

TEMat

divulgación de trabajos de estudiantes de matemáticas

volumen 1
julio de 2017

<http://temat.anemat.com/volumen/2017/>

Una iniciativa de la
Asociación Nacional de Estudiantes de Matemáticas



Información editorial

Editor jefe

Isaac Sánchez Barrera, Barcelona Supercomputing Center (BSC)

Editor asociado

Alberto Espuny Díaz, University of Birmingham

Editores

Javier Martínez Perales, BCAM - Basque Center for Applied Mathematics
Garazi Muguruza Lasa, Universidad Complutense de Madrid

Comité editorial

Pablo Manuel Berná Larrosa, Universidad de Murcia
Francesc Gispert Sánchez, ALGANT, Universität Regensburg
David González Moro
Ander Lamaison Vidarte, Berlin Mathematical School
Alejandra Martínez Moraian, Universidad de Alicante
Víctor Manuel Ortiz Sotomayor, Universitat Politècnica de València
Eva Primo Tàrraga, Universitat de València
Juan Miguel Ribera Puchades, Universidad de La Rioja
Israel Pablo Rivera Ríos, Universidad del País Vasco / Euskal Herriko Unibertsitatea
Lucía Rotger García, Universitat de les Illes Balears

Publica

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

publicaciones@anemat.com

contacto@anemat.com

Diseño de portada: Roberto Berná Larrosa, rbernalarroso@gmail.com

TEMat, divulgación de trabajos de estudiantes de matemáticas – volumen 1 – julio de 2017
e-ISSN: 2530-9633

© 2017 Asociación Nacional de Estudiantes de Matemáticas.

© 2017 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.

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.

Revisiones externas

En este volumen han colaborado realizando revisiones externas:

Carlos Bejines López, Universidad de Navarra

Alba Delgado Calvache, Universitat Politècnica de Catalunya

Tomás Sanz-Perela, Universitat Politècnica de Catalunya y Barcelona Graduate School of Mathematics

Alejandro Suárez Hernández, Institut de Robòtica i Informàtica Industrial CSIC-UPC

Raúl C. Volpe, Universitat de València

Índice general

Carta del presidente de la ANEM	VII
Carta del editor jefe	IX
«Robótica topológica», de David Mosquera Lois	1
«La integral de Henstock-Kurzweil y el segundo teorema fundamental del cálculo», de Javier Martínez Perales	15
«Progresiones aritméticas de colores», de Alberto Espuny Díaz	25
«El teorema de Müntz-Szász sobre la aproximación de funciones continuas», de Daniel Eceizabarrena, Alejandro Mas Mas, Francisco Mengual Bretón y María Soria Carro	31
«Influencia de los tamaños de clase en grupos finitos», de Víctor Manuel Ortiz Sotomayor	45
«Introducción a la lógica difusa y sus aplicaciones», de David Lobo Palacios	53
«Aproximación avariciosa en espacios de Banach», de Pablo M. Berná	69
«Problemas directos e inversos en combinatoria aditiva: las desigualdades de Plünnecke», de Alberto Espuny Díaz	79
«El efecto de Talbot: de la óptica a la ecuación de Schrödinger», de Daniel Eceizabarrena	91

Carta del presidente de la ANEM

Desde la Asociación Nacional de Estudiantes de Matemáticas estamos muy orgullosos de poder presentar este proyecto. Hace año y medio ya que la idea de *TEMat* empezó a gestarse en la Asamblea General de la ANEM de Oviedo en marzo del 2016. Desde aquel momento de ilusión por la concepción de tal idea (casi descabellada en aquel momento) los editores han hecho un grandísimo trabajo promoviendo y construyendo la revista. Por ello, antes de seguir, queremos agradecer todo su desinteresado trabajo, ya que sin ellos esta gran idea se habría quedado en solo eso, una idea.

En matemáticas, así como en cualquier otra ciencia, los investigadores necesitan recurrir a artículos publicados en revistas de élite para poder dar a conocer los resultados de sus proyectos. Asimismo, necesitan publicar para poder obtener subvenciones, sin las cuales, desgraciadamente, es imposible dedicarse a la ciencia hoy en día. Durante los últimos años, hemos podido ver cómo en España la edad media de los autores de artículos en estas revistas ha ido aumentando. Esto significa que dentro de diez o veinte años, cuando nuestros grandes investigadores se jubilen dejando sitio a los que llegamos desde la universidad, el nivel de publicaciones bajará drásticamente al no estar igual de experimentados, dejando a España en la cola de la investigación.

Desde la ANEM no queremos que esto suceda, más bien al contrario: queremos fomentar la participación de los estudiantes en este mundo de la publicación y la ciencia en general. El problema estaba en que la mayoría de revistas que existen en nuestro país (y fuera de él) no se adecúan al perfil de un estudiante, de modo que nos quedamos sin poder publicar. Es más, si un investigador no tiene publicaciones previas, difícilmente se le permite publicar nada. De este modo, vemos *TEMat* como el medio perfecto para acabar con este bucle e incentivar la participación de los estudiantes y recién graduados en las publicaciones científicas.

Así nos pusimos manos a la obra y, casi año y medio después, se ha materializado esa idea pionera, la revista de divulgación de trabajos de estudiantes de matemáticas, de nuestros compañeros, donde podemos enviar nuestros primeros artículos y podemos encontrar ideas para nuestros trabajos de fin de grado o máster. Una revista que todos podemos consultar al tener temáticas (habitualmente) relacionadas con nuestros estudios, avanzando así en los mismos, y que algún compañero que estudia lo mismo que nosotros habrá preparado. Una revista hecha por y para los estudiantes de matemáticas: *TEMat*.

Gracias a ella, universitarios de toda España vamos a poder adentrarnos en el mundo de las publicaciones, lo cual, junto con el Boletín ANEM-RSME, nos da acceso a todo tipo de noticias y artículos de nuestro interés, que nos servirán para poder mejorar como estudiantes y como futuros matemáticos.

Así que adelante. Os animamos a participar en esta revista: vuestra revista.

Andrés Mateo Piñol,
presidente de la ANEM.

Almería, julio de 2017.

Carta del editor jefe

Este ha sido un duro año de trabajo con unos resultados que se pueden considerar muy buenos. Cuando hace un año y medio salió la propuesta de montar una revista científica por y para estudiantes, no puedo negar que me pareció una idea descabellada. Pero todo cambió unos meses después, tras la celebración del XVII *Encuentro Nacional de Estudiantes de Matemáticas* en Barcelona, momento en el que la propuesta no solo me pareció interesante sino realizable. Y así ha sido, aquí está la prueba de ello, por lo que no puedo más que felicitar a quienes tuvieron la idea y agradecer que la siguieran defendiendo.

El contenido del primer número de *TEMat* no aparece de la nada, detrás de estas más de cien páginas hay muchas horas de trabajo de mucha gente: editores, comité editorial, autores, revisores externos y diseñador de la portada. Hay que agradecer, y mucho, a toda esta gente que ha decidido colaborar de forma desinteresada, invirtiendo su tiempo en un proyecto cuya única carta de presentación era, hasta ahora, su página web y la difusión que pudieran hacer el equipo de *TEMat* y la Junta de la ANEM. También hay quien ha colaborado con la revista sin saberlo, como son los autores de los paquetes de \LaTeX y programas que se han usado para montarla. Me extendería bastante más con los agradecimientos, siendo más concreto, pero no es cuestión de hacer que esta carta sea el contenido principal de la revista.

Espero que este sea el primer paso para conseguir que *TEMat* se convierta en una revista de referencia para los estudiantes de matemáticas, tanto presentes como futuros, en lengua castellana. Por el momento, me aventuro a decir que el paso ha sido en la buena dirección.

Isaac Sánchez Barrera,
editor jefe de *TEMat*.

Barcelona, julio de 2017.

TEMat

Este trabajo colaboró con una microcharla durante el XVII *Encuentro Nacional de Estudiantes de Matemáticas*, celebrado en Barcelona en julio de 2016.



Robótica topológica

✉ David Mosquera Lois
Universidade de Santiago de
Compostela
david.mosquera@rai.usc.es

Resumen: Imaginemos un conflicto bélico. Sería deseable que, dadas dos localizaciones, una nave no tripulada (dron) calculase una ruta para trasladar ayuda humanitaria entre ambos emplazamientos. Y, en caso de que alguno de los emplazamientos fuese atacado, que el dron modificase su ruta adecuadamente y de forma autónoma, sin necesidad de ser controlado por humanos. Supongamos ahora un conjunto de robots en una fábrica. Desearíamos ser capaces de programarlos para que realicen unas tareas de forma coordinada, es decir, sin colisionar entre ellos ni entorpecerse unos a otros. En un futuro próximo desearíamos ser capaces de construir carteros automáticos, a los cuales se les proporcionarían dos localizaciones: un punto de recogida y un punto de entrega, y que ellos se encargasen del reparto. Además, sería interesante que si, por algún motivo, alguna de las localizaciones variase sensiblemente, entonces la ruta o recorrido del cartero también variase sensiblemente. Con mayor grado de generalidad, el problema de planificar el movimiento de un robot autónomo consiste en proporcionarle unas tareas a realizar para que las ejecute sin intervención humana. Las situaciones presentadas previamente son ejemplos. En las siguientes páginas estudiaremos las inestabilidades que aparecen en los algoritmos planificadores de movimientos implementados en los robots usando ideas de la topología y la topología algebraica. Además, veremos que las peculiaridades en el comportamiento del sistema dependen de las propiedades homotópicas y topológicas del espacio de configuración del robot, el cual definiremos.

Palabras clave: topología algebraica, topología aplicada, robótica.

MSC2010: 55-02.

Recibido: 11 de febrero de 2017.

Aceptado: 26 de abril de 2017.

Agradecimientos: Trabajo dirigido por Enrique Macías Virgós y realizado durante el disfrute de una beca de colaboración para la iniciación a la investigación en el Departamento de Matemáticas.

Referencia: MOSQUERA LOIS, David. «Robótica topológica». En: *TEMat*, 1 (2017), págs. 1-14. ISSN: 2530-9633.
URL: <http://temat.anemat.com/articulo/2017-p1/>.

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

1. Introducción

En el último siglo ha habido una tendencia a automatizar actividades y procesos que antes eran realizados por humanos. Por ejemplo, las cadenas de producción en fábricas de coches, antaño llenas de trabajadores y ahora sustituidos por brazos robóticos. Otros ejemplos de tal fenómeno son la incorporación de pilotos automáticos en aviones o la aparición de coches autónomos, a los cuales solo hay que indicarles los lugares a los que los pasajeros desean ir y ellos calculan las rutas y conducen. Para automatizar un proceso y que un robot sea capaz de realizarlo es necesario proporcionarle un algoritmo, es decir, una sucesión finita de instrucciones que le permita realizar tal actividad sin necesidad de tomar ninguna decisión no contemplada en las instrucciones. Además, es deseable que los algoritmos sean estables, y explicaremos tal necesidad con un ejemplo. Imaginemos que estamos volando en un vuelo automatizado con un piloto automático desde Madrid al aeropuerto Charles de Gaulle de París y durante el trayecto se produce un accidente en tal aeropuerto y por ello nos desvían a otro de los aeropuertos de París, el Paris-Orly. La estabilidad del algoritmo se traduce en que, si ambos aeropuertos están próximos, entonces las rutas aéreas para llegar a uno u otro deben ser parecidas.

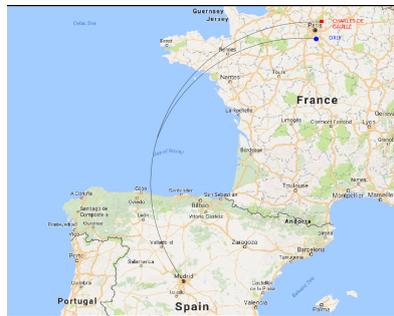


Figura 1: Mostramos las dos rutas asumiendo que el algoritmo es estable. Como los dos aeropuertos están próximos, entonces los caminos para llegar a cada uno de ellos son parecidos.

En las sucesivas páginas intentaremos formalizar matemáticamente las situaciones expuestas, introducir algunos ejemplos más y estudiar bajo qué circunstancias existen los algoritmos estables. Para tal fin utilizaremos técnicas, ideas y heurísticas de la topología.

2. Formulación matemática del problema

Sea X el conjunto de todos los estados o configuraciones posibles de un sistema mecánico (un conjunto de objetos que interactúan siguiendo las leyes de la física, como, por ejemplo, los robots de una fábrica o un avión desplazándose). Una vez dotemos a X de una topología diremos que X es el espacio de configuraciones del sistema. Enunciamos la siguiente definición un poco más restrictiva pero adecuada para trabajar en lo que sigue:

Definición 1. Un espacio de configuraciones de un sistema, o simplemente espacio de configuraciones X , es un espacio topológico conexo por caminos. ◀

Ejemplo 1. Un brazo robótico consiste en un conjunto de barras unidas por articulaciones que les permiten rotar. Si las articulaciones permiten rotación plana, $X = \mathbb{S}^1 \times \dots \times \mathbb{S}^1$ (figura 2), y si permiten rotación espacial, $X = \mathbb{S}^2 \times \dots \times \mathbb{S}^2$. Nótese que permitimos auto intersecciones. El caso sin auto intersecciones es un problema muy distinto y cabe entenderlo como un caso particular del ejemplo 2. ◀

Ejemplo 2 (el espacio de configuraciones «usual»). Sea Y un espacio topológico y sea $X = F(Y, n)$ el subconjunto del producto cartesiano $Y \times \dots \times Y$ (n veces) que contiene a las n -tuplas (y_1, y_2, \dots, y_n) tales que $y_i \neq y_j$ si $i \neq j$. El espacio X representa el sistema de configuraciones de n objetos moviéndose en el espacio Y de tal forma que no se produzcan colisiones. Los casos más relevantes son el caso $Y = \mathbb{S}^m$ (los brazos robóticos en los que no se permiten auto intersecciones), el caso en que Y es un grafo (por ejemplo,

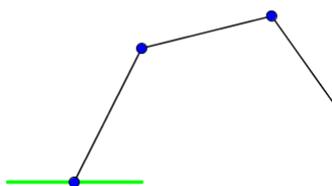


Figura 2: Representación de brazo articulado plano con tres articulaciones y tres barras.

un conjunto de vehículos desplazándose por carreteras), o cuando $Y = \mathbb{R}^m$ (vemos un caso particular en el ejemplo 3).

Ejemplo 3. Un caso particular del ejemplo 2 es el problema de evasión de asteroides [8, capítulo 8, sección 1.4.1, pág. 364], en el cual un objeto en dos o tres dimensiones ha de evitar colisionar con otros objetos móviles.

Recordamos que un grafo finito G es una terna (V, E, δ) formada por

- un conjunto finito $V \neq \emptyset$, llamado conjunto de vértices,
- un conjunto finito E , llamado conjunto de aristas, y
- una aplicación $\delta: E \rightarrow V \times V$, que asocia a cada arista sus extremos.

Ejemplo 4. Otro caso particular del ejemplo 2 es el problema del robot repartidor, el cual describimos a continuación. Imaginemos un conjunto de ciudades, las cuales numeramos, y una empresa de recogida y entrega de paquetes. La empresa dispone de rutas comerciales entre las ciudades. Representamos las ciudades y rutas en un grafo, correspondiendo las primeras a los vértices y las segundas, a las aristas.

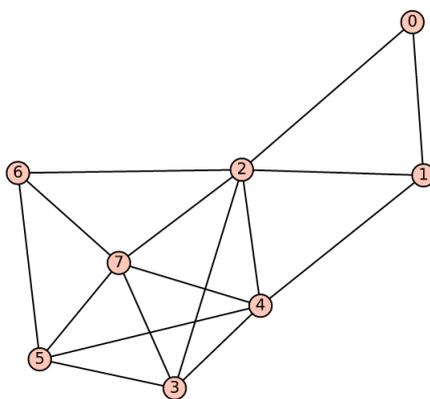


Figura 3: Grafo asociado al problema del robot repartidor.

