

TEMat

divulgación de trabajos de estudiantes de matemáticas

volumen 6
mayo de 2022

<https://temat.es/volumen/2022/>

<http://www.anem.es/>

Una iniciativa de la
Asociación Nacional de Estudiantes de Matemáticas



Publica



Asociación Nacional de Estudiantes de Matemáticas
Plaza de las Ciencias, 3
Despacho 525, Facultad de Ciencias Matemáticas
Universidad Complutense de Madrid
28040 – Madrid

temat@temat.es
contacto@anem.es

Colabora



Real Sociedad Matemática Española
Plaza de las Ciencias, 3
Despacho 525, Facultad de Ciencias Matemáticas
Universidad Complutense de Madrid
28040 – Madrid

Diseño de portada: Roberto Berná Larrosa, rberナルarrosa@gmail.com

TEMat, divulgación de trabajos de estudiantes de matemáticas – volumen 6 – mayo de 2022

e-ISSN: 2530-9633

<https://temat.es/>

© 2022 Asociación Nacional de Estudiantes de Matemáticas.

© 2022 los autores de los artículos.

© Salvo que se indique lo contrario, el contenido de esta revista está disponible bajo una licencia Creative Commons Reconocimiento 4.0 Internacional.

Equipo

Editores jefe

Alberto Espuny Díaz, Technische Universität Ilmenau
Isaac Sánchez Barrera, Universitat Politècnica de Catalunya

Editores adjuntos

Gregorio Martínez Sempere, Universidad de Sevilla
Pablo Nicolás, Universitat Politècnica de Catalunya

Edición

Fernando Ballesta Yagüe, Universidad Autónoma de Madrid
Emilio Domínguez Sánchez, Universidad de Murcia
Álvaro González Hernández, University of Warwick
Alejandra Martínez Moraian, Universidad de Alcalá

Comité editorial

Pablo Manuel Berná Larrosa, Universidad CEU San Pablo
Miguel Camarasa Buades, Basque Center for Applied Mathematics
Domingo García Rodríguez (representante de la RSME), Universitat de València
Álvaro González Hernández, University of Warwick
Midwar López Huapaya (representante de la ANEM), Universidad de Cantabria
Elena López Navarro, Universitat Politècnica de València
Pablo Oviedo Timoneda, University of Birmingham
Martí Roset Julià, McGill University
Lucía Rotger García, Universidad de La Rioja
Paula Segura Martínez, Universitat Politècnica de València
Victor Sotomayor, Centro Universitario EDEM - Valencia

Revisiones externas

En este volumen han colaborado realizando revisiones externas:

Alba Cervera-Lierta (Barcelona Supercomputing Center - Centro Nacional de Supercomputación), Antonio Galbis (Universitat de València), Elies Gil-Fuster (Freie Universität Berlin), Jorge Herrera de la Cruz (Universidad CEU San Pablo), Marta Latorre Balado (Universidad Rey Juan Carlos), Víctor Manero (Universidad de Zaragoza), Javier Martínez Perales (Universidad de Málaga), Álvaro Ortega González (University College London), Óscar Rivero (University of Warwick), Alberto Rodríguez Arenas (Universitat Jaume I), Salvador Segura Gomis (Universitat d'Alacant), Juan Adrián Vargas (Universitat d'Alacant), Álvaro del Valle Vílchez (Universidad de Sevilla).

Sobre TEMat

TEMat es una revista de divulgación de trabajos de estudiantes de matemáticas publicada sin ánimo de lucro por la Asociación Nacional de Estudiantes de Matemáticas. Se busca publicar trabajos divulgativos de matemáticas, escritos principalmente (pero no exclusivamente) por estudiantes, de todo tipo: breves reseñas, introducciones a temas de investigación complejos, o artículos explicando las bases e incluso algún pequeño resultado de trabajos desarrollados por estudiantes.

TEMat persigue el doble objetivo de dar visibilidad a la calidad y diversidad de los trabajos realizados por estudiantes de matemáticas en los centros españoles a la vez que permite a los estudiantes publicar sus primeros artículos, familiarizándose así con el proceso de redacción, revisión y corrección que va asociado a la actividad investigadora.

Se contemplan para su publicación artículos escritos en castellano de todas las áreas de las matemáticas, incluyendo álgebra, análisis, ciencias de la computación, combinatoria, educación matemática, estadística, geometría, teoría de números y cualquier otra área de las matemáticas (puras y aplicadas), así como aplicaciones científicas o tecnológicas en las que las matemáticas jueguen un papel central.

Índice general

Carta de la presidenta de la ANEM	VII
«La geometría taxicab: un mundo donde los círculos son cuadrados», de Aniol Cortada Garcia	1
«Métodos numéricos aplicados al teorema de Gauss sobre la distribución de los números primos», de Ferran Arnau Segarra	17
«Fundamentos de la computación cuántica», de Vicente López Oliva	31
«La ecuación $\bar{\delta}$ y el teorema de Runge», de Melanie Fumero Padrón	49
«La conjetura de Collatz», de Alejandro Gil Asensi	65
«Aritmética y barajas de cartas», de Francisco Albuquerque Picado	83

Carta de la presidenta de la ANEM

La revista *TEMat* llega a su volumen número 6, un número perfecto, abriendo nuevos horizontes. Desde 2017, este proyecto único en el panorama internacional ha sacado una revista escrita por y para estudiantes, fomentando así la participación en el mundo de la publicación y la ciencia.

TEMat nos ofrece la oportunidad de publicar nuestros primeros artículos y la posibilidad de conocer de primera mano el funcionamiento de una revista. Es una herramienta que ponemos al alcance del estudiantado ya no solo para publicar, algo que puede ayudar en subvenciones para investigar, sino que también nos ofrece una variedad de temáticas donde podemos encontrar ideas para trabajos de fin de grado o máster.

El crecimiento de *TEMat* es notorio y prueba de ello son los dos volúmenes de *TEMat monográficos* publicados en el pasado año, así como el interés de ciertas corporaciones en incluir nuestra revista en sus catálogos.

No podría terminar sin agradecer el incansable esfuerzo de todos los miembros del comité editorial. Gracias a todas las personas que trabajan desinteresadamente en este proyecto, desde los autores a los editores, pasando por el comité editorial. Sin todas estas personas no podríamos disfrutar de este maravilloso volumen cargado de nuevos e interesantes artículos. Además, te animo a ti, lector, a que participes en esta maravillosa iniciativa, ya sea enviando tus artículos o dando un pasito al frente participando en el comité editorial.

Clara Martínez Martínez,
presidenta de la ANEM.

Murcia, mayo de 2022.

TEMat

Este trabajo fue galardonado con el primer premio en la edición de 2021 del Premi Poincaré, entregado por la Facultat de Matemàtiques i Estadística de la Universitat Politècnica de Catalunya.



Premi Poincaré

al millor treball de recerca de Batxillerat en Matemàtiques

18a edició 2021

La geometría taxicab: un mundo donde los círculos son cuadrados

✉ Aniol Cortada Garcia^a
Universitat Politècnica de Catalunya
aniolcortadagarcia@gmail.com

Resumen: Nuestra intuición nos tiene acostumbrados a medir según la geometría euclidiana clásica, donde la distancia más corta entre dos puntos es la línea recta. Sin embargo, hay ocasiones en que la línea recta no es una trayectoria posible, como por ejemplo al considerar el trayecto de un taxi por una ciudad. En este trabajo se plantea estudiar la geometría que se deriva de esta nueva manera de medir, la denominada geometría taxicab. Después de introducir el concepto de métrica, se aborda el estudio de diferentes elementos geométricos con la métrica de Manhattan. Se obtienen resultados sorprendentes: las circunferencias son cuadrados, las elipses pueden ser octógonos, los triángulos equiláteros pueden no ser semejantes y el valor de π (entendido como la razón entre el perímetro y el diámetro de una circunferencia) es 4. Todos estos elementos se ilustran haciendo uso del programa GeoGebra. Como aplicación práctica se estudian dos problemas de geometría urbana que dan respuesta a dónde situar equipamientos urbanos de una manera más eficiente. Finalmente, se proponen generalizaciones de la métrica de Manhattan que abren todo un mundo de posibilidades.

Abstract: People are used to measuring according to classical Euclidean geometry, where the shortest distance between two points is the straight line. However, sometimes the straight line is not a possible path, for instance when a taxi goes through a city. In this work, we study the geometry derived from this new way of measuring, called the taxicab geometry. After introducing the concept of metric, the study of different geometric elements is approached by means of the Manhattan metric. Surprising results are obtained: circumferences are squares, ellipses can be octagons, equilateral triangles may not be similar and the value of π (understood as the ratio between the perimeter and the diameter of a circumference) is 4. All these elements are illustrated by means of GeoGebra program. For practical purposes, we study two urban geometry problems concerning the efficient location of urban facilities. Finally, we propose some generalizations of the Manhattan metric that open a new world of possibilities.

Palabras clave: geometrías no euclidianas, geometría taxicab, métrica de Manhattan, GeoGebra, Voronoi.

MSC2020: 51-99.

Recibido: 27 de agosto de 2021.

Aceptado: 18 de abril de 2022.

Agradecimientos: Queremos agradecer a la Facultat de Matemàtiques i Estadística de la UPC y a la revista *TEMat* la oportunidad de publicar este artículo como parte del Premi Poincaré 2021.

Referencia: CORTADA GARCIA, Aniol. «La geometría taxicab: un mundo donde los círculos son cuadrados». En: *TEMat*, 6 (2022), págs. 1-15. ISSN: 2530-9633. URL: <https://temat.es/articulo/2022-p1>.

^aEste trabajo se lleva a cabo a partir del Treball de Recerca de 2.º de Bachillerato desarrollado en el INS Cendrassos de Figueres.

1. Introducción

Cuando se ha de medir una distancia, de entrada, estamos acostumbrados a hacerlo según la geometría euclidiana. Sin embargo, hay ocasiones en que esto no tiene por qué ser lo más conveniente, ni siquiera lo que más se ajuste a la realidad. Ocurre en campos avanzados como la cosmología moderna o en situaciones cotidianas de la vida urbana, y medir de una manera u otra puede ser muy relevante. Valga como ejemplo el caso de James Robbins en 2002 (publicado en el *New York Times*). Este fue detenido por vender drogas a unos 900 pies de una escuela, distancia inferior a 1000 pies, lo cual era considerado como un agravante. Su abogado argumentó que la distancia no podía ser tomada en línea recta ya que en la ciudad nos desplazamos por las calles doblando esquinas. Siguiendo las calles había unos 1200 pies, con lo cual la pena sería menor. Sin embargo, el juez consideró que la distancia se había de calcular como lo haría Euclides y no con la métrica de Manhattan, de manera que fue condenado a una pena de seis a doce años de prisión.

La geometría taxicab¹ es una de esas geometrías que se adapta más fielmente a ciertas situaciones de la vida cotidiana, como sería el ámbito urbano. En este trabajo, estudiaremos esta geometría ayudándonos del programa GeoGebra para modelizar diferentes escenarios. Este no sirve para demostrar teoremas matemáticos, pero sí es muy útil para mostrar y entender muchos conceptos.

En primer lugar, estudiaremos qué es una métrica, analizando varios ejemplos y concluyendo con la métrica de Manhattan.

En el siguiente apartado estudiaremos, haciendo uso de la métrica de Manhattan, diferentes elementos geométricos que se definen a partir del concepto de distancia: mediatrices, triángulos y cónicas. Esto nos permitirá ver lo diferentes que son la geometría taxicab y la euclidiana.

A continuación nos plantearemos dos aplicaciones prácticas de la geometría taxicab a problemas de geometría urbana. De hecho, se trata de dos problemas clásicos de gran utilidad práctica: el punto de Fermat (el lugar que hace que la suma de las distancias a tres vértices sea mínima) y los diagramas de Voronoi (cómo dividir el plano según las regiones más cercanas a unos ciertos puntos).

Finalmente, en el último apartado, estudiaremos algunas generalizaciones que se pueden hacer de la métrica de Manhattan. La primera corresponde a considerar esta métrica en el espacio en lugar del plano. La segunda consiste en cambiar los movimientos perpendiculares a los ejes propios de la geometría taxicab por otros que permitan también desplazamientos diagonales, como sería el caso de los movimientos de las damas chinas. Por último, veremos la generalización a enrejados triangulares de cualquier ángulo.

2. Métricas

Una métrica es el concepto matemático que corresponde a la idea tan común e intuitiva de distancia. En este apartado veremos la definición y algunos ejemplos de métricas que nos mostrarán cómo cambiando la manera de medir distancias se construyen geometrías muy diferentes.

2.1. Definición

Una métrica sobre un conjunto S es una función distancia, $d: S \times S \rightarrow \mathbb{R}$, que para todo $P, Q, R \in S$ satisface las siguientes condiciones:

- (i) $d(P, Q) \geq 0$,
- (ii) $d(P, Q) = 0$ si y solo si $P = Q$,
- (iii) $d(P, Q) = d(Q, P)$ y
- (iv) $d(P, R) \leq d(P, Q) + d(Q, R)$.

¹En diferentes artículos y libros podemos leer indistintamente geometría taxicab, geometría del taxista, métrica de Manhattan, métrica L_1 , geometría urbana... En este trabajo hablaremos de geometría taxicab cuando nos refiramos a la geometría desde un punto de vista global y de métrica de Manhattan cuando hablemos concretamente de la función distancia.

Hay muchas funciones que satisfacen estas propiedades y cualquiera que lo haga será una métrica. Esto conduce a casos muy curiosos, como la métrica de Chebyshev (también llamada «del ajedrez», porque corresponde a los movimientos del rey), la métrica de Hamming (con aplicaciones en teoría de códigos), la métrica discreta, la familia de métricas L^p , etc. A continuación analizaremos algunos ejemplos en \mathbb{R}^2 .

2.2. Ejemplos

2.2.1. Métrica euclidiana

La métrica euclídea sobre el plano cartesiano es la que se ha utilizado desde siempre. Es la distancia intuitiva que encontramos a partir del teorema de Pitágoras. Dados dos puntos del plano $P(p_1, p_2)$ y $Q(q_1, q_2)$, la distancia euclídea viene dada por la expresión

$$d(P, Q) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2}.$$

2.2.2. Métrica del ascensor

Si pensamos en un conjunto de rascacielos colocados en línea, la distancia entre dos pisos que estuvieran en el mismo edificio sería el valor absoluto de la diferencia de coordenadas verticales, o sea, el recorrido que haría el ascensor. Si estuvieran en edificios diferentes, la distancia sería lo que se baja en el primer ascensor, más lo que se camina entre los edificios, más lo que se sube en el segundo ascensor. Por ejemplo, en la figura 1 la distancia entre los dos puntos sería $d(P, Q) = 2 + |4 - 1| + 3 = 8$. Así, podemos ver que la métrica viene definida por la expresión

$$d(P, Q) = \begin{cases} |q_2 - p_2| & \text{si } p_1 = q_1, \\ |p_2| + |q_1 - p_1| + |q_2| & \text{si } p_1 \neq q_1. \end{cases}$$

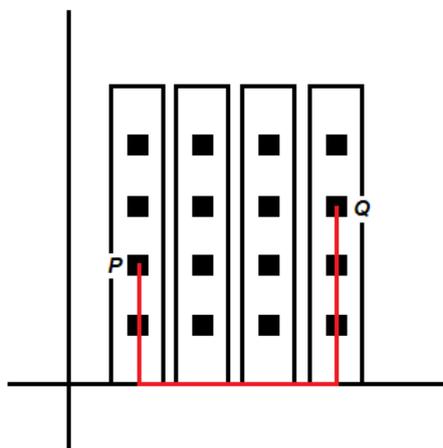


Figura 1: Métrica del ascensor.

Para hacernos una idea de cómo es la geometría con esta métrica, podemos representar las circunferencias en esta métrica, es decir, el conjunto de puntos que equidistan una distancia r de un centro C . Con la ayuda de GeoGebra, podemos ver que las circunferencias de esta métrica tienen forma de cuadrado con un punto encima (siempre que el radio r sea mayor que la altura p_2), como podemos ver en la figura 2.

2.2.3. Métrica de correos

También conocida como la métrica del mensajero, la métrica de correos se define de la siguiente manera: la distancia entre dos puntos es la que resulta de sumar la distancia euclídea de cada punto al origen. Es

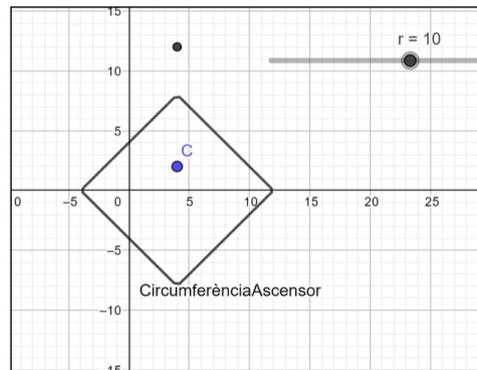


Figura 2: Circunferencia en la métrica del ascensor.

decir, para ir del punto P al punto Q , primero vamos de P al origen y, después, del origen a Q (es el recorrido que haría una carta, va del remitente a la oficina y de la oficina al destinatario, y de ahí el nombre). Viene dada por la expresión

$$d(P, Q) = \begin{cases} 0 & \text{si } P = Q, \\ \sqrt{p_1^2 + q_1^2} + \sqrt{p_2^2 + q_2^2} & \text{si } P \neq Q. \end{cases}$$

Es necesario pedir que $d(P, P) = 0$ por la propia definición de métrica. Este requisito tiene mucho sentido intuitivamente: no es de esperar que una persona que quiera enviarse una carta a su propia casa lo vaya a hacer usando el servicio de correos.

Haciendo uso de GeoGebra podemos analizar cómo son las circunferencias. Se observa el hecho curioso de que, independientemente de donde esté el centro, las circunferencias están siempre centradas en el origen. Además, pueden tener el centro en el interior o en el exterior. En esta métrica, las circunferencias sí tienen forma circular (suponiendo siempre que el radio r es mayor que la distancia euclídea de P al origen).

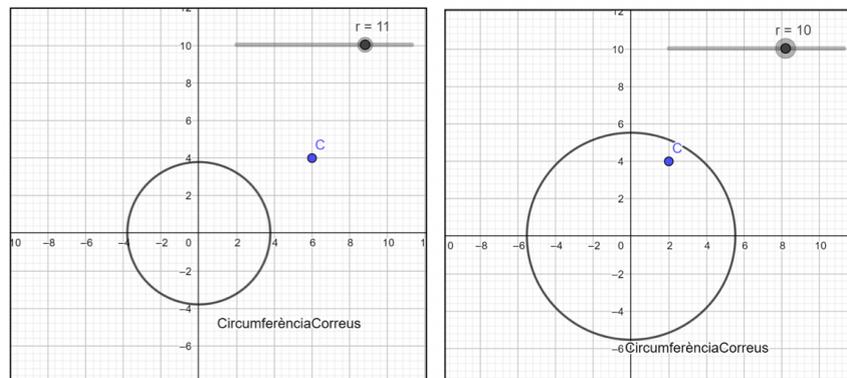


Figura 3: Circunferencias en la métrica de correos.

2.2.4. La métrica de Manhattan

El primero en estudiar esta métrica fue el matemático alemán Hermann Minkowski (1864-1909). Sin embargo, el primero en llamarla «taxicab» fue el matemático austríaco Karl Menger en 1952 en una exposición en el Museo de Ciencia e Industria de Chicago.

El nombre «métrica de Manhattan» viene del diseño en cuadrícula de las calles del barrio neoyorquino, que hace que el camino más corto que un coche puede tomar entre dos puntos de la ciudad no sea la línea recta (que atravesaría los edificios) sino las correspondientes variaciones siguiendo las diferentes calles.

Por ejemplo, en la figura 4, la distancia taxicab de P a Q por cualquiera de los caminos sería 9 (a diferencia de la Euclídea, que sería $\sqrt{45}$). Vale la pena destacar que el camino más corto en esta métrica no es único. De hecho, en el ejemplo anterior podemos trazar $9!/(6! \cdot 3!) = 84$ caminos diferentes de la misma longitud mínima.



Figura 4: Métrica de Manhattan.

La expresión matemática de la métrica de Manhattan (también distancia taxicab, o distancia L_1) es

$$d(P, Q) = |p_1 - q_1| + |p_2 - q_2|$$

para dos puntos $P(p_1, p_2)$ y $Q(q_1, q_2)$ cualesquiera de \mathbb{R}^2 .

3. Elementos de la métrica de Manhattan

En este apartado estudiaremos diferentes elementos geométricos que se definen a partir del concepto de distancia. De esta manera podremos apreciar cómo cambian por el hecho de elegir una métrica u otra.

3.1. Mediatriz

En geometría euclídea, la mediatriz es el lugar geométrico de los puntos $R(x, y)$ que equidistan de dos puntos dados, P y Q . Se trata de un elemento geométrico muy útil en diversas situaciones; por ejemplo, cuando se delimitan regiones según áreas de influencia, como veremos más adelante con los diagramas de Voronoi. En la geometría taxicab, la ecuación de la mediatriz de P y Q viene dada por la expresión

$$d(P, R) = |x - x(P)| + |y - y(P)| = |x - x(Q)| + |y - y(Q)| = d(R, Q),$$

donde $x(P)$, $y(P)$, $x(Q)$ e $y(Q)$ son las coordenadas de P y Q .

Como podemos observar en la figura 5, la mediatriz en el caso general tiene forma de línea quebrada, a diferencia de lo que ocurre en la geometría euclídea. Se observan además dos casos particulares en los que la forma cambia: uno cuando los puntos están alineados con los ejes de coordenadas (como el caso euclidiano), y el otro cuando los puntos están alineados según un ángulo de 45° (en este caso, la mediatriz es un segmento seguido de dos regiones no acotadas).

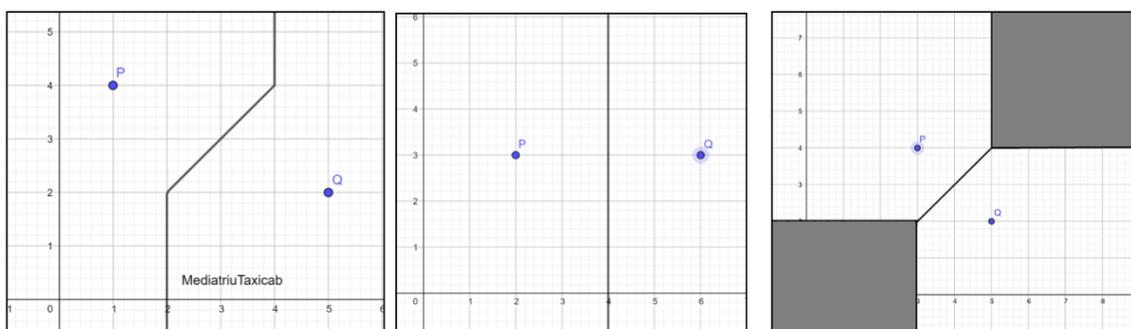


Figura 5: Mediatrices en la geometría taxicab.

3.2. Criterios de congruencia de triángulos

En este apartado veremos que los criterios de congruencia de triángulos en la geometría euclidiana fallan todos en la taxicab. Este punto es muy interesante y, como resalta Krause en su libro² [3], es la única diferencia a nivel axiomático entre las geometrías euclidiana y taxicab.

Criterio (LLL; Lado-Lado-Lado). *Si dos triángulos tienen iguales los tres lados, entonces son congruentes.*

En la figura 6 vemos un ejemplo de dos triángulos equiláteros de lado igual a 4 unidades que, sin embargo, no son congruentes, incumpliendo el criterio LLL.

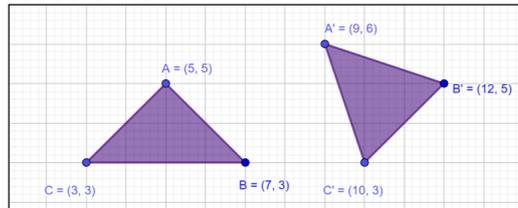


Figura 6: Violación del criterio LLL.

Criterio (LAL; Lado-Ángulo-Lado). *Si dos triángulos tienen iguales dos lados y el ángulo que determinan, entonces son congruentes.*

En el ejemplo de la figura 7, podemos observar que los lados CA , CB , $C'A'$ y $C'B'$ tienen todos la misma longitud (4 unidades) y que los ángulos que determinan son iguales (90°), pero los triángulos no son congruentes, violando el criterio LAL.

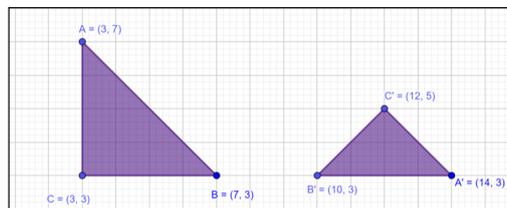


Figura 7: Violación del criterio LAL.

Criterio (ALA; Ángulo-Lado-Ángulo). *Si dos triángulos tienen iguales un lado y los ángulos adyacentes, entonces son congruentes.*

En la figura 8 tenemos el caso de dos triángulos con los lados AC y $A'C'$ de longitud 4 unidades y los ángulos adyacentes iguales a 45° y que, sin embargo, no son congruentes, en contra de lo que afirma el criterio ALA.

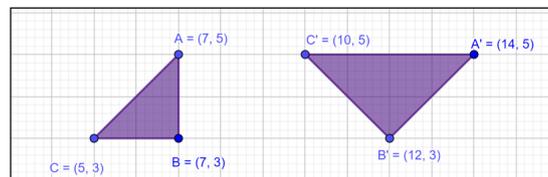


Figura 8: Violación del criterio ALA.

²El libro de Krause se trata de prácticamente el único libro específico sobre la métrica taxicab. Además, tiene un enfoque muy original basado en el estudio de la misma a partir de ejercicios propuestos.

Cabe destacar que, a diferencia de lo que sucede en la métrica euclídea, en la métrica de Manhattan la distancia no es invariante por rotaciones, y eso explica que los criterios de congruencia de triángulos que se cumplen en la geometría euclídea no se verifiquen en la geometría taxicab.

3.3. Cónicas

En la geometría euclidiana, las cónicas son curvas que se obtienen de la intersección de la superficie de un cono con un plano. Por otra parte, tienen una definición muy sencilla como lugar geométrico en términos de la función distancia, lo que las hace muy adecuadas para investigar cómo variarían cuando cambiemos de métrica.

3.3.1. Circunferencia

La circunferencia es el lugar geométrico de los puntos del plano que equidistan de un punto fijo, llamado centro. La distancia del centro a cualquier punto de la circunferencia se denomina radio.

En la geometría taxicab, la ecuación de circunferencia de centro $C(a, b)$ y radio r vendrá dada por la expresión

$$d((x, y), C) = |x - a| + |y - b| = r.$$

Con la ayuda de GeoGebra podemos observar que la circunferencia tiene forma cuadrada. Todos los puntos sobre el cuadrado se encuentran a distancia r del centro C .

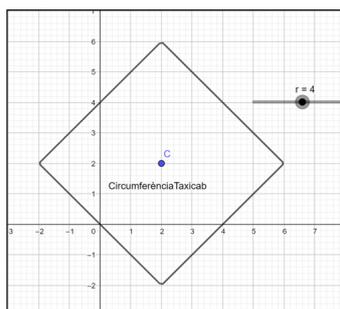


Figura 9: Circunferencia usando la métrica de Manhattan.

Observamos que la distancia entre vértices contiguos para la geometría taxicab es $2r$ o, de forma equivalente, el diámetro de la circunferencia d . En este tipo de geometría, el cociente entre el perímetro de una circunferencia y su diámetro viene dado por

$$\frac{4D}{D} = 4.$$

Dicho cociente para la geometría euclidiana, el cual es una constante universal en matemáticas, recibe el nombre π y su valor es aproximadamente 3,141 59 (las primeras aproximaciones son del año 1900 a. C.). Este resultado muestra que, en cierto sentido, podríamos decir que $\pi = 4$ en la geometría taxicab.

3.3.2. Elipse

La elipse es el lugar geométrico de los puntos P del plano tales que la suma de sus distancias a dos puntos fijos F y F' llamados focos es constante (e igual al eje mayor, $2a$), es decir:

$$d(P, F) + d(P, F') = 2a.$$

En el caso de la métrica taxicab, la ecuación de la elipse de focos A y B viene dada por la expresión

$$|x - x(A)| + |y - y(A)| + |x - x(B)| + |y - y(B)| = 2a.$$

Después de introducir esta fórmula en el programa GeoGebra, vemos que la figura que se genera tiene forma de polígono. Concretamente, en el caso general la elipse tiene forma de octógono (véase la figura 10). Sin embargo, si movemos los focos vemos que no siempre tenemos octógonos. Cuando los dos focos están sobre una recta paralela a alguno de los ejes de coordenadas, la elipse adquiere forma de hexágono. Por otra parte, en el caso extremo en que los focos se encuentran a la distancia máxima $2a$, la elipse se convierte en un rectángulo o un cuadrado, incluyendo todos los puntos interiores (es decir, ya no sería una curva).

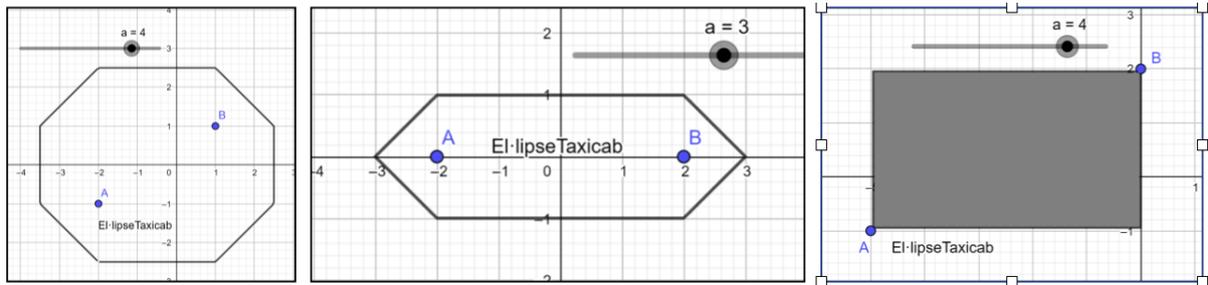


Figura 10: Elipses en la geometría taxicab.

3.3.3. Hipérbola

La hipérbola es el lugar geométrico de los puntos del plano tales que la diferencia entre sus distancias (en valor absoluto) a dos puntos fijos llamados focos es constante,

$$|d(P, F) - d(P, F')| = 2a.$$

En el caso de la distancia Manhattan, la ecuación de la hipérbola de focos A y B es

$$||x - x(A)| + |y - y(A)| - |x - x(B)| - |y - y(B)|| = 2a.$$

Podemos ver el resultado habitual en la figura 11.

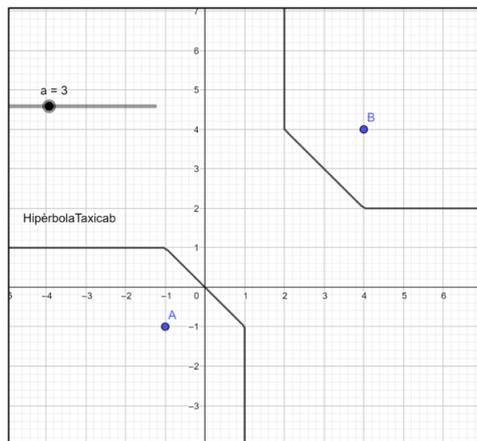


Figura 11: Hipérbola en la geometría taxicab.

Por otra parte, tenemos diferentes casos degenerados en los que la hipérbola deja de ser una curva. Dependiendo de la relación entre las distancias horizontales y verticales de los focos con el valor del parámetro $2a$ se dan varias situaciones: podemos tener líneas paralelas o cambiar parte de las curvas por todo un cuadrante infinito (ver figura 12).

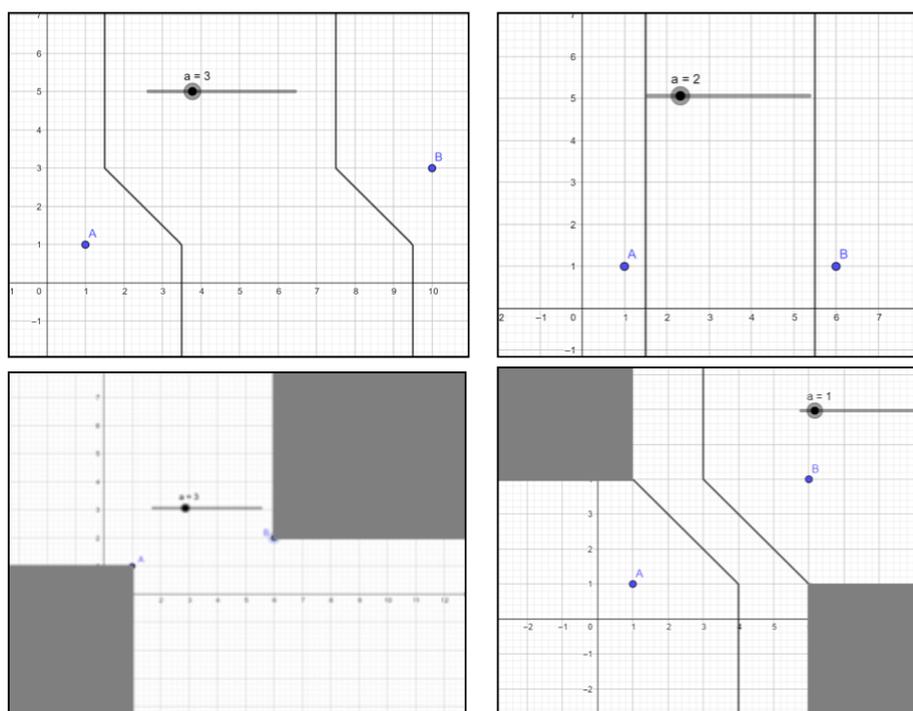


Figura 12: Casos degenerados de hipérbolas en la geometría taxicab.

3.3.4. Parábola

La parábola es el lugar geométrico de los puntos del plano que equidistan de un punto fijo llamado foco y de una recta llamada directriz. Se puede ver que, para la métrica de Manhattan, la ecuación de la parábola de directriz r y foco F viene dada por la expresión

$$|x - x(F)| + |y - y(F)| = d(P, F) = d(P, r) = \min\{|x - x_r|, |y - y_r|\},$$

donde x_r es la abscisa del punto de la recta r con la misma ordenada que $P(x, y)$ e y_r es la ordenada del punto de la recta r con la misma abscisa que $P(x, y)$.

La parábola tiene diferente orientación dependiendo de la pendiente m de la directriz. En la figura 13, vemos que la parábola se abre hacia arriba en el caso $|m| < 1$ (o hacia abajo si el foco está por debajo de la directriz) y que se abre hacia la derecha cuando $|m| > 1$ (o hacia la izquierda si el foco está a la izquierda de la directriz).

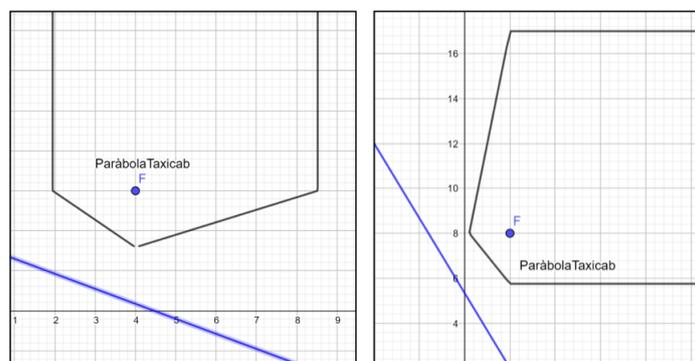


Figura 13: Parábolas en la geometría taxicab.

Finalmente, en la figura 14 vemos los casos $m = -1$, $m = 0$ (directriz horizontal) y $m = \infty$ (directriz vertical). En los casos de directrices con $|m| = 1$, las parábolas tienen tres lados en lugar de cuatro y las ramas infinitas son perpendiculares entre ellas.

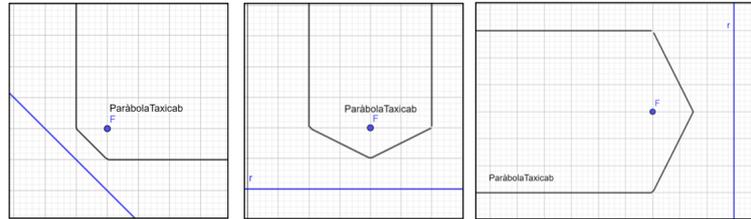


Figura 14: Parábolas en la geometría taxicab con directrices diagonal, horizontal y vertical.

4. Aplicaciones a la geometría urbana

La geometría taxicab puede ser más adecuada que la euclídea para el estudio de ciertas situaciones en la geometría urbana. Como ejemplo, analizaremos dos problemas clásicos. El primero consiste en, dados tres puntos, encontrar cuál es el lugar donde la suma de las distancias desde este lugar a los puntos es mínima. Esto sería útil a la hora de colocar un equipamiento de la manera más eficiente posible. Por ejemplo, si hay tres tiendas de una misma cadena, sería el punto donde pondríamos el almacén de distribución de los productos de las tres tiendas. El segundo consiste en dividir una región según las zonas más cercanas a ciertos puntos. Por ejemplo, si hay tres parques de bomberos en una región y hay un incendio, permitiría saber cuál de los tres parques debería encargarse de apagarlo.

4.1. El punto de Fermat

En una carta, Pierre de Fermat (1601-1665) propuso un problema como reto a E. Torricelli (1608-1647), un discípulo de Galileo: encontrar el punto tal que la suma de sus distancias a los vértices de un triángulo fuera mínima. La búsqueda de ese punto, conocido como punto de Fermat, es un problema clásico de geometría euclidiana que, sin embargo, tiene una solución bastante compleja. Veremos que en el caso de la geometría taxicab no es así.

Observemos la figura 15. Dado un triángulo cualquiera, trazamos por cada vértice las paralelas a los ejes de coordenadas. Siempre encontraremos dos de estas rectas que intersecan dentro del triángulo o, en el caso extremo, en un vértice. Este es el punto de Fermat en la geometría taxicab.

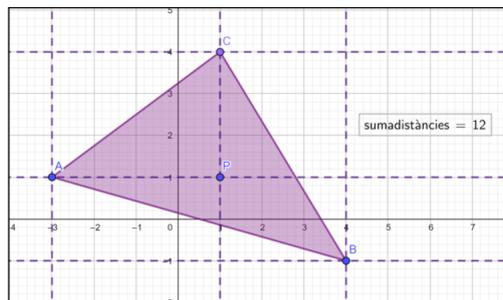


Figura 15: Punto de Fermat.

La demostración es muy intuitiva. Si nos situamos en este punto (al que llamaremos P) y hacemos un desplazamiento vertical de una unidad, nos alejamos una unidad de dos vértices y nos acercamos una al otro. Por tanto, en el cómputo global la suma de distancias se incrementa una unidad. Lo mismo ocurre para el caso horizontal. Concluimos, pues, que un movimiento en cualquier dirección incrementará la suma de distancias a los vértices, por lo cual el punto P es el que hace mínima esta suma.

Podríamos aplicar la construcción anterior al siguiente caso. Consideremos los tres centros de atención primaria (CAP) del Eixample de Barcelona: CAP Casanova, CAP Roger de Flor i CAP Passeig de Sant Joan (puntos verdes de la figura 16) y supongamos que quisiéramos situar un almacén de equipamiento médico de la manera más eficiente posible. Esto correspondería al Punto de Fermat (de color amarillo en la figura 16).

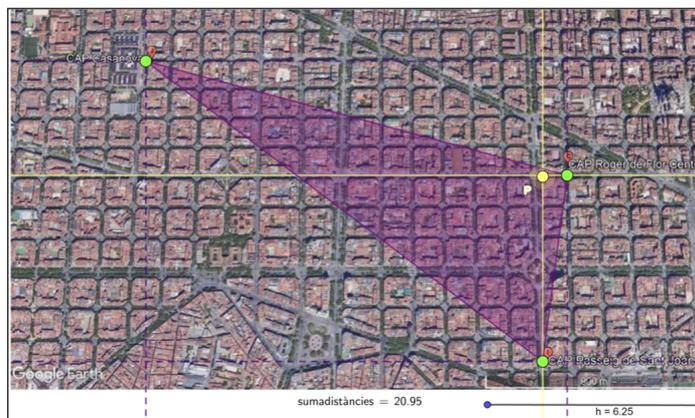


Figura 16: Punto de Fermat en una región de Barcelona.

4.2. Diagramas de Voronoi

Gueorgui Voronoi (1868-1908) fue un matemático ruso conocido por las denominadas regiones de Voronoi. Estas son unas construcciones geométricas que permiten configurar una partición del plano asociada a n puntos, de modo que a cada punto se le asigna una región formada por todos los puntos que son más cercanos a él que a los demás.

La manera de construir un diagrama de Voronoi es uniendo los puntos entre sí y trazando las mediatrices de los segmentos de la unión. Las intersecciones de estas mediatrices determinan una serie de polígonos alrededor de los puntos de manera que engloban las zonas más cercanas.

Actualmente, los diagramas de Voronoi tienen infinidad de aplicaciones en diferentes campos de estudio: gráficos por computadora, epidemiología, geofísica, meteorología... Muchas de estas aplicaciones se diseñan utilizando la métrica euclidiana, pero, como hemos visto en este trabajo, en algunos casos puede ser más adecuado el uso de otras métricas, como la de Manhattan.

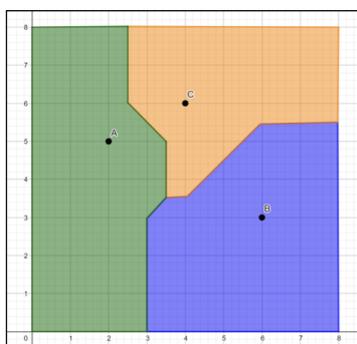


Figura 17: Diagrama de Voronoi con la métrica de Manhattan.

Para construir el diagrama de Voronoi (por ejemplo, de tres puntos) con la métrica de Manhattan, lo que tenemos que hacer es trazar las mediatrices de cada pareja de puntos. Las mediatrices dan lugar a diferentes regiones del espacio, no necesariamente disjuntas dado que una mediatriz puede no ser una

curva. Cada región representa el conjunto de coordenadas que están a menor o igual distancia de un punto que del opuesto. Una vez tenemos las regiones de esta forma, obtenemos el diagrama de Voronoi tomando sus intersecciones, lo que asegura que las coordenadas en las regiones obtenidas están a menor o igual distancia de un punto que del resto. Se puede ver un ejemplo en la figura 17.

Finalmente, consideremos un ejemplo práctico con los tres CAP del apartado anterior. Ahora se trataría de buscar qué CAP queda más cerca de cualquier punto del Eixample y dividir la región en tres zonas de influencia. El resultado lo podemos ver en la figura 18. El punto amarillo tendría los tres CAP a la misma distancia taxicab.

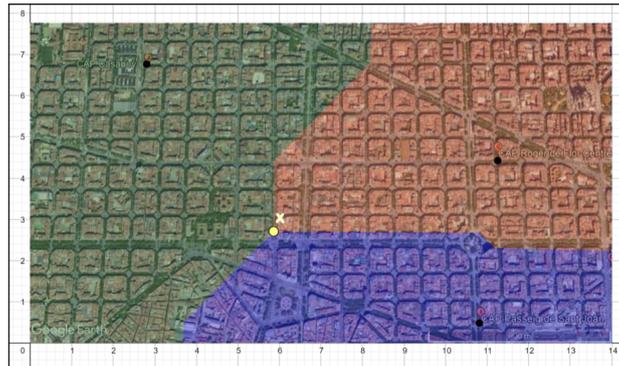


Figura 18: Diagrama de Voronoi en una región de Barcelona.

5. Generalizaciones

Finalmente, en este último apartado estudiaremos algunas generalizaciones que se pueden hacer de la métrica de Manhattan. La primera corresponde al caso tridimensional (también se podría hacer para n dimensiones, aunque la intuición geométrica desaparece). La siguiente es el caso en que nos moviéramos por el enrejado taxicab añadiendo movimientos por líneas con un ángulo de 45° , es decir, que siguiéramos los movimientos de las damas chinas. Finalmente, estudiamos la generalización por un enrejado con ángulos de 60° y toda una familia de métricas que generalizan la métrica de Manhattan por enrejados triangulares con cualquier ángulo.

5.1. Manhattan 3D

La métrica de Manhattan es muy fácil de generalizar a tres dimensiones. La fórmula para la distancia sería

$$d(P, Q) = |p_1 - q_1| + |p_2 - q_2| + |p_3 - q_3|,$$

para dos puntos $P(p_1, p_2, p_3)$ y $Q(q_1, q_2, q_3)$ cualesquiera de \mathbb{R}^3 .

Del mismo modo que la geometría taxicab se interpretaba mediante el recorrido de un taxi por las calles de Manhattan, la generalización a tres dimensiones se puede interpretar como el recorrido que haría un dron entre los rascacielos de Manhattan. Además de desplazarse a través del plano horizontal, también se podría mover en el eje vertical.

Análogamente a los apartados anteriores, entenderemos una esfera como el lugar geométrico de los puntos del espacio que equidistan de un punto fijo llamado centro. Esta distancia fija es el radio. En la geometría taxicab 3D, la ecuación de la esfera de centro $C(a, b, c)$ y radio r viene dada por la expresión

$$d((x, y, z), C) = |x - a| + |y - b| + |z - c| = r.$$

Ayudándonos de GeoGebra, podemos representar la superficie anterior y ver qué forma adopta. Como hemos visto anteriormente, la forma de la esfera dependerá de la distancia que estamos considerando y, en el caso de la métrica de Manhattan 3D, adquiere forma de octaedro, como podemos observar en la figura 19.

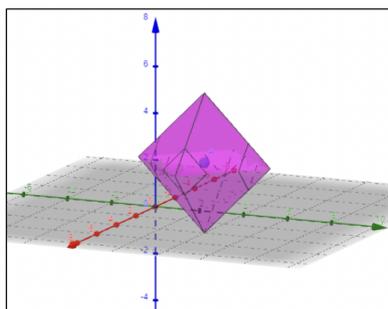


Figura 19: Esfera en la geometría taxicab 3D.

