una cota superior para $\|F''\|$. El problema que presenta esta forma de actuar es que que no siempre es posible determinar tal dominio. Una alternativa mejor y más elegante consiste en suavizar dicha condición por una del tipo

$$\|F''(x)\| \leq \omega(\|x\|), \quad x \in \Omega, \tag{20}$$

donde $\omega : \mathbb{R}_+ \cup \{0\} \to \mathbb{R}_+ \cup \{0\}$ es una función real continua monótona y tal que $\omega(0) \geq 0$ (véanse [5] y [7]).

Volviendo de nuevo a las condiciones tradicionales de convergencia de Newton-Kantorovich, una gran cantidad de diferentes variantes han aparecido a lo largo de los últimos años a la hora de estudiar la convergencia semilocal del método de Newton. Por ejemplo, cambiar la condición (17) por (10) o por la siguiente generalización (véanse [9], [19] y las referencias aquí citadas):

$$\|F'(x) - F'(y)\| \leq K\|x-y\|^q, \quad q \in [0,1], \quad x, y \in \Omega. \tag{21}$$

Se dice entonces que F' es (K,q)-Hölder continua en Ω. Observamos que si $K = L$ y $q = 1$, (21) se reduce a (10).

En la práctica, la verificación de las condiciones (10) y (21) también es difícil en algunos problemas, ya que se encuentran ciertas dificultades técnicas, de manera que el número de ecuaciones que se pueden resolver mediante el método de Newton es limitado. En particular, no se puede analizar la convergencia del método a una solución de ecuaciones en las que aparecen sumas de operadores que satisfacen (10) o (21) indistintamente. Por ejemplo, si consideramos (18) con $H_i'(x(t))$, para $i = 1, 2, \ldots, m$, siendo (K_i, q_i)-Hölder continua en Ω, el correspondiente operador $F : C[a,b] \to C[a,b]$,

$$[F(x)](s) = x(s) - u(s) - \sum_{i=1}^{m} \int_a^b G_i(s,t) H_i(x(t))\,dt, \quad s \in [a,b],$$

no satisface ni (10) ni (21) en $C[a,b]$ al considerar, por ejemplo, la norma del máximo. En este caso,

$$\|F'(x) - F'(y)\| \leq \sum_{i=1}^{m} K_i \|x-y\|^{q_i}, \quad K_i \geq 0, \quad q_i \in [0,1], \quad x, y \in \Omega.$$

Para resolver este tipo de ecuaciones y suavizar las condiciones (10) y (21), podemos considerar la generalización

$$\|F'(x) - F'(y)\| \leq \widetilde{\omega}(\|x - y\|), \quad x, y \in \Omega, \tag{22}$$

donde $\widetilde{\omega} : \mathbb{R}_+ \to \mathbb{R}_+$ es una función real continua no decreciente y tal que $\widetilde{\omega}(0) \geq 0$ (véanse [6] y [7]). De esta manera, $\widetilde{\omega}(z) = \sum_{i=1}^{m} K_i z^{q_i}$ para el caso particular anterior.

Obviamente, las condiciones (17), (10) y (21) son casos particulares de (20) y (22), ya que (20) se reduce a (17) si $\omega(z) = M$ y (22) se reduce a (10) y (21) si, respectivamente, $\widetilde{\omega}(z) = Lz$ y $\widetilde{\omega}(z) = Kz^q$.

Terminamos destacando que la utilización de sucesiones mayorizantes para asegurar la convergencia del método de Newton bajo las condiciones (20) o (22) es difícil, por no decir prácticamente imposible. De aquí surge nuestro principal interés por desarrollar la técnica alternativa a la de las sucesiones mayorizantes, dada en la sección 3, de manera que el método de Newton converja siempre que se cumpla (20) o (22).

Referencias

[1] F. Aguiló y A. Miralles, Consideraciones geométricas acerca del método de Newton, *La Gaceta de la RSME* **7** (2004), n.º 1, 247–260.

[2] S. Banach, *Théorie des opérations linéaires*, Monografie Matematyczne, Varsovia, 1932.

[3] J. L. Chabert et al., *A History of Algorithms: from the Pebble to the Microchip*, Springer-Verlag, Berlín-Heidelberg, 1999.

[4] S. Chandrasekhar, *Radiative transfer*, Dover, Nueva York, 1960.

[5] J. A. Ezquerro y M. A. Hernández, On an application of Newton's method to nonlinear operators with w-conditioned second derivative, *BIT* **42** (2002), n.º 3, 519–530.

[6] J. A. Ezquerro y M. A. Hernández, Generalized differentiability conditions for Newton's method, *IMA J. Numer. Anal.* **22** (2002), 187–205.

[7] J. A. Ezquerro y M. A. Hernández, On the R-order of convergence of Newton's method under mild differentiability conditions, *J. Comput. Appl. Math.* **197** (2006), n.º 1, 53–61.

[8] Gallica-Math: OEuvres complètes, Biblioteca numérica Gallica de la Bibliothèque Nationale de France, `http://mathdoc.emath.fr/OEUVRES/`

[9] M. A. Hernández, The Newton method for operators with Hölder continuous first derivative, *J. Optim. Theory Appl.* **109** (2001), 631–648.

[10] L. V. Kantorovich, On Newton's method for functional equations, *Dokl Akad. Nauk SSSR* **59** (1948), 1237–1240 (en ruso).

[11] L. V. Kantorovich, The majorant principle and Newton's method, *Dokl Akad. Nauk SSSR* **76** (1951), 17–20 (en ruso).

[12] J. H. Mathews, Bibliography for Newton's method, `http://math.fullerton.edu/mathews/n2003/Newton'sMethodBib.html`

[13] J. M. McNamee, A bibliography on roots of polynomials: Newton's method, http://www1.elsevier.com/homepage/sac/cam/mcnamee/02.htm

[14] O. Neugebauer y A. Sachs, *Mathematical cuneiform texts*, American Oriental Society, New Haven, Conn., 1945.

[15] A. Ostrowski, Über die Konvergenz und die Abrundungsfestigkeit des Newtonschen Verfahrens, *Rec. Math.* **2** (1937), 1073–1095.

[16] A. Ostrowski, Über einen Fall der Konvergenz des Newtonschen Näherungsverfahrens, *Rec. Math.* **3** (1938), 254–258.

[17] B. T. Polyak, Newton-Kantorovich method and its global convergence, *J. Math. Sciences* **133** (2006), n.º 4, 1513–1523.

[18] J. L. Varona, Graphic and numerical comparison between iterative methods, *Math. Intelligencer* **24** (2002), n.º 1, 37–46.

[19] T. Yamamoto, Historical developments in convergence analysis for Newton's and Newton-like methods, *J. Comput. Appl. Math.* **124** (2000), 1–23.

[20] L. Yau y A. Ben-Israel, The Newton and Halley methods for complex roots, *Amer. Math. Monthly* **105** (1998), n.º 9, 806–818.

[21] T. J. Ypma, Historical development of the Newton-Raphson method, *SIAM Review* **37** (1995), n.º 4, 531–551.

J. A. Ezquerro, J. M. Gutiérrez, M. Á. Hernández, N. Romero y M. J. Rubio, Dpto. de Matemáticas y Computación, Universidad de La Rioja, 26004 Logroño

Correo electrónico: {jezquer,jmguti,mahernan,natalia.romero,mjesus.rubio}@unirioja.es

Página web: http://www.unirioja.es/dptos/dmc/prienol/