El robot cartero de la empresa recibe dos localizaciones, un punto de recogida y un punto de entrega, y el algoritmo que tiene implementado debe encontrar un camino siguiendo las rutas comerciales. En este caso, el espacio de configuraciones es el grafo asociado a las ciudades y rutas comerciales. Y la estabilidad del algoritmo se traducirá en que si hay dos pares de ciudades donde se recoge y entrega un paquete que «no distan mucho», entonces las rutas comerciales seguidas han de ser parecidas. La idea de distancia entre pares de ciudades podremos formalizarla usando topología.

Ejemplo 5. El ejemplo introductorio del avión y las travesías aéreas también puede verse como un caso particular del ejemplo 2 donde $X = \mathbb{S}^2$, o bien siendo X un casquete esférico; por ejemplo, la región del mapa terrestre visible en la ilustración 1.

Ejemplo 6. Imaginemos una barra rotando sobre su punto medio en el espacio \mathbb{R}^{m+1} . Entonces el espacio de configuraciones se corresponde con el espacio proyectivo real $\mathbb{R}P^m$.

Hasta ahora hemos visto que el problema de planificación de movimientos en el espacio de configuraciones X consiste en diseñar un algoritmo estable que acepte como *input* dos estados o configuraciones del sistema $(A, B) \in X \times X$ y devuelva un camino continuo en X que comience en el estado inicial A y termine en la configuración final B . Además, es pertinente preguntarse si podremos construir el algoritmo, pues, por definición, el espacio de configuraciones X es conexo por caminos. Con el objetivo de dar una respuesta a tal problema introducimos algunas definiciones.

Definición 2. Sean X e Y espacios topológicos. Se denota por $C(X, Y)$ el conjunto de aplicaciones continuas de X a Y . Si C es un subconjunto compacto de X y U es un subconjunto abierto de Y , definimos

$$[C, U] := \{f \in C(X, Y) : f(C) \subset U\}.$$

Los conjuntos $[C, U]$ forman una subbase de una topología en $C(X, Y)$ a la que llamamos topología compacto-abierto. ◀

Observación 1. La motivación de la definición anterior es construir una forma de medir la proximidad de caminos (por ejemplo, rutas aéreas en el ejemplo introductorio o rutas comerciales en el caso del repartidor del ejemplo 4). ◀

Definición 3. Denotamos por PX el espacio de todas las aplicaciones continuas de $I = [0, 1]$ en X con la topología compacto-abierto. Nos referiremos a este espacio como el espacio de caminos PX . Denotamos por $\pi : PX \rightarrow X \times X$ la aplicación que asocia a cada camino $\gamma \in PX$ sus puntos inicial y final: $\pi(\gamma) = (\gamma(0), \gamma(1))$. ◀

Observación 2. Pensando la definición previa en el contexto del repartidor estamos diciendo que, dada una ruta, la aplicación nos devuelve las localizaciones de recogida y entrega de los paquetes. ◀

Dados dos caminos «parecidos», sus extremos no pueden variar mucho, es decir, la aplicación π es continua. Para probarlo comenzaremos con un lema (cuya prueba puede consultarse en el libro de Lee [9, teorema 3.27]) que simplificará la demostración:

Lema 1. Dado el siguiente diagrama conmutativo, donde $X_1 \times \dots \times X_n$ está dotado de la topología producto y $\pi_i : X_1 \times \dots \times X_n \rightarrow X_i$ es la proyección sobre X_i ,

$$\begin{array}{ccc} & X_1 \times \dots \times X_n & \\ & \nearrow f & \downarrow \pi_i \\ Y & \xrightarrow{f_i} & X_i \end{array}$$

se tiene que f es continua si y solo si lo son todas las f_i .

Proposición 2. La aplicación $\pi : PX \rightarrow X \times X$, $\pi(\gamma) = (\gamma(0), \gamma(1))$, es continua.

Demostración. En virtud del lema anterior es suficiente comprobar que $\tilde{\pi}_1 : PX \rightarrow X$, $\tilde{\pi}_1(\gamma) = \gamma(0)$, y $\tilde{\pi}_2 : PX \rightarrow X$, $\tilde{\pi}_2(\gamma) = \gamma(1)$, son continuas. Comenzamos por la continuidad de la primera aplicación. Sea U un abierto en X y veamos que $\tilde{\pi}_1^{-1}(U)$ es abierto en PX (con la topología compacto-abierto). Se tiene que

$$\tilde{\pi}_1^{-1}(U) = \{\gamma \in PX : \gamma(0) \in U\} = [\{0\}, U]$$

es un abierto subbásico pues $\{0\} \subset [0, 1]$ es compacto. La continuidad de la segunda aplicación se argumenta de forma análoga. ■

En términos de las definiciones precedentes, formalizamos la idea de algoritmo de planificación de movimientos como una aplicación continua $s : X \times X \rightarrow PX$ tal que $\pi \circ s = \text{id}_{X \times X}$. Por tanto, el problema de planificación de movimientos consiste en encontrar una sección continua de π . La continuidad de la sección se traducirá en estabilidad del algoritmo. Nuestro objetivo en lo que sigue será determinar bajo qué condiciones existirá una sección continua, es decir, un algoritmo planificador de movimientos estable. Veamos algunos ejemplos más antes de abordar el estudio del caso general.

Ejemplo 7. Sea $X = \mathbb{S}^1$. Sean A y B los estados inicial y final, respectivamente, $A = e^{i\theta_0}$, $B = e^{i\theta_1}$, $0 \leq \theta_0, \theta_1 \leq 2\pi$. Definamos un par de algoritmos de planificación de movimientos como sigue:

Algoritmo 1.

Consideremos la sección $s: X \times X \rightarrow PX$ dada por $s(A, B)(t) = e^{i[t\cdot\theta_1 + (1-t)\cdot\theta_0]}$. Visualmente, el algoritmo funciona como sigue:

- Si $\theta_1 > \theta_0$, entonces nos movemos a velocidad constante en sentido contrario a las agujas del reloj de A hasta B .
- Si $\theta_1 = \theta_0$, entonces no nos movemos.
- Si $\theta_1 < \theta_0$, entonces nos movemos en el sentido horario a velocidad constante de B hasta A .

Algoritmo 2.

El segundo algoritmo lo damos en forma de pseudocódigo:

- Si $\theta_1 \geq \theta_0$, entonces procedemos como en el algoritmo 1.
- En otro caso ($\theta_1 < \theta_0$), dividimos el desplazamiento en dos procesos:
 1. Nos desplazamos en sentido antihorario y velocidad constante hasta $\theta = 0$.
 2. Nos desplazamos en sentido antihorario y velocidad constante desde $\theta = 0$ hasta θ_1 . ◀

Ejemplo 8. Supongamos que en el ejemplo anterior hubiese un número finito de posiciones relevantes en $X = \mathbb{S}^1$, es decir, que un robot cuyos movimientos queremos programar entrase en una rotonda (con un número finito de salidas y entradas). Si el robot entrase por una de ellas que no coincidiese con $\theta_0 = 0$, entonces el primer algoritmo sería una fuente de accidentes pues nuestro robot podría circular en sentido contrario. En este caso, el segundo algoritmo evitaría accidentes pues respetaría las normas de tráfico. ◀

Nótese que en ambos casos los caminos son continuos, pero eso no implica que lo sean los algoritmos planificadores de movimientos (las secciones). Un algoritmo planificador de movimientos será continuo si parejas iniciales y finales próximas (A, B) , (A', B') producen caminos $s(A, B)$, $s(A', B')$ próximos (con la topología compacto-abierto). Por ejemplo, en el segundo algoritmo una sensible variación de los puntos final e inicial puede ocasionar que los caminos devueltos por el algoritmo sean muy distintos (píntense $s(A, B)$ y $s(A, B')$ sobre la figura 4).

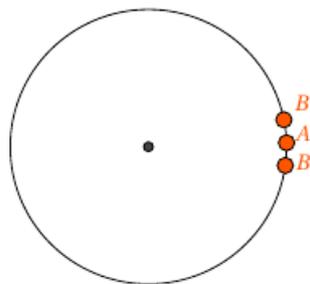


Figura 4: Circunferencia y tres puntos sobre ella, los cuales ilustran la inestabilidad del segundo algoritmo.

La importancia de la continuidad en el algoritmo de planificación de movimientos reside en que una sección no continua podría ocasionar inestabilidades en el comportamiento del robot, como acabamos de comprobar en el ejemplo anterior. Por lo tanto, nuestro siguiente objetivo será caracterizar las situaciones en las que el algoritmo planificador de movimientos será continuo. Trataremos de relacionar la continuidad de la aplicación sección con la topología del espacio de configuraciones. Veamos un ejemplo más que nos informará de lo que sucede en el caso general.

Ejemplo 9. Sea X un subespacio de \mathbb{R}^n con la topología relativa. Decimos que X es un espacio estrellado respecto a $A_0 \in X$ si para todo $A \in X$ el segmento

$$AA_0 = \{tA + (1-t)A_0 : t \in [0, 1]\}$$

está contenido en X . En un espacio X estrellado respecto a $A_0 \in X$ y dados los estados inicial y final (A, B) podríamos definir el siguiente algoritmo:

1. Recorrer el segmento desde A hasta A_0 a velocidad constante.
2. Recorrer el segmento desde A_0 hasta B a velocidad constante.

Podemos formalizar el algoritmo de manera sencilla definiendo $s: X \times X \rightarrow PX$ como sigue:

$$s(A, B) = \begin{cases} (1 - 2t)A + 2tA_0 & \text{si } 0 \leq t \leq \frac{1}{2}, \\ (2t - 1)B + (2 - 2t)A_0 & \text{si } \frac{1}{2} \leq t \leq 1. \end{cases}$$

Intuitivamente, parece que la sección $s: X \times X \rightarrow PX$ es continua. De todas formas, la demostración de este hecho es algo laboriosa y la probaremos en un teorema más general. Nótese que los espacios estrellados son contráctiles y el algoritmo que hemos propuesto utiliza la contractibilidad del espacio. Además, hemos visto que en el caso $X = \mathbb{S}^1$ no hemos encontrado ningún algoritmo continuo y \mathbb{S}^1 no es contráctil. ◀

Observación 3. Recordamos que, informalmente, un espacio es contráctil si se puede deformar de forma continua a uno de sus puntos. Formalmente, un espacio topológico X es contráctil si la aplicación identidad $\text{id}_X: X \rightarrow X$ es homótopa a una aplicación constante $c: X \rightarrow X$, $c(A) = A_0$. ◀

3. Definición de la complejidad topológica

En los últimos ejemplos que presentamos se apreciaba una relación entre la existencia de secciones continuas y la contractibilidad del espacio. Con el resultado siguiente formalizaremos esa relación. Daremos la idea de una posible demostración y omitiremos muchas comprobaciones que pueden consultarse con todo detalle en el trabajo de Mosquera Lois [10].

Teorema 3 (Farber [1]). *Dado el espacio de configuraciones X , existirá una sección continua global $s: X \times X \rightarrow PX$ de π si y solo si X es contráctil.*

Idea de la demostración. Denotemos $I = [0, 1]$.

Supongamos que existe una sección continua global: $s: X \times X \rightarrow PX$,

$$s(A, B) = \gamma: [0, 1] \rightarrow X \times X, \quad \gamma(0) = A, \quad \gamma(1) = B.$$

Y veamos que X es contráctil, es decir, que podemos construir una homotopía entre la aplicación identidad $\text{id}_X: X \rightarrow X$ y una aplicación constante $c: X \rightarrow X$, $c(A) = A_0$. Definimos

$$H: X \times I \rightarrow X, \quad H(A, t) = s(A, A_0)(t).$$

Comprobemos que H es la homotopía buscada:

1. Se puede demostrar que H es continua (ver el trabajo de Mosquera Lois [10]).
2. $H(A, 0) = A$, $H(A, 1) = A_0 = c(A) \quad \forall A \in X$, y, por tanto, se tiene que $H_0 = \text{id}_X$ y $H_1 = c$.

Recíprocamente, supongamos que X es contráctil, es decir, que existe una homotopía $H: X \times I \rightarrow X$ tal que $H_0 = \text{id}_X$, $H_1 = c$. Construiremos una sección continua global de $\pi: PX \rightarrow X \times X$.

Definimos $s: X \times X \rightarrow PX$,

$$s(A, B) = \gamma: I \rightarrow X, \quad \gamma(t) = \begin{cases} H(A, 2t), & t \leq 1/2, \\ H(B, 2 - 2t), & t \geq 1/2. \end{cases}$$

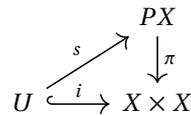
Se tiene que $s: X \times X \rightarrow PX$ es una sección de π en el sentido conjuntista, pues $\pi \circ s = \text{id}_{X \times X}$.

Resta comprobar que $s: X \times X \rightarrow PX$ es continua (de nuevo, ver el trabajo de Mosquera Lois [10]). En consecuencia, hemos encontrado una sección continua global de π . ■

Observación 4. Nótese que la demostración nos proporciona un método para construir las secciones continuas globales sobre espacios contráctiles. Además, una vez vista la demostración del caso general queda probada la continuidad de $s: X \times X \rightarrow PX$ en el ejemplo 9. ◀

Como hemos expuesto, la condición suficiente de la demostración garantiza que si el espacio es contráctil, entonces nuestro problema está resuelto. La condición necesaria nos motiva a plantearnos la siguiente cuestión: en caso de que el espacio no fuese contráctil, ¿podríamos considerar un recubrimiento de $X \times X$ de forma que sobre cada miembro del recubrimiento exista una sección continua? Para responder a esa pregunta definimos un concepto a medida que introducimos a continuación.

Definición 4. Dado un espacio de configuraciones X , llamamos abierto de Farber a un conjunto abierto $U \subset X \times X$ tal que exista una sección continua $s: U \rightarrow PX$ de π sobre U , es decir, $\pi \circ s = i$ donde $i: U \rightarrow X \times X$ es la inclusión. Equivalentemente, que el siguiente diagrama sea conmutativo:



Definición 5. Dado un espacio de configuraciones X , definimos la complejidad topológica de la planificación de movimientos en X como el menor número entero positivo $TC(X) = k$ de abiertos de Farber necesarios para recubrir $X \times X$. Si no existe ningún entero positivo k en las condiciones expuestas, entonces definimos $TC(X) = \infty$. ◀

Dado un recubrimiento por abiertos de Farber

$$X \times X = U_1 \cup U_2 \cup \dots \cup U_k$$

y secciones continuas $s_i: U_i \rightarrow PX, i \in \{1, 2, \dots, k\}$, sobre cada uno de los conjuntos del recubrimiento, podremos construir un algoritmo planificador de movimientos definido en todo el espacio de configuraciones como sigue. Dados dos estados (A, B) buscamos el abierto de Farber U_i con el menor índice i tal que $(A, B) \in U_i$ y obtenemos un camino $s_i(A, B)$. Nótese que con esta construcción no podemos garantizar la estabilidad del algoritmo. Supongamos dos pares de estados (A, B) y (A', B') arbitrariamente próximos de forma que $(A, B) \in U_1$ y $(A', B') \in U_2$ (véase la figura 5). Entonces, los caminos $s_1(A, B)$ y $s_2(A', B')$ podrán ser arbitrariamente distintos pues las secciones $s_1|_{U_1 \cap U_2}$ y $s_2|_{U_1 \cap U_2}$ son en general distintas. Es decir, en cierto modo, la complejidad topológica de un espacio mide el grado de inestabilidad de los algoritmos de planificación de movimientos (más estables) definidos en tal espacio. Si la complejidad es uno, entonces existe un algoritmo de planificación de movimientos estable definido en el espacio, en otro caso no existe ningún algoritmo de planificación de movimientos estable definido en todo el espacio. Y a mayor complejidad, más inestables serán los algoritmos.