5.2. Manhattan triangular o geometría de las damas chinas

Al final de su libro [3], Krause plantea la posibilidad de estudiar nuevas métricas similares a la de Manhattan pero sobre parrillas triangulares. Más adelante se definieron y estudiaron. Concretamente, se investigó la distancia que permite movimientos diagonales además de los horizontales y verticales. Se la llamó métrica de las damas chinas. En el juego de las damas chinas, las piezas se pueden mover verticalmente (norte y sur), horizontalmente (este y oeste) y diagonalmente (nordeste, noroeste, sureste y suroeste), y de ahí el nombre de la métrica.

La definición de la métrica de las damas chinas no es muy obvia de entrada:

$$d(P, Q) = \max\{|p_1 - q_1|, |p_2 - q_2|\} + (\sqrt{2} - 1) \min\{|p_1 - q_1|, |p_2 - q_2|\}.$$

El camino más corto en la métrica de las damas chinas corresponde a avanzar por la diagonal hasta llegar a la misma altura del punto donde queremos terminar y después desplazarnos horizontal o verticalmente. Examinando con un caso particular como el de la figura 20, vemos que $\max\{|p_1 - q_1|, |p_2 - q_2|\} = 5$ y corresponde al lado largo de la cuadrícula, mientras que $\min\{|p_1 - q_1|, |p_2 - q_2|\}$ y corresponde al lado corto. Por lo tanto, la distancia entre P y Q en la métrica de las damas chinas sería

$$5 + (\sqrt{2} - 1) \cdot 3 = 2 + 3\sqrt{2},$$

que equivale a avanzar tres diagonales de longitud y dos unidades horizontales.

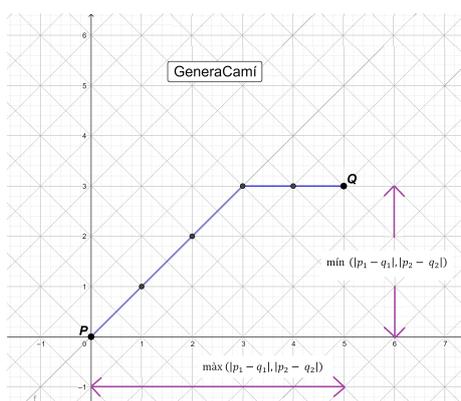


Figura 20: Ejemplo para la métrica de las damas chinas.

Para estudiar qué forma tienen las circunferencias en esta métrica, procederemos como en los apartados anteriores. En este caso, la expresión de la métrica complica la construcción. Una circunferencia de centro $C(a, b)$ y radio r en la métrica de las damas chinas vendrá dada por la expresión habitual $d(P, C) = r$, poniendo la fórmula de la distancia que estamos considerando.

En este caso la representación con GeoGebra se complica un poco y para llevarla a cabo hemos de trabajar en 8 regiones distintas. Terminada la construcción, vemos que la circunferencia con la métrica de las damas chinas es un octógono (figura 21).

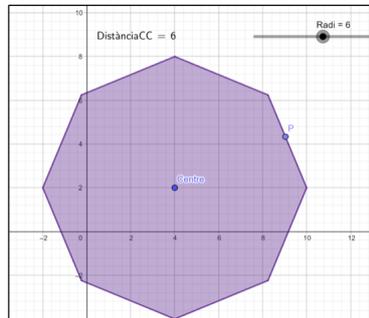


Figura 21: Circunferencia en la métrica de las damas chinas.

5.3. Alfa distancia

Tomando como ejemplo la geometría de las damas chinas, se pueden definir otras geometrías triangulares. Por ejemplo, nos podríamos mover a lo largo de la superposición de dos parrillas triangulares: una formada por triángulos con ángulos internos de 60° , y otra de iguales características pero rotada un ángulo de 90° . Podríamos hablar de este tipo de geometría como geometría del triángulo equilátero. En este caso, la fórmula de la distancia entre dos puntos sería

$$d(P, Q) = \max\{|p_1 - q_1|, |p_2 - q_2|\} + (2 - \sqrt{3}) \min\{|p_1 - q_1|, |p_2 - q_2|\}.$$

En la figura 22 podemos ver la circunferencia con la geometría del triángulo equilátero. Vemos que la circunferencia es un octógono también, pero de forma más cuadrada que la anterior.

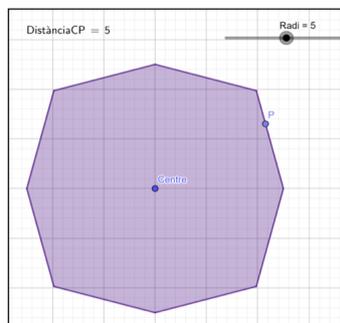


Figura 22: Circunferencia en la métrica del triángulo equilátero.

Pero aquí no acaban las generalizaciones: existen otras que constituyen todo un mundo. De hecho, se puede definir para cada valor de α una distancia (llamada α -distancia) dada por la fórmula

$$d(P, Q) = \max\{|p_1 - q_1|, |p_2 - q_2|\} + (\sec(\alpha) - \tan(\alpha)) \min\{|p_1 - q_1|, |p_2 - q_2|\}.$$

En el caso $\alpha = 0^\circ$ tenemos como caso particular la métrica de Manhattan; en el caso $\alpha = 45^\circ$ tenemos la métrica de las damas chinas, y en el caso $\alpha = 60^\circ$, la geometría del triángulo equilátero. Para los otros valores de α obtenemos una generalización de la geometría del triángulo equilátero, donde ambas parrillas ahora están formadas por triángulos isósceles de ángulos internos α , α y $180^\circ - 2\alpha$. En la figura 23, observamos que las circunferencias correspondientes son octógonos, evolucionando desde el cuadrado taxicab (que correspondería a $\alpha = 0^\circ$) y redondeándose a medida que aumenta el ángulo hasta ser prácticamente un cuadrado cuando $\alpha = 89^\circ$ (el caso $\alpha = 90^\circ$ no está bien definido). Las cuatro circunferencias corresponden a 1° , 30° , 60° y 89° , respectivamente.

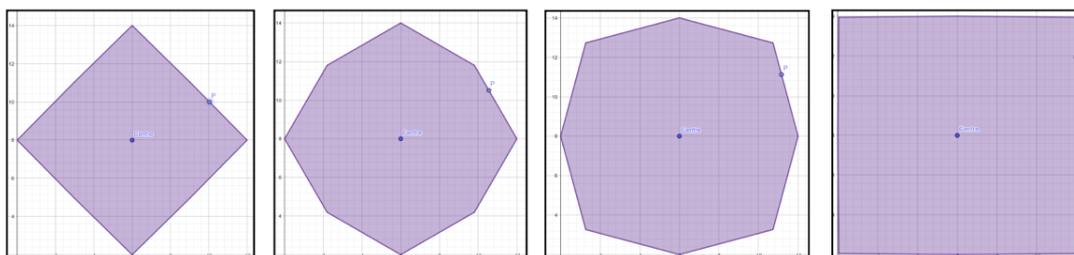


Figura 23: Circunferencias en la métrica de la alfa-distancia.

6. Conclusiones

En este trabajo hemos podido constatar cómo la geometría, que se inicia como una disciplina práctica para medir, evoluciona hasta el punto de ser una disciplina completamente abstracta. Esto conlleva la aparición de nuevas geometrías como la taxicab. Esta geometría, desde el punto de vista axiomático, se diferencia muy poco de la habitual euclidiana, y en cambio es mucho más adecuada en situaciones de geometría urbana, por ejemplo. Hemos podido ver que en esta nueva geometría la distancia más corta ya no es la línea recta y que, además, no hay solo un único camino más corto. Las circunferencias pasan a ser cuadrados, las elipses son hexágonos u octógonos, puede haber triángulos equiláteros de lados iguales pero diferentes entre ellos, el valor de π sería 4... Es una nueva geometría donde la intuición nos engaña constantemente. Como aplicación práctica de esta geometría, es muy interesante el estudio del punto de Fermat y de los diagramas de Voronoi. Muestran claramente que dentro de una ciudad con las calles cuadrículadas (como Barcelona o Manhattan) la geometría taxicab es la que se debe considerar, ya que no nos movemos en línea recta atravesando edificios, sino siguiendo las calles. Finalmente, hay toda una familia de generalizaciones que son muy interesantes y dan pie a seguir investigando sobre los diferentes elementos geométricos en estas nuevas métricas.

Referencias

- [1] GÓMEZ, Joan. *Cuando las rectas se vuelven curvas. Las geometrías no euclídeas*. El mundo es matemático 4. Barcelona, ES: RBA, 2010. ISBN: 978-84-9867-856-7.
- [2] JANSSEN, Christina. *Taxicab Geometry: Not the Shortest Ride Across Town*. Creative Component. Iowa State University, 2017. URL: <https://web.archive.org/web/20111216052147/www.math.iastate.edu/thesisarchive/MSM/JanssenMSMSS07.pdf>.
- [3] KRAUSE, Eugene F. *Taxicab Geometry. An Adventure in Non-Euclidean Geometry*. Nueva York, US: Dover Publications, 1986. ISBN: 978-0-486-25202-5.
- [4] SABATINI, Marco. «La geometria del taxi». En: *MATerials MATemàtics 2007 (2007)*, artículo 4. ISSN: 1887-1097. URL: <https://mat.uab.cat/web/matmat/wp-content/uploads/sites/23/2020/05/v2007n04.pdf>.
- [5] TIAN, Songlin. «Alpha-distance – A generalization of chinese checker distance and taxicab distance». En: *Missouri Journal of Mathematical Sciences* 17.1 (2005), págs. 35-40. ISSN: 0899-6180. <https://doi.org/10.35834/2005/1701035>.

TEMat

Este trabajo obtuvo una mención en la edición de 2020 del Premi Poincaré, entregado por la Facultat de Matemàtiques i Estadística de la Universitat Politècnica de Catalunya.



Métodos numéricos aplicados al teorema de Gauss sobre la distribución de los números primos

✉ Ferran Arnau Segarra^a
Universitat Autònoma de Barcelona
ferran.arnau@gmail.com

Resumen: El origen de este artículo está en cuestionarse qué pasaría si aplicásemos la función inversa de una función que tuviera como expresión la fórmula de Gauss, planteada en su teorema sobre la distribución de los números primos, con el objetivo de comprobar si de esta forma se podría describir o aproximar su comportamiento. Intentar responder esta pregunta implica la utilización de métodos numéricos.

Abstract: The aim of this article is to explore the possibility of describing the behaviour of prime numbers through the inverse of a function whose expression is the formula that Gauss came up with in his theorem about the distribution of prime numbers. Addressing this question demanded the use of numerical methods.

Palabras clave: función inversa, teorema sobre la distribución de los números primos, métodos numéricos.

MSC2020: 11A41, 11Y11.

Recibido: 27 de septiembre de 2020.

Aceptado: 22 de agosto de 2021.

Agradecimientos: Quisiera agradecer a mi tutora del Trabajo de Investigación de Bachillerato por su inestimable ayuda, a mis padres por el apoyo tanto logístico como emocional y al jurado de los Premios Poincaré por darme la oportunidad de publicar mi trabajo en esta revista.

Referencia: ARNAU SEGARRA, Ferran. «Métodos numéricos aplicados al teorema de Gauss sobre la distribución de los números primos». En: *TEMat*, 6 (2022), págs. 17-30. ISSN: 2530-9633. URL: <https://temat.es/articulo/2022-p17>.

^aEste trabajo fue realizado como Trabajo de investigación de Bachillerato en el Institut d'Àlella.

1. Introducción

Encontrar el siguiente número primo es algo que actualmente trae de cabeza a muchos matemáticos, pero, ¿por qué ese interés?

El comportamiento de estos números, aparentemente errático, esconde algunos patrones que parecen no ser exclusivos de estos números. Recientes descubrimientos vinculan el comportamiento de ciertos elementos a escala atómica con los números primos. Parece ser que hay ciertos patrones en los niveles energéticos de los átomos grandes que comparten propiedades muy parecidas con ciertos patrones de los números primos [2].

Además, gracias al teorema fundamental de la aritmética, sabemos que los números primos son algo semejante a un ladrillo en el edificio de los números enteros positivos, y esta última propiedad lleva quizás a las aplicaciones más destacadas. El ejemplo más relevante hoy en día es el papel de los números primos en criptografía. Cada transacción bancaria, o cualquier operación que requiera de un cifrado muy seguro, estará basada en estos números. La idea inicial es relativamente simple: si tomamos dos primos muy grandes, multiplicarlos es un trabajo relativamente sencillo, pero una vez combinados, es muy costoso determinar cuáles han sido los dos primos originales. Por ejemplo, si se tiene el número 999 962 000 357, llegar a la conclusión de que es resultado de multiplicar 999 979 y 99 983 requiere de algo más que de una calculadora. Si los dos primos anteriores fueran extremadamente grandes, ni siquiera con un ordenador muy potente sería realista intentar factorizar o «descifrar» el número resultante. Con esta idea se generan las claves de los sistemas de cifrado modernos.

Pero encontrar números primos grandes no es tarea sencilla, sobre todo porque, como se ha mencionado anteriormente, estos parecen estar repartidos sin orden aparente entre los naturales. Los estudios actuales solo hablan de aproximaciones al comportamiento de los números primos. Esto se debe a que no se conoce ninguna fórmula que pueda ser calculada eficientemente con la propiedad de que genere la sucesión de números primos.

En el siguiente apartado se estudiará una de estas aproximaciones, y se pondrá la base para los apartados siguientes, definiendo una función que aproxima el comportamiento de los números primos. En los otros apartados se tratará de encontrar una inversa de dicha función, que acabará llevando a una familia de funciones, y se estudiará cómo estas funciones pueden o no ser usadas para obtener el k -ésimo número primo. Además, se introducirá una nueva variable, $C(k)$, que indicará la calidad de dicha aproximación. En los apartados finales se contemplará la adición de un nuevo parámetro, m , en la estructura de la familia de funciones, que llevará a plantear un estudio completamente nuevo relacionado con los valores de $C(k)$.

2. La aproximación de Gauss

Uno de los intentos de aproximar el comportamiento de los números primos data del siglo XVIII y su autor es considerado uno de los mejores matemáticos de su tiempo. Johann Carl Friederich Gauss (30 de abril de 1777-23 de febrero de 1855) fue un matemático muy prolífico [5], conocido por la genialidad de sus razonamientos. Con solo 14 años planteó una conjetura en la que se basa este artículo. En dicha conjetura, Gauss propone una fórmula que aproxima la cantidad de números primos anteriores o iguales a un número dado. Dicha fórmula es $x/\ln(x)$, donde x es el número del cual se quiere conocer la cantidad de números primos anteriores a él.

Esta fórmula se obtiene a través de un proceso empírico que involucra una tabla de números primos y una tabla de logaritmos. Gauss también definió una función $\pi(x)$, a través de la cual se obtendría el número exacto de números primos anteriores a un número dado. Este estudio le llevó a proponer el cuadro 1 [3].

De esta tabla se puede observar, por ejemplo, que, de entre los cien primeros números naturales, uno de cada cuatro es primo, o que, de entre los primeros mil números naturales, uno de cada seis es primo. Esta afirmación no se ha de interpretar de manera literal, ya que se trata de una aproximación.

Pero la conclusión más importante que Gauss extrajo de esta tabla fue que los valores de la última columna, correspondientes a los valores de $x/\pi(x)$, crecían un poco más de dos unidades respecto a la fila anterior (es decir, cada vez que se multiplica por 10). Este comportamiento fue de inmediato relacionado con el número e , un número irracional ($e \approx 2,718 28$) que es la base de los logaritmos neperianos.

Cuadro 1: Valores de $\pi(x)$ y $x/\pi(x)$.

x	$\pi(x)$	$x/\pi(x)$	x	$\pi(x)$	$x/\pi(x)$
10	4	2,5	10^5	9592	10,4
10^2	25	4	10^6	78 498	12,7
10^3	168	6,0	10^7	664 579	15,0
10^4	1229	8,1	10^8	5 761 455	17,4

Así pues, se puede definir la aproximación

$$\frac{x}{\pi(x)} \approx \ln(x).$$

Si se despeja, se obtiene que

$$\pi(x) \approx \frac{x}{\ln(x)}.$$

Esta conjetura fue posteriormente demostrada y enunciada como un teorema.

Teorema 1. Sea $\pi : \mathbb{R} \rightarrow \mathbb{N}$ la función que denota la cantidad de números primos que son menores o iguales a x ,

$$\pi(x) = |\{p > 0 \text{ primo} : p \leq x\}|.$$

Entonces, se cumple que

$$\lim_{x \rightarrow +\infty} \frac{\pi(x)}{x/\ln(x)} = 1.$$

La demostración de este teorema es bastante compleja, y fue demostrado formalmente por dos matemáticos de forma independiente: Jacques Hadamard y Charles-Jean de la Vallée Poussin en el año 1896. La demostración completa se puede encontrar en el capítulo xxii de *Introduction à la théorie des nombres* [4]. Esta expresión se puede tomar como una aproximación de la información buscada, y la pregunta inmediata sería: ¿cómo de buena es dicha aproximación? Para responder a esta pregunta podemos utilizar el cuadro 2, que incluye la fiabilidad de la aproximación.

Cuadro 2: Error de la fórmula de Gauss para los valores estudiados.

x	$\pi(x)$	$x/\ln(x)$	Error relativo ^a de $x/\ln(x)$
10	4	4,3429	0,0857
10^2	25	21,7147	0,1314
10^3	168	144,7648	0,1383
10^4	1229	1085,7362	0,1165
10^5	9592	8685,8896	0,0944

^aEl error relativo está definido como $\left| \frac{\text{aproximación obtenida}}{\text{magnitud real}} - 1 \right|$.

A medida que x aumenta, la diferencia entre $\pi(x)$ y $x/\ln(x)$ se va haciendo cada vez mayor. Sin embargo, se sigue del teorema anterior que el error relativo tiende a cero cuando x tiende a infinito. Esta tendencia se puede observar en la figura 1.

3. Aplicación de la función inversa

Procedamos a analizar los elementos de la función $\pi(x)$, y qué es exactamente lo que representan.

Pongamos un ejemplo. Si calculásemos $\pi(8)$, el resultado tendría que ser 4, ya que hay 4 números primos anteriores a 8: 2, 3, 5 y 7. Recordemos que $\pi(x)$ nos da la cantidad de números primos menores o iguales a un número dado, por lo que, si x es un número primo, $\pi(x)$ aumentará su valor en una unidad respecto al valor de x anterior. De esta forma $\pi(2)$ sería igual a 1, $\pi(3)$ sería igual a 2, etc. En otras palabras, cuando x es un número primo, el valor de $\pi(x)$ es su posición como número primo.

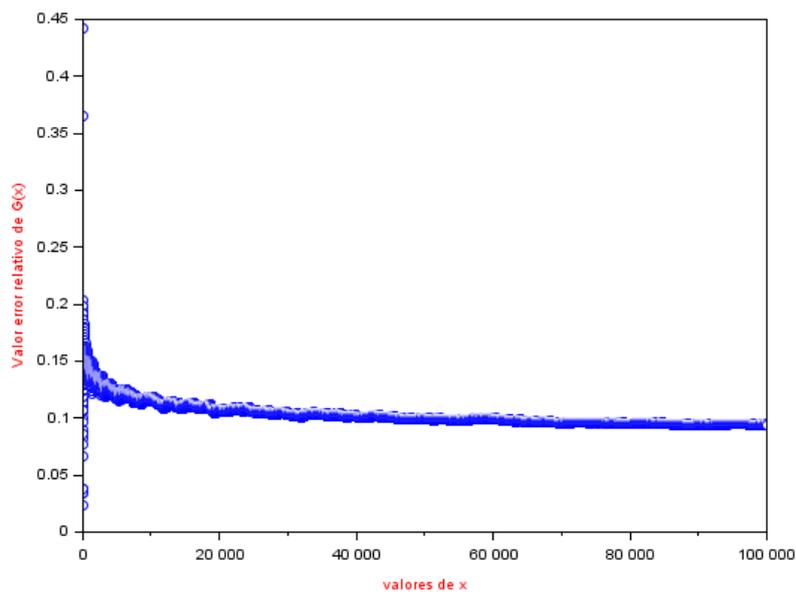


Figura 1: Comportamiento del error relativo de $G(x)$ para los primeros 100 000 valores de x .

De momento, tal y como está planteado el problema, la función $\pi(x)$ nos permite, dado un número primo, obtener la posición que este ocupa en una lista ordenada de los mismos. Pero, ¿y si invirtiéramos los papeles de x y $\pi(x)$? Es decir, si el dato fuera la posición, y el resultado fuera el número primo correspondiente a dicha posición. En otras palabras, cambiar los papeles de x y de y . Este procedimiento no tendría ningún sentido si lo aplicáramos a $\pi(x)$ directamente ya que, por ejemplo, $\pi(8)$, $\pi(9)$ y $\pi(10)$ tienen la misma imagen, y si invirtiéramos los papeles de x y de y , llegaríamos a la conclusión de que el cuarto número primo es a la vez 8, 9 y 10, cosa que es absurda. Entonces, lo ideal sería intercambiar los papeles de x y de y en la expresión de Gauss. Seguidamente se procederá a ver la viabilidad de dicha acción.

La herramienta matemática que invierte los papeles de x y de y es la función inversa, así que primero procedamos a definir una nueva función $G(x) = x/\ln(x)$. Esta función tiene una asíntota vertical en $x = 1$, luego 1 no sería del dominio. Calculando la derivada de $G(x)$, $G'(x) = (\ln(x) - 1)/\ln(x)^2$, se observa que $G(e) = e$ y $G'(e) = 0$, luego puede haber un extremo relativo en $x = e$. Como $\lim_{x \rightarrow 1^+} G(x) = +\infty$, ha de decrecer hasta $x = e$ y $\lim_{x \rightarrow +\infty} G(x) = +\infty$, se deduce que es un mínimo. Además, la derivada es siempre positiva en $(e, +\infty)$, luego se puede afirmar que es absoluto. Este último hecho también demuestra la inyectividad de $G(x)$ en $(e, +\infty)$, haciendo conceptualmente posible esta inversión de papeles en dicho intervalo. Nótese que se ha excluido el 2 de este conjunto, aunque su exclusión tiene un efecto mínimo en futuros resultados. Dado que $G(e) = e$, y a partir de $x = e$, $G(x)$ es continua y creciente, la imagen de $G(x)$ será $(e, +\infty)$.

Entonces, definimos $G(x) = x/\ln(x)$, con dominio $(e, +\infty)$, para poder hablar de su inversa. Procedamos a calcular dicha inversa:

$$y = \frac{x}{\ln(x)} \implies y \cdot \ln(x) = x.$$

Se puede observar que la expresión resultante es una función implícita. Para empezar, hay dos variables. Se puede solventar este problema examinando qué representa cada una de ellas. La y es la posición del número primo y esta posición se determina *a priori*, por lo que podemos definir y como un parámetro libre, al cual se le da un valor inicial. Para evitar confusiones, substituiremos la y por el parámetro k . Por lo tanto, de los resultados anteriores, se puede deducir que $k \in (e, +\infty)$. De hecho, al representar una posición, dicho parámetro solo podría ser un valor natural por lo que, en contexto, $k \in [3, +\infty)$.

Pero de la expresión anterior no se puede despejar x . Solo se puede obtener que x es igual a otra expresión en función de x , cosa poco práctica para trabajar. Si se transforma la igualdad anterior en una expresión igualada a cero, la x que buscaríamos sería la que anula dicha expresión. Es decir,

$$k \cdot \ln(x) - x = 0.$$

Recordemos que x representaba el número del cual queremos saber la cantidad de primos anteriores a él, o en nuestro caso, el número primo del cual queremos conocer la posición que ocupa en una hipotética lista ordenada de números primos. Por lo que, en teoría, el número primo de posición k es aquel que anula la expresión encontrada.

En resumen, ahora en vez de una sola función tenemos una familia de funciones con la estructura $F_k(x) = k \cdot \ln(x) - x$, donde el supuesto número primo en la posición k será un cero de cada respectiva $F_k(x)$, de tal manera que sea x_k dicho cero, $G^{-1}(k) = x_k$ (recordemos que $k \geq 3$). Para encontrar dicho cero, es necesario recurrir a métodos numéricos para calcular ceros de funciones. Nótese que se ha utilizado el artículo «un» en vez de «el» para referirse al cero. Dicha cuestión se tratará en el siguiente apartado.

No hemos de perder de vista que estamos trabajando con aproximaciones. Obviamente, para $k = 5$, el cero de $F_5(x)$ no será exactamente el quinto número primo, por lo que aparece un error que estudiaremos más adelante.

3.1. Análisis de $F_k(x)$

3.1.1. Características generales

El dominio de $F_k(x)$ lo determina la función logarítmica, dado que los otros elementos de $F_k(x)$ son una constante y una resta del valor de x . Por lo tanto, el dominio de $F_k(x)$ será el del logaritmo neperiano, es decir, $(0, +\infty)$, y la función es continua en su dominio.

Dado que, para todo k , $\lim_{x \rightarrow 0} F_k(x) = -\infty$, podemos afirmar que todas las funciones de la familia tienen una asíntota vertical en $x = 0$. Si se estudia el otro extremo del dominio de $F_k(x)$, vemos que las imágenes tienden también a menos infinito, $\lim_{x \rightarrow +\infty} F_k(x) = -\infty$.

La convexidad de la función no cambia. La primera derivada de $F_k(x)$ es $k/x - 1$, cosa que indica un posible extremo relativo cuando $x = k$. Teniendo en cuenta la asíntota vertical en $x = 0$ y que $\lim_{x \rightarrow +\infty} F_k(x) = -\infty$, podemos asegurar que se trata de un máximo absoluto. Dicha afirmación se confirma cuando se observa que, si $x < k$, la derivada es positiva, y si $x > k$, la derivada es negativa.

La segunda derivada, $-k/x^2$, asegura que no hay ningún punto de inflexión, ya que nunca se anula (recordemos que $k \neq 0$).

3.1.2. Ceros de la función

Este es un apartado clave del análisis de $F_k(x)$, ya que nos dará la información más relevante para la investigación. En esta sección se analizará cuándo se anula $F_k(x)$, y cuántas veces lo hace.

Veamos que $F_k(x)$ tiene como máximo dos ceros.

El parámetro k tiene un efecto similar al de un factor de traslado, dado que, cuando aumenta el valor de k , los valores de $F_k(x)$ también aumentan (a partir de cierto x), y gráficamente tenemos funciones similares dispuestas como se puede observar en la figura 2. Experimentalmente se puede comprobar que el número de ceros varía entre 0 y 2 observando que $F_1(x)$ y $F_2(x)$ no cortan el eje X, pero a partir de $F_3(x)$ las gráficas ya cortan dos veces el eje X.

En efecto, si k es mayor que un determinado valor, la función ha de cortar al menos una vez el eje X para alcanzar el valor de $F_k(x)$ en $x = k$, y, tal y como nos indica el límite $\lim_{x \rightarrow +\infty} F_k(x) = -\infty$, es necesario que $F_k(x)$ vuelva a cortar el eje X. Dicho de otra forma, para $x < k$, $F_k(x)$ es creciente, en $x = k$ hay un máximo absoluto y para $x > k$ la función es decreciente. Por lo tanto, uno de los dos ceros será menor que k , y el otro será mayor. Resumiendo, tenemos el siguiente resultado.

Proposición 2. Sea $k \in \mathbb{R}_{\geq 0}$. Consideramos la función $F_k(x) = k \ln(x) - x$. Dicha función tiene las siguientes propiedades.

- (i) Si $k < e$, entonces $F_k(x) \neq 0$ para todo $0 \leq x$.
- (ii) Si $k = e$, $F_k(x) = 0$ si y solo si $x = e$.
- (iii) Si $k > e$, $F_k(x) = 0$ tiene exactamente dos soluciones $x_0, x_k \geq 0$ tales que $x_0 < e$ y $x_k > e$. Además, $G(x_k) = k$.

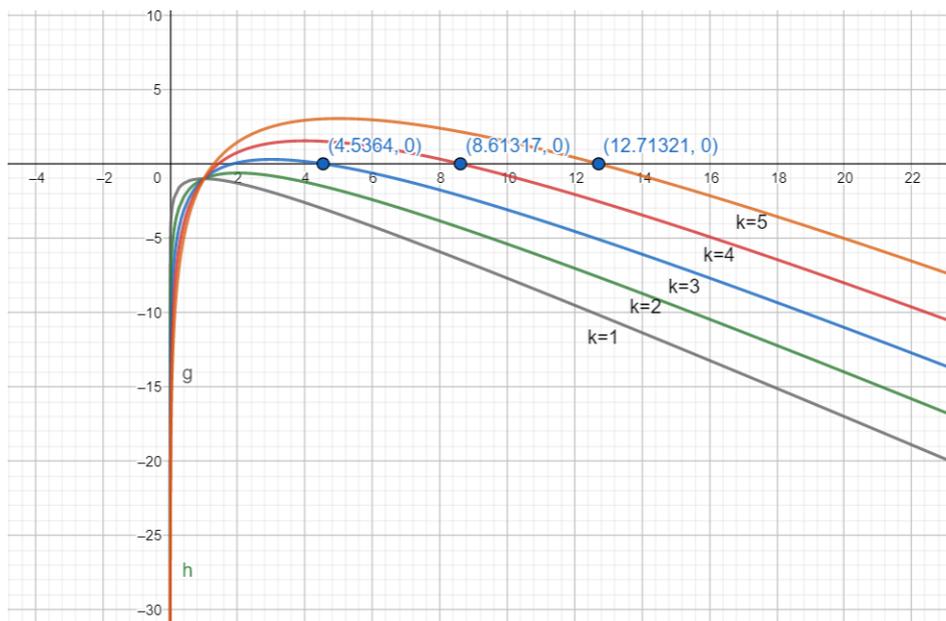


Figura 2: Ilustración de algunas de las primeras gráficas consecutivas de $F_k(x)$.

Demostración. Encontrar los ceros de $F_k(x)$ es equivalente a resolver la ecuación $k = x/\ln(x)$. Como la expresión de la derecha de la igualdad es $G(x)$, a efectos prácticos se está buscando la intersección de $y = k$ y $G(x)$. Como se ha visto que en $x = e$ hay un mínimo, automáticamente se demuestra (I). Cuando $k = e$, $y = k$ representa la tangente de $G(x)$ en $x = e$, luego solo interseca una vez $G(x)$, luego $F_e(x)$ se anula si y solo si $x = e$, demostrando así (II). Al ser $G(x)$ continua, $\lim_{x \rightarrow 1^+} G(x) = +\infty$ y $\lim_{x \rightarrow +\infty} G(x) = +\infty$, en el intervalo $(1, e)$, $G(x)$ asume todos los valores de $(e, +\infty)$, y en $(e, +\infty)$ también, pero en el primer intervalo los valores son decrecientes, y en el segundo son crecientes, luego para $k > e$, $y = k$ interseca dos veces con $G(x)$, luego $F_k(x)$ tendrá dos ceros, demostrando así (III). ■

Para las funciones que cortan dos veces al eje X, ¿cuál de los dos ceros nos da la información relevante para la solución del problema? Esta pregunta tiene una respuesta clara: el número primo será mayor o igual que la posición asignada, ya que los elementos de la sucesión de los números primos son naturales que distan como mínimo en una unidad, luego podemos deducir que el cero que nos interesa es el segundo, ya que es el que tiene el mismo comportamiento que los números primos.

3.1.3. Método óptimo a utilizar para encontrar los ceros de la función

El método óptimo es aquel que se adapta mejor a la función, o, simplemente, al entorno del cero de la función.

La función, después del máximo absoluto en $x = k$, tiene una tendencia descendente, y según la información que nos da la segunda derivada, que no presenta cambio de signo, se puede esperar que la recta tangente en un punto $x_0 > k$ aproxime muy bien la tendencia de $F_k(x)$, permitiendo que el corte de dicha tangente con el eje X sea muy cercano al cero de $F_k(x)$. El método que utiliza tangentes para poder aproximar ceros es el de Newton-Raphson, o método de la tangente [1].

El cuadro 3 muestra que, comparado con otros métodos conocidos, el método de la tangente es más efectivo para distintos valores de k (con una precisión requerida de 6 decimales).

El hecho de que los parámetros iniciales estén más o menos cerca del cero puede afectar al número de pasos necesarios para obtener la aproximación con la precisión requerida, pero en este caso estaríamos hablando de un efecto muy pequeño sobre los datos de la tabla anterior, por lo que podemos calificarlos de válidos. En las pruebas, el método de bisección [1] y el de la *regula falsi* [6] tienen los mismos intervalos; sin embargo, el de la secante [1] no, ya que para poder diferenciarlo del de la *regula falsi* se han escogido imágenes que tengan el mismo signo. Por último, el parámetro inicial del método de la tangente es el valor medio de los intervalos de la secante.

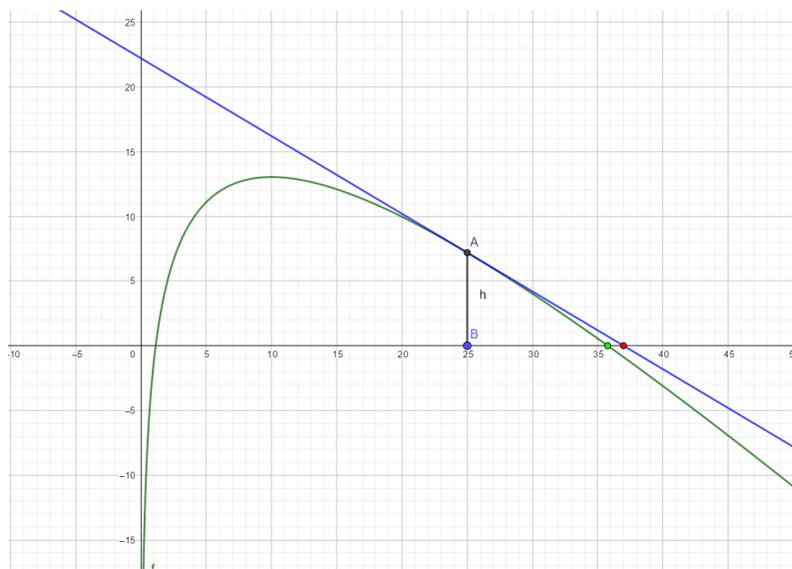


Figura 3: Ejemplo de una primera iteración del método de la tangente aplicada a $F_{10}(x)$.

Cuadro 3: Rapidez con la que algunos de los métodos más conocidos convergen hacia el cero de $F_k(x)$ en comparación con el método de la tangente.

	Bisección		Regula falsi		Secante		Tangente	
k	Parámetros iniciales	Número de pasos	Parámetros iniciales	Número de pasos	Parámetros iniciales	Número de pasos	Parámetros iniciales	Número de pasos
10	$x_0 = 20$ $x_1 = 45$	26	$x_0 = 20$ $x_1 = 45$	6	$x_0 = 11$ $x_1 = 25$	6	$x = 18$	4
10^2	$x_0 = 250$ $x_1 = 840$	30	$x_0 = 250$ $x_1 = 840$	6	$x_0 = 146$ $x_1 = 350$	6	$x = 248$	4
10^3	$x_0 = 5500$ $x_1 = 12\,000$	34	$x_0 = 5500$ $x_1 = 12\,000$	6	$x_0 = 2521$ $x_1 = 5500$	6	$x = 4010$	4
10^4	$x_0 = 60\,000$ $x_1 = 150\,000$	38	$x_0 = 60\,000$ $x_1 = 150\,000$	6	$x_0 = 28\,000$ $x_1 = 70\,000$	6	$x = 49\,000$	4

3.1.4. Sobre la implementación de los métodos numéricos en este estudio

Cada método ha sido programado en C++. El usuario necesita introducir los datos iniciales para que el programa haga la primera iteración, y el mismo programa ya va aumentando el valor de k , adecuando los parámetros iniciales cada vez que se trabaja con un nuevo valor de k .

El método escogido finalmente es el de la tangente y el programa realizado permite guardar la aproximación de cada cero en una lista para su posterior disposición y manipulación.

4. Error en las aproximaciones obtenidas con $F_k(x)$

Por comodidad, en este apartado se trabajará con las siguientes definiciones. Para $k \geq 1$, P_k será el k -ésimo número primo. Para $k \geq 3$, definimos \tilde{x}_k como la aproximación de x_k obtenida con el método de la tangente.

En este apartado, además de estudiar el error relativo cometido en la aproximación, se añadirá un nuevo elemento: una constante. Dicha constante es el valor por el que se debería multiplicar la aproximación para obtener el número primo correcto, y se la denominará $C(k)$: si $k \in \mathbb{Z}_{\geq 3}$, $C(k) = P_k / \tilde{x}_k$, donde x_k está definido en la proposición 2.

Si se consiguiera modelizar estos valores, sería sencillo implementarlos en el proceso de obtención de la aproximación, y de esta forma reducir el error cometido,

$$\begin{cases} k \ln(x) - x = 0, \\ xC(k) \approx R_k. \end{cases}$$

En el cuadro 4 se puede observar el comportamiento del error y de $C(k)$ para valores comprendidos entre 3 y 16 (recordemos que, para $k = 1$ y $k = 2$, $F_k(x)$ no corta el eje X).

Cuadro 4: Comportamiento del error para ciertos valores de k .

Valor de k	\tilde{x}_k	R_k	$C(k)$	Error relativo
3	4,5364	5	1,1021	0,0927
4	8,6131	7	0,8127	0,2304
5	12,7132	11	0,8652	0,1557
6	16,9988	13	0,7647	0,3076
7	21,4649	17	0,7919	0,2626
8	26,0934	19	0,7451	0,3733
9	30,8672	23	0,7451	0,3420
10	35,7715	29	0,8107	0,2335
11	40,7938	31	0,7599	0,3159
12	45,9238	37	0,8056	0,2411
13	51,1525	41	0,8015	0,2476
14	56,4727	43	0,7614	0,3133
15	61,8773	47	0,7595	0,3165
16	67,3610	53	0,7868	0,2709

El error está relacionado con $C(k)$ de manera que si esta se va acercando a 1, el error será más pequeño, ya que si en algún momento $C(k)$ fuera 1 significaría que la aproximación misma corresponde al valor real. El cambio de $C(k)$ y del error cada vez que k aumenta su valor en uno es muy pequeño y se puede observar que tiene un comportamiento oscilante, no tiene una progresión lineal.

Sin embargo, si se estudian los valores de $C(k)$ para valores de k grandes, se aprecia que la tendencia general es ascendente, sin llegar a alcanzar el valor 1, por lo que el error va, aparentemente, disminuyendo, tal y como se puede observar en el cuadro 5.

Cuadro 5: Comportamiento del error para ciertos valores de k .

Valor de k	\tilde{x}_k	R_k	$C(k)$	Error relativo
9987	116 505,2633	104 623	0,898 01	0,113 57
9988	116 518,0228	104 639	0,898 04	0,113 52
9989	116 530,7824	104 651	0,898 05	0,113 51
9990	116 543,5421	104 659	0,898 02	0,113 55
9991	116 556,3019	104 677	0,898 08	0,113 48
9992	116 569,0619	104 681	0,898 01	0,113 56
9993	116 581,8219	104 683	0,897 93	0,113 66
9994	116 594,5821	104 693	0,897 92	0,113 68
9995	116 607,3423	104 701	0,897 89	0,113 71
9996	116 620,1027	104 707	0,897 84	0,113 77
9997	116 632,8632	104 711	0,897 78	0,113 85
9998	116 645,6238	104 717	0,897 73	0,113 91
9999	116 658,3845	104 723	0,897 68	0,113 97
10 000	116 671,1453	104 729	0,897 64	0,114 02

Se sigue apreciando el cambio lento del error y de $C(k)$, incluso de forma más exagerada. También se conservan las oscilaciones de los valores de $C(k)$ y del error, pero son extremadamente más pequeñas.

Viendo el comportamiento de $C(k)$, es tentador pensar que este crecimiento tan lento significa que acabará convergiendo a 1, cosa que implicaría que para valores muy grandes de k el error relativo sería muy pequeño. Presentamos este resultado.

Proposición 3. *Para todo $k \geq 3$ consideramos x_k y P_k , definidos anteriormente. Entonces, se cumple que*

$$\lim_{k \rightarrow +\infty} \frac{x_k}{P_k} = 1.$$

Para demostrar este resultado se utilizará el siguiente lema.

Lema 4. *Se tiene que*

$$\lim_{k \rightarrow +\infty} \frac{P_k}{k \ln(k)} = 1.$$

Demostración. Por el teorema 1, tenemos que

$$\lim_{x \rightarrow +\infty} \frac{\pi(x)}{x/\ln(x)} = 1.$$

Como $\lim_{k \rightarrow +\infty} P_k = +\infty$, entonces

$$\lim_{k \rightarrow +\infty} \frac{\pi(P_k)}{P_k/\ln(P_k)} = 1.$$

Por definición de $\pi(x)$, $\pi(P_k) = k$ para todo $k \geq 1$. Por lo tanto,

$$\lim_{k \rightarrow +\infty} \frac{k \ln(P_k)}{P_k} = 1.$$

Como $\ln(x)$ es continua, se tiene que

$$\lim_{k \rightarrow +\infty} \ln\left(\frac{k \ln(P_k)}{P_k}\right) = \ln(1) = 0.$$

Desarrollando y dividiendo por $\ln(P_k)$ se obtiene que

$$\lim_{k \rightarrow +\infty} \left(\frac{\ln(k)}{\ln(P_k)} + \frac{\ln(\ln(P_k))}{\ln(P_k)} - 1 \right) = 0.$$

Usando l'Hôpital, se puede observar que $\lim_{k \rightarrow +\infty} \ln(\ln(x))/\ln(x) = 0$ y, como $\lim_{k \rightarrow +\infty} P_k = +\infty$, entonces

$$\lim_{k \rightarrow +\infty} \frac{\ln(\ln(P_k))}{\ln(P_k)} = 0.$$

De aquí se deduce que

$$\lim_{k \rightarrow +\infty} \frac{\ln(k)}{\ln(P_k)} = 1.$$

Usando esta última igualdad y que $\lim_{k \rightarrow +\infty} k \ln(P_k)/P_k = 1$, se completa la demostración del lema. ■

Demostración de la proposición 3. Para $k \geq 3$ se tiene que

$$k = G(x_k) = \frac{x_k}{\ln(x_k)},$$

lo que implica que

$$1 = \frac{x_k}{k \ln(x_k)}.$$

Tomando logaritmos, dividiendo por $\ln(x_k)$ y tomando límite cuando k tiende a infinito se deduce que

$$\lim_{k \rightarrow +\infty} \frac{\ln(k)}{\ln(x_k)} = 1,$$

donde se ha usado que $\lim_{x \rightarrow +\infty} x_k = +\infty$. Combinando las dos últimas igualdades se obtiene que

$$\lim_{k \rightarrow +\infty} \frac{x_k}{k \ln(k)} = 1.$$

Utilizando entonces el lema 4 acaba la prueba. ■

Observación. En la investigación se ha trabajado con una aproximación de x_k dada por el método de la tangente, a la que se ha denominado \tilde{x}_k (para $k \geq 3$). Dicha aproximación se ha obtenido con un programa con un error mínimo definido, de tal forma que $|x_k - \tilde{x}_k| \leq 1 \cdot 10^{-7}$. Luego el resultado anterior también se aplica a \tilde{x}_k , ya que

$$\lim_{k \rightarrow +\infty} \frac{\tilde{x}_k}{P_k} = \lim_{k \rightarrow +\infty} \frac{x_k \pm 10^{-7}}{P_k} = 1.$$

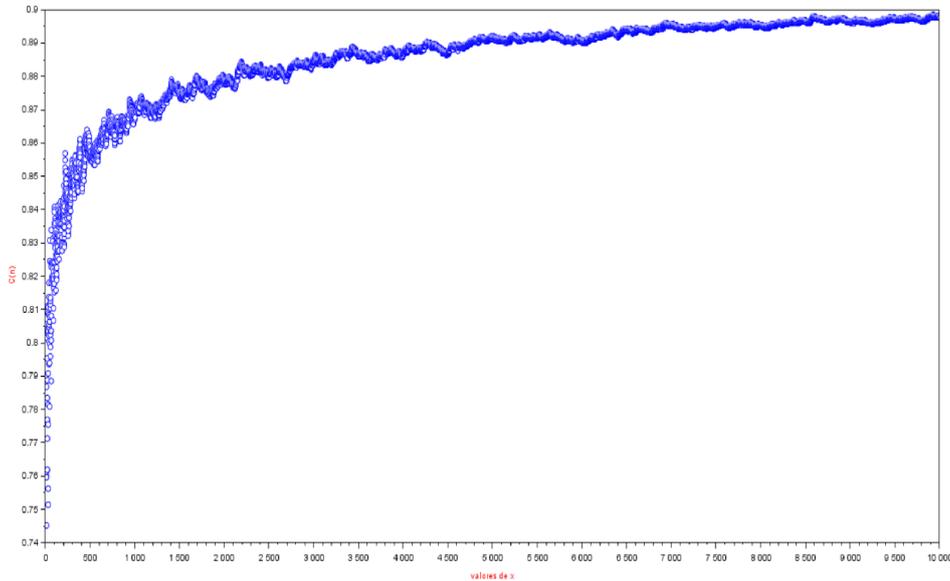


Figura 4: Comportamiento de $C(k)$ para los primeros 10 000 valores de x .

5. Conclusiones iniciales

Teniendo en cuenta los resultados obtenidos en el apartado anterior, podemos decir que, aunque en cierto modo se haya conseguido hablar de la inversa de $G(x)$, en sí misma la aproximación obtenida con el método descrito en este artículo no es muy buena. Es verdad que el error relativo tiene una tendencia decreciente y se ha demostrado que $C(k)$ tiende a 1 (cosa que implica que el error relativo tiende a 0), pero si se mira la diferencia entre el valor obtenido y el valor real, se puede observar que dicha diferencia es muy grande y aumenta cada vez que k lo hace. Por ejemplo, para $k = 10\,000$ el número primo correspondiente a esta posición es 104 729, mientras que la parte entera de la aproximación que se obtiene es 116 671. Erro por más de 12 000 unidades. Es decir, el error relativo tiende a cero, pero el error absoluto va creciendo.

Sin embargo, sí que se generan más preguntas respecto al problema original. ¿Qué pasaría si estudiásemos los valores de $C(k)$? ¿A qué se debe su comportamiento tan reacio al crecimiento? Dichas preguntas serán la base de un nuevo estudio.