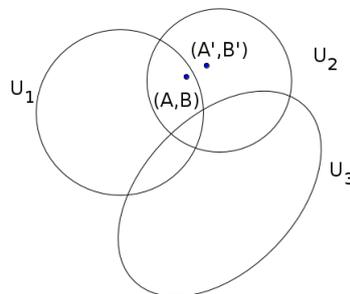


Figura 5: Discontinuidad de un algoritmo planificador de movimientos correspondiente a la familia de abiertos de Farber $\{U_i\}_{i=1}^k$.

Ejemplo 10. Si X es un espacio estrellado, por ejemplo como en el ejemplo 9, o en particular en el caso de espacios convexos, entonces $TC(X) = 1$. ◀

Teorema 4 (Farber [1]). Sea $X = \mathbb{S}^1$, entonces $TC(X) = 2$.

Demostración. La circunferencia no es contráctil; por tanto, $TC(\mathbb{S}^1) > 1$. Encontraremos un recubrimiento por abiertos de Farber $\{U_1, U_2\}$ con sus respectivas secciones continuas $s_i: U_i \rightarrow PX$. Sea $U_1 \subset \mathbb{S}^1 \times \mathbb{S}^1$ el conjunto de pares $\{(A, B): A \neq -B\}$ y sea $U_2 \subset \mathbb{S}^1 \times \mathbb{S}^1$ el conjunto de pares $\{(A, B): A \neq B\}$. Véanse las figuras 6 y 7. En la ilustración 8 mostramos sobre el toro «usual» los diagramas anteriores. Las curvas cerradas azules (cuya intersección es vacía) corresponden a los conjuntos complementarios a U_1 y U_2 . Por tanto, la unión de U_1 y U_2 cubre el toro.

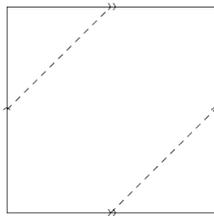


Figura 6: El abierto $U_1 \subset X \times X$.

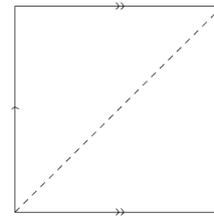


Figura 7: El abierto $U_2 \subset X \times X$.

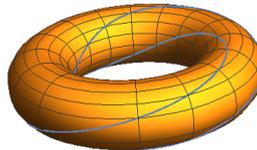


Figura 8: El recubrimiento en el toro «usual».

Se tiene que ambos conjuntos U_1 y U_2 son abiertos. Definimos la aplicación $s_1: U_1 \rightarrow PX$ que lleva A a B con velocidad constante a lo largo del arco más corto que conecta ambos puntos (que es único). Nótese que no es posible extender esta sección de forma continua a los pares de puntos antipodales, pues no habría unicidad del arco más corto conectando ambos puntos y en el momento que eligiésemos un arco en lugar de otro perderíamos la continuidad de la sección. Para definir $s_2: U_2 \rightarrow PX$ fijamos una orientación en la circunferencia, por ejemplo, el sentido de giro antihorario. Entonces, la aplicación llevará A a B a velocidad constante en el sentido de giro antihorario. De nuevo, no es posible extender la sección de forma continua al toro. ■

Observación 5. Nótese que la familia de abiertos formada por $U_1 = U \times U$ y $U_2 = V \times V$, donde $U = \mathbb{S}^1 - \{1\}$ y $V = \mathbb{S}^1 - \{-1\}$, no es de Farber para \mathbb{S}^1 (aunque son contráctiles) pues no recubre el toro, ya que quedan dos puntos sin cubrir. Véase la figura 9.

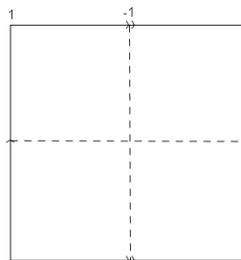


Figura 9: Ejemplo de familia que no es un recubrimiento.

Observación 6. Más adelante veremos cómo calcular la complejidad topológica de la circunferencia sin necesidad de dar algoritmos planificadores de movimientos.

4. Cálculo de la complejidad topológica

A continuación enunciaremos una propiedad fundamental de la complejidad topológica: su invarianza homotópica. Tal propiedad sugiere que las técnicas propias de la topología algebraica son adecuadas para abordar el estudio de la TC. También enunciaremos algunos resultados sobre cotas para la complejidad topológica. No nos pararemos demasiado en las demostraciones ni entraremos en explicaciones muy profundas. De nuevo remitimos al lector interesado al trabajo de Mosquera Lois [10].

Definición 6. Sean X e Y espacios topológicos. Decimos que X e Y son espacios equivalentes por homotopía o que tienen el mismo tipo de homotopía si existen aplicaciones continuas $f: X \rightarrow Y$, $g: Y \rightarrow X$ tales que $f \circ g \simeq \text{id}_Y$ y $g \circ f \simeq \text{id}_X$. ◀

Teorema 5 (Farber [1]). Sean X e Y espacios con el mismo tipo de homotopía. Entonces, $TC(X) = TC(Y)$.

Ejemplo 11. Sea $X = \mathbb{R}P^1$. Puesto que X es homeomorfo a S^1 , entonces $TC(X) = 2$. ◀

El cálculo de la complejidad topológica de los espacios proyectivos reales es un problema abierto y está relacionado con algunos problemas aparentemente muy distintos, tal y como pone de manifiesto el siguiente resultado de Farber, Tabachnikov y Yuzvinsky [5].

Teorema 6 (Farber, Tabachnikov y Yuzvinsky [5]). Si $n \notin \{1, 3, 7\}$, entonces $TC(\mathbb{R}P^n)$ coincide con el menor natural k tal que existe una inmersión del espacio proyectivo $\mathbb{R}P^n$ en \mathbb{R}^{k-1} .

4.1. Cotas superiores

A continuación introducimos los CW-complejos, unos espacios topológicos construidos por bloques. Nos centraremos en tales espacios por dos motivos: primero, porque es viable obtener cotas para la complejidad, y segundo, porque son lo bastante generales para cubrir la mayoría de los espacios que nos interesan en las aplicaciones. Nuestra exposición será breve; para más información sobre los CW-complejos, consúltense los libros de Lee [9] o Hatcher [7].

Definición 7. Llamaremos n -celda abierta e a un espacio topológico homeomorfo a la bola abierta unitaria B^n , y llamaremos n -celda cerrada \bar{e} a un espacio topológico homeomorfo a la bola cerrada unitaria \bar{B}^n . ◀

Definición 8. Dado un espacio topológico no vacío X , sean \bar{e} una n -celda cerrada con $n \geq 1$ y $\varphi: \partial\bar{e} \rightarrow X$ una aplicación continua. Definimos el espacio de adjunción $X \cup_{\varphi} \bar{e}$ obtenido de $X \sqcup \bar{e}$ al identificar $\partial\bar{e}$ con $\varphi(\partial\bar{e})$, y diremos que cualquier espacio homeomorfo a este se obtiene de X al añadir una n -celda a X mediante φ . ◀

Definición 9. Sea $X_0 \subset X_1 \subset \dots \subset X_{n-1} \subset X_n$ una sucesión finita de espacios topológicos satisfaciendo las dos condiciones siguientes:

1. X_0 es un espacio discreto no vacío, y
2. para cada $i \geq 1$, X_i se obtiene a partir de X_{i-1} adjuntando una colección finita (que puede ser vacía) de i -celdas.

Entonces, decimos que el conjunto $X = \bigcup_i X_i$ dotado de la topología cociente es un CW-complejo finito. ◀

Ejemplo 12. Partimos de dos puntos. En el primer paso añadimos 1-celdas (intervalos). En el segundo paso añadimos 2-celdas para obtener un cilindro. Véase la figura 10. ◀



Figura 10: Construcción de un CW-complejo finito.

Muchos espacios topológicos se pueden dotar de estructuras celulares, como, por ejemplo, las esferas, los grafos, los espacios proyectivos, las variedades diferenciables compactas y las superficies compactas.

Ejemplo 13. Un grafo finito es un CW-complejo finito con celdas de dimensiones cero y uno. Las celdas de dimensión cero son los vértices y las celdas de dimensión uno son las aristas. Además, el grafo es conexo si lo es como espacio topológico. ◀

Ejemplo 14. La esfera S^n puede realizarse como un CW-complejo finito con una celda de dimensión cero y una celda de dimensión n . La aplicación característica (de adjunción) para la celda de dimensión n es $q: \overline{\mathbb{B}}^n \rightarrow S^n$, que envía $\partial\overline{\mathbb{B}}^n$ a un punto. ◀

Ejemplo 15 (Hatcher [7]). Sea X una superficie compacta orientable de género g . Entonces X puede construirse como un CW-complejo finito usando una celda de dimensión cero, $2g$ celdas de dimensión uno y una celda de dimensión dos. ◀

Introducimos una proposición que nos hará falta más adelante y que proporcionará información sobre los productos de CW-complejos finitos.

Proposición 7. Sean X, Y dos CW-complejos finitos. Entonces $X \times Y$ admite una estructura de CW-complejo finito.

La demostración del resultado es constructiva y muestra que las celdas del nuevo espacio serán los productos de celdas de X con celdas de Y . Es decir, si las celdas de X e Y son $\{e_j^k\}$ y $\{e_i^l\}$, respectivamente, donde los superíndices indican la dimensión y los subíndices indexan cada una de las celdas en esa dimensión, entonces las celdas de $X \times Y$ serán $\{e_j^k \times e_i^l\}$. Usaremos este hecho más adelante.

Ejemplo 16. Introducimos una ilustración del resultado anterior para $X = S^1$, la circunferencia, e $Y = B^2$, la bola cerrada de dimensión dos. El producto es el toro sólido, al cual podemos dotar de una estructura celular dada por la proposición 7. Ilustramos la estructura celular en la figura 11. ◀

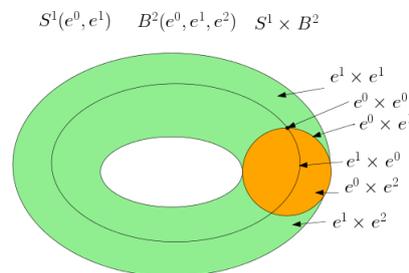


Figura 11: Estructura celular en el toro sólido obtenida como producto de estructuras celulares.

El siguiente resultado nos permite calcular cotas superiores para la complejidad topológica de CW-complejos.

Teorema 8 (Mosquera Lois [10]). Sea X un CW-complejo finito y conexo por caminos. Entonces,

$$TC(X) \leq l(X \times X),$$

donde $l(X \times X)$ representa el número de dimensiones en las que $X \times X$ tiene celdas.

4.2. Cotas inferiores

El resultado anterior nos proporciona cotas superiores para la complejidad. Las cotas inferiores se obtienen usando varias técnicas, siendo la más notable la teoría de la cohomología, la cual consiste en asociar objetos algebraicos (grupos, anillos, módulos o álgebras, por ejemplo) a espacios topológicos de forma que los primeros permitan obtener información sobre los segundos. Véase el trabajo de Mosquera Lois [10] para una exposición detallada de los resultados.

5. Ejemplos y cálculos

A continuación exponemos algunos resultados relativos a la complejidad topológica de algunos espacios y la relacionaremos con los ejemplos iniciales.

5.1. Grafos

Recordamos que un ciclo en un grafo G es la imagen de una aplicación continua $\alpha: [0, 1] \rightarrow G$ inyectiva en $(0, 1)$ y tal que $\alpha(0) = \alpha(1) = v$, siendo v un vértice del grafo. Un grafo se dice que es un árbol si es conexo y no tiene ciclos.

Teorema 9 (Farber [3]). *Sea X un grafo finito y conexo con un número finito de ciclos. Se tiene que*

- *si X es un árbol, entonces $TC(X) = 1$;*
- *si X contiene exactamente un ciclo, entonces $TC(X) = 2$, y*
- *si X contiene más de un ciclo, entonces $TC(X) = 3$.*

Demostración. En virtud del ejemplo 13 sabemos que la estructura CW-complejo del grafo solo contiene celdas de dimensiones uno y cero. Por lo tanto, $l(X \times X) = 3$ y el teorema 8 nos garantiza que $TC(X) \leq 3$. Si X es un árbol, entonces por definición no contiene ningún ciclo y, en consecuencia, es contráctil. El teorema 3 garantiza entonces que $TC(X) = 1$. Si X contiene exactamente un ciclo, entonces es equivalente por homotopía a la circunferencia y la invarianza por homotopía de la complejidad topológica garantiza que $TC(X) = 2$. Si X contiene más de un ciclo, entonces la teoría de cohomología nos permite obtener la cota $TC(X) \geq 3$. ■

Este resultado nos informa acerca de la existencia de algoritmos planificadores de movimientos en el ejemplo 4.

5.2. Esferas

Lema 10. *Sea $X = \mathbb{S}^n$. Entonces $2 \leq TC(X) \leq 3$.*

Demostración. En virtud del ejemplo 14 sabemos que la estructura de CW-complejo de la esfera solo contiene celdas de dimensiones cero y n . Por lo tanto, el teorema 8 nos garantiza que $TC(X) \leq 3$. El resto se deduce del teorema 3 y que la esfera \mathbb{S}^n no es contráctil. ■

Y con un poco más de trabajo (véanse las obras de Farber [1] o Mosquera Lois [10] para una demostración alternativa) se tiene el siguiente teorema.

Teorema 11 (Farber [1]). *Sea $X = \mathbb{S}^n$. Entonces,*

- *si n es impar, $TC(X) = 2$, y*
- *si n es par, $TC(X) = 3$.*

5.3. Brazos articulados

En esta sección volvemos sobre el ejemplo 1 y calcularemos la complejidad topológica del problema de planificación de movimientos de un brazo robótico con n articulaciones y m grados de libertad de movimiento en cada articulación.

Teorema 12 (Farber [1]). *Sea $X = \mathbb{S}^m \times \cdots \times \mathbb{S}^m$, donde $m \geq 1$. Entonces,*

$$TC(X) = \begin{cases} n + 1 & \text{si } m \text{ es impar,} \\ 2n + 1 & \text{si } m \text{ es par.} \end{cases}$$

Demostración. Comenzaremos probando que la complejidad topológica de X está acotada superiormente por $n + 1$ en el caso de que las esferas sean de dimensión impar y por $2n + 1$ en el caso de que las esferas sean de dimensión par. De la desigualdad del producto para la complejidad topológica (teorema 8) y el cálculo de la complejidad topológica para \mathbb{S}^m se deduce lo que queremos probar. La desigualdad en el otro sentido se obtiene con técnicas cohomológicas. ■

También podemos calcular la complejidad topológica de brazos robóticos en los que no permitimos auto intersecciones (lo cual se aproxima más a la realidad). La clave está en el siguiente lema:

Lema 13. *El espacio de configuraciones de un brazo robótico con 2 articulaciones y $m \geq 1$ grados de libertad de movimiento en cada articulación en el que no permitimos auto intersecciones presenta el tipo de homotopía de una esfera de dimensión m . Más formalmente:*

$$X = \{(A, B) \in \mathbb{S}^m \times \mathbb{S}^m : A \neq B\}$$

tiene el tipo de homotopía de \mathbb{S}^m .