6. Planteamiento de un nuevo estudio a partir de los resultados obtenidos

Gauss no fue el único que exploró la estructura $x/\ln(x)$. Adrien-Marie Legendre, matemático francés, propuso un cambio en la fórmula de Gauss e introdujo una constante restando al logaritmo neperiano en el denominador, concretamente $-1,083\,66$ (de ahora en adelante la estructura de Legendre será referida como $L(x)$, donde $L(x) = x/(\ln(x) - 1,083\,66)$). Esta modificación aparentemente reducía el error relativo en gran medida, para los valores iniciales, como se puede apreciar en el cuadro 6.

Sin embargo, se puede apreciar que, mientras el error relativo utilizando $G(x)$ va decreciendo, el error relativo usando $L(x)$ va creciendo, luego cabe la posibilidad de que, para valores muy grandes de x , con $G(x)$ se obtengan mejores aproximaciones que con $L(x)$, aunque es algo que no sabemos.

Cuadro 6: Comparación entre las estructuras de Gauss y de Legendre.

x	$G(x)$	$L(x)$	$\pi(x)$	Error relativo de $G(x)$	Error relativo de $L(x)$
28 422 049	1 656 038,33	1 767 648,555	1 766 023	0,062 27	0,000 92
59 462 017	3 321 742,97	3 535 788,263	3 532 046	0,059 54	0,001 05
91 483 589	4 990 466,67	5 304 008,298	5 298 069	0,058 05	0,001 12
124 143 001	6 661 124,08	7 072 351,595	7 064 092	0,057 04	0,001 16
157 272 821	8 333 000,39	8 840 601,529	8 830 115	0,056 29	0,001 18
190 776 457	10 005 786,4	10 608 739,169	10 596 138	0,055 71	0,001 18
224 600 587	11 679 798,4	12 377 296,964	12 362 161	0,055 19	0,001 22
258 687 353	13 354 272,2	14 145 605,831	14 128 184	0,054 77	0,001 23
293 014 009	15 029 649,8	15 914 234,165	15 894 207	0,054 39	0,001 26
327 545 129	16 705 403,7	17 682 702,852	17 660 230	0,054 06	0,001 27
362 267 359	18 381 841,1	19 451 397,428	19 426 253	0,053 76	0,001 29
397 151 353	20 058 323,4	21 219 693,734	21 192 276	0,053 50	0,001 29
432 194 897	21 735 389,3	22 988 201,069	22 958 299	0,053 26	0,001 30
467 376 781	23 412 566,3	24 756 455,327	24 724 322	0,053 05	0,001 29
502 699 721	25 090 446,0	26 525 109,235	26 490 345	0,052 84	0,001 31
538 148 867	26 768 720,1	28 293 862,735	28 256 368	0,052 64	0,001 32
573 710 677	28 447 096,3	30 062 429,528	30 022 391	0,052 47	0,001 33

El cambio de Legendre es más interesante por lo que representa más que por lo que consigue. Legendre deja claro que un factor restando en el denominador afecta, y mucho, al resultado final. Aplicando el mismo proceso a $L(x)$ que a $G(x)$, definimos $F_k^L(x)$, y denotamos sus ceros como x_k^L , y la aproximación de sus ceros como \tilde{x}_k^L . La aproximación obtenida tendría sus propias constantes, o valores de $C(k)$. Además, el hecho de que inicialmente con $L(x)$ se obtenga un error relativo tan bajo parece indicar que las aproximaciones de $F_k^L(x)$ podrían ser, inicialmente al menos, notablemente mejores que las de $F_k(x)$. Véase, pues, la diferencia entre las dos aproximaciones para valores de x relativamente pequeños (cuadro 7), donde se confirma que para ciertos valores de x , con $F_k^L(x)$ se obtienen mejores aproximaciones. ¿Qué pasaría si definiéramos otro parámetro real libre, m , donde $L(x)$ y $G(x)$ serían casos particulares ($m = -1,083 66$ y $m = 0$, respectivamente)?.

Cuadro 7: Comparación entre las aproximaciones obtenidas con $F_k(x)$ y $F_k^L(x)$.

x	\tilde{x}_k	\tilde{x}_k^L	R_k	$C(k)$ de $F_k(x)$	$C(k)$ de $F_k^L(x)$
9987	116 505,2633	104 606,8800	104 623	0,898 01	1,000 15
9988	116 518,0228	104 618,4599	104 639	0,898 04	1,000 19
9989	116 530,7824	104 630,0400	104 651	0,898 05	1,000 20
9990	116 543,5421	104 641,6201	104 659	0,898 02	1,000 16
9991	116 556,3019	104 653,2003	104 677	0,898 08	1,000 22
9992	116 569,0619	104 664,7807	104 681	0,898 01	1,000 15
9993	116 581,8219	104 676,3611	104 683	0,897 93	1,000 06
9994	116 594,5821	104 687,9417	104 693	0,897 92	1,000 04
9995	116 607,3423	104 699,5224	104 701	0,897 89	1,000 01
9996	116 620,1027	104 711,1031	104 707	0,897 84	0,999 96
9997	116 632,8632	104 722,6840	104 711	0,897 78	0,999 88
9998	116 645,6238	104 734,2650	104 717	0,897 73	0,999 83
9999	116 658,3845	104 745,8461	104 723	0,897 68	0,999 78
10 000	116 671,1453	104 757,4274	104 729	0,897 64	0,999 72

Con el nuevo parámetro, la estructura a invertir sería $x/(\ln(x) + m)$, que denotaremos como $G_m(x)$. La función $G_m(x)$ tiene una asíntota en $x = e^{-m}$. La derivada de $G_m(x)$ es $(\ln(x) + m - 1)/(\ln(x) + m)^2$, donde se observa que e^{1-m} anula la derivada, luego en $x = e^{1-m}$ puede haber un extremo relativo, y haciendo un estudio similar al de $G(x)$ se llega a la conclusión de que es un mínimo absoluto. Por tanto, a partir de $x = e^{1-m}$ la derivada será siempre positiva, y, dado que $G_m(x)$ es continua, también aseguramos su inyectividad en el intervalo $(e^{1-m}, +\infty)$. Como $G_m(e^{1-m}) = e^{1-m}$, estableceremos el dominio y la imagen de $G_m(x)$ como $(e^{-m}, +\infty)$.

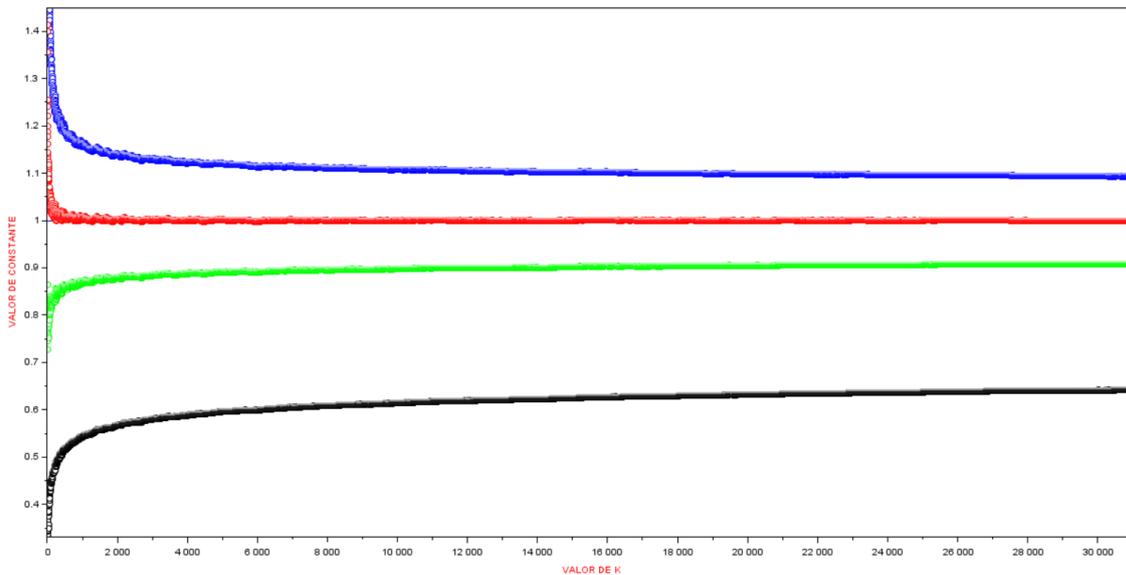


Figura 5: Representación de $C(k)$ para $m = 0$ (verde), $m = -1,083\ 66$ (rojo), $m = -2$ (azul) y $m = 5$ (negro).

Sea entonces $F_k^m(x) = k(\ln(x) + m) - x$, donde el supuesto número primo en la posición k -ésima será un cero de cada respectiva $F_k^m(x)$, de tal manera que, si x_k^m es dicho cero, $G_m^{-1}(k) = x_k^m$. De la misma forma que con $G(x)$, $k \in (e^{1-m}, +\infty)$, donde $k \geq \lfloor e^{1-m} \rfloor + 1$ ($\lfloor x \rfloor$ representa la parte entera de x), y, al representar una posición, k será siempre natural.

Cabe destacar también que, para cada valor de m , $C(k)$ variará. La figura 5 muestra las gráficas de $C(k)$ para diferentes valores de m . Nótese como todas las gráficas tienen una tendencia aparente a acercarse hacia 1, en especial con $m = -1,083\ 66$.

Corolario 5. Para todo $k \geq \lfloor e^{1-m} \rfloor + 1$ consideramos x_k^m y P_k definidos anteriormente. Entonces,

$$\lim_{k \rightarrow +\infty} \frac{x_k^m}{P_k} = 1.$$

Demostración. Es análoga a la de la proposición 3. Se sigue cumpliendo que $\lim_{k \rightarrow +\infty} \ln(k)/\ln(x_k^m) = 1$, y se puede seguir aplicando el lema 4 ya que se cumplía independientemente del parámetro m . ■

Observación. Como el error para calcular las aproximaciones de x_k^m es el mismo que el de x_k , se cumple que $|x_k^m - \tilde{x}_k^m| \leq 1 \cdot 10^{-7}$, luego con el mismo razonamiento que con \tilde{x}_k , se tiene que $\lim_{k \rightarrow +\infty} \tilde{x}_k^m/P_k = 1$. ◀

Si se observan a la escala de la figura 5, parece que correspondan a gráficas de una función que bien podría ser racional. Se ve una clara diferencia entre los valores de m negativos (cuyos valores iniciales son superiores a 1) y los valores de m positivos y 0 (cuyos valores iniciales son inferiores a 1). Estos últimos tienen tendencia ascendente mientras que para los primeros la tendencia es decreciente.

Si se amplía la escala, se puede observar que la apariencia continua que se tiene «desde lejos» es falsa. Nos encontramos con una distribución caótica de puntos, pero con algunas características notables como que, para distintos valores de m , los puntos tienen la misma distribución en su forma para los mismos valores de k . Por ejemplo, como muestran las figuras 6 y 7 para $m = -2$ y $m = 0$.

Este hecho se puede corroborar si se observa el cuadro 8. Se puede observar que, a medida que x va creciendo, la diferencia entre los valores de $C(k)$ de cada respectiva m tiene un cambio muy lento, como si hubiera realmente un factor de traslado.

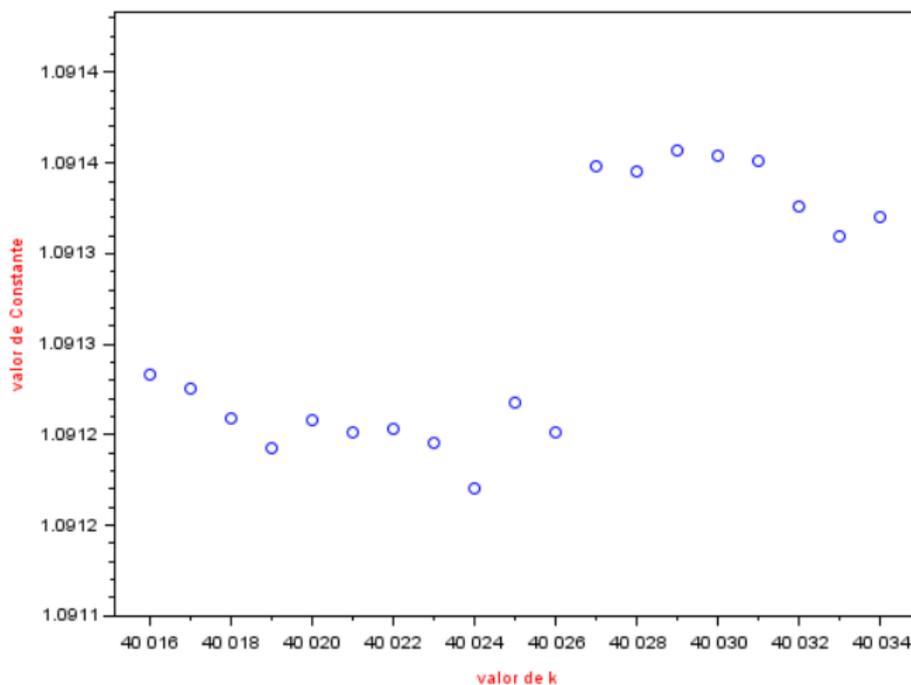


Figura 6: $C(k)$ para $m = -2$ para valores de k entre 40 016 y 40 034.

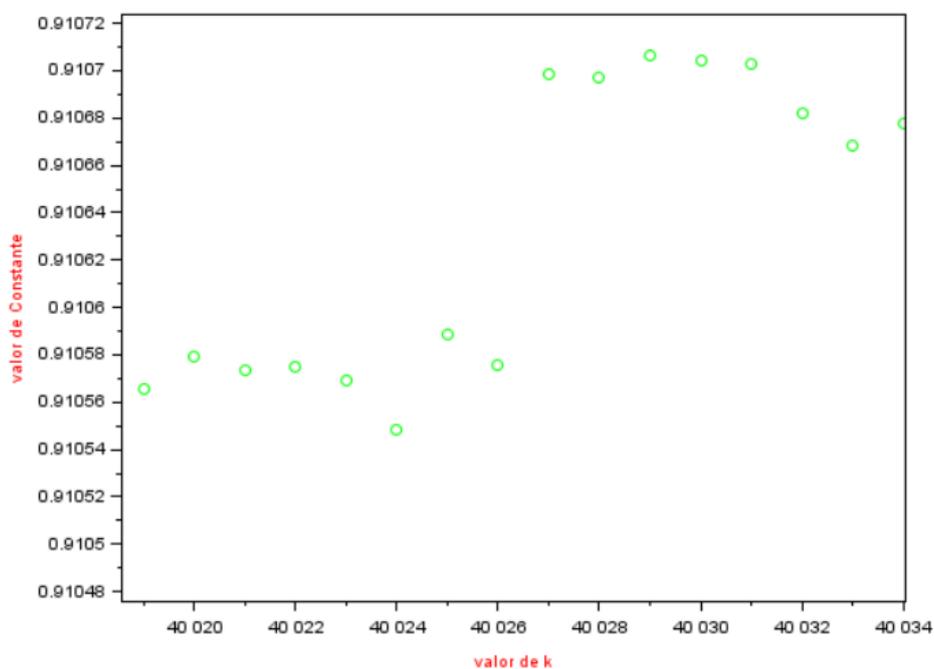


Figura 7: $C(k)$ para $m = 0$ para valores de k entre 40 020 y 40 034.

Cuadro 8: Diferencia, en valor absoluto, entre los valores respectivos de $C(k)$ para $m = -2$ y $m = 0$.

k	Diferencia entre los respectivos valores de $C(k)$	k	Diferencia entre los respectivos valores de $C(k)$
10 000	0,209 767	100 000	0,165 663
10 001	0,209 770	100 001	0,165 663
10 002	0,209 776	100 002	0,165 664
10 003	0,209 755	100 003	0,165 664
10 004	0,209 753	100 004	0,165 666
10 005	0,209 739	100 005	0,165 667
10 006	0,209 734	100 006	0,165 665

Referencias

- [1] AGUILÓ GOST, Francesc; BOADAS ELVIRA, Joan, y GARRIGA VALLE, Ernest. *Temes de càlcul*. Temes clau. Barcelona, ES: Universitat Politècnica de Catalunya. Servei de Publicacions, 1991. ISBN: 978-84-7653-116-7.
- [2] CAMACHO MEDINA, José de Jesús. «Números Primos: Enigmas y Aplicaciones». En: *MasScience* (25 de jun. de 2019). URL: <https://www.massscience.com/2019/06/25/numeros-primos-enigmas-y-aplicaciones/>.
- [3] GRACIÁN, Enrique. *Los números primos. Un largo camino al infinito*. El mundo es matemático. Barcelona, ES: RBA, 2010. ISBN: 978-84-473-6625-5.
- [4] HARDY, Godfrey Harold y WRIGHT, Edward Maitland. *Introduction à la théorie des nombres*. Trad. por Sauvageot, F Con pról. de Goldstein, Catherine. 5.ª ed. Berlín, Heidelberg, DE: Springer, 2007. ISBN: 978-3-540-64332-6.
- [5] MORENO CASTILLO, Ricardo. *Gauss. El príncipe de los matemáticos*. 1.ª ed. La matemática en sus personajes. Madrid, ES: Nivola, 2018. ISBN: 978-84-15913-38-2.
- [6] WEISSTEIN, Eric W. *Method of False Position*. MathWorld—A Wolfram Web Resource. URL: <https://mathworld.wolfram.com/MethodofFalsePosition.html> (visitado 06-2020).

TEMat

Fundamentos de la computación cuántica

✉ Vicente López Oliva
Universitat Jaume I
al341918@uji.es

Resumen: La computación cuántica es un nuevo paradigma que permite abordar algunos problemas intratables con los ordenadores actuales. El impacto de los futuros ordenadores cuánticos puede ser enorme en múltiples áreas como la criptografía, la simulación cuántica o el aprendizaje automático. En este artículo describimos los fundamentos de la computación cuántica, incluyendo tanto las bases matemáticas que la sustentan como las propiedades de la mecánica cuántica que aprovecha para lograr toda su potencia computacional. En concreto, introducimos conceptos como la superposición, el entrelazamiento o el paralelismo cuántico. También mostramos cómo las transformaciones de estados cuánticos se implementan en forma de puertas con las que se construyen circuitos y crean algoritmos cuánticos. Para ilustrar estos últimos mostramos el funcionamiento de un algoritmo cuántico sencillo pero sorprendente, que nos permite la teleportación de estados cuánticos.

Abstract: Quantum computing is a new paradigm that makes it possible to tackle some problems which are intractable with today's computers. The impact of future quantum computers could be enormous in many areas such as cryptography, quantum simulation and machine learning. In this paper we describe the fundamentals of quantum computing, including both the mathematical foundations that underpin it and the properties of quantum mechanics that it harnesses to achieve its full computational power. In particular, we introduce concepts such as superposition, entanglement and quantum parallelism. We also show how quantum state transformations are implemented in the form of gates with which to build circuits and create quantum algorithms. To illustrate these, we show the implementation of a simple but surprising quantum algorithm that allows us to teleport quantum states.

Palabras clave: computación cuántica, cúbit, entrelazamiento, espacio vectorial complejo, algoritmos cuánticos, teleportación cuántica.

MSC2020: 81P68, 68Q12.

Recibido: 2 de marzo de 2021.

Aceptado: 8 de julio de 2021.

Agradecimientos: Quiero mostrar mi agradecimiento al profesor Ximo Gual por darme la oportunidad de realizar el TFG [11], y al profesor José Manuel Badía, sin cuya ayuda no habría sido posible desarrollar el TFG y este artículo no habría visto la luz.

Referencia: LÓPEZ OLIVA, Vicente. «Fundamentos de la computación cuántica». En: *TEMat*, 6 (2022), págs. 31-47. ISSN: 2530-9633. URL: <https://temat.es/articulo/2022-p31>.

© Este trabajo se distribuye bajo una licencia Creative Commons Reconocimiento 4.0 Internacional <https://creativecommons.org/licenses/by/4.0/>

1. Introducción

En la actualidad, tenemos computadores capaces de realizar tareas extremadamente complejas y costosas en segundos¹. No obstante, existen problemas para los que no se conoce ninguna solución con coste polinómico y que se clasifican en la clase NP [12, cap. 3]. Las soluciones a estos problemas tienen un coste exponencial, que los hace intratables para la computación clásica.

Por ejemplo, la simulación realista del comportamiento de sistemas cuánticos sencillos, tales como pequeñas moléculas o reacciones químicas, teniendo en cuenta los efectos de la mecánica cuántica, tiene un coste espacial y temporal tan elevado que no es posible abordarla con los ordenadores clásicos actuales ni previsible en el futuro. Debido a ello, en 1982 el físico Richard Feynman propuso la creación de ordenadores que sacaran provecho de ese mismo tipo de propiedades cuánticas para poder llevar a cabo este tipo de simulaciones [6].

Las propiedades que muestran los sistemas cuánticos han sorprendido por igual a físicos, matemáticos y todo tipo de científicos desde que fueron descubiertas, debido a que contradicen el comportamiento habitual de los objetos macroscópicos que estamos acostumbrados a observar a nuestro alrededor. A pesar de ello, propiedades como el entrelazamiento o la superposición se han podido aprovechar para diseñar algoritmos cuánticos que podrán resolver en horas problemas tan importantes como la factorización de números enteros, que se encuentra en la base de gran parte de los sistemas criptográficos actuales [9, cap. 4]. De hecho, la computación cuántica ha permitido definir nuevas clases de complejidad, tales como BQP, que incluye los problemas que pueden resolverse con una probabilidad acotada de error usando circuitos cuánticos de tamaño polinómico. Lo llamativo es que esta clase incluye algunos de los problemas NP antes mencionados, tales como el de factorización [12, cap. 1].

Se están realizando grandes avances en la construcción de ordenadores cuánticos, aunque faltan años para poder superar las diferentes dificultades tecnológicas a las que nos enfrentamos tales como, por ejemplo, el tratamiento de errores [12, cap. 10]. No obstante, el aumento de potencia de cálculo asociado a estos ordenadores puede suponer grandes avances en campos con un enorme potencial económico y social, como pueden ser los campos de la química [19], la farmacología [7] o el aprendizaje automático [17, cap. 1], entre otros.

2. Unidad básica de información

2.1. Definición de cúbit

Sabemos que la computación clásica, la que opera en los ordenadores que podemos encontrar en nuestro día a día, utiliza una unidad básica de información que llamamos bit. Estos bits representan estados y son utilizados no solo para hacer funcionar el propio ordenador, sino también para almacenar información. Empecemos por definir qué es un bit.

Definición 1. Un **bit** $b \in \{0, 1\}$ es la unidad mínima de información empleada tanto en los computadores clásicos como en la teoría de la información clásica. ◀

Es importante tener en cuenta que en cada momento un bit solo puede estar en uno de sus dos estados posibles, 0 o 1. Dicho de otro modo, ambos estados son excluyentes. La composición de múltiples bits acaba dando lugar a la información con la que trabaja nuestro ordenador. Por ejemplo, para representar la letra Q haciendo uso de los bits clásicos, utilizaremos el conjunto de bits 01011001, según la representación del código ASCII. El homólogo al bit clásico y el que es la unidad de información básica en computación cuántica es el bit cuántico o cúbit.

Podemos dar una primera definición de cúbit de forma análoga al bit de la siguiente forma.

Definición 2. Un **cúbit** ψ es la unidad mínima de información cuántica empleada tanto en los computadores cuánticos como en la teoría de la información cuántica. ◀

¹<https://top500.org>

Un cúbit, al igual que un bit clásico, puede estar en dos estados básicos que se representan utilizando una notación especial. La *notación de Dirac*, también denominada *notación bra-ket*, fue propuesta por primera vez por Dirac [4]. Haremos uso de la notación *ket* para representar el vector v de la forma $|v\rangle = [v_1, \dots, v_n]^T$ y la notación *bra* para expresar el vector w^\dagger , que es el transpuesto conjugado de w , de la forma $\langle w| = [\bar{w}_1, \dots, \bar{w}_n]$. Utilizando esta notación, el producto escalar de w^\dagger y v vendría representado como $\langle w|v\rangle$. Esta notación es utilizada en mecánica cuántica para hacer referencia a vectores en un espacio de Hilbert y está justificada porque un espacio de este tipo y su dual son isomorfos. De este modo, cada *bra* corresponde exactamente a un *ket* y viceversa [12].

La diferencia esencial respecto a los bits clásicos es que los bits cuánticos aprovechan propiedades de la mecánica cuántica para conseguir en algunas circunstancias un aumento enorme en la capacidad de información que pueden almacenar y procesar. La propiedad que explotan para lograrlo es la *superposición de estados*, que hace que un cúbit no esté restringido a sus dos estados básicos, sino que puede estar en una combinación lineal de ambos. Esto da lugar a un conjunto infinito de estados posibles en los que puede estar el cúbit. Más formalmente, los estados de un cúbit se definen de la siguiente forma.

Definición 3. Sean $\alpha, \beta \in \mathbb{C}$ con $|\alpha|^2 + |\beta|^2 = 1$. Definimos el **estado de un cúbit** ψ sobre la base estándar $\{|0\rangle, |1\rangle\}$ como

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle,$$

donde α y β se denominan **amplitudes de estado**. Es habitual representar el estado en forma vectorial como $[\alpha, \beta]^T$, lo que nos permite ver que la base estándar viene dada por $\{|0\rangle, |1\rangle\}$. ◀

Según la definición anterior, un cúbit puede estar en una combinación lineal de estados básicos. Sin embargo, las amplitudes que determinan dicha combinación no pueden ser obtenidas de forma simultánea. Solo podemos obtenerlas de forma experimental bajo ciertas condiciones utilizando técnicas como la tomografía cuántica. Para acceder a la información del cúbit, es necesario realizar una *medida*. La medida de un cúbit en la base estándar nos dará un único valor 0 o 1, «colapsando» el estado de nuestro cúbit en uno de los estados básicos con probabilidad $|\alpha|^2$ de medir un 0 y $|\beta|^2$ de medir un 1. El hecho de que la probabilidad total de medir uno de los dos estados sea 1 explica la condición impuesta en la definición ($|\alpha|^2 + |\beta|^2 = 1$). Una vez el estado ha colapsado, no importa cuántas veces volvamos a realizar la medida en la misma base, el resultado siempre será el mismo.

La medición de estado de un cúbit puede interpretarse geoméricamente como la proyección del vector asociado $|\psi\rangle$ sobre alguno de los vectores de una determinada base (por ejemplo, $\{|v_1\rangle, |v_2\rangle\}$). En concreto, la probabilidad de obtener el estado $|v_i\rangle$ tras realizar la medida viene dada por $p(|v_i\rangle) = |\langle v_i|\psi\rangle|^2$.

Hemos visto que el cúbit suele representarse como un vector. En concreto como un vector complejo en un espacio de Hilbert. Vamos a concretar brevemente esta definición.

Definición 4. Un **espacio vectorial complejo** es un espacio vectorial cuyas propiedades se extienden al dominio de los complejos. ◀

Si añadimos el producto escalar $\langle \cdot | \cdot \rangle$ y una propiedad adicional a un espacio vectorial, obtenemos un espacio de Hilbert, que podemos definir de la siguiente forma.

Definición 5. Un **espacio de Hilbert** es un espacio vectorial completo cuya norma procede de un producto escalar. ◀

Que el espacio vectorial sea completo significa que cualquier secuencia de Cauchy de vectores del mismo converge a algún vector que también pertenece al espacio [10]. El producto escalar dota de una métrica al espacio vectorial, lo que nos permite definir la longitud de los vectores usando la siguiente norma:

$$(1) \quad \|\psi\| = \sqrt{\langle \psi | \psi \rangle}.$$

Una vez definido el concepto de espacio de Hilbert, podemos dar una nueva definición tanto de cúbit como de los estados que puede adoptar.

Definición 6. Un **cúbit** es un espacio de Hilbert de dos dimensiones. ◀

Definición 7. Un **estado puro de un cúbit** es cualquier vector unitario (de longitud 1) en el espacio de Hilbert. ◀

A partir de (1) y de la definición 3, podemos ver que la longitud de todo vector representando un estado puro viene dada por

$$\|\psi\| = \sqrt{\bar{\alpha} * \alpha + \bar{\beta} * \beta} = \sqrt{|\alpha|^2 + |\beta|^2} = 1,$$

donde $\bar{\alpha}$ y $\bar{\beta}$ son las amplitudes complejas conjugadas y $*$ representa el producto de números complejos.

La unitariedad de los vectores representando estados cuánticos, y no solo los basados en un solo cúbit sino también en varios, es fundamental puesto que, como hemos comentado anteriormente, las componentes de dichos vectores están asociadas a las probabilidades de cada uno de los estados básicos y estas siempre tienen que sumar 1. Podemos ver más detalles en el libro de Loceff [10].

Tengamos en cuenta que los estados puros son aquellos de los que conocemos toda la información. Sin embargo, también existen los estados mezcla, de los que no se dispone de toda la información. Estos se suelen definir mediante una combinación de distintos estados puros con una determinada distribución de probabilidad y se representan mediante matrices de densidad [12, cap. 2]. En lo que resta de artículo, nos centraremos en las propiedades y funcionamiento de los estados puros.

2.2. Esfera de Bloch

En este apartado, vamos a ver una representación alternativa de los cúbits que nos ayudará a visualizar sus estados y las transformaciones que hagamos sobre ellos. Según la definición 3, un cúbit viene dado por dos números complejos, lo que quiere decir que tenemos cuatro valores reales, que no pueden representarse gráficamente de forma sencilla. No obstante, si revisamos la definición con más detalle nos daremos cuenta de que podemos representar un cúbit en un espacio de tres dimensiones aprovechando ciertas restricciones sobre los valores posibles de ambos números complejos. Para llegar a esta representación comenzaremos por usar la forma polar para las dos amplitudes:

$$\alpha = r_0 e^{i\phi_0}, \quad \beta = r_1 e^{i\phi_1}.$$

A partir de ellas podemos representar nuestro cúbit como

$$(2) \quad |\psi\rangle = r_0 e^{i\phi_0} |0\rangle + r_1 e^{i\phi_1} |1\rangle.$$

Si extraemos $e^{i\phi_0}$ como factor común, obtenemos que

$$|\psi\rangle = e^{i\phi_0} (r_0 |0\rangle + r_1 e^{i(\phi_1 - \phi_0)} |1\rangle),$$

donde ϕ_0 se denomina *fase global* y $(\phi_1 - \phi_0)$ se denomina *fase relativa*.

La fase global, tal y como la hemos definido, cumple la siguiente propiedad.

Propiedad 8. *La fase global de un cúbit no altera la probabilidad de medir sus estados base.*

Demostración. Sea ψ un cúbit con $\alpha, \beta \in \mathbb{C}$ de forma que $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$. Si le aplicamos una fase global $c = e^{i\phi}$, obtenemos que

$$|c\psi\rangle = c\alpha|0\rangle + c\beta|1\rangle.$$

Dado que $|c| = |e^{i\phi}| = 1$, para todo ϕ , las probabilidades de medir los estados base $|0\rangle$ y $|1\rangle$, respectivamente, son

$$\begin{aligned} |0\rangle &= |c\alpha|^2 = |c|^2 |\alpha|^2 = |\alpha|^2, \\ |1\rangle &= |c\beta|^2 = |c|^2 |\beta|^2 = |\beta|^2. \end{aligned} \quad \blacksquare$$

Esto hace que los efectos de la fase global no sean observables mediante ningún experimento que mida el estado del cúbit.

Por el contrario, la fase relativa sí puede medirse si elegimos la base adecuada y, además, su influencia será fundamental para la implementación de los algoritmos cuánticos.

Por otro lado, teniendo en cuenta que $|\alpha|^2 + |\beta|^2 = 1$ y su representación en (2),

$$|\alpha|^2 + |\beta|^2 = |r_0 e^{i\phi_0}|^2 + |r_1 e^{i\phi_1}|^2 = |r_0|^2 |e^{i\phi_0}|^2 + |r_1|^2 |e^{i\phi_1}|^2 = 1.$$

Luego, dado que r_0 y r_1 son números reales, tenemos que $r_0^2 + r_1^2 = 1$. Por lo tanto, podemos encontrar un ángulo $\theta/2$ de forma única tal que

$$r_0 = \cos\left(\frac{\theta}{2}\right), \quad r_1 = \sin\left(\frac{\theta}{2}\right).$$

Esto nos permite reescribir un cúbit como

$$|\psi\rangle = \cos\left(\frac{\theta}{2}\right)|0\rangle + e^{i\phi} \sin\left(\frac{\theta}{2}\right)|1\rangle,$$

que tiene únicamente dos parámetros reales, cuyos valores cumplen $0 \leq \phi < 2\pi$ y $0 \leq \theta/2 < \pi/2$ (véase el libro de Sutor [17]). Podemos, pues, obtener una representación de los cúbits como puntos en la superficie de una esfera en un espacio \mathbb{R}^3 que podemos ver en la figura 1. Esta representación esférica es la que se conoce como *esfera de Bloch*. Cada estado de un cúbit viene dado de forma única por dos ángulos en esta esfera, que corresponden a la latitud (θ) y la longitud (ϕ) que utilizamos, por ejemplo, para representar la posición de un objeto en la esfera terrestre.

Podemos ver que en la esfera los dos estados que definen la base estándar, $|0\rangle, |1\rangle$, que son ortogonales en el espacio de Hilbert, se representan como antipodales. Haber elegido $\theta/2$ como ángulo definido por r_0 y r_1 permite que cada estado puro quede representado por un único punto en la superficie de la esfera y que el ángulo que nos da la latitud sea θ . De hecho, mientras los estados puros quedan representados sobre la superficie, los estados mezcla, de los que hemos hablado anteriormente, se corresponden a los puntos en el interior de la misma.

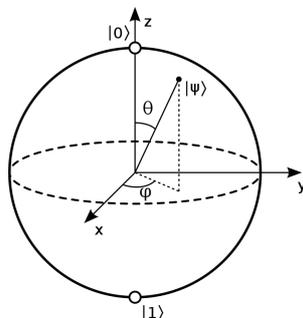


Figura 1: Esfera de Bloch para un cúbit arbitrario $|\psi\rangle$.

2.3. Transformaciones de estados cuánticos

Una vez definida la unidad básica de información cuántica, veamos cómo manipularla. En computación cuántica, para manipular el estado de los cúbits se usan transformaciones lineales, que vienen definidas de la siguiente forma.

Definición 9. Sean dos espacios vectoriales V y W y sea una función $T: V \rightarrow W$. Diremos que T es una **transformación lineal** si cumple las siguientes propiedades:

- (i) para todo $v \in V$ y todo $w \in W$, $T(v + w) = T(v) + T(w)$;
- (ii) para todo $v \in V$ y todo $c \in \mathbb{C}$, $T(cv) = cT(v)$. ◀

Más concretamente, al operar con cúbits, se utilizan transformaciones lineales unitarias que vienen dadas por matrices unitarias². Recordemos su definición.

Definición 10. Sea $M \in \mathbb{C}^{n \times n}$. Diremos que M es una **matriz unitaria** si cumple que $MM^\dagger = I$, siendo M^\dagger la transpuesta conjugada de M . ◀

Las matrices unitarias tienen propiedades que permiten realizar modificaciones sucesivas de los cúbits, manteniendo siempre un estado cuántico puro conforme a la definición 7. La primera de estas propiedades es la siguiente.

Propiedad 11. Sean $M, M' \in \mathbb{C}^{n \times n}$ dos matrices unitarias. Entonces, MM' es una matriz unitaria.

Esta propiedad implica que podemos componer sucesivas transformaciones unitarias en una nueva transformación unitaria. La segunda propiedad importante es que las matrices unitarias preservan el producto escalar.

Propiedad 12. Sea $M \in \mathbb{C}^{n \times n}$ una matriz unitaria y sean $v, w \in \mathbb{C}^n$. Entonces, $\langle Mv | Mw \rangle = \langle v | w \rangle$.

Esto implica que el módulo del vector asociado a los estados de los cúbits se mantiene inalterado, preservando su longitud unidad. Podemos ver más detalles, por ejemplo, en el libro de Yanofsky y Mannucci [18].

En los sistemas clásicos, para modificar la información, disponemos de puertas como la AND o como la OR que nos sirven para operar con los bits [17, cap. 2]. En computación cuántica, tenemos sus homólogos, las puertas cuánticas, que pueden definirse del siguiente modo.

Definición 13. Una **puerta cuántica** es un operador que actúa sobre cúbits y está representado por una matriz unitaria. ◀

Esto implica que, dada una puerta cuántica G , siempre vamos a poder encontrar otra puerta G' de forma que $GG' = \mathbb{I}$ (en particular, con $G' = G^\dagger$, por ser G unitaria). Así pues, podemos decir que toda puerta cuántica es *reversible*, es decir, podemos deducir los valores de entrada conociendo solamente los valores de salida, al contrario de lo que ocurre con algunas puertas clásicas [10, cap. 5].

Veamos algunas de las puertas cuánticas sobre un cúbit más utilizadas. La primera puerta que vamos a describir es la conocida como *puerta de Hadamard* y se representa como H . Esta puerta es una de las más importantes, ya que es la que nos permite conseguir un cúbit en un estado de superposición partiendo de un cúbit que se encontraba en uno de los estados básicos $\{|0\rangle, |1\rangle\}$. Si aplicamos la puerta sobre los estados básicos, obtenemos los siguientes estados:

$$|+\rangle = H|0\rangle = \frac{|0\rangle + |1\rangle}{\sqrt{2}}, \quad |-\rangle = H|1\rangle = \frac{|0\rangle - |1\rangle}{\sqrt{2}}.$$

De este modo, al aplicar la puerta de Hadamard sobre los estados básicos, obtenemos una superposición equiprobable de los mismos. Esto es, si medimos cualquiera de los estados resultantes, obtendremos los estados $|0\rangle$ o $|1\rangle$ con un 50 % de probabilidad. Notemos que $\{|+\rangle, |-\rangle\}$ forman también una base del espacio de Hilbert, que se conoce como *base de Hadamard* o base Pauli X . Podemos ver en la figura 2 que la aplicación de la puerta de Hadamard, como la del resto de puertas cuánticas, corresponde con una rotación del vector asociado al estado cuántico en la esfera de Bloch. La puerta de Hadamard tiene asociada la siguiente matriz unitaria:

$$H = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

Si se aplica una puerta de Hadamard sobre un elemento de la base de Hadamard, por ejemplo sobre el estado $|+\rangle$, se tiene que

$$H|+\rangle = H \frac{1}{\sqrt{2}} (|0\rangle + |1\rangle) = \frac{1}{\sqrt{2}} (H|0\rangle + H|1\rangle) = \frac{1}{2} (|0\rangle + |1\rangle + |0\rangle - |1\rangle) = |0\rangle,$$

²Recordemos que, dada una transformación lineal, podemos encontrar una matriz que represente dicha transformación y viceversa (véase el libro de Loeff [10, cap. 5]), luego podemos hablar de ellas indistintamente.

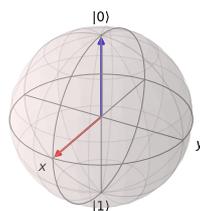


Figura 2: Efecto al aplicar Hadamard sobre el estado $|0\rangle$.

donde se puede observar que las fases del estado $|0\rangle$ se han sumado, mientras que las fases del estado $|1\rangle$ se han cancelado. Este fenómeno se denomina *interferencia* y en algunos casos actúa de forma constructiva y en otros destructiva. Muchos algoritmos cuánticos aprovechan esta propiedad y transforman los estados cuánticos para lograr incrementar las probabilidades asociadas a los estados que son solución del problema y reducir o eliminar las probabilidades del resto de estados.

Las siguientes puertas que introducimos son las *puertas de Pauli*, que son los generadores del grupo especial unitario $SU(2)$. Dado que este grupo es isomorfo con el grupo de rotación $SO(3)$, las tres puertas de Pauli nos permiten representar cualquier rotación en tres dimensiones. De hecho, sus nombres corresponden a los ejes sobre los que cada una de ellas realiza una rotación: X , Y y Z . Sus matrices unitarias asociadas son las siguientes:

$$X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad Y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

La puerta X corresponde a una rotación de 180° respecto al eje x de la esfera, tal y como puede verse en la figura 3a, donde la flecha azul representa el estado anterior a la transformación del estado $|0\rangle$ y la flecha roja el estado posterior. Esta puerta se suele denominar también NOT porque transforma el estado $|0\rangle$ en $|1\rangle$ y viceversa. También se denomina *bit-flip* debido a que, dado un cúbit en superposición $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, lo transforma en $X|\psi\rangle = \beta|0\rangle + \alpha|1\rangle$, es decir, intercambia las amplitudes de los estados base.

La puerta Z corresponde a una rotación de 180° respecto al eje z . Su efecto sobre los estados básicos es dejar inalterado el estado $|0\rangle$ y convertir el estado $|1\rangle$ en $-|1\rangle$. Estas transformaciones no modifican la representación de ambos estados en la esfera de Bloch, tal y como puede observarse en la figura 3b. La mejor forma de ver el efecto de esta puerta es aplicarla sobre los estados de la base de Hadamard. El estado $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$ lo transforma en $Z|+\rangle = (|0\rangle - |1\rangle)/\sqrt{2} = |-\rangle$ y viceversa. Con este ejemplo podemos ver que esta puerta aplica una fase relativa adicional de 180° al estado, cambiando el signo $+$ por $-$ y viceversa, de ahí que la puerta Z se denomine también *phase-flip*.

Por último, la puerta Y corresponde a una rotación de 180° respecto al eje Y . Su efecto sobre los estados básicos es convertir el estado $|0\rangle$ en $i|1\rangle$ y el estado $|1\rangle$ en $-i|0\rangle$. Podemos ver que, si aplicamos una puerta Y sobre un estado cualquiera en superposición, $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, su efecto es el siguiente:

$$Y|\psi\rangle = -i\beta|0\rangle + i\alpha|1\rangle = -i(\beta|0\rangle - \alpha|1\rangle) \equiv \beta|0\rangle - \alpha|1\rangle.$$

La última equivalencia viene dada por el hecho ya comentado con anterioridad de que dos estados cuánticos que solo se diferencien por una fase global de módulo unidad son indistinguibles. Por tanto, podemos ver que la puerta Y combina un *bit-flip* con un *phase-flip*.

La figura 3 representa el efecto de aplicar las puertas de Pauli sobre el estado $|0\rangle$. Debemos notar que las figuras 3a y 3b son iguales, aunque hemos llegado al mismo estado a través de la rotación alrededor dos ejes diferentes, x e y . Notemos también que, en la figura 3c, el estado $|0\rangle$ no se ha visto afectado por la transformación Z , dado que esta realiza una rotación sobre el eje z en el que se sitúa el vector.

Para ver mejor la diferencia entre las puertas X e Y , vamos a aplicarlas sobre el estado resultante de la figura 2 (es decir, sobre el estado $|+\rangle$), obteniendo como resultado las esferas de la figura 4. En este caso, podemos ver como es la puerta X la que deja invariante el cúbit, ya que gira sobre el mismo eje en que se

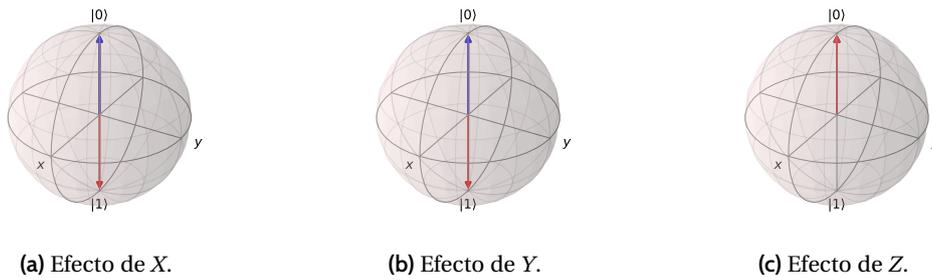


Figura 3: Efectos de las puertas de Pauli sobre el estado $|0\rangle$. En azul el estado inicial y en rojo el final.

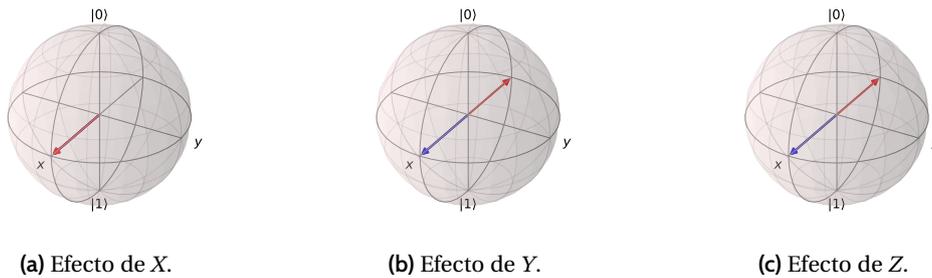


Figura 4: Efectos de las puertas de Pauli sobre el estado $|+\rangle$. En azul el estado inicial y en rojo el final.

encuentra, mientras que las puertas Y y Z confluyen en el mismo resultado, aunque rotando alrededor de dos ejes distintos.

Podemos generalizar las rotaciones alrededor del eje z mediante la puerta R_ϕ^Z , asociada a la siguiente matriz unitaria:

$$(3) \quad R_\phi^Z = \begin{bmatrix} 1 & 0 \\ 0 & e^{i\phi} \end{bmatrix}.$$

Esta puerta realiza una rotación de ángulo ϕ alrededor de z, lo que supone un cambio de fase en el estado del cúbit. Podemos ver que la puerta Z es un caso particular de esta puerta ($Z = R_\pi^Z$). Otros casos particulares de esta puerta son $S = R_{\pi/2}^Z$ y $T = R_{\pi/4}^Z$, cuyos efectos sobre el estado $|+\rangle$ pueden verse en la figura 5.

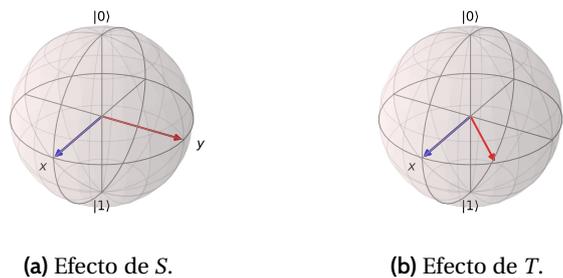


Figura 5: Efectos de las rotaciones sobre el estado $|+\rangle$. En azul, el estado inicial y en rojo, el final.

3. Sistemas cuánticos

3.1. Combinar múltiples cúbits

Al igual que disponer de un ordenador clásico con un único bit no es de mucha utilidad, un ordenador cuántico con un único cúbit tampoco, debido a la cantidad limitada de información que se puede manejar.