Demostración. Definimos $f: \mathbb{S}^m \rightarrow X$, donde $f(A) = (A, -A)$, y $g: X \rightarrow \mathbb{S}^m$, donde $g(A, B) = A$. Se tiene que $g \circ f = \text{id}_{\mathbb{S}^m}$ y que $(f \circ g)(A, B) = (A, -A)$. Construiremos una homotopía entre $f \circ g$ y la identidad de X . Puesto que $B \neq A$, existe una única geodésica entre $-A$ y B . Denotemos tal geodésica por $\alpha: [0, 1] \rightarrow \mathbb{S}^m$. Definimos ahora la homotopía $H: X \times [0, 1] \rightarrow X$ donde $H(A, B, t) = (A, \alpha(t))$. Se tiene que H es continua, pues cada una de las componentes lo son y, en consecuencia, hemos terminado. ■

Teorema 14 (Mosquera Lois [10]). *Sea X el espacio de configuraciones del brazo robótico con 2 articulaciones y $m \geq 1$ grados de libertad de movimiento en cada articulación en el que no permitimos auto intersecciones. Entonces,*

$$TC(X) = \begin{cases} 2 & \text{si } m \text{ es impar,} \\ 3 & \text{si } m \text{ es par.} \end{cases}$$

Demostración. Se sigue de concatenar el lema 13 y los resultados previos sobre la complejidad topológica de las esferas. ■

Teorema 15 (Mosquera Lois [10]). *Sea X el espacio de configuraciones del brazo robótico con $2n$, $n \geq 1$, articulaciones y $m \geq 1$ grados de libertad de movimiento en cada articulación en el que no permitimos auto intersecciones. Entonces,*

$$TC(X) = \begin{cases} n + 1 & \text{si } m \text{ es impar,} \\ 2n + 1 & \text{si } m \text{ es par.} \end{cases}$$

Demostración. Aplicamos el lema 13 y obtenemos que nuestro espacio es equivalente por homotopía a $X = \mathbb{S}^m \times \dots \times \mathbb{S}^m$. En virtud del teorema 12, hemos terminado. ■

5.4. Gimbal-lock y $SO(3)$

Consideremos $SO(3)$, el grupo de rotaciones en \mathbb{R}^3 . Más formalmente y con mayor grado de generalidad:

$$SO(n) = \{A \in M_{n,n}(\mathbb{R}) : AA^T = I, \det(A) = 1\}.$$

Se tiene que $SO(n)$ es un espacio topológico si lo dotamos de la topología relativa que hereda de \mathbb{R}^{n^2} . Es además un grupo topológico, una variedad diferenciable y un grupo de Lie. La importancia de $SO(3)$ reside en sus aplicaciones en física y en el diseño de algoritmos. Es homeomorfo a \mathbb{RP}^3 .

Teorema 16 (Farber [2]). *Sea $X = SO(3)$. Entonces, $TC(X) = 4$.*

Una suspensión cardán (*gimbal* en inglés) es un mecanismo destinado a permitir la rotación de un objeto sobre un eje. Un sistema de suspensión cardán consiste en un conjunto de suspensiones cardán que interactúan entre sí. Se utiliza en los sistemas de navegación de aviones, drones y otros aparatos [6]. Su finalidad reside en controlar la actitud de tales vehículos. Nos centraremos en el sistema formado por tres suspensiones, cada una montada sobre la anterior mediante un sistema de rotación ortogonal (véase la figura 12).

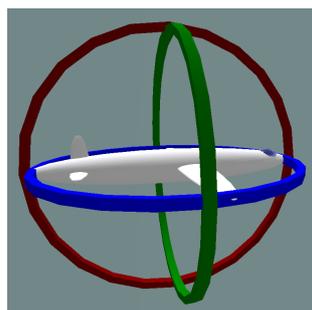


Figura 12: Sistema de navegación de un dron formado por tres *gimbals*. © ⓘ ⓘ CC BY-SA 3.0 MathsPoetry, https://commons.wikimedia.org/wiki/File:No_gimbal_lock.png.

Se denomina *gimbal-lock* a la pérdida de un grado de libertad de movimiento en un sistema de suspensiones de cardán. Tal fenómeno se produce tras una sucesión de rotaciones que culmina en que dos de los ejes de rotación estén en el mismo plano (véase la figura 13).

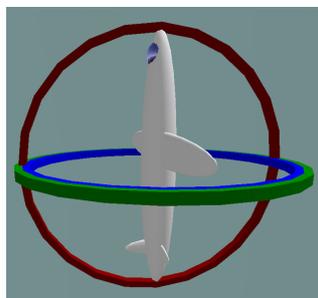


Figura 13: *Gimbal-lock*. © ⓘ ⓘ CC BY-SA 3.0 MathsPoetry, https://commons.wikimedia.org/wiki/File:Gimbal_lock.png.

El *gimbal-lock* se produce debido a que no existe un algoritmo de planificación de movimientos estable global en $SO(3, \mathbb{R})$, el espacio de rotaciones en el espacio euclidiano de dimensión tres. Un caso famoso de *gimbal-lock* se produjo en una de las misiones Apollo de la NASA [6].

6. El estado actual del tema

En los últimos diez años se han publicado muchos artículos y trabajos sobre la complejidad topológica en la robótica. Actualmente es uno de los temas que despierta interés entre investigadores en topología algebraica, geometría diferencial y topología o geometría aplicadas en general. Para una introducción asequible a la temática, con demostraciones detalladas, recomendamos el trabajo de Mosquera Lois [10]. Las referencias de Farber [1-5] son clásicas e inician el estudio de la temática; en concreto, la definición de complejidad topológica se introdujo en el año 2003 [1].

Referencias

- [1] FARBER, Michael. «Topological complexity of motion planning». En: *Discrete and Computational Geometry* 29.2 (2003), págs. 211-221. <https://doi.org/10.1007/s00454-002-0760-9>.
- [2] FARBER, Michael. «Instabilities of robot motion». En: *Topology and its Applications* 140.2-3 (2004), págs. 245-266. <https://doi.org/10.1016/j.topol.2003.07.011>.
- [3] FARBER, Michael. «Topology of robot motion planning». En: *Morse theoretic methods in nonlinear analysis and in symplectic topology*. Springer, 2006, págs. 185-230. https://doi.org/10.1007/1-4020-4266-3_05.

- [4] FARBER, Michael. *Invitation to topological robotics*. European Mathematical Society, 2008.
- [5] FARBER, Michael; TABACHNIKOV, Serge y YUZVINSKY, Sergey. «Topological robotics: motion planning in projective spaces». En: *International Mathematics Research Notices* 34 (2003), págs. 1853-1870. <https://doi.org/10.1155/S1073792803210035>.
- [6] HANSON, Andrew J. *Visualizing Quaternions*. Morgan Kaufmann, 2006.
- [7] HATCHER, Allen. *Algebraic Topology*. Cambridge University Press, 2002.
- [8] LATOMBE, Jean-Claude. *Robot motion planning*. Vol. 124. Springer, 2012.
- [9] LEE, John. *Introduction to Topological Manifolds*. 2nd ed. Vol. 940. Grad. Texts in Math. Springer, 2011.
- [10] MOSQUERA LOIS, David. *Complejidad topológica en la robótica*. Trabajo Final de Grado. Universidade de Santiago de Compostela, 2016. <https://doi.org/10.5281/zenodo.581060>.

TEMat

La integral de Henstock-Kurzweil y el segundo teorema fundamental del cálculo

✉ Javier Martínez Perales
BCAM - Basque Center for Applied
Mathematics
javicemarpe@gmail.com

Resumen: En este artículo repasamos brevemente la definición y los problemas de la integral de Riemann y definimos un concepto de integral con el cual es posible superar las deficiencias de la integral de Riemann. Esta nueva integral es la integral de Henstock-Kurzweil que, con una definición muy similar a la de la integral de Riemann, es capaz de superar tanto a esta integral como a la de Lebesgue en muchos sentidos.

Concretamente, veremos cómo, utilizando esta integral, es posible recuperar cualquier función derivable a partir de su derivada. Para ello, introduciremos el concepto de medidor y, para ilustrar su utilidad, veremos por el camino cómo los medidores nos permiten probar algunos resultados clásicos.

Palabras clave: Riemann, Henstock-Kurzweil, integral, teorema fundamental del cálculo, teorema de Cousin, medidor.

MSC2010: 26A39, 26A42.

Recibido: 3 de marzo de 2017.

Aceptado: 2 de abril de 2017.

Agradecimientos: Me gustaría agradecer a los profesores Francisco Javier Martín Reyes y José Ángel Peláez Márquez de la Universidad de Málaga por su orientación y motivación en el desarrollo de mis trabajos de fin de Grado y Máster, respectivamente, además de por su dedicación a la hora de aconsejarme y ayudarme para poder comenzar en el mundo de la investigación.

Quiero agradecer también a Ana, Antonio, Belén, Claudia, Dani (López y Amado), Guille, Horten, Iván, Jorge, José, Juanfran, Juanmi, Juli, Lorena, Lucía, Manu, Natalia, Toni y Valentina, además de a mis hermanos y abuelos por su amistad, por estar conmigo y por ser mi soporte.

En especial, quiero dedicar este pequeño trabajo a mi padre y a mi madre, por dárme todo.

Referencia: MARTÍNEZ PERALES, Javier. «La integral de Henstock-Kurzweil y el segundo teorema fundamental del cálculo». En: *TEMat*, 1 (2017), págs. 15-23. ISSN: 2530-9633. URL: <http://temat.anemat.com/articulo/2017-p15/>.

1. Introducción

Algo después de la aparición de la teoría de integración de Lebesgue, los matemáticos Arnaud Denjoy [2] en 1912 y Oskar Perron [6] en 1914 desarrollaron nuevos procesos de integración que generalizan el de Lebesgue y resuelven el problema del Teorema Fundamental del Cálculo: toda función derivable se puede recuperar a partir de su derivada mediante estos procesos de integración. Se puede encontrar una exposición de ambas teorías en el libro de Gordon [3].

En los años 50, los matemáticos Jaroslav Kurzweil [5] y Ralph Henstock [4] encontraron, independientemente, un proceso de integración más sencillo que el de Lebesgue, constituyendo una pequeña modificación en la técnica de integración de Riemann pero que da lugar a una teoría de integración aún más general que la de Lebesgue. Este método resulta ser equivalente a los propuestos por Denjoy y Perron (de nuevo se puede encontrar una prueba de ello en el libro de Gordon [3]). La (al menos aparente) complejidad de los procesos de integración de Denjoy y Perron hace aún más interesante, por su sencillez, el estudio de la integral de Henstock-Kurzweil.

En este artículo vamos a ver cómo, con muy poco esfuerzo, la integral de Henstock-Kurzweil es capaz de resolver el problema de recuperar una función a partir de su derivada. Todos los resultados que encontraremos aquí pueden consultarse en varios libros. Entre ellos, uno de los que tratan el tema de manera muy elemental es el libro de Bartle [1], en el cual se basa fundamentalmente esta exposición.

2. Notación y preliminares

A lo largo del texto, I representará un intervalo compacto $[a, b]$, con $a \leq b$, y su longitud será $|I| = b - a$. Un concepto fundamental a la hora de definir la integral de Henstock-Kurzweil es el de partición de un intervalo. Dado un intervalo I , diremos que una familia finita $\mathcal{P} = \{I_i\}_{i=1}^n$ es una **partición** de I si está formada por una familia de intervalos **no solapados** (esto es, con interiores disjuntos) cuya unión es I . Denotaremos por $\mathcal{P}(I)$ al conjunto de todas las particiones del intervalo I .

Podemos ordenar siempre los I_i de una partición de manera que $\sup I_i = \inf I_{i+1}$ para cada i entre 1 y $n - 1$, de modo que siempre los escogeremos de este modo. Además, obsérvese que, dada una partición $\mathcal{P} = \{I_i\}_{i=1}^n$ del intervalo $[a, b]$, obtenemos una colección finita y ordenada de puntos de la recta real,

$$a = x_0 \leq x_1 \leq x_2 \leq \dots \leq x_n = b.$$

Dichos puntos son los extremos de los subintervalos de la partición. Claramente, dada una colección finita y ordenada de puntos como la anterior, obtenemos una partición del intervalo $[a, b]$. En definitiva, podemos dar una partición de dos formas: definiendo los intervalos que la forman o dando los extremos de estos de manera ordenada.

Una vez partido el intervalo I con una partición $\mathcal{P} = \{I_i\}_{i=1}^n$, necesitamos una forma de tener en cuenta los valores de la función a integrar en cada subintervalo de la partición. Para cada I_i , elijamos $t_i \in I_i$. Decimos que el punto t_i es una **etiqueta** de I_i . Al par (I_i, t_i) le llamaremos **intervalo etiquetado** y diremos que $\dot{\mathcal{P}} = \{(I_i, t_i)\}_{i=1}^n$ es una **partición etiquetada** de I . El punto sobre el símbolo \mathcal{P} significará que la partición está etiquetada. Es claro que una partición \mathcal{P} de I puede ser etiquetada de infinitas maneras, pues podemos elegir como etiqueta de un subintervalo de la partición cualquiera de sus puntos.

A continuación, vamos a recordar brevemente la integral de Riemann, para lo que, primero, es conveniente dar una manera de medir cómo de fina es una partición. Dadas dos particiones \mathcal{P} y \mathcal{Q} del intervalo I , diremos que \mathcal{Q} es **más fina** que \mathcal{P} si $\mathcal{P} \subset \mathcal{Q}$, entendiendo \mathcal{P} y \mathcal{Q} como las colecciones ordenadas de los extremos de sus intervalos. Se define la **norma** de una partición $\mathcal{P} = \{I_i\}_{i=1}^n$ del intervalo I como el número real $\|\mathcal{P}\| \equiv \max\{|I_i|\}_{i=1}^n$. La norma de una partición etiquetada se define como la norma de la partición cuyos intervalos conforman dicha partición etiquetada. Nótese que la norma de la partición \mathcal{P} proporciona una manera de controlar las longitudes de cada uno de los subintervalos de \mathcal{P} , esto es, proporciona una manera de medir la afinación de la partición. Es fácil encontrar particiones finas en este sentido, es decir, dado $\delta > 0$, es fácil encontrar una partición \mathcal{P} de I con norma $\|\mathcal{P}\| < \delta$.

Si $f : I \rightarrow \mathbb{R}$ es una función y $\dot{\mathcal{P}} = \{(I_i, t_i)\}_{i=1}^n$ una partición etiquetada de I , se define la **suma de Riemann de f asociada a $\dot{\mathcal{P}}$** como el número real

$$S(f, \dot{\mathcal{P}}) \equiv \sum_{i=1}^n f(t_i)|I_i|.$$

Como ya sabemos, el concepto de la integral de una función se define a partir de un cierto proceso de paso al límite. El proceso de paso al límite que se use es determinante a la hora de definir la integral. El enfoque de Riemann de la integral es el de tomar el límite de las sumas de Riemann cuando la norma de la partición se aproxima a cero. Como vemos, este método no tiene en cuenta cómo se comporta la función a integrar, con lo que, como sabemos, las funciones integrables mediante este método van a tener que verificar unas buenas propiedades muy concretas.

A continuación detallamos un poco el proceso de paso al límite que se utiliza para la integral de Riemann.

Definición 1. Una función **acotada** $f : [a, b] \rightarrow \mathbb{R}$ es **integrable en el sentido de Riemann** en $[a, b]$ si existe $A \in \mathbb{R}$ tal que, para cada $\varepsilon > 0$, podemos encontrar un $\delta > 0$ para el cual, si $\dot{\mathcal{P}}$ es una partición etiquetada de $[a, b]$ con $|I_i| \leq \delta$ para cada $i = 1, \dots, n$, entonces $|S(f, \dot{\mathcal{P}}) - A| < \varepsilon$. ◀

El enfoque de Henstock-Kurzweil es algo distinto. La definición resultante es formalmente similar a la de la integral en el sentido de Riemann. Sin embargo, como ya veremos, el resultado al modificar el método de paso al límite es considerablemente más general que el obtenido al considerar el método de Riemann. A grandes rasgos, lo que marca la diferencia es que el método para escoger las particiones finas permite tener en cuenta el comportamiento de la función a integrar, lo que dará lugar a una mayor clase de funciones integrables. Es más, la integral que se obtiene resulta ser más general que la integral de Lebesgue.

Como ya hemos comentado, las funciones integrables en el sentido de Riemann deben tener unas propiedades muy específicas, con lo que muchas funciones relativamente sencillas quedan fuera de la teoría.

Ejemplo 1. Un ejemplo de función no integrable en el sentido de Riemann es la función de Dirichlet, $\chi_{\mathbb{Q} \cap [0,1]}$, esto es, la función característica de los racionales del intervalo $[0, 1]$, que vale 1 en los números racionales de $[0, 1]$ y 0 en los irracionales de $[0, 1]$. Veámoslo.

Tenemos que probar que, para todo $A \in \mathbb{R}$, hay un $\varepsilon_A > 0$ tal que, para cada $\delta > 0$, existe una partición etiquetada $\dot{\mathcal{P}}_\delta$ con $\|\dot{\mathcal{P}}_\delta\| < \delta$ y

$$|S(\chi_{\mathbb{Q} \cap [0,1]}, \dot{\mathcal{P}}_\delta) - A| > \varepsilon_A.$$

Sean $A \in \mathbb{R} \setminus \{1\}$ y $\varepsilon_A = |1 - A|/2$. Sea $\delta > 0$. Consideremos una partición etiquetada $\dot{\mathcal{P}}$ con norma menor que δ y etiquetada de tal manera que todas las etiquetas t_i de sus intervalos sean un número racional (esto puede hacerse por la densidad de los racionales, por ejemplo). Se tiene, entonces, que $S(\chi_{\mathbb{Q} \cap [0,1]}, \dot{\mathcal{P}}) = \sum_{I_i \in \dot{\mathcal{P}}} |I_i| \chi_{\mathbb{Q} \cap [0,1]}(t_i) = \sum_{I_i \in \dot{\mathcal{P}}} |I_i| = |[0, 1]| = 1$, de modo que

$$|S(\chi_{\mathbb{Q} \cap [0,1]}, \dot{\mathcal{P}}) - A| = |1 - A| > |1 - A|/2 = \varepsilon_A.$$

Si $A = 1$, entonces elegimos $\varepsilon_A = 1/2$. Para cualquier $\delta > 0$ tomamos una partición $\dot{\mathcal{P}}$ con $\|\dot{\mathcal{P}}\| < \delta$ de manera que las etiquetas sean todas irracionales (siempre puede hacerse por la densidad de los irracionales). En este caso tenemos, entonces, que $S(\chi_{\mathbb{Q} \cap [0,1]}, \dot{\mathcal{P}}) = 0$, de manera que

$$|S(\chi_{\mathbb{Q} \cap [0,1]}, \dot{\mathcal{P}}) - A| = 1 > 1/2 = \varepsilon_A.$$

Por tanto, $\chi_{\mathbb{Q} \cap [0,1]}$ no es integrable en el sentido de Riemann ya que hemos encontrado particiones etiquetadas con norma tan pequeña como se quiera cuyas sumas de Riemann valen 0 y otras para las cuales valen 1.

Esto se debe a que, en el proceso de paso al límite, podemos elegir particiones finas etiquetadas de la forma que queramos, lo que, en última instancia, permitirá un mal comportamiento de las sumas de Riemann de una función buena asociadas a una partición fina. ◀

3. El concepto de medidor

Como hemos podido comprobar, la función de Dirichlet no es integrable en el sentido de Riemann. Ahora bien, teniendo en cuenta la idea intuitiva de la integral de una función no negativa (el área bajo la gráfica), es razonable pensar que la función de Dirichlet tenga integral nula (no hay área bajo la gráfica de la función de Dirichlet). En efecto, al introducir la integral de Lebesgue, se comprueba que esto es así para dicha integral. Así pues, si la función de Dirichlet fuese integrable Riemann, el candidato a ser su integral sería 0. Vamos a comprobar que, modificando sutilmente la manera en que medimos la afinación de las particiones en el proceso al límite, se consigue hacer que las sumas de Riemann se vayan a 0.

Sea $\varepsilon > 0$. Como el conjunto de los números racionales del intervalo $[0, 1]$ es numerable, podemos tomar una enumeración, $\{r_k\}_{k \in \mathbb{N}}$, de estos. Consideremos una partición etiquetada $\mathcal{P} = \{(I_i, t_i)\}_{i=1}^n$ de manera que, si $t_i = r_k$ para algún $k \in \mathbb{N}$, entonces se tenga $I_i \subset [t_i - \varepsilon/2^{k+1}, t_i + \varepsilon/2^{k+1}]$. Según cuál sea la etiqueta de un intervalo etiquetado hay dos posibilidades.

- Si $t_i \in \mathbb{R} \setminus \mathbb{Q}$, entonces $f(t_i) = 0$ y el par $(I_i, f(t_i))$ no contribuye a las sumas de Riemann.
- Si $t_i = r_k$ para algún $k \in \mathbb{N}$, entonces $f(t_i) = 1$ y la longitud del intervalo I_i es menor que $\frac{\varepsilon}{2^k}$ por la elección de los intervalos.

Para una tal partición \mathcal{P} tenemos, por tanto, que

$$|S(f, \mathcal{P})| = \left| \sum_{i=1}^n f(t_i) |I_i| \right| \leq \sum_{k=1}^{\infty} \varepsilon/2^k = \varepsilon.$$

Por tanto, si estas particiones fuesen todas las particiones finas entre las que buscamos para la definición de la integral de Riemann, entonces f sería integrable y su integral sería 0. Notemos que, a la hora de escoger particiones lo suficientemente finas, nos hemos fijado en los valores de la función en las etiquetas, de forma que, en aquellas etiquetas donde la función toma el valor 1 (es decir, en aquellas que contribuyen a las sumas de Riemann), los intervalos quedan contenidos dentro de intervalos pequeños controlados, en cierto modo, por la respectiva etiqueta. Hemos conseguido, así, que las sumas de Riemann sean pequeñas.

Esto motiva la búsqueda de un criterio de elección de particiones finas que permita tener en cuenta el comportamiento de la función en las etiquetas de los intervalos de la partición de manera que en las sumas de Riemann solo se tengan en cuenta los valores relevantes de la función, haciendo menor la contribución a la suma de los valores menos relevantes. En el enfoque de Henstock-Kurzweil se permite una mayor libertad a la hora de escoger el criterio de elección de particiones finas, dando la posibilidad de elegir un criterio tan restrictivo que podamos englobar aquellas etiquetas en las que los valores de la función son poco representativos (por ejemplo, los valores no nulos de una función que se anula salvo en un conjunto de medida de Lebesgue cero) en intervalos lo suficientemente pequeños como para que su contribución a las sumas de Riemann sea despreciable.

Esto se consigue, como veremos, introduciendo el concepto de medidor, que controlará la afinación de una partición etiquetada de un intervalo I exigiendo que cada intervalo etiquetado (I_i, t_i) quede contenido en un intervalo $\overline{B}(t_i, \delta_i) = [t_i - \delta_i, t_i + \delta_i]$ que depende de la etiqueta t_i , como hemos hecho con el ejemplo de la función de Dirichlet.

Definición 2. Una función $\delta : I \rightarrow \mathbb{R}$ es un **medidor** en I si $\delta(t) > 0$ para cada $t \in I$. Dados un medidor δ en I y un punto $t \in I$, el intervalo alrededor de t **controlado por el medidor** δ es el intervalo $\overline{B}(t, \delta(t)) = [t - \delta(t), t + \delta(t)]$. ◀

Diremos, en este caso, que el intervalo $\overline{B}(t, \delta(t))$ es el intervalo **δ -controlado** por t aunque, si no hay lugar a confusión, diremos, simplemente, que $\overline{B}(t, \delta(t))$ es el intervalo controlado por t .

Un medidor δ en I permite dar una mejor medida de la afinación de una cierta partición etiquetada, \mathcal{P} , del intervalo I que la que da el criterio de la norma considerado a la hora de definir la integral de Riemann. La manera exacta de dar esta medida es la que se recoge en la siguiente definición.

Definición 3. Sea $\mathcal{P} = \{(I_i, t_i)\}_{i=1}^n$ una partición etiquetada de I . Sea δ un medidor en I . Decimos que \mathcal{P} es **δ -fina** si $I_i \subset [t_i - \delta(t_i), t_i + \delta(t_i)]$ para cada $i = 1, \dots, n$, esto es, si cada subintervalo I_i de la

partición está contenido en el intervalo δ -controlado por el punto t_i (diremos, entonces, que el intervalo I_i está δ -controlado por t_i). Cuando $\dot{\mathcal{P}}$ sea δ -fina, diremos de ella que es una partición **subordinada** a δ y escribiremos $\dot{\mathcal{P}} \ll \delta$. ◀

La definición de medidor no es muy restrictiva. Esto será muy útil en adelante, pues permitirá una gran libertad a la hora de elegir la medida de afinación adecuada (un medidor δ adecuado) para que las particiones finas según el medidor elegido (particiones δ -finas) tengan asociadas unas sumas de Riemann con un buen comportamiento, obteniendo, así, una mayor clase de funciones integrables.

Veamos, a modo de ejemplo, cómo se entiende la medida de afinación de un medidor constante (la que se toma para la integral de Riemann) en términos de particiones δ -finas.

Ejemplo 2. Sea $\delta > 0$. La aplicación $\delta : I \rightarrow \mathbb{R}$ definida por $\delta(t) = \delta$ para cada $t \in I$ es un medidor en I . Llamamos **medidor constante** a este medidor. Obsérvese que, si una partición $\dot{\mathcal{P}}$ de I es δ -fina, entonces $\|\dot{\mathcal{P}}\| \leq 2\delta$. En efecto, si $\dot{\mathcal{P}}$ es δ -fina, se tiene que $I_i \subseteq \overline{B}(t_i, \delta)$ para cada $i = 1, \dots, n$, de donde $|I_i| \leq |\overline{B}(t_i, \delta)| = 2\delta$ para cada $i = 1, \dots, n$, con lo que $\|\dot{\mathcal{P}}\| \leq 2\delta$. Además, se tiene que, si $\dot{\mathcal{P}}$ es una partición con $\|\dot{\mathcal{P}}\| < \delta$, entonces $\dot{\mathcal{P}}$ es δ -fina. ◀

A continuación proporcionamos algunos métodos para construir medidores a partir de medidores ya existentes.

Ejemplo 3. Si δ_1 y δ_2 son dos medidores en el intervalo I , entonces la función $\delta : I \rightarrow \mathbb{R}$ definida por $\delta(t) = \min\{\delta_1(t), \delta_2(t)\}$ es un medidor en I . Es claro que toda partición δ -fina de I es, a la vez, δ_1 -fina y δ_2 -fina. Diremos que δ es el **refinamiento** de los medidores δ_1 y δ_2 . Análogamente, podemos definir el refinamiento de cualquier familia finita de medidores en I . ◀

Está claro que un medidor definido en I es también un medidor en cualquiera de los subintervalos compactos de I . Siempre que se dé este caso tenemos la siguiente proposición.

Proposición 1. Sean $c \in (a, b)$ y δ un medidor en $[a, b]$. Si $\dot{\mathcal{P}}_1$ es una partición δ -fina de $[a, c]$ y $\dot{\mathcal{P}}_2$ es una partición δ -fina de $[c, b]$, entonces $\dot{\mathcal{P}}_1 \cup \dot{\mathcal{P}}_2$ es una partición δ -fina de $[a, b]$.

Demostración. Basta usar la definición. Como $\dot{\mathcal{P}}_1$ es δ -fina, se tiene que cada I_i^1 de $\dot{\mathcal{P}}_1$ está δ -controlado por la correspondiente etiqueta, t_i^1 . Como $\dot{\mathcal{P}}_2$ es δ -fina, se tiene que cada I_i^2 de $\dot{\mathcal{P}}_2$ está δ -controlado por la correspondiente etiqueta, t_i^2 .

Al ser $\dot{\mathcal{P}}_1 \cup \dot{\mathcal{P}}_2$ una partición de $[a, b]$ formada por cada uno de los intervalos etiquetados de las particiones $\dot{\mathcal{P}}_1$ y $\dot{\mathcal{P}}_2$, es claro que cada uno de los intervalos que la conforman está δ -controlado, de donde se obtiene que $\dot{\mathcal{P}}_1 \cup \dot{\mathcal{P}}_2$ es δ -fina. ■

Dado un medidor constante δ , es trivial encontrar particiones δ -finas. Ahora bien, dado un medidor arbitrario, δ , en $[a, b]$, ¿existe siempre una partición etiquetada de $[a, b]$, $\dot{\mathcal{P}} = \{(I_i, t_i)\}_{i=1}^n$, cuyas etiquetas controlen su correspondiente subintervalo etiquetado? Esto es, dado un medidor δ definido en $[a, b]$, ¿existen particiones δ -finas de $[a, b]$?

El resultado que da respuesta general a esta pregunta se sigue, de manera sencilla, de la propiedad arquimediana de los números reales, que asegura que, dados $y \in \mathbb{R}$ y $x > 0$, siempre existe $n \in \mathbb{N}$ con $nx > y$. Es conocido el hecho de que esta propiedad de \mathbb{R} se sigue del axioma del supremo.

La respuesta afirmativa a la pregunta anterior es el contenido del siguiente teorema, que se debe a Pierre Cousin.

Teorema 2 (teorema de Cousin). Sea δ un medidor en $[a, b]$. Entonces, existe una partición δ -fina de $[a, b]$.

Demostración. Lo probaremos por reducción al absurdo.

Supongamos que $[a, b]$ no tiene ninguna partición δ -fina. Sea $c = \frac{a+b}{2}$ y consideremos los subintervalos

$$[a, c] \quad \text{y} \quad [c, b].$$

Según lo visto en la proposición 1, si los dos intervalos $[a, c]$ y $[c, b]$ tuviesen una partición δ -fina, entonces su unión sería una partición δ -fina de $[a, b]$. Así pues, alguno de los subintervalos $[a, c]$ o $[c, b]$ no tiene

ninguna partición δ -fina. Sea cual sea este intervalo, renombrémoslo como $I^1 \equiv [a_1, b_1]$. Sea $c_1 = \frac{a_1+b_1}{2}$ y consideremos los subintervalos

$$[a_1, c_1] \quad \text{y} \quad [c_1, b_1].$$

Como antes, uno de los subintervalos no tiene ninguna partición δ -fina. Sea $I^2 = [a_2, b_2]$ este subintervalo.

Procediendo de este modo, obtenemos una sucesión de intervalos compactos encajados $\{I^n\}_{n \in \mathbb{N}}$, cada uno con longitud $|I^n| = \frac{b-a}{2^n}$, sin particiones δ -finas.

El teorema de los intervalos encajados asegura que existe un único punto x en la intersección de todos los intervalos I^n . Sin embargo, como $\delta(x) > 0$, por la propiedad arquimediana, existe $p \in \mathbb{N}$ tal que

$$|I^p| = \frac{b-a}{2^p} < \delta(x).$$

Por tanto, $I^p \subset [x - \delta(x), x + \delta(x)]$. Además, el par $\{(I^p, x)\}$ es una partición (formada por un solo intervalo) δ -fina de I^p , en contra de la elección de I^p . Esta contradicción prueba que $[a, b]$ tiene alguna partición δ -fina. ■

Como consecuencia del teorema de Cousin puede probarse, por ejemplo, el siguiente teorema clásico.