En concreto, un cúbit nos permite almacenar dos números complejos correspondientes a sus amplitudes. Es por ello que se hace necesario combinar cúbits para poder conseguir sistemas que permitan trabajar con mayores cantidades de información. Los espacios vectoriales asociados a los cúbits se combinan usando el producto tensorial. En concreto, utilizaremos el producto de Kronecker, que podemos definir para vectores complejos de la siguiente forma.

Definición 14. Sean $A \in \mathbb{C}^n$ y $B \in \mathbb{C}^m$ de la forma

$$A = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}, \quad B = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}.$$

Se define el **producto de Kronecker para vectores** como

$$A \otimes B = \begin{bmatrix} a_1 \cdot B \\ \vdots \\ a_n \cdot B \end{bmatrix} = \begin{bmatrix} a_1 b_1 \\ \vdots \\ a_1 b_m \\ \vdots \\ a_n b_1 \\ \vdots \\ a_n b_m \end{bmatrix}.$$

Esto es, para llevar a cabo el producto, multiplicamos cada uno de los elementos del primer vector por todos los del segundo.

Veamos un ejemplo sencillo de cómo combinar dos cúbits. Supongamos que tenemos los cúbits representados por los siguientes vectores: $|\psi_1\rangle = [\alpha_1, \beta_1]^T$ y $|\psi_2\rangle = [\alpha_2, \beta_2]^T$. Entonces, el sistema cuántico definido por ambos es

$$(4) \quad |\psi_1\rangle \otimes |\psi_2\rangle = \begin{bmatrix} \alpha_1 \\ \beta_1 \end{bmatrix} \otimes \begin{bmatrix} \alpha_2 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} \alpha_1 \alpha_2 \\ \alpha_1 \beta_2 \\ \beta_1 \alpha_2 \\ \beta_1 \beta_2 \end{bmatrix}.$$

Este producto se puede aplicar sobre los vectores de la base del espacio vectorial. Por ejemplo, si utilizamos la base de cálculo $\{|0\rangle, |1\rangle\}$ de un cúbit, la combinación de todas las parejas posibles formadas con sus elementos da lugar a la base análoga para un sistema de dos cúbits. Así, dicha base se construye como $\{|0\rangle \otimes |0\rangle, |0\rangle \otimes |1\rangle, |1\rangle \otimes |0\rangle, |1\rangle \otimes |1\rangle\}$, y suele expresarse de forma compacta como $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$. Podemos comprobar que dichos vectores tienen dimensión 2^2 y son

$$|00\rangle = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad |01\rangle = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad |10\rangle = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \quad \text{y} \quad |11\rangle = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

Así pues, los cúbits se combinan con el producto tensorial conforme al siguiente teorema.

Teorema 15. Sean dos sistemas cuánticos independientes Q y Q' , representados respectivamente por los espacios de Hilbert \mathbb{V} y \mathbb{V}' . Entonces, el sistema cuántico resultante al combinar Q y Q' tendrá el producto tensorial $\mathbb{V} \otimes \mathbb{V}'$ como espacio de los estados.

El estado resultante de combinar dos cúbits usando (4) se denomina *estado separable*, ya que, dado el estado $[\alpha_1 \alpha_2, \alpha_1 \beta_2, \beta_1 \alpha_2, \beta_1 \beta_2]^T$, podemos obtener los estados (las amplitudes) de los dos cúbits combinados. Sin embargo, en la gran mayoría de los casos, no es posible separar dichos estados. Por ejemplo, el estado $|\psi\rangle = [1/\sqrt{2}, 0, 0, 1/\sqrt{2}]^T$ no puede separarse en dos cúbits independientes. Cuando esto ocurre se dice que el sistema cuántico está en un *estado no separable*, más habitualmente denominado *estado entrelazado*.

Definición 16. Diremos que un sistema cuántico $|\psi\rangle$ está en un **estado entrelazado** o no separable si no existen $|\psi_1\rangle, |\psi_2\rangle$ de forma que $|\psi\rangle = |\psi_1\rangle \otimes |\psi_2\rangle$.

El entrelazamiento es uno de los fenómenos más llamativos y difíciles de explicar de la mecánica cuántica y también es una de las propiedades que son la base del funcionamiento de la mayoría de los algoritmos cuánticos. Esto se debe a la siguiente propiedad.

Propiedad 17. *Sea un sistema cuántico $|\psi\rangle$ en un estado entrelazado de dos cúbits $|\psi_1\rangle$ y $|\psi_2\rangle$. Entonces, cualquier cambio sobre el estado del cúbit $|\psi_1\rangle$ modificará el estado del cúbit $|\psi_2\rangle$ independientemente de la distancia entre ellos.*

Intentemos explicarlo con un ejemplo. Supongamos que tenemos dos cúbits entrelazados en el siguiente estado:

$$|\psi\rangle = \left[\frac{1}{\sqrt{2}}, 0, 0, \frac{1}{\sqrt{2}} \right]^T = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle),$$

y que hemos separado los sistemas físicos que almacenan ambos cúbits una enorme distancia; por ejemplo, tenemos uno en la Tierra y otro en Júpiter. Sabemos que si medimos en la Tierra el primer cúbit y obtenemos el estado $|0\rangle$, el segundo estará necesariamente en ese mismo estado $|0\rangle$, y lo mismo ocurrirá si al medir el estado del primer cúbit obtenemos $|1\rangle$. Además, el colapso del estado del segundo cúbit se producirá de modo simultáneo al producido al medir el primero, independientemente de la distancia y de que el límite de la velocidad de la luz impide que se dé una comunicación instantánea entre la Tierra y Júpiter. El entrelazamiento da lugar, pues, a un fenómeno «misterioso» que Albert Einstein denominó «*acción fantasmal a distancia*». Esto es, sabemos que el resultado de medir cada cúbit por separado es totalmente aleatorio y que existe una probabilidad del 50 % de que lo midamos en cada uno de los dos estados básicos. Sin embargo, existe una correlación total entre los estados de ambos cúbits que hace que el estado del segundo cúbit quede, en ese caso, definido con un 100 % de probabilidad. Así pues, el estado del sistema de dos cúbits entrelazados solo puede entenderse completamente como un estado global y no como dos componentes manipulables por separado. No obstante, el entrelazamiento no permite la transmisión de información más rápida que la luz. Si medimos el cúbit en la Tierra y obtenemos un $|0\rangle$, no podremos saber si ha sido un resultado aleatorio o es porque se ha medido primero el cúbit en Júpiter y se ha obtenido ese mismo valor. Tal y como veremos al describir la teleportación cuántica, solo usando un canal clásico para comunicar esa información podremos saber quién lo midió primero [1, 5].

En (4) hemos visto que el producto tensorial de dos sistemas de \mathbb{C}^2 resulta en un nuevo sistema en \mathbb{C}^4 . Esto es, combinar dos cúbits duplica el tamaño del espacio vectorial. En general, el crecimiento del tamaño de los sistemas cuánticos viene dado por la siguiente propiedad.

Propiedad 18. *Sea $A \in \mathbb{C}^n$ y sea $B \in \mathbb{C}^m$. Entonces, el producto tensorial $A \otimes B$ pertenece al espacio \mathbb{C}^{nm} .*

Sabemos que los cúbits nos permiten almacenar información codificada en las amplitudes. Un cúbit nos permite almacenar dos estados posibles al mismo tiempo, cada uno de ellos con una cierta probabilidad. En base a la propiedad anterior asociada al producto tensorial, dos cúbits nos permitirán almacenar $2 \times 2 = 2^2$ estados y, en general, n cúbits nos permitirán almacenar 2^n estados. Aumentar en uno la cantidad de cúbits disponibles duplica la cantidad de información que podemos almacenar. Esto es, la información almacenable y utilizable en los algoritmos aumenta exponencialmente con el número de cúbits. En contraste, en el caso de los bits clásicos, la información almacenable aumenta solo linealmente con el número de bits.

3.2. Transformaciones con múltiples cúbits

En el apartado 3.1 vimos que las transformaciones de los estados de un cúbit se asocian a matrices complejas de tamaño 2×2 . Dado que los estados de n cúbits se representan mediante vectores de tamaño 2^n , las matrices asociadas a sus transformaciones serán de tamaño $2^n \times 2^n$. Estas transformaciones se suelen implementar mediante la combinación de puertas cuánticas de uno y dos cúbits.

Por ejemplo, supongamos que tenemos un sistema cuántico de dos cúbits dado por $|\psi\rangle = |\psi_1\rangle \otimes |\psi_2\rangle$ y queremos aplicar una transformación de Hadamard a cada uno de ellos. En ese caso, en base a las propiedades del producto tensorial (véase el libro de Scherer [15]), obtenemos que

$$H|\psi_1\rangle \otimes H|\psi_2\rangle = (H \otimes H)(|\psi_1\rangle \otimes |\psi_2\rangle) = (H \otimes H)|\psi\rangle = H^{\otimes 2}|\psi\rangle.$$

Por tanto, $H^{\otimes 2} = H \otimes H$ es una transformación sobre dos cúbits que tiene el efecto de aplicar una Hadamard sobre el primer cúbit y otra sobre el segundo. De forma análoga, si se quiere aplicar la transformación de Hadamard únicamente sobre el primer cúbit, aplicaremos a ambos cúbits la transformación dada por $H \otimes I_2$, siendo I_2 la matriz identidad en \mathbb{C}^2 . Para poder realizar estas operaciones, es necesario extender la definición del producto de Kronecker a matrices complejas.

Definición 19. Sean dos matrices complejas A y B de la forma

$$A = \begin{bmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{bmatrix}, \quad B = \begin{bmatrix} b_{1,1} & \cdots & b_{1,t} \\ \vdots & \ddots & \vdots \\ b_{p,1} & \cdots & b_{p,t} \end{bmatrix}.$$

Se define el producto de Kronecker como

$$A \otimes B = \begin{bmatrix} a_{1,1} \cdot B & \cdots & a_{1,n} \cdot B \\ \vdots & \ddots & \vdots \\ a_{m,1} \cdot B & \cdots & a_{m,n} \cdot B \end{bmatrix} = \begin{bmatrix} a_{1,1}b_{1,1} & \cdots & a_{1,1}b_{1,t} & a_{1,2}b_{1,1} & \cdots & a_{1,n}b_{1,t} \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{1,1}b_{p,1} & \cdots & a_{1,1}b_{p,t} & a_{1,2}b_{p,1} & \cdots & a_{1,n}b_{p,t} \\ a_{2,1}b_{1,1} & \cdots & a_{2,1}b_{1,t} & a_{2,2}b_{1,1} & \cdots & a_{2,n}b_{1,t} \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{m,1}b_{1,1} & \cdots & a_{m,1}b_{1,t} & a_{m,2}b_{p,1} & \cdots & a_{m,n}b_{p,t} \end{bmatrix}.$$

Esta nueva definición permite obtener la matriz asociada a $H \otimes H$:

$$H \otimes H = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \otimes \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}.$$

Si aplicamos esta transformación $H^{\otimes 2}$ al estado $|00\rangle$ obtenemos un estado especial y muy utilizado para iniciar algoritmos cuánticos:

$$H^{\otimes 2}|00\rangle = H^{\otimes 2}|0\rangle_2 = \frac{1}{\sqrt{2^2}} \sum_{x \in \{0,1\}^2} |x\rangle_2 = \frac{1}{2}(|00\rangle + |01\rangle + |10\rangle + |11\rangle).$$

Podemos ver que, de modo análogo a lo que ocurre con la puerta de Hadamard aplicada a un cúbit, obtenemos una superposición equiprobable de todos los estados básicos de dos cúbits. Este mismo comportamiento puede extenderse a n cúbits usando la transformación $H^{\otimes n}$:

$$(5) \quad H^{\otimes n}|0\rangle_n = \frac{1}{\sqrt{2^n}} \sum_{x \in \{0,1\}^n} |x\rangle_n.$$

Las transformaciones con varios cúbits descritas hasta el momento se aplican individualmente sobre cada cúbit y mantienen el sistema en un estado separable. Para poder obtener toda la potencia dada por el entrelazamiento de sistemas cuánticos es necesario utilizar alguna puerta que permita, tal y como veremos en el próximo apartado, entrelazar varios cúbits. La más conocida de estas es la puerta NOT controlada o CNOT. El primero de los cúbits sobre los que se aplica se denomina *control* y el segundo *objetivo*. Esta puerta tiene como efecto aplicar una puerta NOT (otra denominación de X) sobre el objetivo cuando el control se encuentra en el estado $|1\rangle$ y de no aplicar ningún cambio (es decir, aplicar la identidad sobre el objetivo) cuando el control se encuentra en el estado $|0\rangle$. Su matriz unitaria es

$$\text{CNOT} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix},$$

y si la aplicamos sobre los cuatro estados básicos de dos cúbits su efecto es el siguiente:

$$\begin{aligned} \text{CNOT}|00\rangle &= |00\rangle, & \text{CNOT}|10\rangle &= |11\rangle, \\ \text{CNOT}|01\rangle &= |01\rangle, & \text{CNOT}|11\rangle &= |10\rangle. \end{aligned}$$

Esta puerta es el equivalente reversible de la puerta XOR, ya que obtiene en el cúbit objetivo el resultado de aplicar esa operación sobre los dos cúbits de entrada. Otras dos puertas importantes de varios cúbits son la SWAP y la Toffoli. La puerta SWAP sobre dos cúbits debe su nombre a que intercambia las amplitudes de ambos estados. Esta puerta es muy utilizada por los compiladores cuando realizan transformaciones sobre los circuitos cuánticos [2]. La puerta de Toffoli sobre tres cúbits es una puerta NOT doblemente controlada. Esto es, aplicará una puerta NOT sobre el tercer cúbit si y solo si los dos primeros cúbits se encuentran en el estado $|1\rangle$. Esta puerta permite la implementación reversible de todas las puertas clásicas: OR, AND, NAND, etc. [18].

3.3. Circuitos cuánticos

El modelo más utilizado para representar los algoritmos cuánticos es el basado en circuitos cuánticos, aunque existen otros como el basado en *quantum annealers* [9, cap. 2]. Como ocurre con los circuitos en los ordenadores clásicos, estos se construyen aplicando en un cierto orden puertas, pero en este caso puertas cuánticas sobre cúbits. Dichos cúbits se agrupan en registros.

Definición 20. Un **registro cuántico** es una colección de cúbits ψ_1, \dots, ψ_n que se utilizan para el cálculo. ◀

Es decir, un registro cuántico define el número y orden de los cúbits. Esto, combinado con la sucesión de puertas cuánticas, nos permite definir un circuito cuántico como sigue.

Definición 21. Un **circuito cuántico** es una sucesión de puertas cuánticas P_1, \dots, P_n que se aplican sobre un registro cuántico. ◀

Para representar un circuito cuántico, se disponen los cúbits del registro cuántico en el eje vertical en orden descendente conforme a su posición del registro. Asociada a cada cúbit del registro, se dibuja una línea horizontal que denota el tiempo de izquierda a derecha. Sobre dichas líneas se dibujan las puertas cuánticas que se aplican sobre uno o varios cúbits.

En la figura 6 podemos ver un ejemplo de circuito cuántico sencillo pero muy importante, puesto que se utiliza para entrelazar los estados de los dos cúbits sobre los que se aplica: $|x\rangle$ y $|y\rangle$. Se suele analizar el funcionamiento de un circuito cuántico estudiando el estado del sistema en determinados momentos, tras la aplicación de algunas de las puertas incluidas. En este ejemplo estudiaremos el estado en tres momentos, dados por $|\varphi_1\rangle$, $|\varphi_2\rangle$ y $|\varphi_3\rangle$.

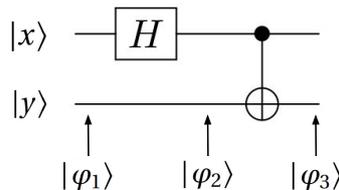


Figura 6: Ejemplo circuito cuántico.

En el estado $|\varphi_1\rangle$, tenemos los cúbits en el estado inicial, es decir, tenemos que $|\varphi_1\rangle = |x\rangle \otimes |y\rangle$. Tras aplicar una puerta de Hadamard (H) sobre el primer cúbit, llegamos al estado

$$|\varphi_2\rangle = H|x\rangle \otimes |y\rangle = (H \otimes I_2)(|x\rangle \otimes |y\rangle).$$

Por último, aplicamos una puerta CNOT usando el primer cúbit como control y el segundo como objetivo. El estado resultante viene dado por

$$|\varphi_3\rangle = \text{CNOT}(H|x\rangle \otimes |y\rangle) = \text{CNOT}(H \otimes I_2)(|x\rangle \otimes |y\rangle).$$

Es interesante observar que las transformaciones aparecen en la ecuación en orden inverso a las puertas en el circuito.

Estudiemos el circuito cuántico para un caso concreto para entender su efecto. Supongamos que el registro cuántico está compuesto por dos cúbits en estado $|0\rangle$. Este es el estado habitual en que suelen iniciarse todos los algoritmos cuánticos. Así pues, el estado inicial será

$$|\varphi_1\rangle = |0\rangle \otimes |0\rangle = |00\rangle.$$

Tras aplicar la puerta de Hadamard sobre el primer cúbit obtenemos

$$|\varphi_2\rangle = H|0\rangle \otimes |0\rangle = \frac{|0\rangle + |1\rangle}{\sqrt{2}} \otimes |0\rangle = \frac{|00\rangle + |10\rangle}{\sqrt{2}}.$$

Por último, aplicamos una puerta CNOT y obtenemos

$$|\varphi_3\rangle = \text{CNOT} \frac{|00\rangle + |10\rangle}{\sqrt{2}} = \frac{|00\rangle + |11\rangle}{\sqrt{2}}.$$

El resultado es precisamente el estado entrelazado comentado en el apartado 3.1. Se trata de uno de los cuatro estados denominados pares de Bell, obtenidos aplicando el circuito anterior sobre cada uno de los cuatro estados básicos de dos cúbits: $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$. Estos estados se utilizan, por ejemplo, en el algoritmo de teleportación cuántica que veremos en el siguiente apartado.

Veamos ahora de dónde surge la potencia computacional de los ordenadores cuánticos. Supongamos para ello que tenemos una función $f: \{0, 1\}^n \rightarrow \{0, 1\}$ que, dada cualquier cadena de n bits, obtiene un bit. Supongamos también que tenemos un circuito cuántico que usa una transformación unitaria U_f para calcular dicha función del siguiente modo:

$$U_f(|x\rangle_n |0\rangle) = |x\rangle_n |f(x)\rangle.$$

Esto es, la transformación deja invariables los primeros n cúbits de entrada y devuelve en el cúbit $n + 1$ el resultado de aplicar la función f sobre los mismos. Puede consultarse, por ejemplo, el libro de Nielsen y Chuang [12] para ver cómo construir esta transformación. Su «potencia» surge cuando el estado de entrada $|x\rangle$ es una superposición equiprobable como la obtenida en (5). En ese caso, teniendo en cuenta que se trata de una transformación lineal, el resultado será

$$U_f \frac{1}{\sqrt{2^n}} \sum_{x \in \{0,1\}^n} (|x\rangle_n |0\rangle) = \frac{1}{\sqrt{2^n}} \sum_{x \in \{0,1\}^n} U_f(|x\rangle_n |0\rangle) = \frac{1}{\sqrt{2^n}} \sum_{x \in \{0,1\}^n} |x\rangle_n |f(x)\rangle.$$

Podemos ver que, aunque la transformación se ha aplicado solo una vez para el estado en superposición, de algún modo la función se ha evaluado simultáneamente para los 2^n estados superpuestos. La posibilidad de almacenar 2^n estados usando solo n cúbits y de poder modificar todos esos estados a la vez se denomina *paralelismo cuántico*. Esa potencia computacional que crece exponencialmente con el número de cúbits es lo que permite abordar con ordenadores cuánticos determinados problemas intratables con ordenadores clásicos. No obstante, para poder aprovechar el paralelismo cuántico, los algoritmos deben modificar los estados superpuestos, de modo que, aprovechando la interferencia entre ellos, se incremente al máximo la probabilidad de obtener finalmente la solución correcta y se decremente la de medir soluciones incorrectas.

4. Algoritmos cuánticos

Siguiendo con el modelo de circuitos cuánticos, los algoritmos cuánticos son circuitos pensados para realizar una tarea o resolver un problema. En este apartado revisaremos uno de los más conocidos y, a pesar de su simplicidad, uno de los más llamativos. Se trata del algoritmo de teleportación cuántica, que ya ha sido utilizado en la práctica desde satélites en órbita [14]. Este algoritmo aprovecha la propiedad de que, dados dos cúbits entrelazados, los cambios sobre uno de los cúbits afectan al otro sin importar la distancia.

En la figura 7 está representado el circuito cuántico del algoritmo. Detallemos los pasos seguidos en el algoritmo. Sea $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ el estado del cúbit que Alice desea enviar a Bob ($|x\rangle$ en el circuito). Para

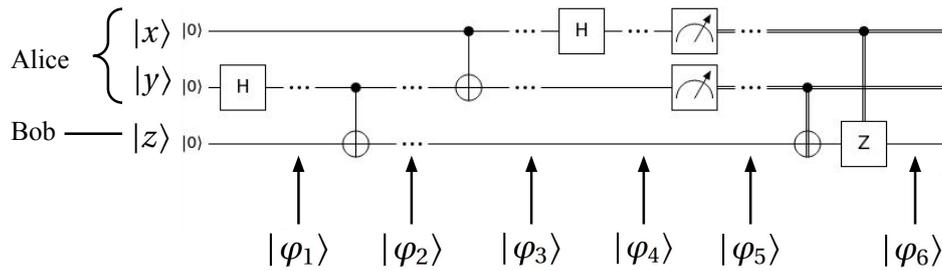


Figura 7: Circuito del algoritmo de teleportación cuántica.

poder enviarlo, necesitan repartirse dos cúbits entrelazados, que serán los cúbits $|y\rangle$ y $|z\rangle$ del circuito. Esto se consigue construyendo con estos cúbits un par de Bell, tal y como se ha descrito en el apartado 3.3. De este modo, en el estado $|\varphi_2\rangle$ los cúbits $|y\rangle$ y $|z\rangle$ se encontrarán en el siguiente estado:

$$|yz\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle).$$

En este punto, Alice se queda con uno de los cúbits entrelazados, $|y\rangle$, y Bob con el otro, $|z\rangle$. Ambos pueden entonces alejarse todo lo que quieran y, en cualquier momento, utilizar los cúbits entrelazados para llevar a cabo la teleportación de $|\psi\rangle$.

Si incluimos el cúbit que queremos teleportar en posesión de Alice, obtenemos el estado

$$|\varphi_2\rangle = |\psi\rangle \otimes |yz\rangle = (\alpha|0\rangle + \beta|1\rangle) \otimes \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) = \frac{1}{\sqrt{2}}(\alpha|000\rangle + \alpha|011\rangle + \beta|100\rangle + \beta|111\rangle).$$

Para iniciar la teleportación, Alice aplica una puerta CNOT sobre sus dos cúbits, transformando el sistema global al estado

$$|\varphi_3\rangle = (\text{CNOT} \otimes I)|\varphi_2\rangle = \frac{1}{\sqrt{2}}(\alpha|000\rangle + \alpha|011\rangle + \beta|110\rangle + \beta|101\rangle),$$

que puede expresarse también como

$$|\varphi_3\rangle = \frac{1}{\sqrt{2}}[\alpha|0\rangle \otimes (|00\rangle + |11\rangle) + \beta|1\rangle \otimes (|10\rangle + |01\rangle)].$$

Lo siguiente que debe hacer Alice es aplicar una puerta H sobre el cúbit $|\psi\rangle$, dejando el sistema en el siguiente estado:

$$\begin{aligned} |\varphi_4\rangle &= (H \otimes I_2) \left(\frac{1}{\sqrt{2}}[\alpha|0\rangle \otimes (|00\rangle + |11\rangle) + \beta|1\rangle \otimes (|10\rangle + |01\rangle)] \right) \\ &= \frac{1}{2} [\alpha(|000\rangle + |011\rangle + |100\rangle + |111\rangle) + \beta(|010\rangle + |001\rangle - |110\rangle - |101\rangle)] \\ &= \frac{1}{2} [|00\rangle(\beta|1\rangle + \alpha|0\rangle) + |01\rangle(\alpha|1\rangle + \beta|0\rangle) + |10\rangle(-\beta|1\rangle + \alpha|0\rangle) + |11\rangle(\alpha|1\rangle - \beta|0\rangle)]. \end{aligned}$$

En este punto, Alice mide sus dos cúbits y puede obtener con igual probabilidad cada uno de los estados básicos. En cada caso, el cúbit de Bob colapsará a un estado diferente dado por el cuadro 1.

Para completar la teleportación, Bob puede tener que realizar algunas transformaciones sobre el estado de sus cúbits que dependerán de los cúbits medidos por Alice. Por ello, Alice debe enviarle el resultado de su medición, y puede hacerlo por un canal clásico, puesto que solo necesita enviar dos bits de información. Las transformaciones que realizará Bob en ese punto para obtener en su cúbit el estado teleportado $|\psi\rangle$ pueden verse en el cuadro 2.

Podemos ver que la teleportación se ha producido y Bob acaba teniendo en todos los casos en el cúbit z el estado $|\psi\rangle$ inicialmente en posesión de Alice. Merece la pena comentar algunos aspectos del proceso.

Cuadro 1: Estado de los cúbits en $|\varphi_5\rangle$, tras la medición de Alice.

cúbits de Alice	cúbit de Bob en $ \varphi_5\rangle$
00	$\alpha 0\rangle + \beta 1\rangle$
01	$\beta 0\rangle + \alpha 1\rangle$
10	$\alpha 0\rangle - \beta 1\rangle$
11	$-\beta 0\rangle + \alpha 1\rangle$

Cuadro 2: Estado de los cúbits en $|\varphi_6\rangle$, tras completarse la teleportación.

cúbits de Alice	cúbit de Bob en $ \varphi_6\rangle$
00	$I(\alpha 0\rangle + \beta 1\rangle) \rightarrow \alpha 0\rangle + \beta 1\rangle$
01	$X(\beta 0\rangle + \alpha 1\rangle) \rightarrow \alpha 0\rangle + \beta 1\rangle$
10	$Z(\alpha 0\rangle - \beta 1\rangle) \rightarrow \alpha 0\rangle + \beta 1\rangle$
11	$ZX(-\beta 0\rangle + \alpha 1\rangle) \rightarrow \alpha 0\rangle + \beta 1\rangle$

Nótese que en ningún momento se ha comunicado información más rápido que la luz, puesto que los bits de Alice se han enviado por un canal clásico. Es cierto que ha sido necesario enviar esos dos bits, pero en realidad se han teleportado las dos amplitudes complejas asociadas a $|\psi\rangle$ con una precisión infinita. Para ello ha sido necesario destruir el estado inicial del cúbit x de Alice cuando esta lo ha medido. También se ha destruido el estado de los dos cúbits inicialmente entrelazados que han usado Alice y Bob como base para lograr la teleportación. Finalmente, en ningún caso estamos hablando de teleportación de materia, sino de información en forma de estado cuántico de un cúbit.

Existen un conjunto de algoritmos cuánticos bien conocidos que sirvieron para demostrar el potencial de la computación cuántica. Estos son el algoritmo de Deutsch-Jozsa [3] que determina si una función es constante o equilibrada, o el algoritmo de Grover [8], que permite buscar información en secuencias de datos no ordenados. Pero sin lugar a dudas, el algoritmo cuántico más conocido y que sirvió para despertar el interés en la computación cuántica es el algoritmo de Shor [16], que permite factorizar grandes números enteros con un coste lineal y puede suponer una amenaza para los sistemas criptográficos actuales en el momento en que podamos construir ordenadores cuánticos con un número suficiente de cúbits [9, cap. 4]. Debido a ello, se están desarrollando nuevos sistemas criptográficos resistentes a la computación cuántica. No obstante, dado que la adaptación de todos los sistemas actuales a la criptografía postcuántica puede llevar más de una década, es urgente no solo desarrollarlos, sino también desplegarlos [9, cap. 4].

Por otro lado, el funcionamiento de los algoritmos comentados se basa en la existencia de ordenadores cuánticos sin errores. En la actualidad, y muy probablemente durante bastantes años, nos encontraremos en la denominada era *Noisy Intermediate-Scale Quantum (NISQ)*, en la que manejaremos unos pocos cientos de cúbits con ruido y puertas con fiabilidad reducida [13]. Para la obtención de un cúbit lógico sin ruido es necesario utilizar sistemas de corrección de errores que involucran cientos de cúbits físicos ruidosos. Dado que queda mucho tiempo para disponer de ordenadores cuánticos con los muchos miles de cúbits necesarios para implementar los algoritmos comentados, nuestra mayor esperanza a corto plazo es el desarrollo de algoritmos eficientes que permitan obtener soluciones aproximadas a problemas con aplicación práctica y rendimiento económico usando pocos cúbits. Entre ellos se encuentran los algoritmos cuánticos variacionales, que permiten utilizar sistemas híbridos que combinan un ordenador clásico con uno cuántico para resolver, por ejemplo, problemas de química cuántica con aplicaciones en farmacología o ciencia de materiales [9, cap. 3].

5. Conclusiones

A lo largo del artículo hemos podido constatar cómo los fundamentos matemáticos de la computación cuántica son sorprendentemente simples. Esto es así a pesar de que algunos de los fenómenos físicos en que se apoya, tales como el entrelazamiento o el efecto de la medida, supusieron un reto para algunas de las mentes más brillantes del siglo pasado, como Albert Einstein y Niels Bohr, y cuya interpretación sigue siendo polémica en la actualidad. Básicamente, la computación cuántica se basa en el álgebra lineal compleja con el uso, menos habitual, del producto tensorial. También hemos podido describir algunas de las propiedades que confieren toda su capacidad y poder computacional a los ordenadores cuánticos. Hemos descrito la superposición de estados cuánticos y el caso particular del entrelazamiento de varios cúbits. Hemos comentado cómo aprovechar el fenómeno de la interferencia. También hemos visto cómo el crecimiento exponencial de la información con el número de cúbits en ciertos casos y la posibilidad de transformar simultáneamente todos los estados en superposición pueden dar lugar a la potencia de cálculo exponencial del paralelismo cuántico para ciertos problemas. Finalmente, hemos visto cómo sacar partido de esas propiedades mediante el uso de circuitos cuánticos para poder llevar a cabo tareas sorprendentes y abordar problemas aparentemente intratables con ordenadores clásicos, tales como la factorización de grandes números enteros.

Referencias

- [1] BELL, John Stewart. «On the Einstein Podolsky Rosen paradox». En: *Physics Physique Fizika* 1.3 (1964), págs. 195-200. ISSN: 0554-128X. <https://doi.org/10.1103/PhysicsPhysiqueFizika.1.195>.
- [2] CHONG, Frederic T.; FRANKLIN, Diana, y MARTONOSI, Margaret. «Programming languages and compiler design for realistic quantum hardware». En: *Nature* 549 (2017), págs. 180-187. ISSN: 1476-4687. <https://doi.org/10.1038/nature23459>.
- [3] DEUTSCH, David y JOZSA, Richard. «Rapid solution of problems by quantum computation». En: *Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences* 439.1907 (1992), págs. 553-558. ISSN: 0962-8444. <https://doi.org/10.1098/rspa.1992.0167>.
- [4] DIRAC, Paul A. M. «A new notation for quantum mechanics». En: *Mathematical Proceedings of the Cambridge Philosophical Society* 35.3 (1939), págs. 416-418. ISSN: 0305-0041. <https://doi.org/10.1017/S0305004100021162>.
- [5] EINSTEIN, Albert; PODOLSKY, Boris, y ROSEN, Nathan. «Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?». En: *Physical Review* 47.10 (1935), págs. 777-780. ISSN: 0031-899X. <https://doi.org/10.1103/PhysRev.47.777>.
- [6] FEYNMAN, Richard P. «Simulating physics with computers». En: *International Journal of Theoretical Physics* 21.6-7 (1982), págs. 467-488. ISSN: 0020-7748. <https://doi.org/10.1007/BF02650179>.
- [7] GERBERT, Philipp y RUER, Frank. *The Next Decade in Quantum Computing—and How to Play*. Boston Consulting Group. 15 de nov. de 2018. URL: <https://www.bcg.com/publications/2018/next-decade-quantum-computing-how-play> (visitado 02-2021).
- [8] GROVER, Lov K. «A fast quantum mechanical algorithm for database search». En: *Proceedings of the twenty-eighth annual ACM symposium on Theory of Computing*. 1996, págs. 212-219. <https://doi.org/10.1145/237814.237866>.
- [9] GRUMBLING, Emily y HOROWITZ, Mark, eds. *Quantum Computing: Progress and Prospects*. National Academies of Sciences, Engineering, and Medicine. Washington, US: The National Academies Press, 2019. <https://doi.org/10.17226/25196>.
- [10] LOCEFF, Michael. *A Course in Quantum Computing*. 2015. URL: https://lapastillaroja.net/wp-content/uploads/2016/09/Intro_to_QC_Vol_1_Loceff.pdf (visitado 02-2021).
- [11] LÓPEZ OLIVA, Vicente. *An introduction to Quantum algorithms*. Trabajo de Fin de Grado. Universitat Jaume I, 2020. URL: <http://hdl.handle.net/10234/191725>.

-
- [12] NIELSEN, Michael A. y CHUANG, Isaac L. *Quantum Computation and Quantum Information. 10th Anniversary Edition*. Cambridge, UK: Cambridge University Press, 2010. <https://doi.org/10.1017/CB09780511976667>.
- [13] PRESKILL, John. «Quantum Computing in the NISQ era and beyond». En: *Quantum 2*, artículo 79 (2018). ISSN: 2521-327X. <https://doi.org/10.22331/q-2018-08-06-79>.
- [14] REN, Ji-Gang; XU, Ping; YONG, Hai-Lin; ZHANG, Liang; LIAO, Sheng-Kai; YIN, Juan; LIU, Wei-Yue; CAI, Wen-Qi; YANG, Meng; LI, Li; YANG, Kui-Xing; HAN, Xuan; YAO, Yong-Qiang; LI, Ji; WU, Hai-Yan; WAN, Song; LIU, Lei; LIU, Ding-Quan; KUANG, Yao-Wu; HE, Zhi-Ping; SHANG, Peng; GUO, Cheng; ZHENG, Ru-Hua; TIAN, Kai; ZHU, Zhen-Cai; LIU, Nai-Le; LU, Chao-Yang; SHU, Rong; CHEN, Yu-Ao; PENG, Cheng-Zhi; WANG, Jian-Yu, y PAN, Jian-Wei. «Ground-to-satellite quantum teleportation». En: *Nature* 549 (2017), págs. 70-73. ISSN: 1476-4687. <https://doi.org/10.1038/nature23675>.
- [15] SCHERER, Wolfgang. *Mathematics of Quantum Computing. An Introduction*. Cham, CH: Springer, 2019. <https://doi.org/10.1007/978-3-030-12358-1>.
- [16] SHOR, Peter W. «Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer». En: *SIAM Review* 41.2 (1999), págs. 303-332. ISSN: 0036-1445. <https://doi.org/10.1137/S0036144598347011>.
- [17] SUTOR, Robert S. *Dancing with qubits. How quantum computing works and how it can change the world*. Birmingham, UK: Packt, 2019. ISBN: 978-1-83882-736-6.
- [18] YANOFSKY, Noson S. y MANNUCCI, Mirco A. *Quantum computing for computer scientists*. Cambridge, UK: Cambridge University Press, 2008. <https://doi.org/10.1017/CB09780511813887>.
- [19] YUAN, Xiao. «A quantum-computing advantage for chemistry». En: *Science* 369 (2020), págs. 1054-1055. ISSN: 1095-9203. <https://doi.org/10.1126/science.abd3880>.

TEMat

La ecuación $\bar{\partial}$ y el teorema de Runge

✉ Melanie Fumero Padrón
Universidad de La Laguna
alu0100828764@ull.edu.es

Resumen: En este trabajo se recoge la relación que existe entre el teorema de aproximación de Runge y la llamada ecuación $\bar{\partial}$: si $\partial/\partial\bar{z} = (\partial/\partial x + i\partial/\partial y)/2$ y K es un subconjunto compacto de un abierto $\Omega \subset \mathbb{C}$, los siguientes enunciados son equivalentes:

- Toda función holomorfa en un entorno de K se puede aproximar uniformemente en K por funciones holomorfas en Ω .
- Para todo $\varepsilon > 0$ y toda función $f \in C_c^\infty(\Omega)$ tal que $K \cap \text{supp } f = \emptyset$ existe una solución $u \in C^\infty(\Omega)$ de la ecuación $\partial u/\partial\bar{z} = f$ en Ω con $\|u\|_K < \varepsilon$.

Abstract: In this work we will show a relation between Runge's approximation theorem and the so called $\bar{\partial}$ problem: if $\partial/\partial\bar{z} = (\partial/\partial x + i\partial/\partial y)/2$ and K is a compact subset of an open set $\Omega \subset \mathbb{C}$, the following are equivalent:

- Any holomorphic function in a neighborhood of K can be uniformly approximated on K by holomorphic functions in Ω .
- For any $\varepsilon > 0$ and any $f \in C_c^\infty(\Omega)$ such that $K \cap \text{supp } f = \emptyset$ there exists a solution $u \in C^\infty(\Omega)$ to the equation $\partial u/\partial\bar{z} = f$ in Ω with $\|u\|_K < \varepsilon$.

Palabras clave: teorema de Runge, ecuación $\bar{\partial}$.

MSC2020: 30E10, 34M03, 35N05, 41A10, 41A20.

Recibido: 19 de abril de 2020.

Aceptado: 18 de junio de 2021.

Referencia: FUMERO PADRÓN, Melanie. «La ecuación $\bar{\partial}$ y el teorema de Runge». En: *TEMat*, 6 (2022), págs. 49-63. ISSN: 2530-9633. URL: <https://temat.es/articulo/2022-p49>.

© Este trabajo se distribuye bajo una licencia Creative Commons Reconocimiento 4.0 Internacional <https://creativecommons.org/licenses/by/4.0/>

1. Introducción

Dado Ω un abierto en el plano complejo \mathbb{C} , denotaremos por $\mathcal{O}(\Omega)$ al espacio de funciones holomorfas en Ω . Dados dos dominios (abiertos y conexos) $\Omega \subsetneq \tilde{\Omega} \subset \mathbb{C}$, la restricción $\mathcal{O}(\tilde{\Omega}) \rightarrow \mathcal{O}(\Omega)$, aunque inyectiva, nunca es sobreyectiva. Basta considerar, para $a \in \tilde{\Omega} \setminus \Omega$, la función $f(z) = 1/(z - a)$: si existiera una extensión \tilde{f} de f a todo $\tilde{\Omega}$, entonces, por el teorema de identidad, $(z - a)\tilde{f}(z) \equiv 1$ en $\tilde{\Omega}$, y esto es una contradicción cuando $z = a$.

Por otro lado, como consecuencia del teorema de Runge, toda función holomorfa en Ω se puede aproximar (uniformemente en compactos de Ω) por funciones holomorfas en $\tilde{\Omega}$ siempre que cualquier componente conexa de $\mathbb{C}_\infty \setminus \Omega$ interseque $\mathbb{C}_\infty \setminus \tilde{\Omega}$ (donde $\mathbb{C}_\infty = \mathbb{C} \cup \{\infty\}$ denota la esfera de Riemann o plano complejo extendido). Que esta condición topológica es necesaria para la aproximación que dicta el teorema de Runge es consecuencia inmediata del ejemplo anterior atendiendo al principio del máximo. Esto significa que el teorema de Runge no se reduce a un mero corolario de un resultado de extensión para funciones holomorfas que, por otro lado, es evidentemente falso.

Para ser más precisos, si $K \subset \mathbb{C}$ es un subconjunto compacto del plano complejo y $\Lambda \subset \mathbb{C}_\infty \setminus K$ un subconjunto del plano extendido, el teorema de Runge establece que cualquier función holomorfa en un entorno de K se puede aproximar uniformemente en K por funciones racionales con polos en Λ ¹. En particular, si $\mathbb{C}_\infty \setminus K$ es conexo, podemos tomar $\Lambda = \{\infty\}$ y concluir que toda función holomorfa en un entorno de K es límite uniforme en K de polinomios².

La idea de usar la ecuación $\bar{\partial}$ para probar el teorema de Runge es la siguiente. Sea f una función holomorfa en un entorno abierto Ω de un compacto $K \subset \mathbb{C}$. Si $\chi \in C_c^\infty(\Omega)$ es una función $C^\infty(\Omega)$ con soporte compacto en Ω y $\chi \equiv 1$ en algún entorno de K , entonces la función $\chi f \in C^\infty(\mathbb{C})$ coincide con f cerca de K pero no es holomorfa en Ω . Ahora se trata de encontrar una función u , pequeña en K , para la que la función $\tilde{f} = \chi f - u$ sea holomorfa en Ω . Esto lleva a considerar el problema no homogéneo para el operador $\partial/\partial\bar{z} = (\partial/\partial x + i\partial/\partial y)/2$, ya que \tilde{f} es holomorfa en Ω precisamente cuando $\partial u/\partial\bar{z} = f\partial\chi/\partial\bar{z}$ en Ω .

La sección 2 contiene los teoremas 1, 2 (este también conocido como desplazamiento de polos) y 5 que desglosan el teorema de Runge aludido anteriormente. De ellos presentamos sus demostraciones clásicas tal y como se suele llevar a cabo cuando se presenta el teorema de Runge en los primeros cursos de variable compleja. En principio, esta prueba no proporciona información sobre el grado de precisión en términos del grado de los polinomios aproximantes (en el caso en que el dominio sea simplemente conexo, es decir, cuando $\mathbb{C}_\infty \setminus \Omega$ es conexo). Para atacar esta cuestión normalmente se usa la teoría de aplicaciones conformes, pero en 1927 G. Szegő, usando métodos elementales (teorema 7), demostró que, si $\partial\Omega$ se reduce a una curva cerrada simple rectificable y $K \subset \Omega$ es compacto con $\text{dist}(K, \partial\Omega) = \delta > 0$, entonces existen constantes $A > 0$ y $0 < \kappa < 1$ y una sucesión de polinomios $\{p_n\}$ donde el grado de p_n es menor o igual a n tal que $\|f - p_n\|_K < A\kappa^n$ (aquí, y en adelante, $\|f\|_K := \sup_{z \in K} |f(z)|$ denotará la norma uniforme de f). Cabe mencionar que, en este resultado, κ depende de δ y la geometría de $\partial\Omega$, pero A también depende de la función f .

La sección 3 recoge los preliminares relativos a la ecuación $\bar{\partial}$ y los resultados que se usarán posteriormente en la sección 5 para presentar la relación que existe entre el problema $\bar{\partial}$ y el teorema de Runge. A saber, los siguientes problemas son equivalentes.

Problema A. Aproximar uniformemente en K cualquier función holomorfa en un entorno de K por funciones holomorfas en Ω . ◀

Problema H. Dada $f \in C_c^\infty(\Omega)$ tal que $K \cap \text{sop } f = \emptyset$ y $\varepsilon > 0$, encontrar una solución $u \in C^\infty(\Omega)$ de $\partial u/\partial\bar{z} = f$ tal que $\|u\|_K < \varepsilon$. ◀

En esta misma sección aparece el primer nexo de unión entre ellos. Haciendo uso de la fórmula de Cauchy-Green (proposición 8), el corolario 10 presenta una solución local del problema $\bar{\partial}$ que, combinado con el teorema de Runge, permite, en el teorema 11, probar la existencia de soluciones globales; es decir,

¹Podría suceder que no se requieran todos los puntos de Λ , es decir, Λ podría contener más puntos de los necesarios para que la aproximación que dicta el teorema de Runge tenga lugar.

²Recuérdese que las funciones racionales con polos en el infinito son precisamente los polinomios.

en cualquier abierto $\Omega \subset \mathbb{C}$, la ecuación $\partial u / \partial \bar{z} = f$ admite soluciones $u \in C^1(\Omega)$ para cualquier dato $f \in C^1(\Omega)$. Para completar la exposición de la sección 5 hemos incluido una discusión sobre el teorema de Szegő basado en las estimaciones L^2 para el problema $\bar{\partial}$ debidas a Hörmander.

De nuevo basado en la fórmula de Cauchy-Green, el teorema 12 en la sección 4 presenta una generalización del teorema de Runge en la que se relaja la hipótesis sobre la función f a aproximar observando que, para ello, es suficiente que $f \in C^1$ en un entorno de K y $\partial f / \partial \bar{z} = 0$ en K . Aquí también recogemos el famoso ejemplo del queso suizo debido a Alice Roth [11], un subconjunto compacto y conexo del disco unidad cerrado con interior vacío sin la propiedad de aproximación de funciones continuas por funciones racionales.

Para finalizar, en la última sección del artículo, el teorema 20 muestra que la ecuación $\partial u / \partial \bar{z} = f \in C_c^1(\mathbb{C})$ admite soluciones $u \in C_c^\infty(\mathbb{C})$ con $\text{sop } u \subset \text{sop } f$ si y solo si f es ortogonal a las funciones racionales con polos en un subconjunto $\Lambda \subset \mathbb{C}_\infty \setminus \text{sop } f$ en las mismas hipótesis que en teorema de Runge. Esto implica el teorema de Runge y cierra este círculo de relaciones.

2. El teorema de Runge

Sea f una función compleja definida en un subconjunto compacto $K \subset \mathbb{C}$. ¿Cuándo podemos aproximar f uniformemente en K por polinomios (en z) o por funciones racionales (cocientes de tales polinomios)? Límites uniformes de funciones racionales con polos fuera de K deben ser continuos y holomorfos en su interior si este no es vacío (en conjuntos abiertos, el límite uniforme en compactos de funciones holomorfas es holomorfo). Así, estas condiciones resultan necesarias para la aproximación de f . El teorema de Runge proporciona condiciones suficientes.

Teorema 1 (Runge). *Sea $K \subset \mathbb{C}$ compacto. Entonces, toda función holomorfa en un entorno de K se puede aproximar uniformemente en K por funciones racionales con polos fuera de K .*

Demostración. La demostración se basa en aproximar la función $z \mapsto 1/(z - a)$ ($a \in \mathbb{C}$) tras descomponer f en «suma» de tales funciones. La prueba que presentamos a continuación corresponde a aquella en el libro de Gamelin [5].

Por hipótesis podemos encontrar un abierto Ω con frontera C^∞ que contiene a K tal que f es holomorfa en un entorno de $\bar{\Omega}$. Por la fórmula integral de Cauchy,

$$f(z) = \frac{1}{2\pi i} \oint_{\partial\Omega} \frac{f(\zeta)}{\zeta - z} d\zeta, \quad z \in \Omega.$$

Ahora descomponemos $\partial\Omega$ en una unión finita de curvas γ_j tales que $\gamma_j \subset \Delta(\zeta_j, r_j)$ y $K \cap \Delta(\zeta_j, r_j) = \emptyset$. Entonces $f = \sum_j f_j$, donde

$$f_j(z) = \frac{1}{2\pi i} \int_{\gamma_j} \frac{f(\zeta)}{\zeta - z} d\zeta, \quad z \notin \gamma_j.$$

La función f_j es holomorfa en $\mathbb{C} \setminus \gamma_j$ ya que admite un desarrollo en serie de potencias decrecientes en $z - \zeta_j$ que converge uniformemente en $|z - \zeta_j| > r_j$ y, en particular, en K . Sumando estos aproximantes vemos que f se puede aproximar uniformemente en K por funciones racionales con polos fuera de K . ■

Cabe mencionar que la misma idea usada para aproximar las funciones $1/(z - \zeta_j)$ en el teorema 1 permite prescribir de antemano los polos de las funciones aproximantes (véase el trabajo de Fumero Padrón [3, Proposición 2.4]), es decir, se tiene el siguiente resultado.

Teorema 2. *Sea $K \subset \mathbb{C}$ un subconjunto compacto del plano complejo y $\Lambda \subset \mathbb{C}_\infty \setminus K$ un conjunto que contenga un punto en cada componente conexa de $\mathbb{C}_\infty \setminus K$. Entonces, cualquier función holomorfa en un entorno de K se puede aproximar uniformemente en K por funciones racionales con polos en Λ .*

Demostración. Utilizaremos un argumento de conexidad. Sean $U \subset \mathbb{C} \setminus K$ una componente conexa cualquiera de $\mathbb{C} \setminus K$ y $a \in U$. Veremos que cualquier función racional con polos en U se puede aproximar uniformemente en K por funciones racionales con polo en a .

Sea V el conjunto formado por aquellos $w \in U$ tales que la función $z \mapsto 1/(z-w)$ se pueda aproximar uniformemente en K por funciones racionales con polos en a . Veremos que V es abierto y cerrado en U , por lo que $V = U$ cuando $V \neq \emptyset$, ya que U es conexo. Que V es cerrado es inmediato de su propia definición: si $w_j \in V$ converge a $w \in U$, entonces $1/(z-w_j) \rightarrow 1/(z-w)$ uniformemente en K , por lo que, por definición de V , $w \in V$ también. Para ver que V es abierto hemos de probar que, si $w_0 \in V$ y $w \in U$ está suficientemente cerca de w_0 , entonces $w \in V$; es decir, que $z \mapsto 1/(z-w)$ se puede aproximar uniformemente en K por funciones polinómicas en $1/(z-w_0)$. En el caso en que $w_0 = \infty$, esto significa que, si $|w|$ es grande, $z \mapsto 1/(z-w)$ se puede aproximar uniformemente en K por polinomios. Para ver esto, desarrollamos $1/(z-w)$ en serie de potencias:

$$\frac{1}{z-w} = -\frac{1}{w} \frac{1}{1-z/w} = -\frac{1}{w} \sum_{k=0}^{\infty} \frac{z^k}{w^k}.$$

Si $|z| \leq M$ en K y $|w| \geq 2M$, la serie anterior está dominada por la serie geométrica $\sum 2^{-k}$ en K , por lo que converge uniformemente en K . Si $w_0 \in \mathbb{C}$, la demostración es esencialmente la misma. Tomemos $0 < \delta < \text{dist}(w_0, K)$. Si $|w - w_0| < \delta/2$, entonces $|w - w_0|/|z - w_0| < 1/2$ para $z \in K$, por lo que la serie geométrica

$$\frac{1}{z-w} = \frac{1}{z-w_0} \frac{1}{1 - \frac{w-w_0}{z-w_0}} = \frac{1}{z-w_0} \sum_{k=0}^{\infty} \frac{(w-w_0)^k}{(z-w_0)^k}$$

converge uniformemente en K . Esto significa que el disco $\Delta(w_0, \delta/2) \subset V$ y V es abierto. Por último, por el teorema 1, $V \neq \emptyset$. ■

En particular, si $\mathbb{C} \setminus K$ es conexo, cualquier función holomorfa en un entorno de K se puede aproximar uniformemente en K por polinomios³ (en este caso basta tomar $\Lambda = \{\infty\}$).

Observación 3. Nótese que el conjunto Λ puede ser infinito; por ejemplo, si $K = \{0\} \cup \bigcup_{j \geq 1} \partial\Delta(1/j, \varepsilon_j)$ con $1/(j+1)^2 > \varepsilon_j \rightarrow 0$ (para que estas circunferencias sean disjuntas). Para este compacto, Λ debe ser del tipo $\Lambda = \{a_j : j \geq 1\} \cup \{\infty\}$ con $a_j \in \Delta(1/j, \varepsilon_j)$, por ejemplo, $a_j = 1/j$. ◀

Como corolario del teorema 2 tenemos la siguiente versión relativa del teorema de Runge.

Teorema 4. Si $K \subset \Omega$ es un subconjunto compacto de una región $\Omega \subset \mathbb{C}$ tal que cualquier componente de $\mathbb{C}_{\infty} \setminus K$ interseque $\mathbb{C}_{\infty} \setminus \Omega$, entonces toda función holomorfa en un entorno de K se puede aproximar uniformemente en K por funciones holomorfas en todo Ω (de hecho, por funciones racionales sin polos en Ω).

Demostración. La hipótesis implica que cada componente de $\mathbb{C}_{\infty} \setminus K$ contiene al menos un punto en $\mathbb{C}_{\infty} \setminus \Omega$ y, por tanto, existe $\Lambda \subset \mathbb{C}_{\infty} \setminus \Omega$ en las hipótesis del teorema 2. Cualquier función racional con polos en Λ es, en particular, holomorfa en Ω . ■

Con esto, el siguiente teorema recoge la versión más general del teorema de Runge [2, Theorem 10.7].

Teorema 5. Las siguientes afirmaciones sobre un subconjunto compacto $K \subset \Omega$ en un abierto $\Omega \subset \mathbb{C}$ son equivalentes:

- (i) $\Omega \setminus K$ no tiene componentes relativamente compactas en Ω .
- (ii) Toda componente de $\mathbb{C}_{\infty} \setminus K$ interseca $\mathbb{C}_{\infty} \setminus \Omega$.
- (iii) Cualquier función holomorfa en un entorno de K se puede aproximar uniformemente en K por funciones racionales sin polos en Ω .
- (iv) Cualquier función holomorfa en un entorno de K se puede aproximar uniformemente en K por funciones holomorfas en Ω .
- (v) Para todo $a \in \Omega \setminus K$ existe $h \in \mathcal{O}(\Omega)$ tal que $|h(a)| > \|h\|_K$.

³Este es el resultado de Runge de 1885. Curiosamente, ese mismo año fue probado el teorema de Weierstrass sobre aproximación polinómica en intervalos.

Demostración. Para ver que (i) implica (ii), sea V una componente acotada de $\mathbb{C}_\infty \setminus K$. Si $V \subset \Omega$, V sería una componente de $\mathbb{C}_\infty \setminus K$ con $\partial V \subset K \subset \Omega$. Que (ii) implica (iii) es el contenido del teorema 4. Trivialmente (iii) implica (iv). Si $V \subset \Omega$ es una componente relativamente compacta de $\Omega \setminus K$ y $a \in V$, la función $f(z) = 1/(z - a)$ es holomorfa en un entorno de K pero no sería uniformemente aproximable en K por funciones holomorfas en Ω : no existe $g \in \mathcal{O}(\Omega)$ tal que $\|f - g\|_K < 1/2$ en K puesto que esto implicaría (por el principio del máximo) que $|1 - (z - a)g(z)| < 1/2$ para todo $z \in V$, lo que es una contradicción en $z = a$. Así, (iv) implica (i).

Por último, esta última idea permite ver que (v) es equivalente a estas cuatro afirmaciones: si $V \subset \Omega \setminus K$ es una componente de $\mathbb{C} \setminus K$ relativamente compacta en Ω y $h \in \mathcal{O}(\Omega)$, entonces, para todo $a \in V$, $|h(a)| \leq \max_{\partial V} |h| \leq \max_K |h|$ si $\bar{V} \subset \Omega$, ya que $\partial V \subset K$. Esto prueba que (v) implica (i). Por otro lado, si $a \in \Omega \setminus K$, (iv) aplicado al compacto $\bar{K} := K \cup \{a\} \subset \Omega$ proporciona una función $h \in \mathcal{O}(\Omega)$ tal que $|1 - h(a)|, \|h\|_K < 1/2$ (basta aproximar la función que vale 0 en un entorno de K y 1 en otro entorno de a disjunto de K). Para esta función tenemos que $|h(a)| > 1/2 > \|h\|_K$. Así, (iv) implica (v). ■

Si denotamos por $\hat{K}_\Omega = \{z \in \Omega : |h(z)| \leq \|h\|_K, \text{ para toda } h \in \mathcal{O}(\Omega)\}$ la envolvente de holomorfía de K respecto a Ω , este teorema implica que \hat{K}_Ω coincide con la unión de K y las componentes relativamente compactas de $\Omega \setminus K$ en Ω [8, Theorem 1.3.3]. Por tanto, (i) o cualquiera de las otras condiciones en este teorema equivale a que $\hat{K}_\Omega = K$. En este caso se dice que K es $\mathcal{O}(\Omega)$ -convexo.

Puesto que cualquier abierto Ω del plano admite un recubrimiento por compactos $\mathcal{O}(\Omega)$ -convexos [13, Theorem 13.3], como consecuencia de este teorema tenemos el siguiente resultado [13, Theorem 13.9].

Teorema 6. Sean $\Omega \subset \mathbb{C}$ un conjunto abierto del plano, Λ un conjunto que interseque cualquier componente de $\mathbb{C}_\infty \setminus \Omega$ y f una función holomorfa en Ω . Entonces, existe una sucesión de funciones racionales $\{r_n\}$ con polos solo en Λ tal que $r_n \rightarrow f$ uniformemente en subconjuntos compactos de Ω .

En el caso especial en el que $\mathbb{C}_\infty \setminus \Omega$ sea conexo, podemos tomar $\Lambda = \{\infty\}$ para concluir que existe una sucesión de polinomios $\{p_n\}$ tal que $p_n \rightarrow f$ uniformemente en subconjuntos compactos de Ω^4 .

Como hemos mencionado en la introducción, ahora estudiaremos el grado de precisión de la aproximación polinómica en el teorema de Runge [7, Theorem 16.6.5] (Szegő, 1927).

Teorema 7. Sean $\Omega \subset \mathbb{C}$ un dominio plano acotado con frontera rectificable tal que $\mathbb{C}_\infty \setminus \Omega$ es conexo, $K \subset \Omega$ un subconjunto compacto y $\delta = \text{dist}(K, \partial\Omega) > 0$. Si f es holomorfa en Ω y continua en $\bar{\Omega}$, entonces existen constantes $A > 0$, $0 < \kappa < 1$ y una sucesión de polinomios $\{p_n\}$ con p_n de grado a lo sumo n tales que

$$(1) \quad \|f - p_n\|_K < A\kappa^n.$$

Aquí, κ depende de Ω y δ , pero A también depende de f .

Demostración. Por la fórmula integral de Cauchy, para $z \in K$

$$(2) \quad f(z) = \frac{1}{2\pi i} \oint_{\partial\Omega} \frac{f(\zeta)}{\zeta - z} d\zeta.$$

Ahora, como en la demostración del teorema 1, dividimos $\partial\Omega$ en una unión finita de arcos γ_j tales que $\gamma_j \subset \Delta(\zeta_j, r_j)$, $K \cap \Delta(\zeta_j, r_j) = \emptyset$ para los que

$$(3) \quad \left| \frac{1}{\zeta - z} - \frac{1}{\zeta_j - z} \right| < \frac{1}{4\varrho}, \quad \zeta, \zeta_j \in \gamma_j, z \in K,$$

donde $\varrho = \text{diám } \partial\Omega$ (el diámetro de $\partial\Omega$, que es finito porque Ω se supone acotado). Por el teorema 6 podemos, para cada j , elegir un polinomio q_j tal que

$$\left| \frac{1}{\zeta_j - z} - q_j(z) \right| < \frac{1}{4\varrho}, \quad z \in K.$$

⁴De hecho, esta propiedad de aproximación caracteriza los dominios simplemente conexos del plano [13, Theorem 13.11].

Así, por (3), para todo $\zeta \in \gamma_j$ y $z \in K$ tenemos que

$$(4) \quad \left| \frac{1}{\zeta - z} - q_j(z) \right| \leq \left| \frac{1}{\zeta - z} - \frac{1}{\zeta_j - z} \right| + \left| \frac{1}{\zeta_j - z} - q_j(z) \right| < \frac{1}{2\varrho}.$$

Ahora definimos

$$(5) \quad Q_m(z) := \sum_j \frac{1}{2\pi i} \int_{\gamma_j} \frac{1 - (1 - (\zeta - z)q_j(z))^m}{\zeta - z} f(\zeta) d\zeta, \quad m = 1, 2, \dots$$

Como para todo $w \in \mathbb{C}$ tenemos que $1 - (1 - w)^m = w \sum_{k=0}^{m-1} (1 - w)^k$, la función racional que aparece en el integrando de cada sumando en (5) es realmente un polinomio en (ζ, z) cuyo grado en z es $(d_j + 1)(m - 1) + d_j = (d_j + 1)m - 1$ si el grado de q_j es d_j . Así, si $d = \max_j d_j$, Q_m es un polinomio de grado a lo sumo $(d + 1)m - 1$.

De (2) y (5), para $z \in K$ tenemos que

$$f(z) - Q_m(z) = \sum_j \frac{1}{2\pi i} \int_{\gamma_j} \frac{(1 - (\zeta - z)q_j(z))^m}{\zeta - z} f(\zeta) d\zeta$$

y, por tanto,

$$|f(z) - Q_m(z)| \leq \frac{2^{-m}}{2\pi} \int_{\partial\Omega} \frac{|f(\zeta)|}{|\zeta - z|} |d\zeta| \leq \frac{\|f\|_{\partial\Omega} \ell(\partial\Omega)}{2\pi\delta} 2^{-m},$$

donde $\ell(\partial\Omega)$ denota la longitud de $\partial\Omega$, ya que, por (4), $|1 - (\zeta - z)q_j(z)| \leq 1/2$ para $\zeta \in \gamma_j$ y $z \in K$. Ahora, para $n = 1, 2, \dots$, sea $m_n = \lfloor n/(d + 1) \rfloor$ el mayor entero menor o igual que $n/(d + 1)$. Entonces, $p_n := Q_{m_n}$ es un polinomio de grado a lo sumo n para el que (1) se verifica con $A = \|f\|_{\partial\Omega} \ell(\partial\Omega)/(\pi\delta)$ y $\kappa = 2^{-1/(d+1)}$. ■

Por lo general, cuando δ decrece, d normalmente crece, por lo que $A \rightarrow \infty$ y $\kappa \rightarrow 1$ cuando $\delta \searrow 0^+$.

3. La ecuación $\bar{\partial}$

La ecuación $\bar{\partial}$ consiste en, dada $f \in C^\infty(\Omega)$ (o en cualquier otro espacio de funciones adecuado), encontrar $u \in C^\infty(\Omega)$ tal que

$$(6) \quad \frac{\partial u}{\partial \bar{z}} = f$$

en Ω . Aquí

$$\frac{\partial u}{\partial \bar{z}} = \frac{1}{2} \left(\frac{\partial u}{\partial x} + i \frac{\partial u}{\partial y} \right),$$

donde

$$\frac{\partial u}{\partial x} = \frac{\partial \alpha}{\partial x} + i \frac{\partial \beta}{\partial x} \quad \text{y} \quad \frac{\partial u}{\partial y} = \frac{\partial \alpha}{\partial y} + i \frac{\partial \beta}{\partial y}$$

si α y β denotan las partes real e imaginaria de u , respectivamente.

Obsérvese que $\partial u / \partial \bar{z} = 0$ equivale a que $\partial u / \partial x = -i \partial u / \partial y$, que a su vez se puede reescribir como

$$\begin{cases} \frac{\partial \alpha}{\partial x} = \frac{\partial \beta}{\partial y}, \\ \frac{\partial \alpha}{\partial y} = -\frac{\partial \beta}{\partial x}, \end{cases}$$

y que reconocemos como las ecuaciones de Cauchy-Riemann. Así, $\partial u / \partial \bar{z} = 0$ en Ω significa que u es holomorfa en Ω .

La importancia de la ecuación no homogénea (6) radica en su uso para la construcción de funciones holomorfas que, por su rigidez (no existen «particiones de la unidad holomorfas»), hacen de ella una

herramienta extremadamente útil. Tras el éxito que ha supuesto, no solo en una, sino en varias variables complejas (donde la naturaleza de (6) es, si cabe, aún mas fundamental), esta idea ha dado lugar al llamado *principio de Oka*: si podemos resolver un problema (de construcción de funciones holomorfas) en la categoría de funciones continuas, entonces también será resoluble en la categoría de funciones holomorfas. Ejemplos de esta técnica se pueden encontrar en las pruebas de los teoremas clásicos de Mittag-Leffler y Weierstrass [8, Theorem 1.4.3 y Theorem 1.5.1; 10, sección 1.4.1].

Que la ecuación $\partial u/\partial \bar{z} = f$ tiene solución en cualquier abierto Ω para cualquier dato $f \in C^\infty(\Omega)$ se puede probar usando el teorema de Runge [8, Theorem 1.4.4; 10, teorema 1.16]. La demostración es consecuencia del corolario 10, que a su vez se sigue de la proposición 16.3.1 del libro de Rudin [12].

Proposición 8 (Cauchy-Green). *Si $u \in C_c^1(\mathbb{C})$, entonces*

$$(7) \quad u(z) = \frac{1}{\pi} \int_{\mathbb{C}} \frac{\partial u/\partial \bar{w}(w)}{z-w} dA(w).$$

Demostración. Por la regla de la cadena (ver la observación 9), en coordenadas polares tenemos que

$$(8) \quad \frac{\partial}{\partial \bar{z}} = \frac{1}{2} e^{i\theta} \left(\frac{\partial}{\partial \varrho} + \frac{i}{\varrho} \frac{\partial}{\partial \theta} \right).$$

Alternativamente, hemos de encontrar dos funciones $A = A(\varrho, \theta)$ y $B = B(\varrho, \theta)$ que cumplan la relación $\partial/\partial \bar{z} = A\partial/\partial \varrho + B\partial/\partial \theta$. Para ello, basta testar esta igualdad con las funciones u y \bar{u} (la función conjugada de u), donde $u(z) = z = \varrho e^{i\theta}$; para la primera tenemos que $e^{i\theta}A + ie^{i\theta}\varrho B = 0$ y, para la segunda, $e^{-i\theta}A - ie^{-i\theta}\varrho B = 1$. Así,

$$\left. \begin{aligned} A + i\varrho B &= 0, \\ A - i\varrho B &= e^{i\theta} \end{aligned} \right\} \implies \begin{cases} A = e^{i\theta}/2, \\ B = ie^{i\theta}/2\varrho. \end{cases}$$

Para $z = 0$, el lado derecho de (7) es igual al límite cuando $\varepsilon \rightarrow 0^+$ de

$$(9) \quad -\frac{1}{2\pi} \int_{\varepsilon}^{\infty} \left(\int_0^{2\pi} \left(\frac{\partial u}{\partial \varrho} + \frac{i}{\varrho} \frac{\partial u}{\partial \theta} \right) d\theta \right) d\varrho$$

y, puesto que la función $\theta \mapsto u(\varrho e^{i\theta})$ es 2π -periódica, la integral de $\partial u/\partial \theta$ es cero y (9) es, atendiendo al teorema de Fubini, igual a

$$-\frac{1}{2\pi} \int_0^{2\pi} \left(\int_{\varepsilon}^{\infty} \frac{\partial u}{\partial \varrho} d\varrho \right) d\theta = \frac{1}{2\pi} \int_0^{2\pi} u(\varepsilon e^{i\theta}) d\theta \xrightarrow{\varepsilon \rightarrow 0^+} u(0),$$

ya que $u(z) = 0$ para $|z|$ suficientemente grande (u tiene soporte compacto) y $u(\varepsilon e^{i\theta}) \rightarrow u(0)$ uniformemente cuando $\varepsilon \rightarrow 0^+$. Esto prueba (7) para $z = 0$.

El caso general se sigue aplicando este a las trasladadas de u : si $a \in \mathbb{C}$ y $u_a(z) = u(a+z)$, de (7) para $z = 0$ tenemos que

$$u(a) = u_a(0) = -\frac{1}{\pi} \int_{\mathbb{C}} \frac{\partial u_a/\partial \bar{w}(w)}{w} dA(w) = -\frac{1}{\pi} \int_{\mathbb{C}} \frac{\partial u/\partial \bar{w}(a+w)}{w} dA(w) = \frac{1}{\pi} \int_{\mathbb{C}} \frac{\partial u/\partial \bar{w}(w)}{a-w} dA(w),$$

que es (7) con $z = a$. ■

Observación 9. Como se ha dicho, (8) también se sigue de la regla de la cadena, a saber,

$$\begin{aligned} \frac{\partial}{\partial \varrho} &= \frac{\partial x}{\partial \varrho} \frac{\partial}{\partial x} + \frac{\partial y}{\partial \varrho} \frac{\partial}{\partial y} = \cos \theta \frac{\partial}{\partial x} + \sin \theta \frac{\partial}{\partial y}, \\ \frac{\partial}{\partial \theta} &= \frac{\partial x}{\partial \theta} \frac{\partial}{\partial x} + \frac{\partial y}{\partial \theta} \frac{\partial}{\partial y} = -\varrho \sin \theta \frac{\partial}{\partial x} + \varrho \cos \theta \frac{\partial}{\partial y}, \end{aligned}$$

por lo que $\partial/\partial x = (\varrho \cos \theta \partial/\partial \varrho - \sin \theta \partial/\partial \theta)/\varrho$ y $\partial/\partial y = (\varrho \sin \theta \partial/\partial \varrho + \cos \theta \partial/\partial \theta)/\varrho$. Así,

$$\begin{aligned} \frac{\partial}{\partial \bar{z}} &= \frac{1}{2} \left(\frac{\partial}{\partial x} + i \frac{\partial}{\partial y} \right) = \frac{1}{2} \left((\cos \theta + i \sin \theta) \frac{\partial}{\partial \varrho} + \frac{(-\sin \theta + i \cos \theta)}{\varrho} \frac{\partial}{\partial \theta} \right) \\ &= \frac{1}{2} \left((\cos \theta + i \sin \theta) \frac{\partial}{\partial \varrho} + \frac{i}{\varrho} (\cos \theta + i \sin \theta) \frac{\partial}{\partial \theta} \right) = \frac{1}{2} e^{i\theta} \left(\frac{\partial}{\partial \varrho} + \frac{i}{\varrho} \frac{\partial}{\partial \theta} \right). \end{aligned} \quad \blacktriangleleft$$

Como consecuencia, tenemos el siguiente corolario.

Corolario 10. Sean $\Omega \subset \mathbb{C}$ abierto y $f \in C_c^1(\Omega)$. Si

$$(10) \quad u(z) = \frac{1}{\pi} \int_{\Omega} \frac{f(w)}{z-w} dA(w), \quad z \in \Omega,$$

entonces $u \in C^1(\Omega)$ y $\partial u/\partial \bar{z} = f$ en Ω .

Demostración. La fórmula (10) se puede escribir como

$$u(z) = \frac{1}{\pi} \int_{\mathbb{C}} \frac{f(z-w)}{w} dA(w)$$

y permite ver que $u \in C^1(\Omega)$ puesto que esta expresión se puede derivar bajo el signo integral, ya que la función $1/z$ es localmente integrable.

Fijemos $a \in \Omega$ y sea $\psi \in C_c^1(\Omega)$ tal que $\psi \equiv 1$ en un entorno $U \Subset \Omega$ ($\bar{U} \subset \Omega$) de a . Si reemplazamos f por $(1-\psi)f$ en (10), la función resultante es holomorfa en U y, por tanto, f se puede reemplazar por ψf para el cómputo de $\partial u/\partial \bar{z}(a)$. Pero derivando bajo el signo integral, de (7) tenemos que

$$\frac{\partial u}{\partial \bar{z}}(a) = \frac{1}{\pi} \int_{\mathbb{C}} \frac{\partial(\psi f)/\partial \bar{w}(a-w)}{w} dA(w) = \frac{1}{\pi} \int_{\mathbb{C}} \frac{\partial(\psi f)/\partial \bar{w}(w)}{a-w} dA(w) = (\psi f)(a) = f(a). \quad \blacksquare$$

Obsérvese que, de hecho, aquí Ω no es relevante: la función u dada en (10) es $C^\infty(\mathbb{C})$ y $\partial u/\partial \bar{z} = f$ en todo el plano, y la segunda parte en la demostración del corolario 10 muestra que (10) sigue siendo solución cuando el segundo miembro tenga sentido; por ejemplo, si Ω es acotado y $f \in C^1(\Omega)$ es acotada [12, Proposition 16.3.2].

Ahora estamos preparados para la demostración del teorema de existencia de soluciones para la ecuación (6) (véase el libro de Hörmander [8, Theorem 1.4.4]).

Teorema 11. Para todo abierto $\Omega \subset \mathbb{C}$ y toda función $f \in C^1(\Omega)$, la ecuación $\partial u/\partial \bar{z} = f$ tiene una solución $u \in C^1(\Omega)$.

Demostración. Sean $K_1 \subset \mathring{K}_2 \Subset \Omega$ dos subconjuntos compactos de Ω y $\psi_1, \psi_2 \in C_c^\infty(\Omega)$ tales que, para $j = 1, 2$, $\psi_j \equiv 1$ en un entorno de K_j . Por el corolario 10, existen dos funciones $u_1, u_2 \in C^1(\Omega)$ tales que $\partial u_j/\partial \bar{z} = \psi_j f$ ($j = 1, 2$), pero, en general, no podemos controlar cuán cerca está u_2 de u_1 en K_1 (el menor de los compactos), aunque, ciertamente, $u_2 - u_1$ es holomorfa en un entorno de K_1 . Así, si además suponemos que K_1 es $\mathcal{O}(\Omega)$ -convexo y $\varepsilon > 0$, por el teorema de Runge podemos encontrar $v \in \mathcal{O}(\Omega)$ tal que $\|u_2 - u_1 - v\|_{K_1} < \varepsilon$. Es decir, reemplazando u_1 por $u_1 + v$ podemos conseguir que $\partial u_j/\partial \bar{z} = \psi_j f$ para $j = 1, 2$ con $\|u_2 - u_1\|_{K_1}$ tan pequeño como se desee.

De este modo, si $\{K_j\}_{j \geq 1}$ es un recubrimiento normal de Ω [13, Theorem 13.3], es decir, un recubrimiento de Ω por compactos $\mathcal{O}(\Omega)$ -convexos, podemos construir una sucesión de funciones $\{u_j\} \subset C^1(\Omega)$ tal que $\partial u_j/\partial \bar{z} = \psi_j f$ (en particular, $u_{j+1} - u_j$ es holomorfa en un entorno de K_{j-1}) y $\|u_{j+1} - u_j\|_{K_j} < 2^{-j}$ para todo $j = 1, 2, \dots$. Por tanto, esta sucesión converge uniformemente en compactos a una función $u = \lim_j u_j = \sum_j (u_{j+1} - u_j) \in C^1(\Omega)$ que satisface $\partial u/\partial \bar{z} = f$ en Ω ya que, por las estimaciones de Cauchy [13, Theorem 10.26], $\|(u_{j+1} - u_j)'\|_K \leq C_K 2^{-j}$ para todo subconjunto compacto $K \subset \Omega$ si j es suficientemente grande. \blacksquare

Nótese que la solución así construida es $C^\infty(\Omega)$ cuando $f \in C^\infty(\Omega)$. Esto también es consecuencia de la elipticidad del operador $\bar{\partial}$.

4. Una generalización del teorema de Runge

Por el teorema de representación de Riesz [13, Theorem 2.14], el dual de $C(K)$ con la métrica uniforme coincide con el espacio de las medidas complejas en K . Esto quiere decir que cualquier funcional lineal acotado en $C(K)$ es de la forma

$$f \in C(K) \longrightarrow \mu[f] := \int_K f d\mu \in \mathbb{C}$$

para alguna medida compleja μ en K .

Sea $\mathcal{R}_\Lambda(K)$ el subespacio de $C(K)$ que consiste en las restricciones a K de las funciones racionales con polos en Λ . El teorema 2 afirma que, si $\Lambda \subset \mathbb{C}_\infty \setminus K$ contiene al menos un punto en cada componente de $\mathbb{C}_\infty \setminus K$, $\mathcal{R}_\Lambda(K)$ es denso en el subespacio de $C(K)$ formado por las restricciones a K de las funciones holomorfas en un entorno de K , y, de acuerdo al teorema de Hahn-Banach [13, Theorem 5.19], esto es equivalente a probar que si una medida compleja μ es «ortogonal» a $\mathcal{R}_\Lambda(K)$ ($\mu \perp \mathcal{R}_\Lambda(K)$), es decir, tal que $\mu[r] = 0$ para toda $r \in \mathcal{R}_\Lambda(K)$, también lo es al subespacio de las restricciones a K de funciones holomorfas en un entorno de K , esto es, $\mu[f] = 0$ para toda función f holomorfa en un entorno de K .

De hecho, la prueba funcional clásica [13, Theorem 13.6] permite relajar las hipótesis en el teorema 2 [4, Theorem 1.1].

Teorema 12. Sean $K \subset \mathbb{C}$ un subconjunto compacto del plano complejo y $\Lambda \subset \mathbb{C}_\infty \setminus K$ un conjunto que contenga un punto en cada componente conexa de $\mathbb{C}_\infty \setminus K$. Si f es C^1 en un entorno de K y $\partial f / \partial \bar{z} = 0$ en K , entonces f se puede aproximar uniformemente en K por funciones racionales con polos en Λ .

Demostración. Sea $\mu \perp \mathcal{R}_\Lambda(K)$. Puesto que f es C^1 en un entorno de K , podemos, sin pérdida de generalidad, suponer que f es $C_c^1(\mathbb{C})$. Por (7) y el teorema de Fubini,

$$(11) \quad \mu[f] = \int_K f(z) d\mu(z) = \int_K \left(\frac{1}{\pi} \int_{\mathbb{C}} \frac{\partial f / \partial \bar{w}(w)}{z-w} dA(w) \right) d\mu(z) = \int_{\mathbb{C} \setminus K} \frac{\partial f}{\partial \bar{w}}(w) h(w) dA(w),$$

donde

$$h(w) := \frac{1}{\pi} \int_K \frac{d\mu(z)}{z-w}, \quad w \notin K.$$

La función h así definida es holomorfa en $\mathbb{C} \setminus K$ y

$$h^{(n)}(w) = \frac{n!}{\pi} \int_K \frac{d\mu(z)}{(z-w)^{n+1}}, \quad w \notin K,$$

para $n = 0, 1, \dots$. Puesto que para cada $a \in \Lambda$ las funciones $z \mapsto 1/(z-a)^n$ pertenecen a $\mathcal{R}_\Lambda(K)$, por hipótesis, $h^{(n)}(a) = 0$ para todo $n = 0, 1, \dots$ y, del teorema de identidad, $h \equiv 0$ en cualquier componente acotada de $\mathbb{C} \setminus K$. De igual manera, $h(\infty) = 0$ y

$$h^{(n)}(\infty) = \frac{d}{dw} \Big|_{w=0} h(1/w) = n \frac{d^{n-1}}{dw^{n-1}} \Big|_{w=0} \frac{1}{\pi} \int_K \frac{d\mu(z)}{zw-1} = -\frac{n!}{\pi} \int_K z^{n-1} d\mu(z) = 0$$

para $n = 1, 2, \dots$. Con todo, $h \equiv 0$ en $\mathbb{C} \setminus K$ y $\mu[f] = 0$. ■

Existe un teorema de aproximación polinómica debido a Mergelyan [13, Theorem 20.5] que afirma que, para compactos K con complementario $\mathbb{C}_\infty \setminus K$ conexo, cualquier función continua en K y holomorfa en su interior puede ser aproximada uniformemente por polinomios. De hecho, si el complementario de K tiene un número finito de componentes, cualquier función continua en K y holomorfa en su interior se puede aproximar uniformemente en K por funciones racionales sin polos en K (esto es consecuencia del teorema 10.4 en el libro de Gamelin [4], que establece que para esto es suficiente con que los diámetros de las componentes de $\mathbb{C} \setminus K$ estén uniformemente acotados inferiormente, es decir, que exista $\delta > 0$ tal que $\text{diám } V \geq \delta$ para cualquier componente conexa V de $\mathbb{C} \setminus K$).

El resultado análogo al de Runge para funciones racionales es falso. Existen compactos K con interior vacío y una función continua f en K que no se puede aproximar uniformemente en K por funciones racionales con polos fuera de K .

Un ejemplo viene dado por el llamado *queso suizo*⁵ S que se obtiene del disco unidad cerrado $\bar{\Delta}$ tras quitar una sucesión de discos abiertos $\{\Delta_j\}$ con clausuras mutuamente disjuntas y tales que $\bar{\Delta} \setminus \bigcup_j \Delta_j$ tiene interior vacío (figura 1).

⁵Introducido en 1938 por la matemática suiza Alice Roth en su tesis doctoral [11, pág. 96], más parece un queso gruyere que, por cierto, recibe su nombre del distrito de Gruyère del cantón de Friburgo (Suiza) donde se fabrica.

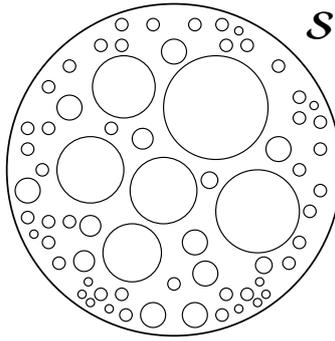


Figura 1: Queso suizo.

Para su construcción, sea $S = \{s_1, s_2, \dots\}$ una sucesión densa en el disco unidad abierto Δ y elijamos $\Delta_j = \Delta(s_j, r_j) \subset \Delta$ cuyo centro es el primer elemento de S que no esté contenido en ningún disco previo Δ_k para $k < j$ para entonces tomar $r_j > 0$ lo suficientemente pequeño para que $\bar{\Delta}_j$ sea disjunto con $\bar{\Delta}_k$ si $k < j$.

Supongamos, además, que para la sucesión de radios así elegida se cumple que $\sum_j r_j^2 < 1$ y sea $S = \bar{\Delta} \setminus \bigcup_j \Delta_j$. S es compacto con interior vacío (por construcción, $S \cap \Delta_j = \emptyset$ para todo j), ya que es cerrado, acotado y S es denso. Si f es cualquier función continua en S , entonces $|\oint_{\partial \Delta_j} f(\zeta) d\zeta| \leq 2\pi \|f\|_{S} r_j$, por lo que la serie $\sum_j \oint_{\partial \Delta_j} f(\zeta) d\zeta$ converge absolutamente. Por el teorema de Cauchy, $\oint_{\partial \Delta} f(\zeta) d\zeta = \sum_j \oint_{\partial \Delta_j} f(\zeta) d\zeta$ si f es racional con polos fuera de S y, por tanto, también si f es límite uniforme de tales funciones. Pero

$$\oint_{\partial \Delta_j} \bar{\zeta} d\zeta = ir_j \int_{-\pi}^{\pi} (\bar{s}_j + r_j e^{-i\theta}) e^{i\theta} d\theta = i\bar{s}_j r_j \int_{-\pi}^{\pi} e^{i\theta} d\theta + 2\pi i r_j^2 = 2\pi i r_j^2$$

y $\oint_{\partial \Delta} \bar{\zeta} d\zeta = 2\pi i$. Puesto que $\sum_j r_j^2 < 1$, la función $f(z) = \bar{z}$ no se puede aproximar uniformemente en S por funciones racionales sin polos en S . Nótese que lo anterior implica que la medida

$$d\mu(z) := \begin{cases} dz & \text{en } \partial \Delta, \\ -dz & \text{en } \bigcup_j \partial \Delta_j \end{cases}$$

es ortogonal al espacio $\mathcal{R}(S)$ de todas las funciones racionales con polos fuera de S y, sin embargo, $\mu[\bar{z}] = 2\pi i(1 - \sum_j r_j^2) \neq 0^6$.

Observación 13. Que S también es conexo es consecuencia de un resultado de topología general que establece que la intersección encajada de compactos conexos sigue siendo compacta y conexa. Como $S = \bigcap_j \bar{\Delta} \setminus D_k$ con $D_k = \bigcup_{j \leq k} \Delta_j$ y $\bar{\Delta} \setminus D_k$ es conexo, S también lo es. La prueba de este resultado topológico va como sigue: si $\{K_j\}_{j \geq 1}$ es una familia de compactos conexos encajados ($K_{j+1} \subset K_j$) en un espacio topológico X y U, V son abiertos disjuntos tales que $K = \bigcap_j K_j \subset W = U \cup V$, entonces $\bigcap_j K_j \cap X \setminus W = \emptyset$, que, por compacidad, implica que $\bigcap_{\ell \leq m} K_{j_\ell} \cap X \setminus W = \emptyset$ ya que $K_j \cap X \setminus W$ es cerrado para $j \geq 1$. Así, $K_{j_m} \cap X \setminus W = \emptyset$ y, por tanto, $K_{j_m} \subset W = U \cup V$. Puesto que K_{j_m} es conexo y $U \cap V = \emptyset$, se sigue que $K_{j_m} \subset U$ o $K_{j_m} \subset V$. Esto implica que $K \subset K_{j_m} \subset U$ o $K \subset K_{j_m} \subset V$ y K es conexo. Por último, K es compacto ya que es intersección de compactos.

Para $j \geq 1$, los conjuntos $K_j := \{z \in \mathbb{C} : x^2/j^2 + y^2 \geq 1, |y| \leq 1\}$ son conexos y encajados pero, sin embargo, $K = \bigcap_j K_j = \{y = \pm 1\}$ no es conexo. Así, en general, la intersección encajada de conexos no es conexa. Esto prueba que, en el resultado anterior, no se puede omitir la hipótesis de compacidad. ◀

⁶La condición $\sum_j r_j^2 < 1$ es necesaria ya que, en caso contrario, S tendría medida cero y cualquier función continua en S se podría aproximar uniformemente en S por funciones racionales [6].

5. Relación entre el problema $\bar{\partial}$ y el teorema de Runge

Anteriormente hemos visto que el teorema de Runge implica la existencia de soluciones para la ecuación $\bar{\partial}$. En esta sección veremos que, recíprocamente, la existencia de soluciones para el problema $\bar{\partial}$ con estimaciones implica el teorema de Runge.

Como antes, consideremos K un subconjunto compacto de un abierto $\Omega \subset \mathbb{C}$. Resulta interesante que los siguientes problemas⁷ son equivalentes.

Problema A ((iv) en el teorema 5). Aproximar uniformemente en K cualquier función holomorfa en un entorno de K por funciones holomorfas en Ω . ◀

Problema H. Dada $f \in C_c^\infty(\Omega)$ tal que $K \cap \text{sop } f = \emptyset$ y $\varepsilon > 0$, encontrar una solución $u \in C^\infty(\Omega)$ de $\partial u / \partial \bar{z} = f$ tal que $\|u\|_K < \varepsilon$. ◀

Problema A implica problema A. Sea $u_0 \in C^\infty$ cualquier solución de $\partial u_0 / \partial \bar{z} = f$ en Ω (teorema 11). Puesto que $K \cap \text{sop } f = \emptyset$, u_0 es holomorfa en $\Omega \setminus \text{sop } f \supset K$ y, para $\varepsilon > 0$ dado, por hipótesis existirá $h \in \mathcal{O}(\Omega)$ tal que $\|u_0 - h\|_K < \varepsilon$. La función $u := u_0 - h$ satisface que $\partial u / \partial \bar{z} = f$ y $\|u\|_K < \varepsilon$. ■

Problema H implica problema A. Sean $g \in \mathcal{O}(U)$ holomorfa en algún entorno $U \Subset \Omega$ de K ($K \subset U \subset \bar{U} \subset \Omega$), $\varepsilon > 0$ y consideremos $\chi \in C_c^\infty(U)$ tal que $\chi \equiv 1$ en un entorno de K . Para $f = g \partial \chi / \partial \bar{z} \in C_c^\infty(\Omega)$ existirá $u \in C^\infty(\Omega)$ tal que $\partial u / \partial \bar{z} = f$ y $\|u\|_K < \varepsilon$. Entonces, la función $h := \chi g - u$ es holomorfa en Ω (ya que g es holomorfa en $U \supset \text{sop } \chi$ y, por tanto, $\partial h / \partial \bar{z} = g \partial \chi / \partial \bar{z} - \partial u / \partial \bar{z} = 0$ en Ω) y $\|g - h\|_K = \|u\|_K < \varepsilon$. ■

Observación 14. Como sabemos, si el problema A tiene solución, entonces $\hat{K}_\Omega = K$ (ver la discusión tras el teorema 5), es decir, cualquier componente de $\Omega \setminus K$ interseca $\mathbb{C} \setminus \Omega$. Esto se puede ver directamente del problema H: si $a \in \Omega \setminus K$ y $\chi \in C_c^\infty(\Omega)$ es tal que $\chi \equiv 1$ en un entorno de a y $\chi \equiv 0$ en algún entorno de K , entonces, para $\tilde{K} = K \cup \{a\}$, se tiene que $\tilde{K} \cap \text{sop } \partial \chi / \partial \bar{z} = \emptyset$ y, por tanto, si relativo al \tilde{K} el problema H tiene solución u con dato $f = \partial \chi / \partial \bar{z}$, entonces $h := \chi - u \in \mathcal{O}(\Omega)$, $|h(a) - 1| < \varepsilon$ y $|h| < \varepsilon$ en K . Así, si $\varepsilon \leq 1/2$, $|h(a)| > 1 - \varepsilon \geq \varepsilon > \|h\|_K$. Puesto que $a \in \Omega \setminus K$ es arbitrario, concluimos que $\hat{K}_\Omega = K$. ◀

El siguiente es un ejemplo directo (trivial desde el punto de vista del problema A) donde el problema H para $\Omega = \mathbb{C}$ no tiene solución.

Ejemplo 15. Sea $\chi \in C^\infty(\mathbb{C})$ tal que $\text{sop } \chi = \{z \in \mathbb{C} : |z| \geq 1/2\}$ y $\chi \equiv 1$ para $|z| \geq 3/4$. Tomando $f = z^{-1} \partial \chi / \partial \bar{z} \in C_c^\infty(\mathbb{C})$ y $K = \mathbb{T} = \{z \in \mathbb{C} : |z| = 1\}$, el problema H no tiene solución. Efectivamente, nótese que $K \cap \text{sop } f = \emptyset$ ya que $\text{sop } \partial \chi / \partial \bar{z} \subset \{z \in \mathbb{C} : 1/2 \leq |z| \leq 3/4\}$ y que cualquier solución del problema $\partial u / \partial \bar{z} = f$ hace que $h = zu - \chi$ sea entera. Pero, por el principio del máximo, tenemos que $1 = |h(0) + 1| \leq \|h + 1\|_{\mathbb{T}} = \|u\|_{\mathbb{T}}$. Esto quiere decir que la solución al problema $\bar{\partial}$ no se puede elegir arbitrariamente pequeña en \mathbb{T} : el problema H no tiene solución si $\varepsilon < 1$. ◀

Tras esta discusión, podemos enunciar el siguiente teorema.

Teorema 16. Sea K un subconjunto compacto de un abierto $\Omega \subset \mathbb{C}$. Los siguientes enunciados son equivalentes.

- Cualquier función holomorfa en un entorno de K se puede aproximar uniformemente en K por funciones holomorfas en Ω .
- Para todo $\varepsilon > 0$ y toda función $f \in C_c^\infty(\Omega)$ tal que $K \cap \text{sop } f = \emptyset$ existe una solución $u \in C^\infty(\Omega)$ de la ecuación $\partial u / \partial \bar{z} = f$ tal que $\|u\|_K < \varepsilon$.
- $\hat{K}_\Omega = K$.

Es decir, podemos añadir la propiedad (b) a la lista de equivalencias en el teorema 5. La relación entre (b) y (c) está implícita en la observación anterior.

Una forma de abordar el problema H es vía el siguiente teorema de Hörmander, que proporciona soluciones al problema (6) con estimaciones L^2 .

⁷Las denominaciones A y H en estos problemas se refieren a Aproximación y Hörmander, respectivamente.

Teorema 17 (Hörmander [8]). *Sea $\phi \in C^2(\Omega)$ una función subarmónica en un dominio $\Omega \subset \mathbb{C}$, es decir, $\Delta\phi = 4\partial^2\phi/\partial z\partial\bar{z} \geq 0$ en Ω . Entonces, para toda $f \in L^2_{\text{loc}}(\Omega)$ existe una solución⁸ $u \in L^2_{\text{loc}}(\Omega)$ de $\partial u/\partial\bar{z} = f$ tal que*

$$\int_{\Omega} |u|^2 e^{-\phi} dA \leq \int_{\Omega} \frac{|f|^2}{\Delta\phi} e^{-\phi} dA.$$

Así, para probar (b) en el teorema 16, todo lo que hemos de hacer es encontrar «pesos» ϕ adecuados. Sirva el siguiente ejemplo como ilustración.