Teorema 3 (teorema de Weierstrass). *Sea $f : [a, b] \rightarrow \mathbb{R}$ una función continua. Entonces, f alcanza su máximo y su mínimo en $[a, b]$.*

Demostración. Lo probaremos para el máximo, lo cual es suficiente. Supongamos que no es cierto. Entonces, dado $s \in [a, b]$, este no es el máximo de f en $[a, b]$, luego existe $X(s) \in [a, b]$ tal que $f(s) < f(X(s))$ y, por la continuidad de f , existe $\delta_s > 0$ con

$$f(t) < f(X(s)) \quad \text{para cada } t \in [s - \delta_s, s + \delta_s].$$

Sea $\delta : [a, b] \rightarrow \mathbb{R}^+$ el medidor definido por $\delta(x) = \delta_x$. Por el teorema de Cousin, existe $\mathcal{P} = \{(I_i, t_i)\}_{i=1}^n$ δ -fina. Sea t_0 con $f(X(t_0)) = \max\{f(X(t_i))\}_{i=1}^n$. Como $X(t_0) \in [a, b]$ y $[a, b] = \bigcup_{i=1}^n I_i$, existe i_0 tal que $X(t_0) \in I_{i_0}$. Pero entonces,

$$f(X(t_0)) < f(X(t_{i_0})),$$

en contra de la elección de t_0 . Por tanto, f tiene un máximo en $[a, b]$. ■

Aún más, podemos probar la equivalencia entre el teorema de Cousin, el axioma del supremo y la propiedad arquimediana. En efecto, recordemos que el axioma del supremo implica la propiedad arquimediana (también es equivalente al principio de los intervalos encajados). Notemos, por otro lado, que la propiedad arquimediana y el principio de los intervalos encajados son lo único que hemos necesitado para probar el teorema de Cousin. El siguiente teorema prueba que el teorema de Cousin implica el axioma del supremo.

Teorema 4 (axioma del supremo). *Sea $S \subset \mathbb{R}$ un conjunto no vacío y acotado. Entonces, S tiene supremo.*

Demostración. Sea S un conjunto no vacío. Como S es acotado, existe un intervalo $[a, b]$ tal que $S \subset [a, b]$. Supongamos que S no tiene supremo. En particular, ningún punto de S puede ser cota superior de S .

- Si $x \in S$, entonces existe $y_x \in S$ tal que $x < y_x$. De lo contrario, si para cada $y \in S$ tenemos $x \geq y$, entonces x sería la mínima cota superior de S , lo cual contradice la hipótesis. Para un tal x definimos $\delta_x = y_x - x > 0$.

Nótese que, si x pertenece a un intervalo $[c, d] \subset [a, b]$ con $d - c < \delta_x$, entonces d no es cota superior de S , pues $d \leq x + \delta_x = x + y_x - x = y_x \in S$ y sabemos que ningún punto de S puede ser cota superior de S .

- Si $x \in [a, b] \setminus S$ y x es cota superior de S , entonces, al no haber un supremo de S , existe otra cota superior de S , z_x , tal que $z_x < x$. En tal caso, definimos $\delta_x = x - z_x$.

Nótese que, si x pertenece a un intervalo $[c, d] \subset [a, b]$ con $d - c < \delta_x$, entonces $z_x = x - \delta_x \leq c$, luego c es cota superior de S .

- Si $x \in [a, b] \setminus S$ y x no es cota superior de S , entonces existe $y_x \in S$ tal que $x < y_x$. Definamos, en tal caso, $\delta_x = y_x - x$. Obsérvese que y_x no es cota superior de S porque ningún punto de S puede ser cota superior de S .

Nótese que, si x pertenece a un intervalo $[c, d] \subset [a, b]$ con $d - c < \delta_x$, entonces d no es cota superior de S , pues $d \leq x + \delta_x = x + y_x - x = y_x \in S$.

En cualquiera de los casos, obsérvese que, o bien $[c, d]$ verifica que c es cota superior de S , o bien verifica que d no es cota superior de S .

Definamos, entonces, el medidor $\delta : [a, b] \rightarrow \mathbb{R}$ mediante $\delta(x) = \delta_x$. Por el teorema de Cousin, existe una partición δ -fina de $[a, b]$, $\dot{\mathcal{P}} = \{([a_{i-1}, a_i], t_i)\}_{i=1}^n$.

Claramente $a_0 = a$ no es cota superior de S . Como $\dot{\mathcal{P}}$ es δ -fina tenemos, por el razonamiento anterior, que, o bien a_0 es cota superior de S , o bien a_1 no es cota superior de S . Por tanto, a_1 no es cota superior de S . Razonando de igual manera, a_2 no es cota superior de S . Este razonamiento es válido para cada $j = 1, \dots, n$, luego ningún a_j es cota superior de S . Esto contradice el hecho de que $a_n = b$ sí que es cota superior de S . Por tanto, S debe tener supremo. ■

4. La integral de Henstock-Kurzweil

Ya estamos preparados para definir el proceso de paso al límite en la definición de la integral. A continuación, damos dos definiciones equivalentes de la integral de Henstock-Kurzweil o integral medidora. La primera de ellas es formalmente muy parecida a la definición de la integral en el sentido de Riemann salvo porque, ahora, usaremos un medidor δ , en lugar de un número $\delta > 0$ (un medidor constante) para controlar la afinación de las particiones.

Definición 4. Una función $f : I \rightarrow \mathbb{R}$ es **integrable en el sentido de Henstock-Kurzweil** (o **HK-integrable**) en I si existe $B \in \mathbb{R}$ tal que, para cada $\varepsilon > 0$, podemos encontrar un medidor δ en I para el cual, si $\dot{\mathcal{P}}$ es una partición etiquetada de I con $|I_i| \leq \delta(t_i)$ para cada $i = 1, \dots, n$, entonces $|S(f, \dot{\mathcal{P}}) - B| < \varepsilon$. ◀

La siguiente es una definición equivalente en términos del concepto de δ -afinación de una partición respecto de un medidor.

Definición 5. Una función $f : I \rightarrow \mathbb{R}$ es **integrable en el sentido de Henstock-Kurzweil** (o **HK-integrable**) en I si existe $C \in \mathbb{R}$ tal que, para cada $\varepsilon > 0$, podemos encontrar un medidor δ en I para el cual, si $\dot{\mathcal{P}}$ es una partición δ -fina de I , entonces $|S(f, \dot{\mathcal{P}}) - C| < \varepsilon$. ◀

El hecho de que las dos definiciones anteriores son equivalentes es un sencillo ejercicio que el lector puede hacer para familiarizarse con el nuevo concepto. Para la integral de Riemann, es sencillo ver que el valor de la integral de una función integrable es único. Para la integral de Henstock-Kurzweil también es sencillo probar esto haciendo uso del refinamiento de un medidor.

Teorema 5 (unicidad de valor de la HK-integral). *Supongamos que $f : [a, b] \rightarrow \mathbb{R}$ es integrable en el sentido de Henstock-Kurzweil. Entonces, el valor de la integral es único.*

Demostración. Sea $\varepsilon > 0$. Supongamos que C' y C'' satisfacen la definición 5. Existen, entonces, dos medidores, δ' y δ'' en $[a, b]$ tales que, si $\dot{\mathcal{P}}$ es a la vez δ' -fina y δ'' -fina, tenemos

$$|S(f, \dot{\mathcal{P}}) - C'| < \varepsilon/2 \quad \text{y} \quad |S(f, \dot{\mathcal{P}}) - C''| < \varepsilon/2.$$

Así pues, si δ es el refinamiento de δ' y δ'' y $\dot{\mathcal{P}}$ es una partición δ -fina de $[a, b]$ (que existe por el teorema de Cousin), entonces, por la desigualdad triangular,

$$|C' - C''| \leq |C' - S(f, \dot{\mathcal{P}})| + |S(f, \dot{\mathcal{P}}) - C''| < \varepsilon/2 + \varepsilon/2 = \varepsilon.$$

Como esto es para cada $\varepsilon > 0$, concluimos que $C' = C''$. ■

Visto esto, notaremos el valor de la integral de una función integrable en el sentido de Henstock-Kurzweil de la forma habitual.

Como es natural por la forma en que se ha definido el concepto de integral en el sentido de Henstock-Kurzweil, la integral de Riemann es un caso particular de esta.

Teorema 6. Sea $f: [a, b] \rightarrow \mathbb{R}$ una función. Si f es integrable en el sentido de Riemann en $[a, b]$, entonces es integrable en el sentido de Henstock-Kurzweil en $[a, b]$. Además, en tal caso, los valores de las integrales coinciden.

Demostración. Basta tener en cuenta que, por definición, f es integrable Riemann en $[a, b]$ con integral R si para cada $\varepsilon > 0$ existe un medidor constante $\delta > 0$ tal que, si $\mathcal{P} \ll \delta$, entonces $|S(f, \mathcal{P}) - R| < \varepsilon$. ■

Conocemos, entonces, una amplia clase de funciones integrables en el sentido de Henstock-Kurzweil: las funciones continuas, las funciones monótonas y acotadas en $[a, b]$ y, en general, las funciones acotadas continuas en casi todo punto según la medida de Lebesgue (esto es, las funciones integrables en el sentido de Riemann).

5. El segundo teorema fundamental del cálculo

A continuación vamos a ver cómo, con las pocas herramientas con las que contamos, estamos ya en condiciones de probar la versión más general del II TFC, que establece que la derivada de una función en un intervalo $[a, b]$ es siempre integrable en el sentido de Henstock-Kurzweil, sin más hipótesis adicionales. Antes de probarlo necesitaremos el siguiente lema técnico cuya prueba, que se sigue fácilmente de la condición de derivabilidad en un punto, se puede encontrar en el libro de Bartle [1].

Lema 7. Sea $F: [a, b] \rightarrow \mathbb{R}$ una función derivable en $t \in [a, b]$. Dado $\varepsilon > 0$, existe $\delta_\varepsilon(t) > 0$ tal que, si $c, d \in [a, b]$ satisfacen $t - \delta_\varepsilon(t) \leq c \leq t \leq d \leq t + \delta_\varepsilon(t)$, entonces

$$|F(d) - F(c) - F'(t)(d - c)| \leq \varepsilon(d - c).$$

Teorema 8 (II teorema fundamental del cálculo). Sea $F: [a, b] \rightarrow \mathbb{R}$ una función derivable en $[a, b]$. Entonces, F' es integrable en el sentido de Henstock-Kurzweil en $[a, b]$ y, además,

$$\int_a^x F'(t) dt = F(x) - F(a) \text{ para cada } x \in [a, b].$$

Demostración. Bastará probarlo para $x = b$, para el resto de puntos de $[a, b]$ es análogo. Sea $\varepsilon > 0$. Consideremos el medidor asociado a $\varepsilon/(b - a)$ por el lema anterior. Sea $\mathcal{P} = \{([x_{i-1}, x_i], t_i)\}_{i=1}^n$ una partición $\delta_{\varepsilon/(b-a)}$ -fina de $[a, b]$. Vamos a probar que $|F(b) - F(a) - S(F', \mathcal{P})| \leq \varepsilon$. Como para cada $i \in \{1, \dots, n\}$, x_{i-1} y x_i verifican

$$t_i - \delta_{\varepsilon/(b-a)}(t_i) \leq x_{i-1} \leq t_i \leq x_i \leq t_i + \delta_{\varepsilon/(b-a)}(t_i),$$

por la propiedad de $\delta_{\varepsilon/(b-a)}$ dada en el lema anterior se tiene que

$$(1) \quad |F(x_i) - F(x_{i-1}) - F'(t_i)(x_i - x_{i-1})| \leq \frac{\varepsilon(x_i - x_{i-1})}{(b - a)} \text{ para cada } i \in \{1, \dots, n\}.$$

Para estimar $F(b) - F(a) - S(F', \mathcal{P})$ utilizaremos la suma telescópica

$$F(b) - F(a) = \sum_{i=1}^n F(x_i) - F(x_{i-1}),$$

de modo que

$$F(b) - F(a) - S(F', \mathcal{P}) = \sum_{i=1}^n [F(x_i) - F(x_{i-1}) - F'(t_i)(x_i - x_{i-1})].$$

Luego

$$|F(b) - F(a) - S(F', \mathcal{P})| \leq \sum_{i=1}^n |F(x_i) - F(x_{i-1}) - F'(t_i)(x_i - x_{i-1})|.$$

Por (1) se tiene, entonces, que

$$|F(b) - F(a) - S(F', \mathcal{P})| \leq \sum_{i=1}^n \frac{\varepsilon(x_i - x_{i-1})}{b - a} = \varepsilon \frac{b - a}{b - a} = \varepsilon. \quad \blacksquare$$

6. Comentarios

Como hemos podido comprobar, la integral de Henstock-Kurzweil resulta ser una integral muy versátil, permitiendo probar el II TFC de una manera muy sencilla, prácticamente a partir de la definición. También es posible probar la versión conocida del I TFC para la integral de Lebesgue e incluso una versión mejorada del II TFC permitiendo algunas excepciones en la condición de derivabilidad.

Si estudiamos la bibliografía disponible, podemos comprobar que la integral de Henstock-Kurzweil es más general que la integral de Lebesgue. En particular, las funciones integrables en el sentido de Lebesgue son precisamente aquellas funciones integrables en el sentido de Henstock-Kurzweil cuyo valor absoluto también es integrable en el sentido de Henstock-Kurzweil (podemos encontrar una prueba en el libro de Bartle [1]). De hecho, se puede probar que existen funciones integrables en el sentido de Henstock-Kurzweil cuyo valor absoluto no es integrable en el sentido de Henstock-Kurzweil. Un ejemplo sencillo puede construirse a partir de una serie condicionalmente convergente. Esto demuestra que la integral de Henstock-Kurzweil es estrictamente más general que la integral de Lebesgue.

Por otro lado, puede probarse que no existen integrales impropias en el sentido de Henstock-Kurzweil, es decir, si podemos definir la integral de Henstock-Kurzweil de una función como el límite de unas integrales de Henstock-Kurzweil de la función en subintervalos del dominio, entonces la función es inmediatamente integrable en el sentido de Henstock-Kurzweil en todo el intervalo. Esto, de nuevo, pone de manifiesto el hecho de que la integral de Henstock-Kurzweil es estrictamente más general que las integrales de Riemann y Lebesgue.

Por último, notemos que no se ha necesitado ninguna teoría de la medida para definir la integral de Henstock-Kurzweil. Es más, a partir de la integral de Henstock-Kurzweil puede definirse la teoría de la medida y la integral de Lebesgue en \mathbb{R} , un programa de estudio que se propone en el libro de Gordon [3] y que se pone en práctica en el de Bartle [1].

Por esta y por algunas razones más, algunos autores están de acuerdo en que el estudio de la integral de Henstock-Kurzweil es una buena forma de introducir el concepto de integral a los nuevos estudiantes de Cálculo, aunque, por otro lado, esta integral no es muy utilizada en investigación, lo cual quizá se deba al hecho de que no es posible definirla en contextos tan abstractos como la integral de Lebesgue, además de que no se conoce una topología natural para el espacio de las funciones integrables en el sentido de Henstock-Kurzweil, con lo que no se dispone, por ejemplo, de teoremas de densidad como los que se tienen para los espacios L^p , que son fundamentales a la hora de desarrollar la teoría de operadores definidos en estos espacios.

Referencias

- [1] BARTLE, Robert G. *A modern theory of integration*. Vol. 32. Graduate Studies in Mathematics. American Mathematical Society, Providence, RI, 2001, págs. xiv+458. ISBN: 0-8218-0845-1. <https://doi.org/10.1090/gsm/032>.
- [2] DENJOY, Arnaud. *Une extension de l' integrale de M. Lebesgue*. 1912.
- [3] GORDON, Russell A. *The integrals of Lebesgue, Denjoy, Perron, and Henstock*. Vol. 4. Graduate Studies in Mathematics. American Mathematical Society, Providence, RI, 1994, págs. xii+395. ISBN: 0-8218-3805-9. <https://doi.org/10.1090/gsm/004>.
- [4] HENSTOCK, Ralph. *Definitions of Riemann type of the variational integrals*. Vol. 11. 1961, págs. 402-418.
- [5] KURZWEIL, Jaroslav. *Generalized ordinary differential equations and continuous dependence on a parameter*. Vol. 7 (82). 1957, págs. 418-449.
- [6] PERRON, Oskar. *Über den Integralbegriff*. 1914.