Ejemplo 18. Sean $r < 1$, $K = \overline{\Delta(0, r)} = \{z \in \mathbb{C} : |z| \leq r\}$ y $f \in C^{\infty}(\mathbb{C})$ con $\bar{\Delta} \cap \text{sop } f = \emptyset$, donde $\Delta = \Delta(0, 1)$ denota el disco unidad. La función $\phi(z) = \ln(1 + |z|^2)$ es estrictamente subarmónica en \mathbb{C} , ya que

$$\Delta\phi(z) = 4 \frac{\partial^2}{\partial z \partial \bar{z}} \ln(1 + |z|^2) = 4 \frac{\partial}{\partial \bar{z}} \left(\frac{\bar{z}}{1 + |z|^2} \right) = \frac{4}{(1 + |z|^2)^2} > 0.$$

Por el teorema 17 aplicado a la función $\phi_m = m\phi$ ($m = 1, 2, \dots$), existe una solución $u = u_m \in C^{\infty}(\mathbb{C})$ de $\partial u/\partial\bar{z} = f$ tal que

$$\int_{\mathbb{C}} \frac{|u(z)|^2}{(1 + |z|^2)^m} dA(z) \leq \frac{1}{4m} \int_{\mathbb{C}} \frac{|f(z)|^2}{(1 + |z|^2)^{m-2}} dA(z) \leq \frac{\|f\|_{\mathbb{C}}^2 |\text{sop } f|}{m2^m},$$

\uparrow
 $\bar{\Delta} \cap \text{sop } f = \emptyset$

donde $|\text{sop } f|$ denota la medida de área de $\text{sop } f$, y, por tanto,

$$\int_{\Delta} |u(z)|^2 dA(z) \leq 2^m \int_{\Delta} \frac{|u(z)|^2}{(1 + |z|^2)^m} dA(z) \leq 2^m \int_{\mathbb{C}} \frac{|u(z)|^2}{(1 + |z|^2)^m} dA(z) \leq \frac{\|f\|_{\mathbb{C}}^2 |\text{sop } f|}{m}.$$

Puesto que u es holomorfa en Δ , por la propiedad del valor medio [5, Chapter III: §4; 10, corolario 2.5] y la desigualdad de Hölder concluimos que

$$\|u\|_K \leq \frac{1}{\sqrt{\pi}\delta} \left(\int_{\Delta} |u(z)|^2 dA(z) \right)^{1/2} \leq \frac{\|f\|_{\mathbb{C}} |\text{sop } f|^{1/2}}{\delta\sqrt{\pi m}} \xrightarrow{m \rightarrow +\infty} 0$$

si $\delta = 1 - r > 0$. De hecho, $\|u\|_K < \varepsilon$ tan pronto como

$$m \geq m_{\delta}(\varepsilon) := \left\lceil \frac{\|f\|_{\mathbb{C}}^2 |\text{sop } f|}{\pi\delta^2\varepsilon^2} \right\rceil + 1. \quad \blacktriangleleft$$

Observación 19. Además, en este ejemplo, como u es holomorfa para $|z|$ suficientemente grande (lo es fuera del soporte de f , que es compacto), tendremos como antes

$$|u(z)| \leq \frac{1}{\pi} \int_{\Delta(z, 1)} |u(w)| dA(w) \leq \frac{1}{\sqrt{\pi}} \left(\int_{\Delta(z, 1)} \frac{|u(w)|^2}{(1 + |w|^2)^m} dA(w) \right)^{1/2} \left(\int_{\Delta(z, 1)} (1 + |w|^2)^m dA(w) \right)^{1/2} \lesssim |z|^m,$$

es decir, u presenta un crecimiento polinómico. Esto implica que la función aproximante $h = u - \chi g$ que se produce en el problema H es un polinomio de grado a lo sumo $m_{\delta}(\varepsilon)$ (h es entera con crecimiento polinómico ya que coincide con u para $|z|$ suficientemente grande). Nótese, además, que en este caso

$$|\text{sop } f| = \left| \text{sop } \frac{\partial \chi}{\partial \bar{z}} \right| \approx 2\pi\delta$$

y

$$\|f\|_{\infty} = \left\| g \frac{\partial \chi}{\partial \bar{z}} \right\|_{\mathbb{C}} \approx \|g\|_K \frac{1}{\delta},$$

⁸En sentido débil. En caso de que $f \in C^{\infty}$, cualquier solución de esta ecuación es automáticamente C^{∞} .

ya que $\text{sop } \partial\chi/\partial\bar{z} \subset \{z \in \mathbb{C} : r < |z| < 1\}$. Por tanto,

$$(12) \quad m_\delta(\varepsilon) \approx \left\lceil \frac{2\|g\|_K^2}{\delta^3\varepsilon^2} \right\rceil + 1.$$

Esto proporciona una estimación del grado del polinomio aproximante en términos de los datos relevantes del problema: el error de aproximación ε y el tamaño δ del entorno de K donde la función a aproximar se supone holomorfa. Como cabe esperar, esta estimación permite ver que el grado del polinomio aproximante es «inversamente» proporcional al tamaño del entorno donde la función originalmente es holomorfa. Obsérvese además que, como es natural, la expresión (12) es homogénea de grado cero en (g, ε) : si un polinomio p aproxima g en K con error ε , entonces, para $\lambda > 0$, el polinomio λp aproxima λg con error $\lambda\varepsilon$ y sus grados son iguales.

Por otro lado, la estimación (12) en general no es óptima. Por ejemplo, si $|a| > 1$ y queremos aproximar $f(z) = 1/(a - z)$ sobre el disco unidad cerrado, podemos truncar el desarrollo de Taylor de f en el origen. Como

$$f(z) = \frac{1}{a} \frac{1}{1 - z/a} = \frac{1}{a} \left(\sum_{k=0}^m \frac{z^k}{a^k} + \sum_{k>m} \left(\frac{z}{a}\right)^k \right)$$

y, para $0 \leq \varrho < 1$,

$$\max_{|w| \leq \varrho} \left| \sum_{k>m} w^k \right| = \max_{|w| \leq \varrho} \frac{|w|^{m+1}}{|1 - w|} = \frac{\varrho^{m+1}}{1 - \varrho},$$

con $\varrho = 1/|a|$ se tiene que

$$\|f - p_m\|_{\bar{\Delta}} = \frac{1}{|a|} \frac{1/|a|^{m+1}}{1 - 1/|a|} = \frac{1}{(|a| - 1)|a|^{m+1}},$$

donde $p_m(z) = \sum_{k=0}^m z^k/a^{k+1}$. Puesto que en este caso $\delta = |a| - 1$ (f es holomorfa en el disco $\Delta(0, |a|)$), se sigue que el error al aproximar f en el disco cerrado por p_m es $\varepsilon = 1/\delta(1 + \delta)^{m+1}$ y, por tanto,

$$m = \frac{\ln(1/\varepsilon\delta)}{\ln(1 + \delta)} - 1 \approx \frac{1}{\delta} \ln(1/\varepsilon\delta)$$

cuando $\delta \rightarrow 0^+$, bastante menor que la estimación $1/\delta^3\varepsilon^2$ dada en (12).

Por otra parte, nótese que en este ejemplo no hemos usado el teorema 17 en toda su generalidad, ya que este permite, de antemano, la elección del peso ϕ . De hecho, con otra elección (adaptada a $\bar{\Delta}$), el teorema 17 proporciona la estimación $m \leq 8 \log(1/\varepsilon)/\delta$ [9] (comparar con el teorema 7). Como se puede comprobar, esta estimación es bastante mejor que $m \approx 1/\delta^2\varepsilon$ obtenida anteriormente por truncación. ◀

6. Soluciones localizadas de la ecuación $\bar{\partial}$

Siguiendo la sugerencia planteada en el libro de Andersson [1, Chapter 3], en esta sección probaremos el problema 18 (que proporciona una versión funcional del teorema de Runge) allí propuesto.

Como hemos visto, por el teorema 11, la ecuación $\partial u/\partial\bar{z} = f$ admite soluciones en cualquier abierto del plano. Por otro lado, esta ecuación con dato $f \in C_c^\infty(\mathbb{C})$ admite soluciones con soporte compacto si y solo si los momentos $\mu_k[f] := \int_{\mathbb{C}} z^k f(z) dA(z) = 0$ para todo $k = 0, 1, \dots$ [10, sección 1.4.2] (obsérvese que esta condición sobre los momentos equivale a que f sea ortogonal a los polinomios, es decir, a las funciones racionales con polo en el infinito). En tal caso, la demostración del teorema 1.27 en el trabajo de Maciá Medina [10] muestra que la solución (que es necesariamente única) se anula en la componente no acotada del complementario del $\text{sop } f$. Esto significa que $\text{sop } u \subset \widehat{\text{sop } f}$ (la envolvente de holomorfia respecto al plano $\widehat{\text{sop } f} = \widehat{\text{sop } f_{\mathbb{C}}}$). A la vista de esto, una cuestión natural es la siguiente: ¿qué condiciones adicionales hay que exigir a f para que $\text{sop } u \subset \text{sop } f$? Como veremos, la respuesta a esta pregunta, sugerida por la demostración del teorema 12 y que a estas alturas quizás no resulte sorprendente, implica el teorema de Runge (observación 21).

Teorema 20. Sean $f \in C_c^\infty(\mathbb{C})$ y $\Lambda \subset \mathbb{C}_\infty \setminus \text{sop } f$ un conjunto con al menos un punto en cada componente de $\mathbb{C}_\infty \setminus \text{sop } f$. Entonces, la ecuación $\partial u / \partial \bar{z} = f$ admite una solución con soporte en $\text{sop } f$ ($\text{sop } u \subset \text{sop } f$) si y solo si $\int_{\mathbb{C}} r(z)f(z) dA(z) = 0$ para toda función racional r con polos en Λ .

Demostración. La condición es necesaria: si u es una solución de $\partial u / \partial \bar{z} = f$ con soporte contenido en $\text{sop } f$ y r es cualquier función racional con polos en Λ , por el teorema de Green

$$\begin{aligned} \int_{\mathbb{C}} r(z)f(z) dA(z) &= \int_{\Omega} r(z) \frac{\partial u}{\partial \bar{z}}(z) dA(z) = \frac{1}{2i} \int_{\Omega} \frac{\partial}{\partial \bar{z}}(ru) d\bar{z} \wedge dz \\ &= \frac{1}{2i} \int_{\Omega} d(ru dz) = \frac{1}{2i} \int_{\partial\Omega} ru dz = 0, \end{aligned}$$

donde Ω es cualquier abierto acotado con frontera C^∞ que contenga $\text{sop } f$ pero no a los polos de r (basta, por ejemplo, considerar un disco lo suficientemente grande que contenga a $\text{sop } f$ al que se eliminan discos centrados en los polos de r suficientemente pequeños para que sean disjuntos entre sí y con $\text{sop } f$).

Recíprocamente, la solución que buscamos debe coincidir con la presentada en el teorema 1.27 del trabajo de Maciá Medina [10], es decir, la dada por (10). Por este teorema, u se anula en la componente no acotada de $\mathbb{C} \setminus \text{sop } f$. Si V es cualquier otra componente de $\mathbb{C} \setminus \text{sop } f$, entonces u es holomorfa en V y, si $a \in \Lambda \cap V$,

$$u(z) = \frac{1}{\pi} \int_{\mathbb{C}} \frac{f(w)}{z-w} dA(w) = -\frac{1}{\pi} \int_{\mathbb{C}} \left(\sum_{n \geq 0} \frac{(z-a)^n}{(w-a)^{n+1}} \right) f(w) dA(w) = \sum_{n \geq 0} \mu_n^a[f](z-a)^n,$$

donde

$$\mu_n^a[f] = -\frac{1}{\pi} \int_{\mathbb{C}} \frac{f(w)}{(w-a)^{n+1}} dA(w), \quad n = 0, 1, \dots$$

si $|z-a| < \text{dist}(a, \text{sop } f)/2$. Puesto que la función racional $w \mapsto 1/(w-a)^{n+1}$ tiene polo en $a \in \Lambda$, por hipótesis $\mu_n^a[f] = 0$ para todo $n = 0, 1, \dots$. El principio de identidad implica que $u \equiv 0$ en V ya que, como se ha dicho, $u \in \mathcal{O}(V)$ y V es conexo (nótese el paralelismo de esta demostración con la del teorema 12, donde u juega el papel de h). ■

Observación 21. Si $K \subset \mathbb{C}$ es compacto y μ es una medida compleja soportada en K , la solución del problema $\partial u / \partial \bar{z} = f$ con dato μ es precisamente la función h que aparece en la demostración del teorema 12. Como se puede comprobar fácilmente, el teorema 20 (con las modificaciones necesarias) sigue siendo válido si el segundo miembro en esta ecuación es una medida (en este caso, la solución u será localmente integrable [10, teorema 1.8]). Así, el teorema 20 implica el de Runge. ◀

Referencias

- [1] ANDERSSON, Mats. *Topics in complex analysis*. Universitext: Tracts in Mathematics 3080. Nueva York, US: Springer, 1997. <https://doi.org/10.1007/978-1-4612-4042-6>.
- [2] BRUNA, Joaquim y CUFÍ, Julià. *Complex analysis*. EMS Textbooks in Mathematics. Zúrich, CH: European Mathematical Society, 2013. <https://doi.org/10.4171/111>.
- [3] FUMERO PADRÓN, Melanie. *Aproximación Compleja*. Trabajo de Fin de Grado. Universidad de La Laguna, 2018. URL: <https://riull.ull.es/xmlui/handle/915/9634>.
- [4] GAMELIN, Theodore W. *Uniform algebras*. Prentice-Hall Series in Modern Analysis. Englewood Cliffs, US: Prentice-Hall, 1969. ISBN: 978-0-13-937805-8.
- [5] GAMELIN, Theodore W. *Complex analysis*. Undergraduate Texts in Mathematics. Nueva York, US: Springer, 2001. <https://doi.org/10.1007/978-0-387-21607-2>.
- [6] HARTOGS, F y ROSENTHAL, A. «Über Folgen analytischer Funktionen». En: *Mathematische Annalen* 104 (1931), págs. 606-610. ISSN: 0025-5831. <https://doi.org/10.1007/BF01457959>.
- [7] HILLE, Einar. *Analytic function theory*. Introductions to Higher Mathematics. Nueva York, US: Ginn y Company, 1962.

-
- [8] HÖRMANDER, Lars. *An introduction to complex analysis in several variables*. 3.^a ed. North-Holland Mathematical Library 7. Ámsterdam, NL: North-Holland, 1990. ISBN: 978-0-444-88446-6.
- [9] LEAR, Daniel. «A quantitative Runge's theorem in Riemann surfaces». En: *Reports@SCM* 1.1 (2014), págs. 15-32. ISSN: 2385-4227. <https://doi.org/10.2436/20.2002.02.2>.
- [10] MACIÁ MEDINA, Víctor J. *Análisis Complejo: la ecuación $\bar{\partial}$ y funciones armónicas en el plano*. Trabajo de Fin de Grado. Universidad de La Laguna, 2017. URL: <https://riull.ull.es/xmlui/handle/915/4265>.
- [11] ROTH, Alice. «Approximationseigenschaften und Strahlengrenzwerte meromorpher und ganzer Funktionen». En: *Commentarii Mathematici Helvetici* 11 (1938), págs. 77-125. ISSN: 0010-2571. <https://doi.org/10.1007/BF01199693>.
- [12] RUDIN, Walter. *Function theory in the unit ball of \mathbb{C}^n* . Classics in Mathematics. Berlín, DE: Springer, 1980. <https://doi.org/10.1007/978-3-540-68276-9>.
- [13] RUDIN, Walter. *Real and complex analysis*. 3.^a ed. Nueva York, US: McGraw-Hill Book Co., 1987. ISBN: 978-0-07-054234-1.

TEMat

La conjetura de Collatz

✉ Alejandro Gil Asensi^a
Universidad de La Rioja
algila@unirioja.com

Resumen: La conjetura de Collatz, también llamada el problema $3x+1$, el problema de Siracusa, el problema de Kakutani, el algoritmo de Hasse o el problema de Ulam, es uno de los problemas matemáticos sin resolver cuyo enunciado es muy fácil de comprender, pero resulta harto complicado abordar una demostración. Concretamente, esta conjetura afirma que si $C(n)$ es la función de Collatz definida por $C(n) = n/2$ si n es par y por $C(n) = 3n + 1$ si n es impar, entonces tras un número finito de iteraciones de $C(n)$ se llega al valor 1 independientemente del valor entero positivo de partida.

Abstract: The Collatz conjecture, also known as the $3x + 1$ problem, the Syracuse problem, Kakutani's problem, Hasse's algorithm or Ulam's problem, is one of these unsolved mathematical problems which is easily stated but extremely difficult to prove. Specifically, the conjecture asserts that if $C(n)$ is the Collatz function defined as $C(n) = n/2$ if n is even, and $C(n) = 3n + 1$ if n is odd, then $C(n)$ reaches 1 in a finite number of iterations, independently of the starting positive integer.

Palabras clave: Collatz, conjetura de Collatz, conjetura $3x + 1$, problema de Ulam.

MSC2020: 11-37.

Recibido: 20 de marzo de 2021.

Aceptado: 13 de febrero de 2022.

Agradecimientos: Quiero agradecer a mis tutores del Trabajo de Fin de Grado, por animarme a enviar este artículo. Agradezco a la ANEM y a todos los que colaboran en esta revista para incentivar el talento matemático y divulgar las matemáticas.

Referencia: GIL ASENSI, Alejandro. «La conjetura de Collatz». En: *TEMat*, 6 (2022), págs. 65-81. ISSN: 2530-9633.
URL: <https://temat.es/articulo/2022-p65>.

^aEl autor estaba afiliado a la Universidad de Alicante cuando se desarrolló la mayor parte del trabajo.

1. La conjetura de Collatz

La conjetura de Collatz, también conocida como problema de Kakutani, conjetura de Ulam, problema de Siracusa o, más genéricamente, como problema $3x + 1$ (o $3n + 1$), es un típico problema matemático fácil de enunciar cuya dificultad reside en ser capaz de demostrar sus implicaciones. El problema consiste en estudiar el comportamiento de las iteraciones de la función de Collatz $C(x)$ (con $x \in \mathbb{N} = \{1, 2, 3, \dots\}$), definida como

$$C(x) = \begin{cases} \frac{x}{2} & \text{si } x \equiv 0 \pmod{2}, \\ 3x + 1 & \text{si } x \equiv 1 \pmod{2}. \end{cases}$$

La conjetura está relacionada con las iteraciones de esta función, aunque en la literatura aparece con más frecuencia la función

$$T(x) = \begin{cases} \frac{x}{2} & \text{si } x \equiv 0 \pmod{2}, \\ \frac{3x + 1}{2} & \text{si } x \equiv 1 \pmod{2}. \end{cases}$$

De tal forma, se ahorra un paso si x es impar. Si se itera en las funciones $C(x)$ y $T(x)$ utilizando, por ejemplo, 6 como semilla, se obtienen las secuencias (6, 3, 10, 5, 16, 8, 4, 2, 1, ...) y (6, 3, 5, 8, 4, 2, 1, ...), respectivamente. Se observa que estas secuencias acaban llegando al bucle (1, 4, 2) en el caso de $C(x)$ y (1, 2) en el caso de $T(x)$.

Se va a utilizar la siguiente terminología en relación con las secuencias generadas por estas funciones.

Definición 1 (trayectoria). Se llama trayectoria de $x \in \mathbb{N}$ a $\Omega(x) = (x, T(x), T^2(x), \dots)$. ◀

Definición 2 (ciclo). Se dice que $\Omega(x)$ es un ciclo de longitud k si $T^k(y) = y$ para todo $y \in \Omega(x)$. Al ciclo (1, 2) se le llama ciclo trivial. ◀

Se puede observar que para cada x solo pueden suceder tres cosas:

1. Existe $k \in \mathbb{N}$ tal que $T^k(x) = 1$. Es decir, *su trayectoria es convergente*.
2. $\Omega(x)$ tiene un *ciclo no trivial* o una *órbita*.
3. $\lim_{k \rightarrow \infty} T^k(x) = +\infty$, o lo que es lo mismo, *su trayectoria es divergente*.

Los conceptos de trayectoria y ciclo se pueden definir tanto para $C(x)$ como para $T(x)$. En el caso de utilizar la función $C(x)$, se llamará ciclo trivial al ciclo (1, 4, 2).

Con esto, podemos formular la conjetura de Collatz, la cual afirma lo siguiente.

Conjetura 3. *Para todo $x \in \mathbb{N}$, existe un entero positivo k tal que $T^k(x) = 1$ y, por ende, toda trayectoria es convergente.*

Normalmente la conjetura de Collatz se formula para todo entero positivo, pero nada impide extender la función $T(x)$ a enteros. Como $T(0) = 0$ y $T(-1) = -1$, aparecen dos nuevos ciclos unipuntuales, pero también aparecen los siguientes: (-5, -7, -10) y (-17, -25, -37, -55, -82, -41, -61, -91, -136, -68, -34). En tal caso, se conjetura que cualquier trayectoria termina por entrar en el ciclo trivial (1, 2) o en alguno de los ciclos anteriores.

Este problema se le atribuye a Lothar Collatz (1910-1990), matemático alemán que realizó importantes aportaciones en análisis funcional, teoría de la aproximación, optimización y ecuaciones diferenciales, entre otros campos. Planteó el problema en la década de 1930, mientras estudiaba el comportamiento de funciones bajo iteraciones. Otros importantes matemáticos como Helmut Hasse (1898-1979), Shizuo Kakutani (1911-2004) o Stanislaw Marcin Ulam (1909-1984) también plantearon, de manera independiente, dicha conjetura y, por ello, figuran como autores en la literatura.

Que el problema lleve casi cien años sin resolverse implica preguntarse por qué la conjetura es tan difícil. Por un lado, cabe destacar que la conjetura no apareció en la literatura hasta los años 70. Por otro lado, el comportamiento bajo iteraciones de este tipo de funciones es muy impredecible. Ejemplo de ello puede verse con los iterados $C^k(27)$ (véase la figura 1).

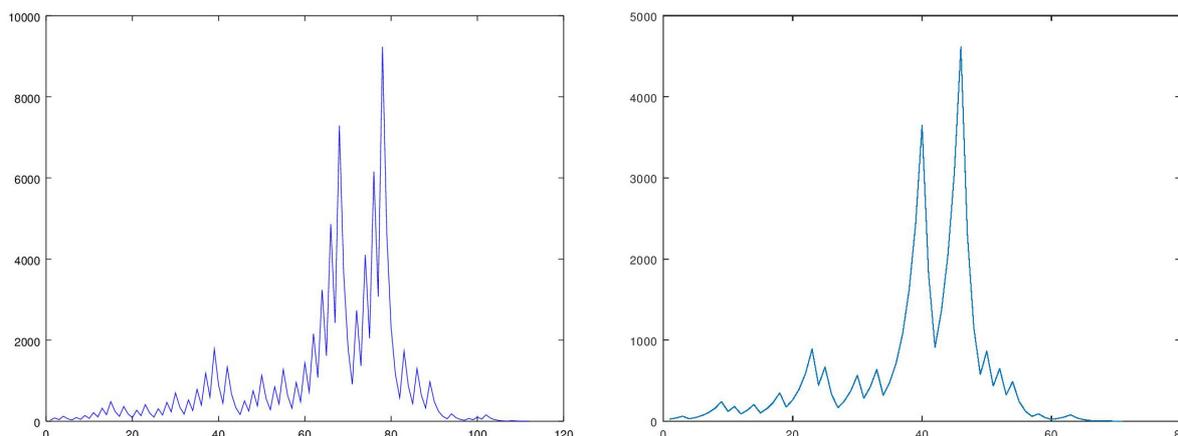


Figura 1: Trayectoria del número 27. A la izquierda, mediante la función $C(x)$ y a la derecha por $T(x)$. Como puede verse, ambas trayectorias tienen una forma parecida a pesar de que la izquierda necesita más pasos para llegar al 1 y, por eso, tiene una forma más punteada. En total, se necesitan 110 pasos con $C(x)$ (i. e., $C^{110}(27) = 1$) y 70 con $T(x)$ (i. e., $T^{70}(27) = 1$). El valor máximo de la trayectoria también es diferente, $C^{77}(27) = 9232$ a la izquierda y $T^{45}(27) = 4616$ a la derecha (justamente la mitad).

Al tratarse de un problema abierto, existen en la literatura diferentes formas de abordarlo. El objetivo a lo largo de este artículo es presentar algunas de las que se han considerado más interesantes y que más se acercan a tener una demostración de la conjetura. En primer lugar, se va a extender el problema a un tipo de funciones más general en la sección 2. En la sección 3 se procederá a presentar conceptos de teoría ergódica y su función junto a cadenas de Markov para el estudio del comportamiento de los iterados módulo m . La sección 4 introduce el concepto de tiempo de parada con la intención de presentar el teorema de Terras, el cual evidencia que la conjetura es cierta para «casi todos» los números (en un sentido de densidad natural que se presentará en la misma sección). Además, se relacionará este resultado con uno mucho más fuerte al que recientemente llegó Terence Tao¹. Finalmente, en la última sección se realizarán algunas reflexiones sobre la dificultad de la conjetura y el estado actual de la misma. También se mencionarán algunos otros avances en otras direcciones diferentes que no se han podido abordar en este artículo. Junto a esto, los lectores podrán consultar la bibliografía para acceder a más información sobre el tema.

2. Generalizaciones del problema $3x + 1$

Antes de que aparecieran, en los años 70, los primeros trabajos relacionados con el denominado problema $3x + 1$, en los años 60 ya se encontraban problemas semejantes. Por ejemplo, el problema al que llegó Murray Klamkin en 1963. Este estaba relacionado con la función $U(n)$ definida como

$$(1) \quad U(n) = \begin{cases} \frac{2n}{3} & \text{si } n \equiv 0 \pmod{3}, \\ \frac{4n-1}{3} & \text{si } n \equiv 1 \pmod{3}, \\ \frac{4n+1}{3} & \text{si } n \equiv 2 \pmod{3}, \end{cases}$$

donde $n \in \mathbb{N}$. El problema consistía en estudiar si los iterados a partir del valor $n = 8$ forman un conjunto infinito o no. Es decir, si la trayectoria de $n = 8$ por la función U es divergente. Este problema continúa sin ser resuelto hasta la fecha, aunque se conjetura que es cierto [10].

¹Terence Tao fue merecedor de la prestigiosa Medalla Fields en 2006, la cual recibió en la vigésima quinta edición del *Congreso Internacional de Matemáticas*, en Madrid [23].

Este hecho sirve para sospechar que la dificultad de la conjetura está muy ligada a la dificultad del estudio de las iteraciones de funciones de esta índole. En esta sección se va a dar una generalización de este tipo de funciones y a estudiar si todas ellas tienen un comportamiento parecido.

2.1. Funciones de Collatz generalizadas

Para generalizar el problema $3x + 1$ se va a recurrir a las llamadas funciones de Collatz generalizadas (véanse las definiciones 4 y 5).

Definición 4 (función admisible). Se dice que una función T es admisible si envía enteros a enteros, es decir, si es de la forma $T: \mathbb{Z} \rightarrow \mathbb{Z}$. ◀

Definición 5 (función de Collatz generalizada). Sea una función admisible T . Si existen un número natural $d \geq 2$ y enteros a_i, b_i , con $i \in \{0, 1, \dots, d-1\}$, tales que

$$(2) \quad T(x) = \frac{a_i x + b_i}{d} \quad \text{si } x \equiv i \pmod{d},$$

entonces se dice que T es una función de Collatz generalizada. ◀

A lo largo de este artículo se mencionarán con frecuencia enteros a_i y b_i con $i \in \{0, 1, \dots, d-1\}$ que siempre se referirán a los que aparecen en la definición anterior.

Proposición 6 (caracterización de admisibilidad). Sea T una función de la forma presentada en (2). Entonces, T es admisible si y solo si $ia_i + b_i \equiv 0 \pmod{d}$ para todo $0 \leq i \leq d-1$.

Demostración. Sea $x \in \mathbb{Z}$ con $x \equiv i \pmod{d}$, de modo que $x - i = kd$ para cierto $k \in \mathbb{Z}$. Entonces,

$$\begin{aligned} \frac{a_i x + b_i}{d} = t \in \mathbb{Z} &\iff a_i(i - i + x) + b_i = dt \\ &\iff ia_i + b_i = dt - (x - i)a_i = dt - dka_i = d(t - ka_i) \\ &\iff ia_i + b_i \equiv 0 \pmod{d}. \quad \blacksquare \end{aligned}$$

Ejemplos de funciones de Collatz generalizadas son la función $T(x)$ del problema $3x + 1$, cuando se toman coeficientes $d = 2$, $a_0 = 1$, $b_0 = 0$, $a_1 = 3$ y $b_1 = 1$, o la función de Klamkin $U(x)$ (definida en (1)) tomando $d = 3$, $a_0 = 2$, $a_1 = a_2 = 4$, $b_0 = 0$, $b_1 = -1$ y $b_2 = 1$. La admisibilidad de $T(x)$ es fácil de ver mediante la caracterización de admisibilidad, ya que $0a_0 + b_0 = 0 \equiv 0 \pmod{2}$ y $1a_1 + b_1 = 4 \equiv 0 \pmod{2}$. Análogamente se puede comprobar la admisibilidad de $U(x)$.

2.2. Funciones de tipo relativamente primo

Dentro de las funciones generalizadas de Collatz están las llamadas de *tipo relativamente primo*, que han sido objeto especial de estudio.

Definición 7 (función de tipo relativamente primo). Una función generalizada de Collatz se dice que es de tipo relativamente primo si se da que

$$(3) \quad \text{mcd}(a_0 a_1 \cdots a_{d-1}, d) = 1. \quad \blacktriangleleft$$

La función $T(x)$ del problema $3x + 1$ tiene $a_0 = 1$, $b_0 = 0$, $a_1 = 3$ y $b_1 = 1$, luego, como $\text{mcd}(1 \cdot 3, 2) = 1$, pertenece a la familia de funciones de tipo relativamente primo. Sin embargo, para la función de Collatz, $C(x)$, $a_0 = 1$, $b_0 = 0$, $a_1 = 6$ y $b_1 = 2$, luego $\text{mcd}(1 \cdot 6, 2) = 2 \neq 1$ y, por tanto, no pertenece a este tipo.

Por otro lado, en cuanto a la función del problema análogo $5x + 1$ definida como

$$T_5(x) = \begin{cases} \frac{x}{2} & \text{si } x \equiv 0 \pmod{2}, \\ \frac{5x + 1}{2} & \text{si } x \equiv 1 \pmod{2}, \end{cases}$$

también es una función de tipo relativamente primo. Sin embargo, parece que el comportamiento de esta función es diferente, ya que se conjetura que existen trayectorias divergentes [10]. Por tanto, todo apunta a que algunas funciones generalizadas de Collatz van a tener trayectorias divergentes y otras no. La siguiente conjetura fue realizada por Matthews y Watts [17, conjeturas (i), (ii) y (iv)]. Indica cuándo va a darse cada casuística.

Conjetura 8 (conjetura sobre los ciclos y las trayectorias divergentes). *En lo referente a las funciones de Collatz generalizadas de tipo relativamente primo, se ha conjeturado lo siguiente.*

- (i) Si $|a_0 a_1 \cdots a_{d-1}| < d^d$, entonces en algún momento todas las trayectorias llegan a un ciclo (y, por tanto, siempre existirá al menos un ciclo).
- (ii) Si $|a_0 a_1 \cdots a_{d-1}| > d^d$, entonces existen trayectorias divergentes.
- (iii) Si la trayectoria $(T^k(n))_{k \geq 0}$, $n \in \mathbb{Z}$, no entra nunca en un ciclo, entonces las iteraciones se distribuyen uniformemente (mód d^α) para todo $\alpha \geq 1$, es decir,

$$\lim_{N \rightarrow \infty} \frac{1}{N+1} \text{Card}\{k \leq N \mid T^k(n) \equiv j \pmod{d^\alpha}\} = \frac{1}{d^\alpha}, \quad \text{para todo } 0 \leq j \leq d^\alpha - 1,$$

donde $\text{Card}\{\cdot\}$ denota el cardinal del conjunto.

El caso $|a_0 a_1 \cdots a_{d-1}| = d^d$ no hace falta cubrirlo puesto que, por la condición (3), esto nunca puede darse. Esta conjetura va acorde a lo mencionado antes (en el problema $3x + 1$). Se obtiene que $|a_0 a_1| = 3 < 2^2$, luego la conjetura nos dice que todas las trayectorias serán cíclicas en algún momento. En cambio, para la conjetura $5x + 1$ se tiene que $|a_0 a_1| = 5 > 2^2$ y, por tanto, deben existir trayectorias divergentes. También cabe resaltar que esta conjetura no depende de los valores que tengan los b_i . Esta conjetura también parece ser un problema intratable [17].

3. Estudio de los iterados

El estudio de los iterados es interesante para una mayor comprensión del comportamiento de $T(x)$. En esta sección se estudiará no solo cómo poder dar una fórmula para el K -ésimo iterado, sino también cómo estos iterados se distribuyen módulo m . Para ello, resulta muy útil plantear modelos de Markov [16].

Para un x cualquiera, se puede dar una fórmula del K -ésimo iterado, pero es más laboriosa. Consideremos el caso más genérico: el de funciones de Collatz generalizadas definidas en (2). Siguiendo la misma notación anterior, definimos $a_K(x) = a_i$ y $b_K(x) = b_i$ si $T^K(x) \equiv i \pmod{d}$ para $0 \leq i < d$, de tal forma que

$$T^{K+1}(x) = \frac{a_K(x)T^K(x) - b_K(x)}{d}.$$

Teorema 9 (expresión del K -ésimo iterado). *Para una función generalizada de Collatz, $T(x)$, el K -ésimo iterado se puede expresar como*

$$(4) \quad T^K(x) = \frac{a_0(x) \cdots a_{K-1}(x)}{d^K} \left(x - \sum_{i=0}^{K-1} \frac{b_i(x)d^i}{a_0(x) \cdots a_i(x)} \right).$$

Si $T^i(x) \neq 0$ para todo $i \geq 0$, también se verifica que

$$(5) \quad T^K(x) = \frac{a_0(x) \cdots a_{K-1}(x)}{d^K} x \prod_{i=0}^{K-1} \left(1 - \frac{b_i(x)}{a_i(x)T^i(x)} \right).$$

Demostración. Se va a demostrar (4) por inducción. En primer lugar, para el caso $K = 1$, se tiene que

$$T(x) = \frac{a_0(x)}{d} \left(x - \frac{b_0(x)}{a_0(x)} \right) = \frac{a_0(x)x - b_0(x)}{d}.$$

Como $a_0(x) = a_i$ y $b_0(x) = b_i$ si $x \equiv i \pmod{d}$, se llega a que

$$T(x) = \frac{a_i x - b_i}{d} \quad \text{si } x \equiv i \pmod{d}.$$

Ahora, supongamos que se da la igualdad (4) para todo $K \leq N$. Veamos que se cumple para $K = N + 1$. Tenemos que

$$T^{N+1}(x) = T(T^N(x)) = \frac{a_0(T^N(x))}{d} \left(T^N(x) - \frac{b_0(T^N(x))}{a_0(T^N(x))} \right).$$

Se puede ver fácilmente de la definición de $a_i(x)$ y $b_i(x)$ que $a_0(T^k(x)) = a_k(x)$ y $b_0(T^k(x)) = b_k(x)$. Por tanto,

$$\begin{aligned} T^{N+1}(x) &= \frac{a_k(x)}{d} \left(\frac{a_0(x) \cdots a_{k-1}(x)}{d^k} \left(x - \sum_{i=0}^{k-1} \frac{b_i(x)d^i}{a_0(x) \cdots a_i(x)} \right) - \frac{b_k(x)}{a_k(x)} \right) \\ &= \frac{a_0(x) \cdots a_{k-1}(x)a_k(x)}{d^{k+1}} \left(x - \sum_{i=0}^{k-1} \frac{b_i(x)d^i}{a_0(x) \cdots a_i(x)} \right) - \frac{b_k(x)}{d} \\ &= \frac{a_0(x) \cdots a_{k-1}(x)a_k(x)}{d^{k+1}} \left(x - \sum_{i=0}^k \frac{b_i(x)d^i}{a_0(x) \cdots a_i(x)} \right). \end{aligned}$$

Con esto, queda demostrado (4). La demostración de (5) se omite por ser similar. ■

Con el objetivo de conocer mejor las iteraciones, se va a estudiar cómo se comportan los iterados módulo m . La idea es conocer cuánto tiempo permanece la función $T(x)$ en la clase de elementos congruentes con i (mód m). Cuando $m = d$, puede estimarse que

$$(6) \quad T^K(x) \sim \frac{1}{d} \left(\prod_{i=0}^{d-1} a_i^{f_i} \right) x,$$

donde $f_i \in [0, 1]$ es la frecuencia con la que los iterados permanecen en la clase de los elementos congruentes con i (mód d) (a la que en la definición 10 denotaremos como $B(i, d)$). En el teorema 14 se dará la expresión de estas frecuencias límites cuando $T(x)$ sea una función de Collatz de tipo relativamente primo. Por (6), cabe esperar que los iterados $T^K(x)$ crezcan o decrezcan geométricamente en base a estas frecuencias.

A pesar de que los iterados $T^K(x)$ quedan determinados por x , se han empleado en la bibliografía muchas ideas propias de la estadística y la probabilidad. Por ejemplo, se puede considerar la sucesión de los iterados $x, T(x), T^2(x), \dots, T^K(x), \dots$ como un proceso estocástico [16], esto es, una sucesión de variables aleatorias. En nuestro caso, se observa que la clase de congruencia del iterado $T^{K+1}(x)$ solo está determinada por la clase en que está la iteración anterior, $T^K(x)$. Esto corresponde a procesos muy concretos conocidos como cadenas de Markov, los cuales se introducirán más adelante. Por tanto, con el objetivo de estudiar las clases de congruencias módulo m de los iterados para órbitas largas, se construirá, a partir de las cadenas de Markov, lo que se conoce como matriz de transición. Esta matriz aporta mucha información del comportamiento de los iterados cuando $K \rightarrow \infty$.

Definición 10. Definimos la clase j (mód m), donde $j \in \{0, 1, 2, \dots, m-1\}$, como

$$B(j, m) := \{x \in \mathbb{Z} \mid x \equiv j \pmod{m}\}. \quad \blacktriangleleft$$

Definición 11 (conjunto ergódico). Diremos que $S \subseteq \mathbb{Z}$ es un conjunto T -invariante módulo m si $T(S) \subseteq S$ y se cumple que $S = B(i_1, m) \cup B(i_2, m) \cup \dots \cup B(i_t, m)$, es decir, que es una unión de t clases de congruencias módulo m . Diremos que S es un conjunto ergódico módulo m si $S \neq \emptyset$ y S es un conjunto T -invariante módulo m minimal. La condición de minimalidad implica que, si existe otro conjunto ergódico módulo m , R , tal que $R \subseteq S$, entonces $R = S$. ◀

El motivo por el que tiene interés estudiar los conjuntos ergódicos módulo m es que la aplicación inversa de una función de Collatz generalizada T , aplicada sobre una clase de congruencias módulo m , resulta ser una unión de clases de congruencias módulo md [3]. Por ejemplo, si T es la función de Collatz del problema $3x + 1$, entonces su aplicación inversa, T^{-1} , satisface que

$$T^{-1}(B(j, m)) = \begin{cases} B(2j, 2m) \cup B\left(\frac{2j-1}{3}\right) & \text{si } m \not\equiv 0 \pmod{3}, \\ B(2j, 2m) & \text{si } m \equiv 0 \pmod{3} \text{ y } j \not\equiv 2 \pmod{3}, \\ B(2j, 2m) \cup B\left(\frac{2j-1}{3}, \frac{2m}{3}\right) & \text{si } m \equiv 0 \pmod{3} \text{ y } j \equiv 2 \pmod{3}. \end{cases}$$

De esta forma, solo es necesario estudiar separadamente los conjuntos tales que $T(S) \subseteq S$ (es decir, los conjuntos ergódicos) para obtener una idea del comportamiento de la función T .

Observación 12. Cada conjunto ergódico módulo m , $S_i^{(m)}$, con $i \geq 1$, tiene intersección vacía con cualquier otro conjunto ergódico. Para verlo, tomemos dos conjuntos ergódicos $S_1^{(m)}$ y $S_2^{(m)}$ diferentes. Como ambos son uniones de clases de congruencias módulo m , la intersección $S_1^{(m)} \cap S_2^{(m)}$ también lo es. Además

$$\begin{aligned} T(S_1^{(m)} \cap S_2^{(m)}) &\subseteq S_1^{(m)}, \\ T(S_1^{(m)} \cap S_2^{(m)}) &\subseteq S_2^{(m)}. \end{aligned}$$

Luego $T(S_1^{(m)} \cap S_2^{(m)}) \subseteq S_1^{(m)} \cap S_2^{(m)}$ y, por tanto, es un conjunto ergódico. Sin embargo, por la condición de minimalidad de $S_1^{(m)}$ y de $S_2^{(m)}$, como $S_1^{(m)} \cap S_2^{(m)} \subseteq S_i^{(m)}$ ($i = 1, 2$), se tiene que

$$S_1^{(m)} = S_1^{(m)} \cap S_2^{(m)} = S_2^{(m)}.$$

Luego los conjuntos ergódicos módulo m son disjuntos. Además, si suponemos que tenemos r de ellos en total, entonces se cumple que

$$\mathbb{Z} = S_0^{(m)} \cup S_1^{(m)} \cup \dots \cup S_r^{(m)},$$

donde $S_1^{(m)}, S_2^{(m)}, \dots, S_r^{(m)}$ son los diferentes conjuntos ergódicos módulo m y $S_0^{(m)}$ es la unión de las clases de congruencias restantes, a veces llamada clases de transición pues representa a aquellas clases que en alguna iteración pueden abandonarla. ◀

Ejemplo 13. Consideremos la función de Collatz $T(x)$ del problema $3x + 1$. Para $m = 2$ solo tenemos las clases de congruencias $B(0, 2)$ y $B(1, 2)$. Para números de la forma $2n$ se tiene que $T(2n) = n$, luego $T(B(0, 2)) = \mathbb{Z}$, así que $T(B(0, 2)) \not\subseteq B(0, 2)$. Por tanto, $B(0, 2)$ no es ergódico por no ser T -invariante. También se puede comprobar que $B(1, 2)$ no es ergódico puesto que $T(B(1, 2)) \not\subseteq B(1, 2)$. Como ni $B(0, 2)$ ni $B(1, 2)$ son ergódicos, el único conjunto que puede serlo es $\mathbb{Z} = B(0, 2) \cup B(1, 2)$.

Consideremos ahora el caso $m = 3$. Las clases de congruencias son $B(0, 3), B(1, 3)$ y $B(2, 3)$. Por tanto, los posibles conjuntos ergódicos serán alguna de estas clases de congruencias, o unión de algunas de ellas. Se puede comprobar que $T(B(i, 3)) \not\subseteq B(i, 3)$ para todo $i \in \{0, 1, 2\}$. Sin embargo, se observa que la unión $S_1^{(3)} = B(1, 3) \cup B(2, 3)$ está formada por los números de la forma $3k + 1$ y $3k + 2$. Estos números pueden tener imágenes $(3k + 1)/2, (3k + 2)/2$ si son pares o $(9k + 4)/2, (9k + 7)/2$ si son impares, que en cualquier caso no son un múltiplo de 3. Luego $T(S_1^{(3)}) \subseteq S_1^{(3)}$ y, por tanto, es un conjunto ergódico (mód 3). Como la intersección de conjuntos ergódicos diferentes tiene que ser vacía, el otro conjunto ergódico posible es $B(0, 3)$, pero no lo es por no ser T -invariante, luego no existe otro conjunto ergódico. Finalmente,

$$\mathbb{Z} = S_0^{(3)} \cup S_1^{(3)},$$

donde $S_0^{(3)} = B(0, 3)$ es el conjunto de clases de transición.

En general, si $3 \nmid m$, el único conjunto ergódico es \mathbb{Z} y, si $3 \mid m$, $\mathbb{Z} - 3\mathbb{Z}$ [3, ejemplo 1.1]. ◀

El siguiente resultado fue obtenido por Matthews y Watts [18] para estimar las frecuencias límites de las clases de congruencias $B(j, m)$.

Teorema 14. Sea $T(x)$ una función de Collatz generalizada de tipo relativamente primo. Sea

$$S = B(i_1, m) \cup B(i_2, m) \cup \dots \cup B(i_t, m)$$

un conjunto ergódico módulo m y llamemos $S' = (B(i_1, m), B(i_2, m), \dots, B(i_t, m))$. Entonces, la frecuencia límite de la componente $B(j, m)$ (con $j \in \{i_1, \dots, i_t\}$) viene dada por

$$(7) \quad \mu_{S'}(B(j, m)) := \lim_{N \rightarrow \infty} \frac{1}{N} \text{Card} \{K \leq N \mid T^k(n) \in B(j, m)\}.$$

Se sabe que esto es cierto en el caso de funciones de tipo relativamente primo, pero permanece sin probarse en el resto de casos, aunque hay evidencia computacional que afirma que esto se sigue dando. Un algoritmo para calcular las frecuencias límite puede encontrarse en el trabajo de Leigh [14].

Para estudiar la distribución de los iterados módulo m es útil utilizar procesos de Markov (o cadenas de Markov). Se trata de procesos estocásticos discretos con m posibles estados, E_1, \dots, E_m , en los que la probabilidad de que se dé un estado depende únicamente del estado anterior. En nuestro caso, cada una de las observaciones $T^K(x)$, con $K = 0, 1, 2, \dots$, estará en el estado i si $T^K(x)$ pertenece a $B(i, m)$. La intención es estudiar cuál es la probabilidad de pasar de un estado a otro en cada una de las iteraciones (a la probabilidad de pasar del estado j al estado i la denotaremos por q_{ij}) y así construir una matriz de transición (o matriz de Markov) que nos permitirá estudiar el comportamiento de los iterados en el límite.

Definición 15 (matriz de Markov). Sea $p_{ij}(m)$ el número de clases de congruencias módulo md en el conjunto $T^{-1}(B(i, m)) \cap B(j, m)$ y sea $q_{ij}(m) = p_{ij}(m)/d$. La matriz $m \times m$

$$Q_T(m) = [q_{ij}(m)] = \begin{pmatrix} q_{00}(m) & q_{01}(m) & \cdots & q_{0(m-1)}(m) \\ q_{10}(m) & q_{11}(m) & \cdots & q_{1(m-1)}(m) \\ \vdots & \vdots & \ddots & \vdots \\ q_{(m-1)0}(m) & q_{(m-1)1}(m) & \cdots & q_{(m-1)(m-1)}(m) \end{pmatrix}$$

se llama matriz de Markov. Los q_{ij} representan la probabilidad de transición del estado j al estado i . ◀

Teorema 16. La matriz de Markov $Q_T(m)$ es una matriz estocástica. Esto es, $q_{ij} \geq 0$ para todo $0 \leq i, j \leq m-1$ y los valores de cada columna suman 1 (algunos autores trabajan con filas en vez de columnas).