TEMat

Este trabajo colaboró con una microcharla durante el XVII *Encuentro Nacional de Estudiantes de Matemáticas*, celebrado en Barcelona en julio de 2016.



Progresiones aritméticas de colores

✉ Alberto Espuny Díaz^a
University of Birmingham
axe673@bham.ac.uk

Resumen: En este artículo presentamos de manera introductoria el clásico teorema de van der Waerden sobre progresiones aritméticas monocromáticas. Nos centramos principalmente en comprender el enunciado, dar algunos ejemplos y presentar algunos de los problemas que lo rodean. En particular, divagamos sobre el valor exacto de los números de van der Waerden y sobre cotas para estos valores. Damos una demostración sencilla de una cota inferior utilizando el método probabilístico. Finalmente, presentamos el teorema de Szemerédi, que generaliza el resultado de van der Waerden, y demostramos el teorema de van der Waerden a partir de esta generalización.

Palabras clave: progresiones aritméticas, van der Waerden, números de van der Waerden, Szemerédi, coloraciones.

MSC2010: 05D10, 05D40.

Recibido: 24 de diciembre de 2016.

Aceptado: 20 de marzo de 2017.

Agradecimientos: Este trabajo estuvo parcialmente financiado por la beca 2015 / COLAB / 00069 del Ministerio de Educación, Cultura y Deporte.

Referencia: ESPUNY DÍAZ, Alberto. «Progresiones aritméticas de colores». En: *TEMat*, 1 (2017), págs. 25-30. ISSN: 2530-9633. URL: <http://temat.anemat.com/articulo/2017-p25/>.

^aEl autor estaba afiliado a la Universitat Politècnica de Catalunya cuando se desarrolló la mayor parte de este trabajo.

1. Introducción

Todos sabemos lo que son las progresiones aritméticas. Se trata de un objeto matemático que ya nos presentaron en el instituto, con una definición simple y fácil de entender y unas propiedades también simples, pero ya interesantes. Esto hace que sea quizá uno de los primeros objetos matemáticos que nos encontramos que dan un poco de juego, más allá de la pura manipulación algebraica.

Como es bien sabido, una progresión aritmética es una sucesión de números a_1, a_2, a_3, \dots tales que la diferencia entre un número y el siguiente es siempre la misma ($a_{i+1} - a_i = d$). De este modo, una progresión aritmética se puede definir con solo dos parámetros: un valor inicial a_1 y una diferencia d . En caso de querer una progresión de una determinada longitud, se puede introducir esta longitud como un tercer parámetro (por ejemplo, k). Por simplicidad, podemos referirnos a progresiones aritméticas de longitud k como k -progresiones aritméticas, o k -PA.

En este artículo presentamos algunos resultados clásicos de combinatoria, que se pueden englobar dentro de la combinatoria aditiva (que estudia la interacción entre operaciones y conjuntos) o de la teoría de Ramsey (que busca condiciones para que haya estructuras en los conjuntos estudiados), en los que se trabaja con progresiones aritméticas de un modo general: no nos interesa conocer cuál es la progresión aritmética, sino solo saber de su existencia y su longitud. Y la forma de trabajar con ellas va a ser coloreando los números naturales.

2. El teorema de van der Waerden

Podemos jugar a un juego en el que nuestro objetivo es colorear los enteros sin formar progresiones aritméticas del mismo color; vamos a empezar considerando progresiones de longitud tres. Consideremos el siguiente ejemplo, en el que coloreamos los primeros números naturales con color rojo o azul. Comenzamos coloreando los primeros ocho naturales:

1 2 3 4 5 6 7 8 9

Por ahora vamos bien, pero supongamos que queremos colorear también el número 9. Podemos elegir hacerlo bien en rojo, o bien en azul. Si lo hacemos en rojo, sin embargo, observamos que $\{1, 5, 9\}$ es una progresión aritmética roja de longitud tres; si lo coloreamos en azul, $\{3, 6, 9\}$ forma una progresión aritmética azul de longitud tres. Así pues, observamos que el ejemplo que hemos elegido no se puede continuar sin formar progresiones aritméticas monocromáticas de longitud tres.

Obviamente, ahora podríamos elegir otro ejemplo y volver a probar. La pregunta natural que nos hacemos es la siguiente: ¿existe alguna forma de colorear los enteros con dos colores de manera que no aparezcan progresiones aritméticas monocromáticas de longitud tres? O más en general: ¿existe alguna forma de colorear los enteros con r colores de manera que no aparezcan progresiones aritméticas monocromáticas de longitud k ? Y la respuesta, negativa, viene dada por el teorema de van der Waerden.

Teorema 1 (van der Waerden, 1927 [14]). *Para todo par de números naturales $k \geq 3$, $r \geq 2$ existe un entero $N = N(r, k)$ tal que toda coloración de los primeros $n \geq N$ números naturales con r colores contiene alguna progresión aritmética monocromática no trivial de longitud k .*

Nótese que en el teorema se habla de progresiones aritméticas no triviales. Una progresión aritmética trivial es una cuya diferencia es cero, es decir, una progresión de la forma a, a, a, a, \dots , y en ese caso siempre hay, trivialmente, progresiones monocromáticas de longitud arbitraria. Para que el resultado tenga interés, estas se deben evitar y, en general, cuando hablemos de progresiones aritméticas nos referiremos a progresiones no triviales.

Existen varias maneras de demostrar este teorema. La más tradicional en el mundo de la combinatoria es una demostración por doble inducción, primero en el número de colores y luego en la longitud de la progresión. Hay otras demostraciones combinatorias sencillas, aunque casi todas usan ideas parecidas. Después hay demostraciones mucho más complicadas, que permiten obtener el teorema de van der Waerden como un simple corolario, pero que necesitan técnicas más avanzadas para demostrar el resultado principal (el teorema de Szemerédi); a esto nos referiremos en la sección 3. Existen también

otras demostraciones no combinatorias; el teorema de Szemerédi tiene demostraciones no combinatorias, que permiten entonces demostrar también el teorema de van der Waerden.

La siguiente pregunta, obviamente, es cuál es el mínimo valor posible de N . A este valor se le conoce como el número de van der Waerden, $W(r, k)$. Volvamos al ejemplo. Hemos visto que podemos colorear hasta el número ocho sin formar 3-PA monocromáticas, de modo que sabemos que $W(2, 3) \geq 9$. En el ejemplo no podemos colorear el nueve sin formar una 3-PA monocromática, pero podría ser que con una coloración diferente de los primeros ocho enteros sí consigamos colorearlo y obtengamos así $W(2, 3) \geq 10$. No es el caso, en realidad, y se puede comprobar que $W(2, 3) = 9$ escribiendo todas las posibilidades, ya que se trata de un ejemplo pequeño (tengamos en cuenta que este es el número de van der Waerden más pequeño que tiene interés, ya que si se considera $r = 1$ o $k < 3$, los valores son triviales). Consideremos ahora otro ejemplo, usando tres colores, en el que también queremos evitar las 3-PA. En este caso coloreamos los números de manera «voraz»:

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22

Al intentar colorear el 22 se puede comprobar fácilmente que no hay manera de hacerlo sin formar una 3-PA monocromática ($\{4, 13, 22\}$, $\{16, 19, 22\}$ o $\{20, 21, 22\}$), de modo que $W(3, 3) \geq 22$. En este caso sí se pueden conseguir coloraciones mejores, y, de hecho, se ha comprobado que $W(3, 3) = 27$.

A pesar de lo simple que es el problema de determinar los valores de $W(r, k)$, resulta ser un problema muy difícil. De hecho, se conocen solo siete valores: $W(2, 3) = 9$ ([2]), $W(2, 4) = 35$ ([2]), $W(2, 5) = 178$ ([11]), $W(2, 6) = 1132$ ([8]), $W(3, 3) = 27$ ([2]), $W(3, 4) = 293$ ([7]) y $W(4, 3) = 76$ ([1]).

El siguiente problema natural es encontrar cotas para el valor de los números de van der Waerden, que permitan, ya que no conocer, sí estimar su valor. Buscamos, por lo tanto, dos tipos de cotas: las inferiores y las superiores. Para la cota inferior, las cotas más ajustadas se pueden conseguir construyendo ejemplos suficientemente grandes en los que no aparezcan k -progresiones aritméticas monocromáticas. Sin embargo, las cotas generales que este método permite obtener son bastante débiles, y cada una de las cotas ajustadas sirve solo para un caso concreto. Existen otros métodos que permiten obtener cotas generales mucho mejores, como, por ejemplo, el método probabilístico. Este método utiliza técnicas probabilísticas y tiene múltiples aplicaciones en demostraciones no constructivas de existencia de objetos matemáticos. Usando en este caso coloraciones aleatorias de los primeros enteros, es un ejercicio sencillo obtener el siguiente resultado.

Proposición 2. *Los números de van der Waerden cumplen la cota*

$$W(r, k) \geq r^{\frac{k-1}{2}} \sqrt{k-1}.$$

Esta cota se puede mejorar usando técnicas probabilísticas más avanzadas, pero por ahora vamos a demostrar este resultado.

Demostración. Sea $[n] = \{1, 2, \dots, n\}$. Supongamos que coloreamos $[n]$ de manera aleatoria, uniformemente y de manera independiente, es decir, cada número se colorea con el i -ésimo color con probabilidad $\frac{1}{r}$ independientemente del color que tengan los demás números. El número de progresiones aritméticas de longitud k que podemos encontrar en $\{1, 2, \dots, n\}$ se puede acotar fácilmente considerando que solo necesitamos dos parámetros para definirlos: su valor inicial y su diferencia. De este modo, podemos tomar como mucho n valores iniciales y $\frac{n}{k-1}$ diferencias (ya que las diferencias que tomemos deben ser lo suficientemente pequeñas como para que al sumarlas $k-1$ veces al valor inicial sigamos teniendo un número menor que $n+1$), lo que supone un total de como mucho $\frac{n^2}{k-1}$ progresiones aritméticas (obviamente, hay muchas menos, pero con esta cota nos basta).

Ahora consideremos una de estas k -PA. La probabilidad de que sea monocromática, con nuestra coloración aleatoria, es $\frac{1}{r^{k-1}}$ (la probabilidad de que sea toda del primer color es $\frac{1}{r^k}$, ya que la probabilidad de que cada número sea de ese color es $\frac{1}{r}$ y los sucesos son independientes, y en total hay r colores). Así pues, la probabilidad de que *alguna* de las progresiones sea monocromática se puede acotar superiormente por la suma de las probabilidades de que cada una lo sea, y teniendo en cuenta la cota sobre el número de k -PA

obtenemos la expresión

$$\begin{aligned} \Pr(\exists \text{ una } k\text{-PA monocromática}) &= \Pr\left(\bigcup_{S \subseteq [n] \text{ } k\text{-PA}} \{S \text{ es monocromática}\}\right) \\ &\leq \sum_{S \subseteq [n] \text{ } k\text{-PA}} \Pr(S \text{ es monocromática}) \\ &= \sum_{S \subseteq [n] \text{ } k\text{-PA}} \frac{1}{r^{k-1}} \leq \frac{n^2}{(k-1)r^{k-1}}. \end{aligned}$$

Ahora bien, si este último valor es menor que 1, eso quiere decir que la probabilidad de que haya una progresión aritmética monocromática es también menor que 1, lo que significa que la probabilidad de que no haya ninguna es positiva. Es decir, que existe alguna manera de colorear los primeros n números de manera que no hay ninguna progresión aritmética monocromática. Y esto querrá decir que el número de van der Waerden $W(r, k)$ tiene que ser mayor que la cantidad de números que hemos coloreado. ¿Para qué valor de n ocurre esto? Podemos comprobar que

$$\frac{n^2}{(k-1)r^{k-1}} < 1 \iff n < r^{\frac{k-1}{2}} \sqrt{k-1},$$

lo que quiere decir que $W(r, k) \geq r^{\frac{k-1}{2}} \sqrt{k-1}$, como queríamos demostrar. ■

Encontrar cotas superiores resulta ser un problema mucho más complicado. Las primeras cotas que se conocieron, proporcionadas por la demostración del teorema, daban funciones que no se pueden definir con recursión primitiva (las funciones recursivas primitivas son un subconjunto de las funciones computables (aquellas para las que existe un algoritmo que las calcula); por dar una idea intuitiva, cualquiera de las funciones habituales o de sus composiciones (por ejemplo, exponenciales de exponenciales de exponenciales...) son funciones recursivas primitivas). La primera cota que se puede obtener con recursión primitiva se debe a Shelah [10], que proporcionó una demostración puramente combinatoria diferente del teorema, y no se encontró hasta el año 1988. La mejor cota general que se conoce en la actualidad la encontró Timothy Gowers [5] usando técnicas analíticas avanzadas, y da lugar a la cota

$$W(r, k) \leq 2^{2^{r^{2^{2^{k+9}}}}}.$$

Es evidente que estas cotas dejan un margen muy amplio, que no permite dar siquiera un orden de magnitud del valor de $W(r, k)$.

Hay otros muchos problemas que se pueden plantear en este ámbito. Por ejemplo, ¿se cumple el teorema de van der Waerden para progresiones aritméticas infinitamente largas? ¿Se puede extender a otros contextos, o generalizar a otras dimensiones? O sobre las progresiones aritméticas que aparecen, ¿son siempre las del color más «frecuente» en la coloración? La respuesta a esa última pregunta, en cierto sentido, es la que presentamos en la siguiente sección.

3. El teorema de Szemerédi

Uno puede pensar que una coloración de los números naturales con r colores es equivalente a tomar una partición de los enteros. Si, por ejemplo, coloreamos los primeros N naturales, entonces obtenemos conjuntos $A_1, \dots, A_r \subseteq \{1, \dots, N\}$, uno por cada color, que son disjuntos y cubren los primeros N naturales. Si sabemos que estos primeros números van a contener una k -PA monocromática, eso quiere decir que existe una progresión aritmética en uno de los conjuntos A_i . Pero ¿en cuál?

Decimos que un conjunto $A \subseteq \{1, 2, \dots, N\}$ tiene densidad positiva si su tamaño es una fracción positiva de N , y que tiene densidad δ si δ es el valor de esa fracción. Unos años después de que van der Waerden presentase su teorema, Erdős y Turán [3] conjeturaron que, para cada valor de k y densidad δ fijos, existe un valor de N suficientemente grande para el cual cualquier conjunto $A \subseteq \{1, \dots, N\}$ de densidad al

menos δ contendrá una k -PA. Es decir, si se colorean suficientes números, entonces habrá progresiones aritméticas de longitud k de cualquiera de los colores que tengan una densidad mayor que δ , y esto será cierto para cualquier δ positivo. Esto implica que, en una partición de \mathbb{N} , todo subconjunto de densidad positiva contiene una k -PA, para cualquier valor de k fijo.