Demostración. Claramente, $q_{ij} \geq 0$ por definición. Además,

$$B(j, m) = \mathbb{Z} \cap B(j, m) = T^{-1}(\mathbb{Z}) \cap B(j, m) = \bigcup_{i=0}^{m-1} [T^{-1}(B(i, m)) \cap B(j, m)].$$

Por lo tanto, $B(j, m)$ es unión disjunta de $\sum_{i=0}^{m-1} p_{ij}(m)$ clases de congruencias módulo md . Por otro lado, como

$$B(j, m) = \bigcup_{k=0}^{d-1} [B(j + km, md)],$$

se tiene que $B(j, m)$ es unión de d clases de congruencias (mód md). Luego se sigue que $\sum_{i=0}^{m-1} p_{ij}(m) = d$ y, finalmente, $\sum_{i=0}^{m-1} q_{ij}(m) = 1$. ■

Observación 17. Si $d \mid m$, entonces

$$q_{ij}(m) = \begin{cases} \frac{1}{d} & \text{si } T(i) \equiv j \pmod{\frac{m}{d}}, \\ 0 & \text{en otro caso.} \end{cases}$$

En caso contrario, el cálculo de los $q_{ij}(m)$ resulta complicado de hallar. ◀

Por ejemplo, para $T(x)$ del problema $3x + 1$ y $m = 3$ se tiene que los estados posibles son $B(0, 3)$, $B(1, 3)$ y $B(2, 3)$. Consideremos $\{0, 3, 6, \dots\}$ los elementos de $B(0, 3)$. Entonces,

$$T(\{0, 3, 6, \dots\}) = \{0, 3, 6, \dots\} \cup \{5, 14, 23, \dots\} = B(0, 3) \cup B(5, 9)$$

y, como $B(5, 9) \subset B(2, 3)$, intuitivamente podemos afirmar que el estado $B(0, 3)$ pasa al estado $B(0, 3)$ con probabilidad $1/2$ o al estado $B(2, 3)$ con la misma probabilidad. Esto puede verificarse computacionalmente. Para $B(1, 3)$,

$$T(\{1, 4, 7, \dots\}) = \{2, 5, 8, \dots\} \cup \{2, 11, 20, \dots\} \subset B(2, 3),$$

luego del estado $B(1, 3)$ siempre se pasa al estado $B(2, 3)$. Por último,

$$T(\{2, 5, 8, \dots\}) = \{1, 4, 7, \dots\} \cup \{8, 17, \dots\},$$

luego el estado $B(2, 3)$ pasa al estado $B(1, 3)$ o $B(2, 3)$ con misma probabilidad. Por tanto,

$$Q_T(3) = \begin{bmatrix} 1/2 & 0 & 0 \\ 0 & 0 & 1/2 \\ 1/2 & 1 & 1/2 \end{bmatrix}.$$

Hasta ahora se ha trabajado con matrices de transición de un paso (una iteración). El objetivo es conocer la matriz de transición en el límite, es decir, calcular $(Q_T(m))^K$ y hacer $K \rightarrow \infty$.

Teorema 18. Sea p_{Kij} el número de clases de congruencias módulo md^K en $T^{-K}(B(i, m)) \cap B(j, m)$. Entonces,

$$[p_{ij}]^K = [p_{Kij}].$$

Además, se satisface que

$$(Q_T(m))^K = \left[\frac{p_{Kij}}{md^K} \right].$$

Continuando con el ejemplo anterior, para el problema $3x + 1$ y $m = 3$ se tiene que

$$(Q_T(3))^k = \begin{bmatrix} \frac{1}{2^k} & 0 & 0 \\ \frac{1}{3} \left(1 + \frac{(-1)^k}{2^{k+1}} \right) - \frac{1}{2^{k+1}} & \frac{1}{3} \left(1 + \frac{(-1)^k}{2^{k-1}} \right) & \frac{1}{3} \left(1 - \frac{(-1)^k}{2^k} \right) \\ \frac{1}{3} \left(2 - \frac{(-1)^k}{2^{k+1}} \right) - \frac{1}{2^{k+1}} & \frac{1}{3} \left(2 - \frac{(-1)^k}{2^{k-1}} \right) & \frac{1}{3} \left(2 + \frac{(-1)^k}{2^k} \right) \end{bmatrix},$$

luego cuando $k \rightarrow \infty$

$$(Q_T(3))^k \rightarrow \begin{bmatrix} 0 & 0 & 0 \\ 1/3 & 1/3 & 1/3 \\ 2/3 & 2/3 & 2/3 \end{bmatrix}.$$

Luego el único conjunto ergódico (mód 3) es $S = \mathbb{Z} - 3\mathbb{Z}$, como ya fue mencionado anteriormente. Esto implica el curioso hecho de que, una vez se llega a un número que no es múltiplo de 3, la trayectoria nunca vuelve a pasar por un número múltiplo de 3 [16].

4. Tiempo de parada

La conjetura $3x + 1$ se ha formulado en la literatura de diversas formas como, por ejemplo, la que presentamos como conjetura 3. En esta sección, se va a reformular la conjetura en términos de un concepto muy importante denominado tiempo de parada. Este concepto, junto a otros que se introducen en esta sección, se pueden definir para una función generalizada de Collatz cualquiera. Sin embargo, durante esta sección denotaremos por $T(x)$ a la función de Collatz del problema $3x + 1$.

Definición 19 (tiempo de parada). Se conoce a $\sigma(n)$ como el tiempo de parada de $n \in \mathbb{N}$ y se define como

$$\sigma(n) := \inf\{k \in \mathbb{N} \mid T^k(n) < n\}.$$

Por convenio, se define $\sigma(n) = +\infty$ si ningún entero positivo k verifica que $T^k(n) < n$. ◀

Por ejemplo, para $n = 7$, se tiene la trayectoria $\Omega(7) = (7, 11, 17, 26, 13, 20, 10, 5, 16, 8, 4, 2, 1)$. Observamos que $k = 7$ es el menor entero positivo que verifica que $T^7(7) = 5 < 7$. Por tanto, $\sigma(7) = 7$.

Definición 20. Llamaremos

$$S(k) := \{n \mid \sigma(n) < k\}$$

al conjunto de naturales con tiempo de parada inferior a k . ◀

Con esto, se puede redefinir la conjetura en términos de tiempos de parada de la siguiente forma [8].

Conjetura 21. *Cualquier número natural tiene tiempo de parada finito.*

La equivalencia de esta conjetura con la conjetura 3 es evidente. Si todo número $n \in \mathbb{N}$ satisface que $T^i(n) = 1$ para cierto i , necesariamente su tiempo de parada debe ser finito. El recíproco puede comprobarse por inducción. Supongamos que la conjetura 3 es cierta para todo número menor o igual que n . Ahora, como todo número tiene tiempo de parada finito, el número $n + 1$ va a llegar a un número x menor o igual que n en $\sigma(n + 1)$ de iteraciones. Aplicando la hipótesis de inducción a este $x \leq n$, se tiene que existe un entero positivo i tal que $T^i(x) = 1$. Por tanto, $T^{\sigma(n+1)+i}(n + 1) = 1$, terminando así de demostrar la equivalencia de las conjeturas.

En lo relativo a este concepto, el resultado más importante es el teorema de Terras [24] (que se verá en el teorema 37). Este teorema presenta una prueba de que el conjunto de números que satisfacen la conjetura es denso en el sentido de densidad natural, que se define a continuación. También mencionaremos brevemente un resultado similar (véase el teorema 38), pero mucho más potente, al que se ha llegado recientemente.

Definición 22 (densidad asintótica o densidad natural). Un subconjunto $A \subseteq \mathbb{N}$ tiene densidad asintótica α si $\text{Card}(A \cap [1, n])/\text{Card}([1, n]) \rightarrow \alpha$ cuando $n \rightarrow \infty$, donde $[1, n] = \{1, \dots, n\}$. Dicho de otra forma, tiene densidad asintótica α si existe el límite

$$\mathbb{D}(A) := \lim_{x \rightarrow \infty} \frac{1}{x} \text{Card}\{n \in A \mid n \leq x\}$$

y es finito con valor α . ◀

El concepto de densidad asintótica o densidad natural de un conjunto se refiere a la proporción de elementos de ese conjunto en el intervalo $[1, n]$ a medida que este intervalo se hace grande. Los siguientes resultados nos conducirán al teorema de Terras (teorema 37), donde se muestra que el conjunto de elementos con tiempo de parada finito tiene densidad asintótica 1; en otras palabras, se prueba que

$$\mathbb{D}(S(k)) = \lim_{x \rightarrow \infty} \frac{1}{x} \text{Card}\{n \in \mathbb{N} \mid n \leq x \text{ y } \sigma(n) < k\} \rightarrow 1$$

cuando $k \rightarrow +\infty$. Para probar el teorema de Terras es necesario definir previamente el siguiente concepto.

Definición 23 (vector de paridad). Se define como el vector de paridad de n a $v(n) := (x_0(n), x_1(n), \dots)$, donde, para cada $0 \leq i < \infty$, se tiene $x_i(n) \in \{0, 1\}$ con $x_i(n) \equiv T^i(n) \pmod{2}$. Se define el vector de paridad k -truncado como $v_k(n) := (x_0(n), x_1(n), \dots, x_{k-1}(n))$. ◀

Por ejemplo, observando $\Omega(7)$, tenemos que $v(7) = (1, 1, 1, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1, \dots)$ y $v_4(7) = (1, 1, 1, 0)$. Similarmente a la representación de los iterados que se dio en el teorema 9, podemos expresar los iterados en función de los vectores de paridad de la siguiente forma.

Teorema 24 (representación de los iterados). *Se satisface la siguiente igualdad para el k -ésimo iterado de n :*

$$T^k(n) = \lambda_k(n)n + \rho_k(n),$$

donde

$$\lambda_k(n) = \frac{3^{x_0(n)+x_1(n)+\dots+x_{k-1}(n)}}{2^k}$$

y

$$\rho_k(n) = \sum_{i=0}^{k-1} x_i(n) \frac{3^{x_{i+1}(n)+\dots+x_{k-1}(n)}}{2^{k-i}},$$

donde los $x_i(n)$ denotan los elementos del vector de paridad de n .

Demostración. Para $k = 1$ tenemos que $\lambda_1 = 3^{x_0(n)}/2$ y $\rho_1(n) = x_0(n)/2$. Entonces, si n es par, $x_0(n) = 0$ y, en tal caso, $T(n) = n/2$. Si, en cambio, n es impar, se tiene que $x_0(n) = 1$ y, por tanto, $T(n) = (3n + 1)/2$. De esta forma se obtiene la definición de T . Ahora, supongamos que la igualdad es cierta hasta $k - 1$.

Caso 1. $T^{k-1}(n) \equiv x_{k-1}(n) \equiv 0 \pmod{2}$. Entonces,

$$T^k(n) = T(T^{k-1}(n)) = \frac{\lambda_{k-1}(n)n + \rho_{k-1}(n)}{2},$$

y, como $x_{k-1}(n) = 0$, se tiene que

$$\frac{\lambda_{k-1}(n)}{2} = \lambda_k(n)$$

y

$$\frac{\rho_{k-1}(n)}{2} = \frac{1}{2} \sum_{i=0}^{k-2} x_i(n) \frac{3^{x_{i+1}(n)+\dots+x_{k-2}(n)}}{2^{k-1-i}} = \sum_{i=0}^{k-1} x_i(n) \frac{3^{x_{i+1}(n)+\dots+x_{k-1}(n)}}{2^{k-i}} = \rho_k(n).$$

Con todo esto, se llega a que

$$T^k(n) = \lambda_k(n)n + \rho_k(n).$$

Caso 2. $T^{k-1}(n) \equiv x_{k-1}(n) \equiv 1$ (mód 2). Entonces,

$$T^k(n) = T(T^{k-1}(n)) = \frac{3\lambda_{k-1}(n)n + 3\rho_{k-1}(n) + 1}{2}.$$

Y ahora, como $x_{k-1}(n) = 1$,

$$\frac{3\lambda_{k-1}}{2} = \lambda_k(n)$$

y

$$\begin{aligned} \frac{3\rho_{k-1}(n) + 1}{2} &= \frac{1}{2} \left[\sum_{i=0}^{k-2} x_i(n) \frac{3^{x_{i+1}(n)+\dots+x_{k-2}(n)}}{2^{k-1-i}} \right] 3^{x_{k-1}(n)} + \frac{1}{2} \\ &= \sum_{i=0}^{k-2} x_i(n) \frac{3^{x_{i+1}(n)+\dots+x_{k-1}(n)}}{2^{k-i}} + x_{k-1}(n) \frac{1}{2} \\ &= \sum_{i=0}^{k-1} x_i(n) \frac{3^{x_{i+1}(n)+\dots+x_{k-1}(n)}}{2^{k-i}} = \rho_k(n), \end{aligned}$$

y de nuevo se llega a que $T^k(n) = \lambda_k(n)n + \rho_k(n)$. ■

Con esta representación de los iterados se observa que una condición indispensable para que $T^k(n) < n$ se cumpla para cierto k es que $\lambda_k(n) < 1$, debido a que $\rho_k(n)$ nunca es negativo. Este hecho motiva la siguiente definición.

Definición 25 (coeficiente del tiempo de parada). Se llama coeficiente del tiempo de parada, $\omega(n)$, al menor k tal que $\lambda_k(n) < 1$. ◀

Proposición 26. Se cumple que $\omega(n) \leq \sigma(n)$.

Demostración. Si $\sigma(n) = +\infty$, la desigualdad se cumple trivialmente, así que supongamos que $\sigma(n) = k$. Entonces, por el comentario anterior, $\lambda_k(n) < 1$, luego $\omega(n) \leq k = \sigma(n)$. ■

Proposición 27. Se satisface que $T^k(n) < n$ si y solo si $\rho_k(n)/(1 - \lambda_k(n)) < n$.

Demostración. Inmediato por el teorema 24. ■

Teorema 28 (periodicidad). Sea $v_k(n)$ un vector de paridad k -truncado. Entonces, se cumple que $v_k(n) = v_k(m)$ si y solo si $n \equiv m$ (mód 2^k).

Demostración. Supongamos que $v_k(n) = v_k(m)$. Notemos que, en ese caso, $\lambda_k(n) = \lambda_k(m)$ y $\rho_k(n) = \rho_k(m)$. Entonces,

$$T^k(n) - T^k(m) = \lambda(n)(n - m) = \frac{3^{x_0(n)+x_1(n)+\dots+x_{k-1}(n)}}{2^k}(n - m)$$

y, como esta expresión tiene que ser un entero, se debe cumplir que $2^k \mid (n - m)$.

Por otro lado, si $n \equiv m$ (mód 2^k), veamos que se cumple que $x_i(m) = x_i(n)$ para todo $i = 0, 1, \dots, k$. El caso $i = 0$ es inmediato por hipótesis. Supongamos que se cumple hasta j y veamos que se cumple para $j + 1$. Como $x_{j+1}(n) \equiv T^{j+1}(n)$ y tanto en la expresión de $\lambda_{j+1}(n)$ como la de $\rho_{j+1}(n)$ solo aparece hasta el término $x_j(n)$, se tiene que $x_{j+1}(n) - x_{j+1}(m) \equiv T^{j+1}(n) - T^{j+1}(m) = \lambda_{j+1}(n)(n - m) \equiv 0$ (mód 2^k). ■

Cabe destacar que este hecho también implica la periodicidad de las funciones $\lambda_k(n)$ y $\rho_k(n)$. Esta propiedad motiva a definir al conjunto

$$[n : k] := \{m \in \mathbb{N} \mid m \equiv n \pmod{2^k}\}$$

y centrarse en el estudio de los elementos de este conjunto.

Teorema 29. *Supongamos que $\omega(n) = k$ y sea $m \in [n : k]$. Entonces, existe $M \in \mathbb{N}$ tal que para todo $m \geq M$ se satisface que $\sigma(m) = k$.*

Demostración. Como $\rho_k(n)$ y $\lambda_k(n)$ son periódicas con periodo 2^k , entonces $t(n) = \rho_k(n)/(1 - \lambda_k(n))$ también lo es. Sea R verificando que $t(n) < n + R2^k$. Entonces, para $r \geq R$ se cumple que

$$t(n + r2^k) = t(n) < n + R2^k \leq n + r2^k,$$

y, por la proposición 27, esto implica que $T^k(n + r2^k) < n + r2^k$. Llamando $m = n + r2^k \geq n + R2^k = M$, se concluye, por tanto, que $\sigma(m) \leq k$. Finalmente, como siempre se verifica que $\sigma(m) \geq \omega(m) = \omega(n) = k$, entonces $\sigma(m) = k$. ■

Definición 30. Se define $P[\omega = k] := \text{Card}\{n \in [1, 2^k] \mid \omega(n) = k\}/2^k$ como la proporción de enteros del intervalo $[1, 2^k]$ con coeficiente de parada igual a k . Análogamente se definen $P[\omega \leq k]$, $P[\omega \geq k]$, $P[\omega < k]$ y $P[\omega > k]$. ◀

Teorema 31. *El límite*

$$F(k) := \lim_{x \rightarrow \infty} \frac{1}{x} \text{Card}\{n \in \mathbb{N} \mid n \leq x \text{ y } \sigma(n) \geq k\}$$

existe y vale $P[\omega \geq k]$ para todo $k \in \mathbb{N}$.

Demostración. Por la periodicidad del coeficiente de parada se satisface que

$$P[\omega = k] = \lim_{m \rightarrow \infty} \frac{1}{m} \text{Card}\{n \leq m \mid \omega(n) = k\}.$$

Por el teorema 29, el conjunto de los números con coeficiente de parada igual a k es igual al de los números con tiempo de parada k salvo, a lo sumo, un conjunto finito de números. Por tanto,

$$P[\omega = k] = \lim_{m \rightarrow \infty} \frac{1}{m} \text{Card}\{n \leq m \mid \sigma(n) = k\}$$

y se sigue que

$$P[\omega \geq k] = \lim_{m \rightarrow \infty} \frac{1}{m} \text{Card}\{n \leq m \mid \sigma(n) \geq k\} = F(k). \quad \blacksquare$$

Se observa que

$$\mathbb{D}(S(k)) = 1 - F(k) = 1 - P[\omega \geq k] = P[\omega < k]$$

cuando $k \rightarrow \infty$. El objetivo de los resultados siguientes es probar que $F(k) \rightarrow 0$ cuando $k \rightarrow \infty$ para concluir que el conjunto $\mathbb{D}(S(k))$ de los números con tiempo de parada finito tiene densidad asintótica 1.

Para hallar $F(k)$ vamos a utilizar que

$$(8) \quad P[\omega \geq k] = \frac{1}{2^k} \text{Card}\{n \in [1, 2^k] \mid \lambda_i(n) > 1 \text{ para todo } 1 \leq i \leq k - 1\}.$$

Como $\lambda_i(n) > 1$ para todo $0 \leq i \leq k - 1$, se sigue que

$$\begin{aligned} \lambda_i(n) = \frac{3^{x_0(n)+x_1(n)+\dots+x_{i-1}(n)}}{2^i} > 1 &\iff 3^{x_0(n)+x_1(n)+\dots+x_{i-1}(n)} > 2^i \\ &\iff (x_0(n) + x_1(n) + \dots + x_{i-1}(n)) \log(3) > i \log(2) \\ &\iff x_0(n) + x_1(n) + \dots + x_{i-1}(n) > i \frac{\log(2)}{\log(3)}. \end{aligned}$$

Con esto, se observa que una forma posible para hallar $P[\omega \geq k]$ es hallar cuántos vectores $v_i(n)$ de longitud $i \leq k - 1$ satisfacen que $\lambda_i(n) > 1$. Por ello, vamos a definir los siguientes conceptos.

Definición 32 (vector admisible). Sea $\gamma = \log(2)/\log(3)$. Un vector $v_k = (v_0, v_1, \dots, v_{k-1})$, con $v_i \in \{0, 1\}$, es admisible si

$$v_0 + v_1 + \dots + v_{i-1} > i\gamma, \quad \text{para todo } 1 \leq i \leq k-1.$$

Si, además, se verifica que

$$v_0 + v_1 + \dots + v_{k-1} > k\gamma,$$

se dice que el vector admisible es activo, y en otro caso se le llama terminal. ◀

Definición 33. Se definen

$$n(a, k) = \begin{cases} \text{número de vectores admisibles de longitud } k \text{ con } a \text{ ceros} & \text{si } a \in [0, k], \\ 0 & \text{si } a \notin [0, k], \end{cases}$$

y

$$c(a, k) = \begin{cases} 1 & \text{si } a < k(1 - \gamma), \\ 0 & \text{si } a \geq k(1 - \gamma). \end{cases}$$

A $n(a, k)$ se le conoce como el coeficiente binomial modificado. ◀

Se observa que, si a es el número de ceros de un vector v_k de longitud k cuyas componentes toman valores 0 o 1, entonces la condición $a < k(1 - \gamma)$ equivale a $b > k\gamma$, siendo $b = k - a$ el número de unos de v_k . Además, si un vector admisible es terminal, entonces $c(a, k) = 0$ pero $c(a, k - 1) = 1$.

Teorema 34. Se cumple que $P[\omega \geq k] = 2^{-k} \sum_{a=0}^k n(a, k) = F(k)$.

Demostración. Trivial, partiendo de la definición de $n(a, k)$ y de la expresión de $P[\omega \geq k]$ dada en (8). ■

Teorema 35. Sean $n(0, 1) = 0, n(1, 1) = 0$. Entonces, se satisface la recurrencia

$$n(a, k + 1) = c(a, k)n(a, k) + c(a - 1, k)n(a - 1, k).$$

Demostración. Basta observar que $c(a, k)n(a, k)$ denota el número de vectores admisibles y activos de longitud k con a ceros, y análogamente para $c(a - 1, k)n(a - 1, k)$. Si tenemos un vector admisible de longitud $k + 1$ con a ceros, eliminando la última componente, será un vector activo con a o $a - 1$ ceros. Por otro lado, los vectores de longitud k activos pueden extenderse añadiendo una componente a vectores admisibles. ■

Corolario 36. Se tiene que $n(a, k) \leq \binom{k}{a}$.

Demostración. Utilizando el hecho de que $c(a, k)$ es una función booleana, por la fórmula de recursión anterior se tiene que $n(a, k + 1) \leq n(a, k) + n(a - 1, k)$ y ahora se puede probar la desigualdad por inducción simplemente del hecho de que $\binom{k+1}{a} = \binom{k}{a} + \binom{k}{a-1}$. ■

Teorema 37 (Terras, 1976). Se satisface que $\lim_{k \rightarrow \infty} F(k) = 0$.

Demostración. Por el teorema 34 se cumple que

$$P[\omega \geq k] = \sum_{a=0}^k \frac{n(a, k)}{2^k} = F(k).$$

Sea v_k un vector admisible de longitud k y sea a el número de ceros y $b = k - a$ el número de unos. Si v_k es activo, entonces $3^b/2^k > 1$, donde $a < k(1 - \gamma)$. En cambio, si es terminal se tiene que $3^b/2^k < 1$, pero $3^b/2^{k-1} < 1$ y entonces $a < (k - 1)(1 - \gamma)$. Juntando todo se tiene que $n(a, k) = 0$ cuando $a > \lfloor k(1 - \gamma) \rfloor$, y con esto y el corolario 36 se llega a que

$$P[\omega \leq k] = \sum_{a=0}^{\lfloor k(1-\gamma) \rfloor} \frac{n(a, k)}{2^k} \leq \sum_{a=0}^{\lfloor k(1-\gamma) \rfloor} \binom{k}{a} \frac{1}{2^k}.$$

Esta última expresión corresponde a $P[S_k \leq \lfloor k(1 - \gamma) \rfloor]$, donde $S_k \sim \text{Bin}(k, 1/2)$ (es decir, sigue una distribución binomial²) puesto que es la suma de k variables independientes que se distribuyen como una Bernoulli con $p = 1/2$. Entonces, normalizando, se tiene que

$$(9) \quad \sum_{a=0}^{\lfloor k(1-\gamma) \rfloor} \binom{k}{a} \frac{1}{2^k} = P[S_k \leq k(1 - \gamma)] = P\left[\frac{S_k - k/2}{\sqrt{k}/2} \leq \frac{k(1 - \gamma) - k/2}{\sqrt{k}/2} = \sqrt{k}(1 - 2\gamma)\right].$$

Aplicando el teorema central del límite [22] se obtiene que

$$\lim_{k \rightarrow \infty} P\left[\frac{S_k - k/2}{\sqrt{k}/2} \leq x\right] = \Phi(x) = \frac{1}{2\sqrt{\pi}} \int_{-\infty}^x e^{-t^2/2} dt.$$

Dado un $\varepsilon > 0$, podemos encontrar $x \in \mathbb{R}$ tal que $\Phi(x) < \varepsilon$ y, como $1 - 2\gamma < 0$, también podemos encontrar un $K > 0$, suficientemente grande, que cumpla que

$$\sqrt{k}(1 - 2\gamma) < x \quad \text{para todo } k \geq K$$

y

$$\frac{S_k - k/2}{\sqrt{k}/2} \leq x \quad \text{para todo } k \geq K.$$

Por lo tanto, finalmente se tiene que

$$P[\omega \leq k] \leq \sum_{a=0}^{\lfloor k(1-\gamma) \rfloor} \binom{k}{a} \frac{1}{2^k} = P\left[\frac{S_k - k/2}{\sqrt{k}/2} \leq \sqrt{k}(1 - 2\gamma)\right] < \varepsilon, \quad \text{para todo } k \geq K,$$

luego

$$F(k) = \lim_{x \rightarrow \infty} \frac{1}{x} \text{Card}\{n \mid n \leq x \text{ y } \sigma(n) \geq k\} \rightarrow 0$$

cuando $k \rightarrow \infty$. ■

En conclusión,

$$1 - F(k) = D(S(k)) = \lim_{x \rightarrow \infty} \frac{1}{x} \text{Card}\{n \leq x \mid \sigma(n) < k\}$$

existe y $D(S(k)) \rightarrow 1$ cuando $k \rightarrow \infty$. Esto nos indica de que «casi todos» los enteros positivos (en un sentido de densidad asintótica) cumplen la conjetura de Collatz.

Recientemente se han realizado avances en un sentido similar al resultado anterior. El célebre matemático Terence Tao demostró en 2019 una versión más fuerte del teorema de Terras. En primer lugar, reemplazó el sentido que antes se ha dado al término de «casi todos», aludiendo a la densidad asintótica, por otro basado en la densidad logarítmica. La densidad logarítmica de un conjunto $A \subseteq \mathbb{N}$ se define como el siguiente límite (si existe):

$$\delta(A) := \lim_{x \rightarrow \infty} \frac{1}{\log x} \sum_{\substack{n \in A \\ n \leq x}} \frac{1}{n}.$$

En su artículo, Tao [21] prueba lo siguiente.

Teorema 38 (Terence Tao, 2019). *Sea $f : \mathbb{N} \rightarrow \mathbb{N}$ una función cualquiera con $\lim_{n \rightarrow \infty} f(n) = +\infty$ y sea $C_{\min}(n) = \min\{C^k(n) \mid k \in \mathbb{N}\}$ el valor mínimo de la trayectoria de n mediante la función de Collatz $C(x)$. Entonces, se tiene que*

$$C_{\min}(n) < f(n),$$

para «casi todo» $n \in \mathbb{N}$ (en el sentido de densidad logarítmica).

²Si $X \sim \text{Bin}(n, p)$, entonces la expresión de una binomial $P[X \leq k]$ viene dada por $P[X \leq k] = \sum_{x=0}^k \binom{n}{x} p^x (1-p)^{n-x}$.

5. Conclusiones

En este artículo se ha pretendido presentar algunos de los conceptos más importantes en relación a la conjetura $3x + 1$ y algunos de los resultados parciales más importantes. Otros campos de estudio no han podido tratarse en este artículo. Por tal motivo se proporciona bibliografía a la que el lector con interés en profundizar en este tema pueda acudir si desea conocer más sobre la conjetura.

- Lagarias [8, 10, 12] presenta una introducción al problema $3x + 1$ tratando su contexto histórico y resultados más relevantes. También recopila en su libro *The ultimate challenge: the $3x + 1$ problem* [11] todas las investigaciones más recientes y relevantes que se han realizado sobre la conjetura.
- Varios autores han tratado de estudiar la conjetura hacia atrás mediante el estudio de la función inversa, $T^{-1}(x)$. A esto se le ha llamado árboles de Collatz y para información al respecto es recomendable consultar el artículo de Applegate y Lagarias [1]. En este aspecto, también se ha tratado el problema desde el punto de vista del álgebra mediante el llamado «semigrupo $3x + 1$ » [2] y se ha llegado a resultados interesantes.
- Se ha tratado de estudiar modelos estocásticos que puedan permitir predecir el comportamiento de los iterados, por ejemplo, en el artículo de Lagarias y Weiss [13].
- Evidencia computacional de la conjetura puede verse en el trabajo de Oliveira e Silva [19], donde se ha comprobado la conjetura hasta $20 \times 2^{58} \approx 5,7646 \times 10^{18}$.
- Avances importantes en el estudio de ciclos se han realizado en el trabajo de Eliahou [5], donde el resultado principal demuestra que, si existen ciclos no triviales, estos deben ser de longitud al menos 17 087 915. A pesar de esto, debemos mencionar que no se conoce ninguna prueba de que deba existir un número finito de ciclos, aunque se cree que esto es cierto.
- Otros autores también han extendido la conjetura a otros espacios como \mathbb{Q} [9], \mathbb{R} [7] o incluso \mathbb{C} [15].
- Por último, debemos hacer mención al trabajo de Conway [4] en cuanto a la indecidibilidad de la conjetura.

La conjetura de Collatz es todavía un problema de actualidad. Es interesante tanto para aficionados a las matemáticas como para expertos debido a su sencillez y su dificultad al mismo tiempo. Todavía hay matemáticos que continúan trabajando en ella. Por citar uno de los últimos intentos realizados, Peter Schorer [20] dio a conocer una nueva vía de demostración en septiembre de 2019, pero desgraciadamente se le encontró un error insalvable. El avance más importante, quizás, sea el realizado por Terence Tao en 2019 [21], el cual presentamos en el teorema 38, y parece que establece una nueva conexión de la conjetura con el área de ecuaciones en derivadas parciales [6].

¿Por qué se trata de un problema tan difícil a pesar de que es muy fácil de enunciar? Lagarias [10] reflexiona sobre la complejidad del problema. Por un lado, los iterados tienen un comportamiento «pseudoaleatorio», es decir, aunque estén perfectamente definidos, parecen comportarse aleatoriamente. Esto hace que, como hemos visto, el problema se conecte con teoría ergódica y sistemas dinámicos; resulta, por tanto, difícil de abordar. Destaca, por otro lado, la indecidibilidad del problema. De hecho, el resultado de Conway [4, teorema 1] nos indica que no existe ningún algoritmo con parámetro de entrada la función $T(x)$ y un $n \in \mathbb{N}$ capaz de decidir en un número finito de pasos si existe i tal que $T^i(n) = 1$ o no. Se puede observar que, construyendo un algoritmo que calcule las iteraciones, en el caso de que en algún momento $T^i(n) = 1$, el algoritmo podría devolver «sí». Sin embargo, si algún entero entrara en una trayectoria divergente, nuestro algoritmo no tendría forma de devolver «no».

En cuanto a la dificultad del problema, el prolífico matemático Paul Erdős (1913-1996) decía que «las matemáticas aún no están preparadas para tales problemas» [8].

Referencias

- [1] APPLGATE, David y LAGARIAS, Jeffrey C. «The distribution of $3x + 1$ trees». En: *Experimental Mathematics* 4.3 (1995), págs. 193-209. ISSN: 1058-6458. <https://doi.org/10.1080/10586458.1995.10504321>.

- [2] APPLGATE, David y LAGARIAS, Jeffrey C. «The $3x + 1$ semigroup». En: *Journal of Number Theory* 117.1 (2006), págs. 146-159. ISSN: 0022-314X. <https://doi.org/10.1016/j.jnt.2005.06.010>.
- [3] BUTTSWORTH, Robert N. y MATTHEWS, Keith R. «On some Markov matrices arising from the generalized Collatz mapping». En: *Acta Arithmetica* 55.1 (1990), págs. 43-57. ISSN: 0065-1036. <https://doi.org/10.4064/aa-55-1-43-57>.
- [4] CONWAY, John Horton. «Unpredictable iterations». En: *The ultimate challenge: the $3x + 1$ problem*. Ed. por Lagarias, Jeffrey C. Providence, US: American Mathematical Society, 2010, págs. 219-224. ISBN: 978-0-8218-4940-8.
- [5] ELIAHOV, Shalom. «The $3x + 1$ problem: new lower bounds on nontrivial cycle lengths». En: *Discrete Mathematics* 118 (1993), págs. 45-56. ISSN: 0012-365X. [https://doi.org/10.1016/0012-365X\(93\)90052-U](https://doi.org/10.1016/0012-365X(93)90052-U).
- [6] HARTNETT, Kevin. «Un gran resultado matemático para un “problema peligroso”». En: *Investigación y ciencia* (18 de dic. de 2019). URL: <https://www.investigacionyciencia.es/noticias/un-gran-resultado-matematico-para-un-problema-peligroso-18125>.
- [7] KONSTADINIDIS, Pavlos. «The real $3x + 1$ problem». En: *Acta Arithmetica* 122 (2004), págs. 35-44. ISSN: 0065-1036. <https://doi.org/10.4064/aa122-1-3>.
- [8] LAGARIAS, Jeffrey C. «The $3x + 1$ problem and its generalizations». En: *The American Mathematical Monthly* 92.1 (1985), págs. 3-23. ISSN: 0002-9890. <https://doi.org/10.2307/2322189>.
- [9] LAGARIAS, Jeffrey C. «The set of rational cycles for the $3x + 1$ problem». En: *Acta Arithmetica* 56.1 (1990), págs. 33-53. ISSN: 0065-1036. <https://doi.org/10.4064/aa-56-1-33-53>.
- [10] LAGARIAS, Jeffrey C. «The $3x + 1$ problem: an overview». En: *The ultimate challenge: the $3x + 1$ problem*. Ed. por Lagarias, Jeffrey C. Providence, US: American Mathematical Society, 2010, págs. 3-30. ISBN: 978-0-8218-4940-8.
- [11] LAGARIAS, Jeffrey C., ed. *The ultimate challenge: the $3x + 1$ problem*. Providence, US: American Mathematical Society, 2010. ISBN: 978-0-8218-4940-8.
- [12] LAGARIAS, Jeffrey C. «Erdős, Klarner, and the $3x + 1$ problem». En: *The American Mathematical Monthly* 123.8 (2016), págs. 753-776. ISSN: 0002-9890. <https://doi.org/10.4169/amer.math.monthly.123.8.753>.
- [13] LAGARIAS, Jeffrey C. y WEISS, Alan. «The $3x + 1$ problem: two stochastic models». En: *The Annals of Applied Probability* 2.1 (1992), págs. 229-261. ISSN: 1050-5164. <https://doi.org/10.1214/aop/1177005779>.
- [14] LEIGH, George M. «A Markov process underlying the generalized Syracuse algorithm». En: *Acta Arithmetica* 46.2 (1986), págs. 125-143. ISSN: 0065-1036. <https://doi.org/10.4064/aa-46-2-125-143>.
- [15] LETHERMAN, Simon; SCHLEICHER, Dierk, y WOOD, Reg. «The $3n + l$ -problem and holomorphic dynamics». En: *Experimental Mathematics* 8.3 (1999), págs. 241-251. ISSN: 1058-6458. <https://doi.org/10.1080/10586458.1999.10504402>.
- [16] MATTHEWS, Keith R. «Generalized $3x + 1$ mappings: Markov chains and ergodic theory». En: *The ultimate challenge: the $3x + 1$ problem*. Ed. por Lagarias, Jeffrey C. Providence, US: American Mathematical Society, 2010, págs. 79-104. ISBN: 978-0-8218-4940-8.
- [17] MATTHEWS, Keith R. y WATTS, Anthony M. «A generalization of Hasse’s generalization of the Syracuse algorithm». En: *Acta Arithmetica* 43.2 (1984), págs. 167-175. ISSN: 0065-1036. <https://doi.org/10.4064/aa-43-2-167-175>.
- [18] MATTHEWS, Keith R. y WATTS, Anthony M. «A Markov approach to the generalized Syracuse algorithm». En: *Acta Arithmetica* 45.1 (1985), págs. 29-42. ISSN: 0065-1036. <https://doi.org/10.4064/aa-45-1-29-42>.
- [19] OLIVEIRA E SILVA, Tomás. «Empirical verification of the $3x + 1$ conjecture and related conjectures». En: *The ultimate challenge: the $3x + 1$ problem*. Ed. por Lagarias, Jeffrey C. Providence, US: American Mathematical Society, 2010, págs. 189-207. ISBN: 978-0-8218-4940-8.

-
- [20] SCHORER, Peter. *A solution to the $3x + 1$ problem*. Technical report. Palo Alto, US: Hewlett-Packard Laboratories.
- [21] TAO, Terence. «Almost all orbits of the Collatz map attain almost bounded values». En: *arXiv e-prints* (2021). arXiv: 1909.03562v4 [math.PR].
- [22] «Teorema del límite central». En: *Wikipedia*. URL: https://es.wikipedia.org/wiki/Teorema_del_l%C3%ADmite_central.
- [23] «Terence Tao». En: *Wikipedia*. URL: https://es.wikipedia.org/wiki/Terence_Tao.
- [24] TERRAS, Riho. «A stopping time problem on the positive integers». En: *Acta Arithmetica* 30.3 (1976), págs. 241-252. ISSN: 0065-1036. <https://doi.org/10.4064/aa-30-3-241-252>.

TEMat

Este trabajo colaboró con una microcharla durante el XVII *Encuentro Nacional de Estudiantes de Matemáticas*, celebrado en Barcelona en julio de 2016.



Aritmética y barajas de cartas

✉ Francisco Albuquerque Picado
Universidade de Lisboa
fmapic@gmail.com

Resumen: A partir del siglo XVIII, con la difusión de las barajas de cartas, se han creado trucos y efectos «mágicos» con estos objetos. Con la llegada del siglo XX, las barajas de cartas se convierten en un accesorio habitual, hasta esencial, para los ilusionistas. Uno de los magos de esta época fue Si Stebbins, quien, durante sus espectáculos, divulgó trucos con una ordenación de la baraja de cartas de póquer (52 cartas, 4 palos) que hoy en día recibe su nombre.

En este artículo, se introduce la ordenación de Si Stebbins y se explora otra ordenación que la precede en más de 350 años, presente en un libro del siglo XVII, escrito por el matemático portugués Gaspar Cardozo de Sequeira. Como conclusión, se estudian las matemáticas que hay detrás de cada una de estas ordenaciones, determinando el caso general, aplicable a un conjunto de n objetos.

Abstract: From the 18th century onwards, with the dissemination of playing cards, “magic” tricks and effects have been created using them. In the 20th century, playing cards became a customary accessory, perhaps even essential, for magicians. One of them, during this time, was Si Stebbins, who during his shows performed tricks with a deck of poker playing cards (52 cards, 4 suits) pre-arranged in a specific order which is now named in his honor.

In this article, we expose Si Stebbins's order and we explore a different arrangement preceding it by more than 350 years, present in a 17th century book written by the Portuguese mathematician Gaspar Cardozo de Sequeira. Concluding the article, we study what mathematical rules allow us to explain how any of both arrangements work in order to have the desired effect, ending with the determination of the general case which is applicable to a set of n objects.

Palabras clave: matemática, matemática recreativa, magia, ilusionismo, álgebra abstracta.

MSC2020: 00A08.

Recibido: 3 de marzo de 2017.

Aceptado: 28 de octubre de 2021.

Agradecimientos: Agradezco a la ANEM por la oportunidad de poder participar con este artículo en *TEMat* y agradezco igualmente a la organización del XVII ENEM por haberme dado a conocer el proyecto y haber aceptado mi charla sobre esta temática.

Por sus sugerencias, agradezco a André Sintra, Anne Elorza Deias, Francisco Clemente García, Joaquim Brugués Mora, Lia Malato Leite, Manuel Cotero, Marta Chmielewska, Pedro J. Freitas y Rui Baptista Moura.

Los agradecimientos finales van para mi familia por su apoyo permanente y para la Associação Ludus que, a través del Circo Matemático, me ha enseñado que la Matemática puede tener una vertiente lúdica y divertida utilizando conceptos fundamentales para cualquier matemático.

Referencia: ALBUQUERQUE PICADO, Francisco. «Aritmética y barajas de cartas». En: *TEMat*, 6 (2022), págs. 83-96. ISSN: 2530-9633. URL: <https://temat.es/articulo/2022-p83>.

© Este trabajo se distribuye bajo una licencia Creative Commons Reconocimiento 4.0 Internacional <https://creativecommons.org/licenses/by/4.0/>

1. Introducción

«Aquí tengo una baraja de cartas. La voy a cortar y barajar cuantas veces quieras. Cuando estés satisfecho, ¡eliges una carta y yo adivinaré cuál es tu carta, a través del uso de las ondas cosmopsicobiológicas!»

Probablemente, en algún punto de nuestras vidas, ya todos habremos escuchado esta conversación o conversaciones parecidas (tal vez sin las citadas ondas) por parte de un familiar, un amigo u otro. Es normal. La magia atrae a la humanidad desde hace milenios. Existen relatos de un truco utilizando bolas y vasos en el Antiguo Egipto, el cual, casi 5000 años después, sigue teniendo éxito hoy en día. Con el paso del tiempo, esta atracción por la magia llevó a que siempre se buscaran nuevas formas y objetos para nuevos efectos, aparentemente, mágicos. Se han creado trucos con cuerdas, con aros metálicos, con animales vivos, y, por fin, con cartas de juego.

Las cartas se destacan, hoy en día, porque muchos de nosotros no conseguiríamos imaginar la magia sin ellas. Pero lo más curioso es que no sabemos ni cómo fueron inventadas ni la historia de su utilización. Sabemos que llegaron a Europa en el siglo XIV (Alemania, Francia, España) y 200 años después, en 1593, se publicó el primer libro sobre juegos de cartas, llamado *Giocchi di carte bellissimi di regola e di memoria*, de Oratio Galazzo. Le siguió otro libro en 1612 llamado *Thesouro de Prudentes* [7], donde el tercer capítulo está dedicado al ilusionismo y muchos de los efectos descritos utilizan cartas.

En el siglo XIX, la magia, hasta entonces solo vista en ferias y eventos privados, pasó a ser parte de actos en grandes salas y auditorios donde se montaron grandes espectáculos llenos de *glamour* y fantasía. A partir del siglo XX, las barajas de cartas se volvieron populares de tal forma que se publicaron y popularizaron innumerables trucos de magia utilizando cartas, popularidad que sigue atrayendo a la gente a este mundo fantástico a día de hoy [3, 4].

2. ¡Orden! ¡Orden en la baraja!

Volvamos a nuestra situación inicial. Un voluntario elige una carta de una baraja de cara hacia abajo, abierta en abanico por un mago. Guarda su carta, no se la enseña a nadie. ¡Pasados unos segundos, el mago consigue adivinar la carta elegida! ¿Cómo lo ha hecho?

En trucos de magia usando cartas, es común que la baraja esté ordenada de una forma particular. En lenguaje especializado, llamamos *stack* a una ordenación particular de la baraja de cartas. Hay varios tipos de *stacks*, pero el caso que veremos ahora es, simultáneamente, una *full stack*, ya que necesita todas las cartas de la baraja, y una *sequential stack*, porque el conjunto de cartas ordenadas forma una secuencia donde cada carta en la baraja nos informa de cuál es la carta siguiente, formando un ciclo [4].

2.1. Métodos de preparación

Ante todo, para clarificar, estudiaremos ordenaciones solo en *barajas de póquer*. Cada carta tiene dos características que la definen totalmente: su palo y su valor. Para nuestros métodos, queremos tener una secuencia de palos y una secuencia de valores cuya conjugación permita recorrer la baraja entera. Como si la baraja entera fuera un circuito cerrado.

2.2. Si Stebbins

Uno de los grandes magos en popularizar una ordenación de la baraja de cartas para sus trucos fue el mago estadounidense Si Stebbins, fotografiado en la figura 1, en su libro *Card Tricks And The Way They Are Performed* [10].

No se conoce mucho sobre su vida. Sabemos que nació en 1867, fue acróbata y payaso de circo, bajo el nombre artístico *Vino*, pero adoptó el nombre Si Stebbins en 1892. Hizo espectáculos en ferias de muestras, una gira con una empresa de automóviles y también tuvo un espectáculo itinerante con el apoyo de su panadería local. Murió a los 83 años [8].



Figura 1: Si Stebbins.

2.2.1. Método de preparación de Si Stebbins

A continuación, se muestra el método de ordenación de Si Stebbins. Primero explicaremos como ordenar la secuencia de palos, seguidamente la secuencia de valores y, al final, describiremos como conseguir la ordenación de toda la baraja.

En cuanto a la secuencia de palos, a lo largo de nuestra *stack*, los palos se repiten cada cuatro naipes y cambian entre Tréboles [♣], Corazones [♥], Picas [♠] y Rombos [♦], siempre por este orden:



A este orden, en lenguaje técnico, los ilusionistas lo llaman el orden «CHaSeD» (*Clubs* [♣], *Hearts* [♥], *Spades* [♠], *Diamonds* [♦]).

En cuanto a la secuencia de valores, los valores describen una progresión aritmética. Al haber 13 cartas por palo, estaremos ordenando los 13 valores de cada una de ellas. O sea, ordenando los números naturales entre 1 y 13. Por convención, asignemos los valores de esta forma:

A	2	3	4	5	6	7	8	9	10	J	Q	K
1	2	3	4	5	6	7	8	9	10	11	12	13

Ahora, por tener una progresión aritmética, para describir esta secuencia de valores es suficiente decir que la diferencia de la progresión es 3. Aclarando, al empezar con 1, le sumamos 3 y le sigue el 4, después le sumamos 3 y tenemos el 7, 10 (7 + 3), 13 (10 + 3), 16 (13 + 3)... Aquí paramos porque llegamos a un obstáculo: 16 no estaba en el conjunto a ordenar. Como procedimiento (adelante será explicado), al llegar a un valor más grande que 13, le sustraemos 13 y el resultado será el valor deseado. Por lo tanto, 16 menos 13 es igual a 3 y podemos seguir la secuencia: sumando 3 a 3 llegamos a 6, y así en adelante. En este caso obtenemos la secuencia

1	4	7	10	13	3	6	9	12	2	5	8	11
---	---	---	----	----	---	---	---	----	---	---	---	----

en la cual nos surgen los 13 valores por orden en un ciclo único. Ahora sustituimos los valores por las cartas correspondientes y obtenemos la siguiente secuencia:

1	4	7	10	13	3	6	9	12	2	5	8	11
A	4	7	10	K	3	6	9	Q	2	5	8	J

Ahora, teniendo una secuencia de palos y una secuencia de valores, solo nos queda emparejarlas término a término y tendremos el orden final de la baraja.

Orden final:

A	4	7	10	K	3	6	9	Q	2	5	8	J
♣	♥	♠	♦	♣	♥	♠	♦	♣	♥	♠	♦	♣
A	4	7	10	K	3	6	9	Q	2	5	8	J
♥	♠	♦	♣	♥	♠	♦	♣	♥	♠	♦	♣	♥
A	4	7	10	K	3	6	9	Q	2	5	8	J
♠	♦	♣	♥	♠	♦	♣	♥	♠	♦	♣	♥	♠
A	4	7	10	K	3	6	9	Q	2	5	8	J
♦	♣	♥	♠	♦	♣	♥	♠	♦	♣	♥	♠	♦

En el caso de que el lector quiera apilar las cartas para poder utilizar la baraja más cómodamente, en este diagrama, una carta está por debajo de otra carta si se encuentra a su izquierda. Es decir, A♣ está por debajo de 4♥, 4♥ está por debajo de 7♠, etc. Al llegar al final de una fila, se pasa a la fila de debajo y las cartas anteriormente apiladas se colocan por debajo de la primera carta de esta nueva fila, empezando a contar a partir de la izquierda.

2.3. Gaspar Cardozo de Sequeira

En 1612, el matemático portugués Gaspar Cardozo de Sequeira publicó el libro *Thesouro de Prudentes* [7], mostrado en la figura 2, considerado uno de los primeros libros en abordar el ilusionismo.

Tampoco se sabe mucho sobre Gaspar Cardozo de Sequeira. Nació en el norte de Portugal, en la villa de Murça [7]. Fue Mestre de Artes por la Universidad de Alcalá (actual Universidad Complutense de Madrid) y fue profesor de Matemática en las ciudades de Lisboa, Coimbra y Porto, habiendo enseñado también en España [5, 6].

La siguiente *stack* aparece en el truco de abertura del capítulo de ilusionismo del libro *Thesouro de Prudentes*.

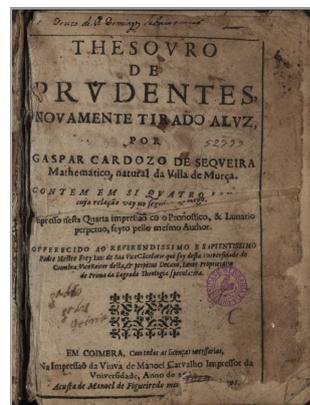


Figura 2: El libro *Thesouro de Prudentes*.

2.3.1. Método de ordenación de Sequeira

El texto original de Sequeira utiliza una baraja de 48 cartas, sin la carta 10. Quitando esta carta de la baraja usual, nos quedamos con la siguiente asignación de valores:

A	2	3	4	5	6	7	8	9	J	Q	K
1	2	3	4	5	6	7	8	9	10	11	12

Para preparar esta *stack*, será más sencillo empezar por la secuencia de valores. En este caso, la diferencia de la progresión aritmética es 5. Empezando con 1, le siguen 6 (1 + 5), 11 (6 + 5), 16 (11 + 5)... Una vez más, 16 no aparece en el conjunto de valores considerado. Para seguir con la secuencia, es suficiente sustraer 12 a 16. De este modo, 16 menos 12 es igual a 4, y recomenzamos el procedimiento hasta que todos los valores hayan sido utilizados. En números, tenemos

1	6	11	4	9	2	7	12	5	10	3	8
---	---	----	---	---	---	---	----	---	----	---	---

Esta secuencia se convierte en cartas de la manera siguiente:

1	6	11	4	9	2	7	12	5	10	3	8
A	6	Q	4	9	2	7	K	5	J	3	8

En el texto original de Sequeira, la secuencia de los cuatro palos es un poco diferente. Empezamos con Tréboles (♣), Diamantes (◇), Picas (♠) y Corazones (♥) y, conforme pasan las cartas, los palos cambian por este orden junto a la secuencia de valores, como en la ordenación de Si Stebbins. Sin embargo, hay un detalle. La única excepción a la regla es que en el paso de la K al 5 no se cambia el palo.

Orden final:

A	6	Q	4	9	2	7	K	5	J	3	8
♣	◇	♠	♥	♣	◇	♠	♥	♥	♣	◇	♠
A	6	Q	4	9	2	7	K	5	J	3	8
♥	♣	◇	♠	♥	♣	◇	♠	♠	♥	♣	◇
A	6	Q	4	9	2	7	K	5	J	3	8
♠	♥	♣	◇	♠	♥	♣	◇	◇	♠	♥	♣
A	6	Q	4	9	2	7	K	5	J	3	8
◇	♠	♥	♣	◇	♠	♥	♣	♣	◇	♠	♥

La forma de apilar las cartas para utilizar la baraja más cómodamente es igual a la forma de apilar las cartas en la ordenación de Stebbins. Podemos poner en práctica nuestras ordenaciones en la próxima sección.

3. Un pequeño truco

El próximo truco se llama *Adivinhação de Sequeira* (Adivinación de Sequeira) y aparece en el libro *Matemagia* [9] de la Associação Ludus.

Efecto:

Un voluntario elige una carta de una baraja de cara hacia abajo, dispuesta en forma de abanico por un mago. Guarda su carta, no la enseña a nadie. ¡Pasados unos segundos, el mago adivina la carta elegida!

Método:

Para poder hacer este truco, es suficiente tener la baraja de cartas preparada usando una de las dos *stacks* que hemos visto arriba.

El mago presenta al voluntario la baraja dispuesta como un abanico (figura 3a). El voluntario retira una carta y el mago corta la baraja en el sitio de la carta elegida, colocando al fondo de la baraja el conjunto de cartas que queda a la derecha de la carta elegida desde la perspectiva del mago (figura 3b).

Observación 1. En la figura 3c, al revés, se enseña la baraja unida nuevamente. La pila azul representa el conjunto de cartas a la derecha de la carta elegida. La pila roja representa las restantes.

¡Así, la carta al fondo de la baraja es la carta anterior a la carta elegida por el voluntario! Al saber cómo se construye la secuencia, sumamos lo que tengamos que sumar al valor de esta carta y pensamos cuál es el palo siguiente, para saber cuál es la carta del voluntario. Se aclara que barajar las cartas puede desordenar la baraja y se debe evitar para que el truco funcione.

Ejemplos:

1. Supongamos que la baraja está ordenada *à la* Si Stebbins: avanzando de 3 en 3 para los valores y ♣, ♥, ♠, ♦ para los palos. Al fondo de la baraja, está un «3♠». Ahora, miramos el valor. Sumamos 3 al valor 3 y nos quedamos con 6. Además, sabemos que ♦ sigue a ♠. Por lo tanto, la carta del voluntario es el 6♦, como aparece en la figura 3d.
2. Supongamos que la baraja está ordenada *à la* Sequeira, es decir, de 5 en 5 para los valores y ♣, ♦, ♠, ♥ para los palos. Al fondo de la baraja, está una «Q♣». Ahora, miramos el valor. La Q vale 11 y sumándole 5 llegamos a 16. 16 es más grande que 12. Sustrayéndole 12, nos quedamos con 4. Este es el valor de la carta elegida. En esta ordenación, el palo de ♦ sigue a ♣. Por lo tanto, la carta elegida es el 4♦.
3. Supongamos que la baraja está ordenada *à la* Sequeira, como en el punto anterior. Al fondo de la baraja, está una «K♠». La K vale 12 y sumándole 5 llegamos a 17. Sustrayéndole 12, nos quedamos con 5. Este es el valor de la carta elegida. En esta ordenación, el palo no cambia en el paso de la K al 5, por lo que seguimos con el palo ♠. Por lo tanto, la carta elegida es el 5♠.



(a) El abanico.



(b) Carta elegida.



(c) Baraja completa.



(d) Carta adivinada.

Figura 3: Distintas fotografías de la baraja.

4. ¿Y las matemáticas? ¡¿Dónde están?!

Ante todo, los siguientes resultados matemáticos se encuentran en cualquier libro introductorio de álgebra abstracta. El texto abajo seguirá el libro *Criptografía e Segurança* de Almeida y Napp [2].

Consideramos \mathbb{Z} , el conjunto de los números enteros, y en él definimos dos relaciones binarias que funcionarán como nuestras herramientas más importantes.

Definición 2 (división). Sean $a \in \mathbb{Z}$ y $n \in \mathbb{Z} \setminus \{0\}$. Decimos que « n divide a » si $a = nb$ para cierto $b \in \mathbb{Z}$. Lo denotamos como $n \mid a$.

Para aclarar esta definición, veamos un ejemplo sencillo. Como $15 = 3 \cdot 5$, existe un entero b tal que $15 = 3b$. Por lo tanto, $3 \mid 15$.

Definición 3 (congruencia). Sean $a, b \in \mathbb{Z}$ y $n \in \mathbb{Z} \setminus \{0\}$. Decimos que « a es congruente con b módulo n », si $n \mid a - b$. Es decir, existe $c \in \mathbb{Z}$ tal que $a = cn + b$. Lo denotamos como $a \equiv b \pmod{n}$.

Con el fin de clarificar esta definición, volvamos a ver algo que hemos usado sin definirlo cuando presentábamos las ordenaciones:

$$16 = 13 \cdot 1 + 3 \iff 16 \equiv 3 \pmod{13}, \quad 16 = 12 \cdot 1 + 4 \iff 16 \equiv 4 \pmod{12}.$$

Estos últimos ejemplos nos dan un primer indicio de como explicar las sustracciones. En realidad, correspondían con hacer congruencias módulo 13 (Stebbins) o módulo 12 (Sequeira). Vemos que, en ambos casos, la congruencia nos indica cómo continuar las progresiones aritméticas al llegar a un valor no considerado.

Proposición 4. Para todo $n \in \mathbb{Z} \setminus \{0\}$, la relación de congruencia módulo n es una relación de equivalencia en \mathbb{Z} . Las clases de equivalencia de esta relación son las clases de congruencia.

Definición 5. Dado $n \in \mathbb{Z} \setminus \{0\}$, \mathbb{Z}_n es el conjunto formado por las clases de congruencia módulo n , que suelen ser representadas como $\{\overline{0}, \overline{1}, \dots, \overline{n-1}\}$. Sin embargo, es una formulación equivalente y nos será más provechoso considerarlas como $\{\overline{1}, \overline{2}, \dots, \overline{n}\}$.

Este conjunto es, de manera natural, un anillo. Definimos la suma y la multiplicación de clases como

$$\overline{a} + \overline{b} = \overline{a + b} \quad \text{y} \quad \overline{a} \overline{b} = \overline{ab}.$$

Con esto presente, veamos dos pequeñas historias. Estas historias sirven como guía visual para lo que se está haciendo con las ordenaciones, así enseñando que hay más aplicaciones que las barajas de cartas.

Primera historia:

Un piso de un museo tiene tres cuartos llenos de esculturas y cuadros. Hay un guardia de seguridad encargado de vigilarlos en turnos de doce horas. Su ronda va de la siguiente forma. En cada cuarto, el vigilante se queda una hora. Empieza en el primer cuarto, sigue al segundo cuarto y acaba en el tercer cuarto. Al acabar su hora en el tercer cuarto, vuelve al primero y todo recomienza. Representamos esta historia con un reloj cuadrangular muy colorido. Definamos que

1. En los cuadrados azules, tenemos las horas en las que el vigilante está vigilando el primer cuarto.
2. En los cuadrados rojos, tenemos las horas en las que el vigilante está observando el segundo cuarto.
3. En los cuadrados naranjas, tenemos las horas en las que el vigilante está guardando el tercer cuarto.

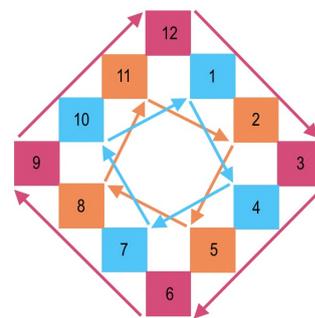


Figura 4: Un reloj cuadrangular esquematizando nuestra historia.

Lo más importante es que *tenemos tres ciclos disjuntos*. Esto es, por ejemplo, hay horas en las que el vigilante nunca está en el segundo cuarto.

Segunda historia:

Supongamos que nos gustan las flores y, entonces, compramos una rosa cada tres días. Representemos los días de la semana con el diagrama numérico de la figura 5.

Además, la florista tiene su tienda abierta todos los días y, por convención, el domingo es el día 1, el lunes es el día 2... y el sábado es el día 7. Las flechas nos indican cuál es la sucesión de días en los que compraremos una rosa.

Ahora vemos que, al comprar una rosa cada tres días, las flechas nos indican solamente un ciclo conexo.

Lo más importante aquí es que *hay un único ciclo disjunto*. Esto es, no hay día de la semana en el que no compremos una flor alguna vez.

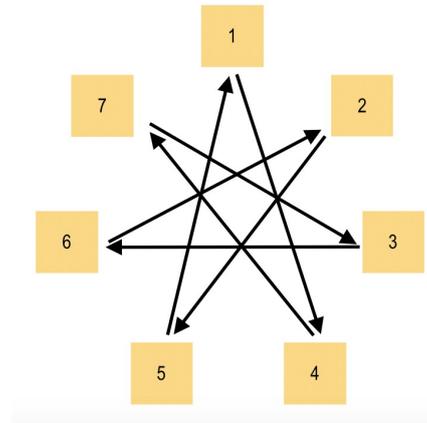


Figura 5: Una representación cíclica de una semana.

Tenemos aquí dos casos cuyos comportamientos distintos se pueden explicar por la teoría de las ecuaciones diofánticas. Empecemos por dos lemas que, al solo servir para demostrar las proposiciones y teoremas que les siguen, también se presentan sin una prueba.

Lema 6 (Bézout). Sean $a, b \in \mathbb{Z} \setminus \{0\}$. Entonces, existen $u, v \in \mathbb{Z}$ tales que

$$au + bv = \text{mcd}(a, b).$$

Lema 7 (Euclides). Sean $a, b \in \mathbb{Z}$ y $n \in \mathbb{Z} \setminus \{0\}$. Si $\text{mcd}(a, n) = 1$ y $n \mid ab$, entonces $n \mid b$.

En este punto, ya hemos visto todas las definiciones y resultados introductorios necesarios para entender lo fundamental de este artículo. Podemos, por fin, presentar y demostrar los resultados clave con el fin de explicar las historias y las ordenaciones de las barajas.

Proposición 8. Sean $a, n \in \mathbb{Z} \setminus \{0\}$ y $c \in \mathbb{Z}$. La ecuación $ax + ny = c$ es resoluble para $x, y \in \mathbb{Z}$ si y solo si $\text{mcd}(a, n) \mid c$.

Demostración. Consideremos $S = \{ax + ny : x, y \in \mathbb{Z}\}$ como el conjunto de las combinaciones lineales de a y n . Denotemos $d := \text{mcd}(a, n)$.

(\Rightarrow) Por el lema de Bézout (lema 6), d es combinación lineal de a y n ; por lo tanto, $d \in S$. De este modo, tomemos $x_0, y_0 \in \mathbb{Z}$ tales que $ax_0 + ny_0 = d$. Multiplicando esta ecuación por $k \in \mathbb{Z}$, tenemos que $a(kx_0) + n(ky_0) = kd$. Por lo tanto, $kd \in S$, es decir, los múltiplos de d están en S .

(\Leftarrow) Recíprocamente, sea $c \in S$. Entonces, $c = ax + ny$ para ciertos $x, y \in \mathbb{Z}$. Como $d \mid a$ y $d \mid n$, tenemos que $d \mid c = ax + ny$. Por lo tanto, c es múltiplo de d .

Por lo tanto, las combinaciones lineales de a y n corresponden a los múltiplos de $\text{mcd}(a, n)$. ■

Teorema 9. Sean $a, n \in \mathbb{Z} \setminus \{0\}$ tales que $d = \text{mcd}(a, n)$. Sea $c \in \mathbb{Z}$ tal que $d \mid c$. Tomemos la ecuación $ax + ny = c$, para $x, y \in \mathbb{Z}$. Supongamos que (x_0, y_0) es una solución particular de la ecuación. Entonces, sus soluciones enteras son exactamente

$$x = x_0 + t \frac{n}{d}, \quad y = y_0 - t \frac{a}{d},$$

donde $t \in \mathbb{Z}$.

Demostración. Sean (x_0, y_0) y (x_1, y_1) soluciones enteras de la ecuación $ax + ny = c$. Entonces,

$$(ax_1 + ny_1) - (ax_0 + ny_0) = a(x_1 - x_0) + n(y_1 - y_0) = c - c = 0.$$

Por lo tanto, $a(x_1 - x_0) = -n(y_1 - y_0)$ y se deduce que

$$(1) \quad \frac{a}{d}(x_1 - x_0) = -\frac{n}{d}(y_1 - y_0),$$

de donde se concluye que

$$\frac{n}{d} \mid \frac{a}{d}(x_1 - x_0).$$

Como $\text{mcd}(a/d, n/d) = 1$, por el lema de Euclides (lema 7), $n/d \mid (x_1 - x_0)$. Así que existe $t \in \mathbb{Z}$ tal que

$$(x_1 - x_0) = t \frac{n}{d} \iff x_1 = x_0 + t \frac{n}{d}.$$

Sustituyendo en la ecuación (1), obtenemos que

$$\frac{a}{d} \left(t \frac{n}{d} \right) = -\frac{n}{d}(y_1 - y_0) \iff \left(\frac{a}{d} t \right) \frac{n}{d} = -\frac{n}{d}(y_1 - y_0) \iff \frac{a}{d} t = -(y_1 - y_0),$$

de lo que se deduce que

$$y_1 = y_0 - t \frac{a}{d}.$$

Por lo tanto, cualquier solución (x, y) tiene la forma

$$x = x_0 + t \frac{n}{d}, \quad y = y_0 - t \frac{a}{d}. \quad \blacksquare$$

Los próximos teoremas serán los más importantes. Nuestra idea es reformular los resultados anteriores a modo de puente entre las congruencias y las progresiones aritméticas.

Teorema 10. Sean $a, n \in \mathbb{Z} \setminus \{0\}$ y $b \in \mathbb{Z}$. Denotamos $d := \text{mcd}(a, n)$.

1. La congruencia $ax \equiv b \pmod{n}$ tiene soluciones, para $x \in \mathbb{Z}$, si y solo si $d \mid b$.
2. Si $d \mid b$, la solución es única módulo n/d .

Demostración. Esta demostración seguirá principalmente lo que hemos estado haciendo hasta ahora, pero reformulado en el lenguaje de las congruencias.

1. La ecuación $ax \equiv b \pmod{n}$ tiene solución si y solo si existe $x_0 \in \mathbb{Z}$ tal que $ax_0 = b + ny_0$, para $y_0 \in \mathbb{Z}$. Por la proposición 8, esto es equivalente a que $d \mid b$, como queremos probar.
2. Supongamos ahora que $ax - ny = b$ es posible y tomemos una solución particular (x_0, y_0) . Por el teorema 9, cualquier solución (x, y) es de la forma

$$x = x_0 + t \frac{n}{d}, \quad y = y_0 - t \frac{a}{d}.$$

Es decir, toda solución x de $ax \equiv b \pmod{n}$ se obtiene como

$$x = x_0 + t \frac{n}{d}.$$

Como

$$x_0 + t \frac{n}{d} \equiv x_0 \pmod{\frac{n}{d}},$$

entonces, cualquier solución es congruente con x_0 módulo n/d y la solución es única módulo n/d . \blacksquare

Este corolario, aunque se obtenga de todo lo que hemos deducido hasta ahora, es bastante útil para explicar las mencionadas historias, y también las ordenaciones de la baraja de póquer ya detalladas.

Corolario 11. Sean $a, n \in \mathbb{Z} \setminus \{0\}$ y $b \in \mathbb{Z}$. La congruencia $ax \equiv b \pmod{n}$ tiene una solución única módulo n , para $x \in \mathbb{Z}$, si y solo si $\text{mcd}(a, n) = 1$.

En particular, si $\text{mcd}(a, n) = 1$, $ax \equiv 1 \pmod{n}$ tiene solución y es única módulo n . A esta solución, a^{-1} , que cumple que $aa^{-1} \equiv 1 \pmod{n}$ la llamamos «inverso módulo n » y decimos que a es «invertible». Por ejemplo, $3 \cdot 5 \equiv 1 \pmod{7}$, por lo que $5 \equiv 3^{-1} \pmod{7}$. El inverso módulo 7 de 3 es 5.

En adelante, relajaremos el formalismo y nos referiremos a las clases $\bar{1}, \bar{2}, \dots, \bar{n}$ por los números $1, 2, \dots, n$. Antes de hacer la última conexión, veamos una concreción de este corolario similar a las tablas de las ordenaciones de la baraja. Considerando la congruencia $3x$ módulo 7, tenemos lo siguiente:

x	1	2	3	4	5	6	7
$3x \pmod{7}$	3	6	2	5	1	4	7

Esta correspondencia es inyectiva, *i. e.*, cada imagen de $3x$ solo tiene un argumento que le corresponde. O sea, para todo b elegido, $3x \equiv b \pmod{7}$ tiene solución única. Podemos representar las imágenes en un ciclo como $(3, 6, 2, 5, 1, 4, 7)$ y se puede comprobar que es el mismo ciclo de la figura 5. No obstante, si examinamos la congruencia $3x$ módulo 12, tendremos la tabla siguiente:

x	1	2	3	4	5	6	7	8	9	10	11	12
$3x \pmod{12}$	3	6	9	12	3	6	9	12	3	6	9	12

Aquí pasan dos cosas indeseadas:

1. No todo valor de $\{1, 2, \dots, 12\}$ es imagen de $3x$. Representando las imágenes en un diagrama, solo se tendrá el ciclo $(3, 6, 9, 12)$, el cual corresponde al ciclo rojo de la figura 4.
2. No hay soluciones únicas: $3 \cdot 3 \equiv 3 \cdot 7 \equiv 9 \pmod{12}$.

En la congruencia $3x \equiv b \pmod{12}$, $\text{mcd}(3, 12) = 3 \neq 1$, por lo que el comportamiento descrito en los dos ítems anteriores queda justificado. Sin embargo, podemos explicar mejor este fenómeno con el próximo teorema, un teorema fundamental para ordenar una secuencia de valores periódicamente, y al cual hemos llamado «teorema de dibujo de estrellas simples». A tal efecto, trabajaremos con \mathbb{Z}_n , donde la relación de congruencia introducida antes en \mathbb{Z} se convierte en una relación de igualdad, *i. e.*, en vez de escribir $a \equiv b \pmod{n}$, escribimos $a = b$ en \mathbb{Z}_n .

Teorema 12 (dibujo de estrellas simples). *Consideremos $q, n \in \mathbb{Z} \setminus \{0\}$ y $x \in \mathbb{Z}$. La aplicación $\theta: \mathbb{Z}_n \rightarrow \mathbb{Z}_n$, $\theta(x) = qx$, es una biyección si y solo si $\text{mcd}(q, n) = 1$.*

Demostración. Sean $q \in \mathbb{Z}$ y $n \in \mathbb{Z} \setminus \{0\}$.

(\Leftarrow) Supongamos que $\text{mcd}(q, n) = 1$. Queremos ver que θ es una biyección.

Para ver que es inyectiva, sean $c, d \in \mathbb{Z}$ y supongamos $\theta(c) = \theta(d)$. De este modo, $qc = qd$, lo que equivale a que $qc \equiv qd \pmod{n}$. Como $\text{mcd}(q, n) = 1$, la congruencia $qx \equiv 1 \pmod{n}$ tendrá solución única (corolario 11). Sea $y_0 (\neq 0)$ esta solución. Multiplicando $qc \equiv qd \pmod{n}$ por y_0 resulta

$$y_0(qc) \equiv y_0(qd) \pmod{n} \iff (y_0q)c \equiv (y_0q)d \pmod{n} \iff c \equiv d \pmod{n}.$$

O sea, en \mathbb{Z}_n , $c = d$. Por lo tanto, θ es inyectiva. Debido a que el tamaño del dominio es igual al tamaño del codominio, por el principio del palomar, θ será biyectiva.

(\Rightarrow) Supongamos que θ es una biyección. Queremos ver que $\text{mcd}(q, n) = 1$.

En particular, existe $x \in \mathbb{Z}_n$ tal que $qx = 1$, *i. e.*, $qx \equiv 1 \pmod{n}$. Así, q es invertible y, por lo tanto, $\text{mcd}(q, n) = 1$. ■

A fin de visualizar las implicaciones de este teorema, definimos un objeto inspirado por la teoría de grafos.

Definición 13 (estrella). Consideremos un polígono de n lados. Al unir sus vértices cada q vértices sucesivamente hasta alcanzar el vértice inicial, se obtendrá una **estrella**. Esta será **simple** si recorremos todo vértice antes de regresar al vértice inicial, y **múltiple** en el caso contrario. Independientemente, diremos que q es la fase de la estrella. ◀

Observación 14. Aunque este nombre pueda parecer confuso, ya que *estrella* se refiere a un concepto muy similar en teoría de grafos, veremos que es apropiado por su visualización. ◀

Como el propósito de esta definición es puramente visual y geométrico, consideremos un polígono de ocho lados y, en él, dos estrellas de fases 3 y 2, respectivamente.

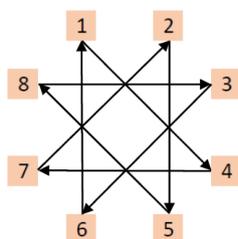


Figura 6: Estrella simple de fase 3.

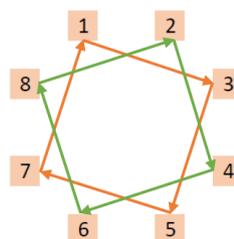


Figura 7: Estrella múltiple de fase 2.

Estos ejemplos muestran que no todas las fases en un conjunto de n puntos generan una estrella simple. En la figura 6, la fase es 3 y nuestra figura nos revela una estrella simple, cosa que no ocurre en la figura 7, en la cual la fase es 2. Así, podemos afirmar algo central que explica el nombre del teorema 12.

Observación 15 (estrellas y aplicaciones). El recorrido de una estrella simple de n puntos y fase q es lo mismo que el trayecto de las imágenes de $\theta: \mathbb{Z}_n \rightarrow \mathbb{Z}_n$, $\theta(x) = qx$, con θ siendo una aplicación biyectiva. Dados $q \in \mathbb{Z}$ y $x \in \mathbb{Z}_n$, definimos qx como $\bar{q}x$, donde \bar{q} es el representante de q en \mathbb{Z}_n . ◀

Con un cálculo rápido, se ve que imágenes sucesivas de θ tendrán diferencia q , como los puntos sucesivos de nuestras estrellas. Igualmente, al ser biyectiva, a cada elemento de \mathbb{Z}_n le asignamos un valor distinto, sin olvidar ningún valor. Esto también ocurre cuando intentamos dibujar una estrella: no queremos olvidar puntos y no queremos que haya más de una flecha saliendo de cada punto.

Observación 16. Utilizando el resultado del teorema 12 y la observación 15, tenemos una estrella simple en la figura 6, porque $\text{mcd}(3, 8) = 1$. Pero, como $\text{mcd}(2, 8) \neq 1$, la estrella de la figura 7 no es simple. ◀

En este momento del artículo, ya tenemos las herramientas para entender por qué en algunos diagramas tenemos solo un ciclo y en otros no. Pero esto no es suficiente para analizar las ordenaciones de las barajas.

Como una carta tiene dos características fundamentales (valor y palo), analizaremos pares de secuencias simultáneamente a lo largo del texto. Así, necesitaremos tener presente un teorema famoso en álgebra abstracta, el teorema del resto chino. A continuación, generalizamos este teorema a módulos no coprimos, al cual hemos llamado «pequeño teorema del resto chino» [1, p. 53-54].

Teorema 17 (del resto chino). Sean $a_1, \dots, a_k \in \mathbb{Z}$ y $n_1, \dots, n_k \in \mathbb{N} \setminus \{0, 1\}$ coprimos dos a dos. Entonces, el sistema de congruencias

$$\begin{cases} x \equiv a_1 & (\text{mód } n_1), \\ x \equiv a_2 & (\text{mód } n_2), \\ \vdots \\ x \equiv a_k & (\text{mód } n_k). \end{cases}$$

tiene una solución única módulo $N = n_1 \cdots n_k$.

Teorema 18 (pequeño teorema del resto chino). Sean $a, b \in \mathbb{Z}$ y $k, p \in \mathbb{N} \setminus \{0, 1\}$. Podemos denotar $D := \text{mcd}(k, p)$ y $M := \text{mcm}(k, p)$. Entonces, si $a \equiv b \pmod{D}$, el sistema de congruencias

$$S = \begin{cases} x \equiv a & (\text{mód } k), \\ x \equiv b & (\text{mód } p). \end{cases}$$

tiene una solución única módulo M .

Demostración. Supongamos que $a \equiv b \pmod{D}$, es decir, que $D \mid a - b$. Utilizando el lema de Bézout (lema 6), podemos escribir D como combinación lineal de k y p . Para ciertos $\alpha, \beta \in \mathbb{Z}$, se tiene que

$$(2) \quad D := \text{mcd}(k, p) = k\alpha + p\beta.$$

Multiplicando la ecuación (2) por $\gamma = (a - b)/\text{mcd}(k, p)$, obtenemos que

$$a - b = k(\alpha\gamma) + p(\beta\gamma).$$

Así, construimos x , una solución del sistema S ,

$$x = a + k(-\alpha\gamma) = b + p(\beta\gamma).$$

Para comprobar que esta solución es única módulo M , podemos tomar x_0 , una solución alternativa del sistema S . De este modo, tenemos que

$$x_0 = a + k\alpha_0 = b + p\beta_0.$$

Por lo tanto, se produce que

$$x - x_0 = k(-\alpha\gamma - \alpha_0) = p(\beta\gamma - \beta_0).$$

Por consiguiente, $x - x_0$ es, simultáneamente, un múltiplo de k y de p . En consecuencia, también será un múltiplo de $\text{mcm}(k, p)$. En conclusión, esta solución es única módulo M . ■

4.1. ¿Y las barajas? ¿Y las historias?

Ahora que sabemos lo que hace falta para tener una biyección, es necesario entender su utilidad. Una biyección en un conjunto finito se llama *permutación* porque su efecto es *reordenar* sus elementos.

Lo que se tiene en la segunda historia es una reordenación. Después de atribuir valor numérico a los días de la semana (1, 2, 3, 4, 5, 6, 7), recorremos sus días según un orden específico y obtenemos la estrella simple (3, 6, 2, 5, 1, 4, 7), como se puede seguir por las flechas en la figura 5.

En las barajas, con nuestras ordenaciones, tenemos una reordenación. Por ejemplo, en la *stack* de Stebbins, después de atribuir valores a las cartas, del ciclo de valores (1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13) pasamos al ciclo de valores (3, 6, 9, 12, 2, 5, 8, 11, 1, 4, 7, 10, 13). Como representa un ciclo, empezando por 1, también puede ser leído como (1, 4, 7, 10, 13, 3, 6, 9, 12, 2, 5, 8, 11) y, volviendo al apartado 2.2.1, podemos comprobar que este ciclo corresponde a los valores en cada línea. Esto se debe al teorema 12, ya que $\text{mcd}(3, 13) = 1$, donde 3 corresponde a la diferencia de la progresión y 13 corresponde a las posiciones. O sea, $\theta: \mathbb{Z}_{13} \rightarrow \mathbb{Z}_{13}$, $\theta(x) = 3x$, es una biyección. Estudiando la progresión de Sequeira, el razonamiento es igual: $\text{mcd}(5, 12) = 1$ y, por lo tanto, $\theta: \mathbb{Z}_{12} \rightarrow \mathbb{Z}_{12}$, $\theta(x) = 5x$, es una biyección.

Para cada *stack*, para los palos, no es difícil formalizar una secuencia con progresiones aritméticas no repitiendo un palo sin haber pasado por todos los anteriores. Sin pérdida de generalidad, asignemos los valores (1, 2, 3, 4) a los palos (Tréboles [♣], Corazones [♥], Picas [♠], Rombos [♦]), respectivamente. Entonces, en la *stack* de Stebbins, nuestra progresión aritmética será la progresión de diferencia 1 en \mathbb{Z}_4 (es decir, {1, 2, 3, 4, 1, 2, ...}) y será una biyección ya que $\text{mcd}(1, 4) = 1$. En la *stack* de Sequeira, nuestra progresión aritmética será la progresión de diferencia 3 en \mathbb{Z}_4 (es decir, {3, 2, 1, 4, 3, 2, ...}), y será una biyección ya que $\text{mcd}(3, 4) = 1$. O sea, siguiendo esto, tendremos una reordenación de cada conjunto de cuatro palos.

Observación 19. Los más atentos ya habréis notado que en esta formulación la secuencia de palos de Sequeira nos da el ciclo (3, 2, 1, 4) en vez de (1, 4, 3, 2) (si hay dudas, basta volver al apartado 2.3.1 y convertir los palos en números). Como se trata de un ciclo, solo hace falta empezar el ciclo por 1. ◀

4.1.1. Casi finalizando...

En lo que respecta a las barajas, hay algo más que añadir. Hemos creado una reordenación de valores y una reordenación de palos independientes entre sí. Lo importante ahora es conjugarlo todo, para entender el motivo de haber recorrido la baraja entera.

Como hemos comentado antes, una carta tiene dos características que la definen totalmente: su valor y su palo. Es decir, podemos ver una carta como un par ordenado (*valor, palo*), *i. e.*, un elemento de $\mathbb{Z}_k \times \mathbb{Z}_p$, en el que k es el número de valores y p es el número de palos. A cada posición de la baraja le corresponde una carta, y así, en cada posición esperamos ver un elemento de la secuencia de valores y un elemento de la secuencia de palos.

Observación 20. Mediante el contexto, puede ser necesario especificar el módulo de cada secuencia. Así, también denotamos el par ordenado (*valor, palo*) por (*valor* (mód k), *palo* (mód p)). ◀

Por ejemplo, supongamos que tenemos una baraja ordenada con la *stack* de Si Stebbins y queremos la posición de la carta 7♠. En esta ordenación, nuestra progresión de valores tiene diferencia 3 y nuestra progresión de palos tiene diferencia 1. Tenemos 13 valores por 4 palos. Como par ordenado, esta carta tiene la representación (7, 3), y es un elemento de $\mathbb{Z}_{13} \times \mathbb{Z}_4$. Así, para descubrir su posición, utilizando el pequeño teorema del resto chino (teorema 18), es necesario resolver

$$S_0 = \begin{cases} 3 \cdot x \equiv 7 \pmod{13} & \text{[VALORES]}, \\ 1 \cdot x \equiv 3 \pmod{4} & \text{[PALOS]}. \end{cases}$$

Esto es equivalente a

$$S_1 = \begin{cases} x \equiv 11 \pmod{13} & \text{[VALORES]}, \\ x \equiv 3 \pmod{4} & \text{[PALOS]}, \end{cases}$$

ya que $3^{-1} \equiv 9 \pmod{13}$ y $3^{-1} \cdot 7 \equiv 9 \cdot 7 \equiv 11 \pmod{13}$. Aplicando el teorema 18, se ve que

$$x \equiv 11 \pmod{52}.$$

Es decir, presuntamente, la carta en cuestión está en la posición 11. Ahora, si vamos a la tabla del apartado 2.2.1, vemos que eso no es verdad. ¿Que ha pasado?

Cuando conjugamos las progresiones aritméticas de diferencia 3 módulo 13 y de diferencia 1 módulo 4, buscamos pares ordenados ($3x \pmod{13}, x \pmod{4}$). Esto significa que en la primera posición, $x = 1$, tenemos el par ordenado (3 (mód 13), 1 (mód 4)) correspondiendo a la carta 3♣. Sin embargo, con la baraja ordenada cara arriba, esta carta está en la octava posición. Esto significa que la *stack* de Stebbins va ocho términos por delante con respecto a las progresiones aritméticas. Así, si sustraemos estos ocho términos, sigue que

$$x - 8 \equiv 11 - 8 \equiv 3 \pmod{52},$$

y así la carta estará en la tercera posición. Podemos comprobar en la tabla del apartado 2.2.1 que la carta 7♠ es la tercera carta en nuestra *stack*.

Aunque podamos calcular la posición de una carta con respecto al principio de la ordenación, lo más importante a tener en cuenta es que *la solución es única en toda la baraja* y lo será independientemente de por dónde empecemos a contar. El teorema 18 aplicado a la baraja de póquer nos dice que cada posición calculada será única módulo $\text{mcm}(13, 4)$, el cual es 52, que es el número de cartas de nuestra baraja.

En el caso de Sequeira, hay un detalle que se explica fácilmente utilizando el teorema 18. A título de ejemplo, consideremos la ordenación de Sequeira, sin hacer la excepción de no cambiar el palo de la K a la 5.

Para simplificar nuestros cálculos y no tener que sustraer términos, empezamos la secuencia de Sequeira con la carta 5♠, la cual está en la primera posición ($x = 1$) entre los pares ($5x \pmod{12}, 3x \pmod{4}$).

VALORES (en \mathbb{Z}_{12})	5	10	3	8	A	6	Q	4	9	2	7	K
PALOS (en \mathbb{Z}_4)	♠	♥	♣	♦	♠	♥	♣	♦	♠	♥	♣	♦
	3	2	1	4	3	2	1	4	3	2	1	4

¿Cual es la 13.^a carta? El próximo término de la secuencia de los valores es 5, el próximo término de la secuencia de los palos es 3. Pero esto corresponde a la carta inicial, y aún no hemos visto la baraja entera. ¿Qué pasa aquí?

Si queremos calcular la posición de una carta en esta ordenación similar a la ordenación de Sequeira, utilizamos la técnica anterior. Supongamos que queremos calcular la posición de la carta $5\spadesuit$. En la ordenación de Sequeira, la progresión de valores tiene diferencia 5 y la progresión de los palos tiene diferencia 3. La carta deseada tiene la representación $(5, 3)$ como un par ordenado de $\mathbb{Z}_{12} \times \mathbb{Z}_4$.

Calculamos su posición resolviendo el sistema

$$S_2 = \begin{cases} 5x \equiv 5 & (\text{mód } 12) & [\text{VALORES}], \\ 3x \equiv 3 & (\text{mód } 4) & [\text{PALOS}]. \end{cases}$$

Después de manipulaciones algebraicas, podemos aplicar el teorema 18 y vemos que

$$x \equiv 1 \pmod{12}.$$

Nuestra carta se encuentra en la primera posición, pero la solución es única módulo $\text{mcm}(12, 4)$, el cual es 12. Es decir, la decimotercera carta será igual a la primera, porque $13 \equiv 1 \pmod{12}$, como hemos comprobado. No obstante, nuestra baraja tiene 48 cartas. Esta secuencia de cartas no recorre la baraja entera. Por esta razón, Sequeira necesitó el paso artificial de eliminar cuatro cambios de palo.

Antes de nuestro último teorema, es necesario mencionar que, en los sistemas de congruencias estudiados, cada congruencia es posible y tiene solución única, debido a los teoremas 11 y 12.

El lema siguiente es un resultado intermedio para ayudar a entender en qué circunstancias no tendríamos el problema de Sequeira.

Lema 21. Sean $a, b \in \mathbb{N} \setminus \{0\}$. Entonces,

$$\text{mcm}(a, b) \text{mcd}(a, b) = ab.$$

Para nuestro trabajo, la posición de cada carta tiene que ser única en toda la baraja. Es decir, queremos que las secuencias de valores y palos solo recomiencen simultáneamente en el final de la baraja. O sea, hace falta que el mínimo común múltiplo del número de elementos en cada secuencia sea igual al número de cartas de la baraja. Por el lema 21, esto solo pasa si

$$\text{mcm}(k, p) = kp \iff \text{mcd}(k, p) = 1.$$

En resumen, para que sepamos cómo ordenar una baraja como Si Stebbins o Gaspar Cardozo de Sequeira, presentamos el último teorema cuya demostración se deducirá de todo lo que hemos visto hasta ahora.

Teorema 22 (teorema de ordenación de barajas). Sean $k, p, d, d' \in \mathbb{N} \setminus \{0, 1\}$. Consideremos una baraja con p palos distintos de k valores distintos cada uno (o sea, de kp cartas distintas).

- Podemos ordenar los valores en progresión aritmética de diferencia d cuando $\text{mcd}(k, d) = 1$, y
- podemos ordenar los palos en progresión aritmética de diferencia d' cuando $\text{mcd}(p, d') = 1$.

La secuencia de cartas recorrerá la baraja completamente cuando $\text{mcd}(k, p) = 1$.

Observación 23 (nota personal). Las características y dificultades de la ordenación de Sequeira son importantes hoy en día. La baraja española viene en 40 (10 valores por palo) o 48 cartas (12 valores por palo). Si consideramos una baraja con 4 palos y nuestra regla es solamente alternar el palo a cada carta como en la *stack* de Stebbins, tendremos el mismo problema. En caso de que consideremos 10 valores por palo, tenemos que $\text{mcd}(4, 10) = 2$ y, por lo tanto, $\text{mcm}(4, 10) = 20 < 40$. Y al considerar 12 valores por palo, volvemos al problema de Sequeira. ◀

Con el último teorema, el teorema de ordenación de barajas, podemos generalizar las preparaciones al respecto del número de cartas, periodo entre cartas y palos, valores a ordenar... ¡Podéis crear incluso las vuestras! Además, este truco puede hacerse con cualquier baraja de cartas. Solo hay que asignar valores adecuadamente a las cartas.

Observación 24. Generalizando el teorema del resto chino (teorema 17) de forma semejante al teorema 18, se pueden incluso contemplar barajas con más que las dos características de valor y palo. Adaptando el teorema de ordenación de barajas, se pueden encontrar más criterios para ordenar estas nuevas barajas. ◀

5. Conclusiones

En conclusión, es importante destacar que aún faltaban casi 200 años para que Gauss publicara su obra *Disquisitiones Arithmeticae* donde se empezó a estudiar con rigor los fundamentos de la teoría de números, la cual explica todo lo que hemos visto en este artículo. Sin embargo, el tema ya era utilizado comúnmente para maravillar a otros.

En este artículo, al tener el ejemplo de dos *stacks*, hemos podido utilizar su «mecanismo base» matemático para entenderlas mejor e incluso generalizarlas. Lo bueno de las matemáticas es que cada teorema, proposición o creación de base matemática trae consigo nuevas cuestiones, nuevos retos: la búsqueda del conocimiento se presenta fructífera.

Nuestro teorema central nos permite tomar cualquier baraja existente y, asignando valores a las cartas, ordenarlas utilizando cada carta en una secuencia única. Esto nos servirá comenzando por poder estudiar y comprender ordenaciones existentes y sus trucos asociados hasta crear nuevas ordenaciones, pensar en nuevas barajas, adaptar trucos existentes e incluso inventar nuevos.

Pero también nos enseña lo que necesitamos para poder dibujar estrellas simples. Esto puede ser útil para cuando se está aburrido. La imaginación no tiene límites, y todo se debe a ver secuencias desde un punto de vista más elevado.

Esta es la importancia de la matemática recreativa: sencillamente, del mismo modo puede motivar a alguien a ser un matemático (es suficiente pensar en el efecto de la columna editorial de Martin Gardner, *Mathematical Games*, en el siglo xx), como puede ayudarnos a visualizar algo que no nos parece trivial, o incluso puede ocultar el próximo teorema a revolucionar el mundo.

Al final, solo nos queda jugar y aprender.

Referencias

- [1] ALBUQUERQUE PICADO, FRANCISCO. *Formas Quadráticas e Testes de Primalidade em "Disquisitiones Arithmeticae"*. Tese de mestrado. Universidade de Lisboa, 2021. URL: <http://hdl.handle.net/10451/48855>.
- [2] ALMEIDA, Paulo J. y NAPP, Diego. *Criptografia e Segurança*. Engebook. Oporto, PT: Publindústria, Edições Técnicas, 2017. ISBN: 978-989-723-210-7.
- [3] BRIDGER, Darren. *The history of magic and the mind*. 10 de abr. de 2010. URL: <http://www.darrenbridger.net/articles/the-history-of-magic-and-the-mind/>.
- [4] DYMENT, Doug. «An Introduction to Full-Deck Stacks». En: *The Deceptionary*. URL: <http://www.deceptionary.com/aboutstacks.html> (visitado 12-01-2020).
- [5] «Gaspar Cardoso de Sequeira». En: *Wikipedia*. URL: https://pt.wikipedia.org/wiki/Gaspar_Cardoso_de_Sequeira (visitado 20-01-2020).
- [6] SANTOS, José Carlos. «A Ilusão Portuguesa». En: *Gazeta de Matemática* 158 (2009), págs. 30-31. URL: <http://gazeta.spm.pt/getArtigo?gid=248> (visitado 25-01-2020).
- [7] SEQUEIRA, Gaspar Cardozo de. *Thesouro de Prudentes*. Coimbra, PT: Na impressão da viuva de Manoel de Carvalho, 1664. URL: https://digitalis.uc.pt/pt-pt/fundo_antigo/thesouro_de_prudentes (visitado 20-02-2020).
- [8] «Si Stebbins». En: *Magicpedia*. URL: http://www.geniimagazine.com/wiki/index.php/Si_Stebbins (visitado 28-01-2020).
- [9] SILVA, Alexandre; FREITAS, Pedro; SILVA, Jorge Nuno, y HIRTH, Tiago. *Matemagia*. Con pról. de Viana, José Paulo. 2.ª ed. Lisboa, PT: Associação Ludus, 2017. ISBN: 978-989-99506-3-4.
- [10] STEBBINS, Si (pseud.) *Card Tricks and the Way they are Performed*. Ca. 1898. URL: <http://www.deceptionary.com/ftp/SStebbins.pdf> (visitado 28-01-2020).

TEMat, volumen 6. Mayo de 2022.

e-ISSN: 2530-9633



Publicado con la colaboración de la
Real Sociedad Matemática Española

© 2022 Asociación Nacional de Estudiantes de Matemáticas.

© 2022 los autores de los artículos.

©  Salvo que se indique lo contrario, el contenido está disponible bajo una licencia Creative Commons Reconocimiento 4.0 Internacional.