En primer lugar, se consiguió demostrar que la conjetura es cierta para progresiones aritméticas de longitud 3 (este resultado se conoce como el teorema de Roth [9], y pasaron más de quince años entre la formulación de la conjetura y este primer resultado). En total pasaron casi cuarenta años hasta que Szemerédi dio una respuesta general positiva a esta conjetura, primero resolviendo el caso $k = 4$ [12] y más tarde el caso general [13], consiguiendo así un teorema que se considera uno de los resultados más relevantes en combinatoria. Una formulación exacta puede ser como la que presentamos aquí:

Teorema 3. *Sea $k \geq 3$ un entero y sea $\delta > 0$ un número real. Entonces, existe un entero $N = N(k, \delta)$ tal que para todo $n \geq N$, todo conjunto $A \subseteq \{1, \dots, n\}$ con $|A| \geq \delta n$ contiene una k -PA no trivial.*

La demostración del teorema de Szemerédi se basa principalmente en el lema de regularidad de Szemerédi (y algunas generalizaciones del mismo), una herramienta de teoría de grafos que ha dado lugar a aplicaciones muy diversas y que aún hoy constituye una de las principales técnicas para resolver problemas en combinatoria. Además de la primera demostración que consiguió Szemerédi, se han encontrado otras demostraciones usando técnicas muy diversas. Por ejemplo, Furstenberg [4] consiguió una demostración utilizando teoría ergódica, y Gowers [5] utilizó técnicas analíticas. Cada una de estas demostraciones ha dado lugar a herramientas que después han sido ampliamente utilizadas en otros contextos. Además, en los últimos años se han conseguido algunas nuevas demostraciones combinatorias.

Nótese que el teorema de van der Waerden se puede recuperar como un corolario del teorema de Szemerédi. De este modo, el teorema de Szemerédi es una generalización directa del de van der Waerden.

Demostración del teorema 1. Consideremos una coloración de los primeros n números naturales, que da una partición de $\{1, \dots, n\}$ en r conjuntos A_1, A_2, \dots, A_r . Por el principio del palomar, existe un color i tal que $|A_i| \geq \frac{n}{r}$, es decir, existe un conjunto con densidad mayor o igual que $\frac{1}{r}$. Ahora, sea $\delta = \frac{1}{r}$ y apliquemos el teorema de Szemerédi; como resultado, para un n suficientemente grande, cualquier subconjunto de $\{1, \dots, n\}$ con densidad al menos δ contendrá una k -progresión aritmética. En particular, A_i contiene una k -progresión aritmética, y todos los elementos de A_i son del mismo color, de modo que esta progresión es monocromática. ■

Una vez el teorema 3 está demostrado, uno se puede volver a hacer muchas preguntas relacionadas. Por ejemplo, uno puede plantearse la posibilidad de encontrar valores del mínimo número $N(k, \delta)$ que satisface el teorema. Parece evidente que este número será, en general, más grande que el número de van der Waerden asociado a los mismo parámetros (con la identificación $r = 1/\delta$), aunque intentar calcularlo, una vez más, resulta un problema muy complejo. Se pueden dar cotas de diversos tipos; una vez más, la mejor cota superior que tenemos se debe a Gowers [5], que dio la cota

$$N(k, \delta) \leq 2^{2^{(\delta^{-1})2^{k+9}}}$$

(de hecho, las cotas para los números de van der Waerden son una consecuencia de estas, sustituyendo δ por $\frac{1}{r}$).

Una forma de reinterpretar el teorema de Szemerédi es afirmar que el tamaño del conjunto $A \subseteq \{1, \dots, n\}$ más grande posible que no contenga k -PA es sublineal en n (asintóticamente). De este modo, una cuestión natural es preguntarse cuál es el mayor tamaño posible de este conjunto. Así, se define $r_k(N)$ como el mayor tamaño posible de un conjunto $A \subseteq \{1, \dots, N\}$ que no contiene k -PA, y existen diversos resultados que acotan el valor de este número.

Finalmente, se pueden considerar muchos problemas relacionados con este tipo de construcciones. Un ejemplo reciente es un aclamado resultado de Green y Tao [6], que demostraron que el conjunto de los números primos contiene progresiones aritméticas arbitrariamente largas.

Referencias

- [1] BEELER, Michael D. y O'NEIL, Patrick E. «Some new van der Waerden numbers». En: *Discrete Math.* 28.2 (1979), págs. 135-146. ISSN: 0012-365X. [https://doi.org/10.1016/0012-365X\(79\)90090-6](https://doi.org/10.1016/0012-365X(79)90090-6).
- [2] CHVÁTAL, Vašek. «Some unknown van der Waerden numbers». En: *Combinatorial Structures and their Applications (Proc. Calgary Internat. Conf., Calgary, Alta., 1969)*. Gordon y Breach, New York, 1970, págs. 31-33.
- [3] ERDŐS, Paul y TURÁN, Paul. «On some sequences of integers». En: *Journal of the London Mathematical Society* 11.4 (1936), págs. 261-264. <https://doi.org/10.1112/jlms/s1-11.4.261>.
- [4] FURSTENBERG, Harry. «Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions». En: *J. Analyse Math.* 31 (1977), págs. 204-256. ISSN: 0021-7670. <https://doi.org/10.1007/BF02813304>.
- [5] GOWERS, William T. «A new proof of Szemerédi's theorem». En: *Geom. Funct. Anal.* 11.3 (2001), págs. 465-588. ISSN: 1016-443X. <https://doi.org/10.1007/s00039-001-0332-9>.
- [6] GREEN, Ben y TAO, Terence. «The primes contain arbitrarily long arithmetic progressions». En: *Ann. of Math. (2)* 167.2 (2008), págs. 481-547. ISSN: 0003-486X. <https://doi.org/10.4007/annals.2008.167.481>.
- [7] KOURIL, Michal. «Computing the van der Waerden number $W(3, 4) = 293$ ». En: *Integers* 12 (2012), Paper No. A46, 13. ISSN: 1553-1732.
- [8] KOURIL, Michal y PAUL, Jerome L. «The van der Waerden number $W(2, 6)$ is 1132». En: *Experiment. Math.* 17.1 (2008), págs. 53-61. ISSN: 1058-6458. URL: <http://projecteuclid.org/euclid.em/1227031896>.
- [9] ROTH, Klaus F. «On certain sets of integers». En: *J. London Math. Soc.* 28 (1953), págs. 104-109. ISSN: 0024-6107. <https://doi.org/10.1112/jlms/s1-28.1.104>.
- [10] SHELAH, Saharon. «Primitive recursive bounds for van der Waerden numbers». En: *J. Amer. Math. Soc.* 1.3 (1988), págs. 683-697. ISSN: 0894-0347. <https://doi.org/10.2307/1990952>.
- [11] STEVENS, R. S. y SHANTARAM, R. «Computer-generated van der Waerden partitions». En: *Math. Comp.* 32.142 (1978), págs. 635-636. ISSN: 0025-5718. <https://doi.org/10.2307/2006173>.
- [12] SZEMERÉDI, Endre. «On sets of integers containing no four elements in arithmetic progression». En: *Acta Math. Acad. Sci. Hungar.* 20 (1969), págs. 89-104. ISSN: 0001-5954. <https://doi.org/10.1007/BF01894569>.
- [13] SZEMERÉDI, Endre. «On sets of integers containing no k elements in arithmetic progression». En: *Acta Arith.* 27 (1975). Collection of articles in memory of JuriíVladimirovič Linnik, págs. 199-245. ISSN: 0065-1036.
- [14] VAN DER WAERDEN, Bartel Leendert. «Beweis einer baudetschen vermutung». Alemán. En: *Nieuw Arch. Wiskd., II. Ser.* 15.2 (1927), págs. 212-216. ISSN: 0028-9825.

TEMat

El teorema de Müntz–Szász sobre la aproximación de funciones continuas

Daniel Eceizabarrena
BCAM - Basque Center for Applied
Mathematics
deceizabarrena@bcamath.org

Alejandro Mas Mas
Universidad Autónoma de Madrid
alejandro.mas@uam.es

Francisco Mengual Bretón
Universidad Autónoma de Madrid
francisco.mengual@uam.es

✉ María Soria Carro
The University of Texas at Austin
maria.soriac@math.utexas.edu

Resumen: El teorema de Weierstrass es un resultado clásico sobre la aproximación de funciones continuas mediante polinomios en intervalos cerrados y acotados de \mathbb{R} . En este artículo tratamos una generalización de dicho teorema que, en vez de polinomios, considera potencias cuyos exponentes satisfacen ciertas propiedades. Este resultado se conoce como el teorema de aproximación de Müntz–Szász. En primer lugar, introducimos teoría básica del análisis real y complejo, que será útil para probar los resultados principales y, a continuación, presentamos el teorema y la prueba dada por Szász.

Palabras clave: Müntz, Szász, aproximación, funciones continuas.

MSC2010: 30B60.

Recibido: 6 de febrero de 2017.

Aceptado: 24 de junio de 2017.

Agradecimientos: Queremos agradecer a José Luis Fernández, catedrático de la Universidad Autónoma de Madrid, por proponernos el tema del artículo y dirigirnos el trabajo.

Referencia: ECEIZABARRENA, Daniel; MAS MAS, Alejandro; MENGUAL BRETÓN, Francisco y SORIA CARRO, María. «El teorema de Müntz–Szász sobre la aproximación de funciones continuas». En: *TEMat*, 1 (2017), págs. 31-44. ISSN: 2530-9633. URL: <http://temat.anemat.com/articulo/2017-p31/>.

1. Breve biografía sobre los autores

1.1. Herman Müntz

Herman Müntz fue una persona muy activa tanto en su faceta de matemático como en lo que se refiere al contexto social más próximo a su propia situación. Además de trabajar en el campo de la teoría de la aproximación, trató durante su vida temas tan diversos como la geometría, ecuaciones en derivadas parciales, ecuaciones integrales y la teoría de números. Asimismo, desarrolló una importante actividad filosófica y divulgativa sobre el judaísmo. Sin embargo, su carrera profesional fue una constante de altibajos y llena de objetivos sin cumplir.

Con el nombre de Chaim Müntz nació en la ciudad de Łódź el 28 de agosto de 1884 en el seno de una familia burguesa judía. En aquella época, la ciudad formaba parte del Reino de Polonia, estado adherido al Imperio Ruso, y experimentaba un crecimiento demográfico sin precedentes en el marco de una potente industrialización. El joven Müntz, que hablaba polaco, alemán y ruso de manera fluida, cursó sus estudios primarios en su ciudad natal antes de trasladarse a Berlín, donde ingresó en la Friedrich-Wilhelms-Universität para estudiar matemáticas, ciencias naturales y filosofía. Durante sus años universitarios, contó con profesores de la talla de Frobenius, Landau y Schwarz, y se graduó en 1906.

La pedagogía fue un eje importante de su vida que empezó a desarrollar después de graduarse, ya que dedicó una parte considerable de sus siguientes años a la enseñanza privada de matemáticas. Paralelamente, combinando su interés por la filosofía y sus orígenes judíos, dedicó una parte importante de su tiempo a promover una reforma en la comunidad judía y una sociedad en la que este colectivo disfrutara de una mayor inclusión. En este contexto, publicó un libro titulado *Wir Juden* en 1907, donde defiende el sionismo socialista, una posición frecuente en la época. A lo largo de su vida escribió otros textos que no fueron publicados como libros por diferentes razones, pero que vieron la luz parcialmente en forma de artículos. Todas estas actividades no le impidieron obtener el título de doctor en 1910, bajo la supervisión de Schwarz y Schottky, con la calificación de *magna cum laude* gracias a un trabajo sobre ecuaciones en derivadas parciales en superficies minimales.

Durante los siguientes años, Müntz intentó conseguir la habilitación que le permitiría obtener un puesto universitario. Lo intentó en emplazamientos tan variados como en Múnich, Jerusalén, Gotinga, Giessen y El Cairo, sin éxito. Por ello, tuvo que compaginar la investigación matemática con otras actividades como la docencia en varias escuelas y la elaboración de reseñas de artículos científicos. De todos modos, y a pesar de su fracaso institucional continuado, el propio Müntz se sentía resarcido por el hecho de que a partir de 1927 colaboró con Albert Einstein. Aunque no publicaron ningún artículo conjuntamente, Einstein agradeció su colaboración en más de una publicación.

La ocasión que con tanto ahínco había perseguido llegó en 1929 cuando la Universidad Estatal de Leningrado le ofreció el cargo de catedrático de matemáticas. A raíz de esto, consiguió un notable reconocimiento en la Unión Soviética, como muestra el hecho de que fue uno de los cuatro enviados al Congreso Internacional de Matemáticas en Zúrich en 1932. En 1934 publicó un libro sobre ecuaciones integrales al que aún se hace referencia. Desgraciadamente, su situación en Rusia evolucionó de la misma manera que el clima político europeo: en 1937 fue abruptamente expulsado de la Unión, muy probablemente por su nacionalidad alemana obtenida en 1919. Müntz aterrizó en Tallín y finalmente en Suecia en 1938. Tan solo siete años después de conseguir su tan ansiado objetivo de estabilizarse en una universidad, vio cómo su logro se desvaneció, y en el intento de recuperar su estatus no volvió a tener éxito. Tampoco consiguió penetrar en el entorno de la comunidad matemática sueca, y trabajó en problemas matemáticos relacionados con la hemodinámica. Habiendo dedicado los últimos años a la enseñanza, murió en 1956 a la edad de 71 años.

1.2. Otto Szász

Otto Szász fue un prolífico matemático que investigó en muchas áreas del análisis matemático. Además de la teoría de la aproximación, sus campos de trabajo fueron, entre otros, las fracciones continuas, series de potencias acotadas, polinomios trigonométricos, series de Fourier y métodos de sumabilidad.

Szász nació en Alsószúcs, en la zona rural de Hungría en el Imperio Austrohúngaro (hoy Dolná Súča, Eslovaquia) el 11 de diciembre de 1884. Realizó sus estudios universitarios en la Universidad de Budapest

y en la Universidad de Gotinga durante los primeros años del siglo xx, donde fue alumno de ilustres matemáticos como Klein, Hilbert, Minkowski, Toeplitz y Herglotz. Después de graduarse, comenzó los estudios de doctorado bajo la dirección de Leopold Fejér y los culminó en 1911. Durante todos estos años, también visitó las universidades de Múnich y París.

Sin ninguna duda, Szász tuvo más éxito que Müntz en lo que a puestos académicos se refiere. Poco después de obtener el título de doctor, se incorporó a la propia Universidad de Budapest como *privatdozent*, un cargo de profesor en las universidades de habla alemana que, aunque sin constar estrictamente entre el personal, permitía a jóvenes investigadores preparar su habilitación. Tras pocos años allí, tuvo la oportunidad de trasladarse a Fráncfort, también como *privatdozent* hasta 1920 y como profesor titular desde entonces hasta 1933. Es en esta época cuando Szász produjo la mayor parte de los más de 120 artículos que publicó durante su vida.

Como en muchos otros casos, la situación política en Alemania le hizo abandonar el país rumbo a los Estados Unidos. Tras pasar por el MIT y por Brown, se asentó en la Universidad de Cincinnati, donde trabajó el resto de su vida. Murió durante unas vacaciones en Suiza el 19 de septiembre de 1952, a la edad de 67 años. Entre sus colaboradores y amistades personales se encontraban ilustres matemáticos como su director Fejér, Landau, Perron y Pringsheim. Cabe mencionar que en 1939 Szász recibió el premio Julius König de la Sociedad Matemática y Física de Hungría, y fue miembro de la American Mathematical Society.

2. El problema de la aproximación

Sin entrar en demasiados detalles, podríamos describir la teoría de la aproximación como un área del análisis en la que se busca aproximar una función, que podría ser *a priori* complicada, mediante ciertas funciones más sencillas y fácilmente manejables. A comienzos del siglo xix, la definición de una función se limitaba a proporcionar una fórmula explícita para el cálculo de sus valores. Sin embargo, hubo un momento en el que la comunidad matemática empezó a describir conjuntos de funciones mediante las propiedades que estas deberían cumplir. Así, se hacía más complicado describir todas las funciones mediante fórmulas. Pensemos por un momento en el claro y sencillo ejemplo de las funciones continuas: podemos encontrar funciones de lo más rebuscadas que difícilmente admitirían una expresión explícita manejable.

Este paso es el que hace que la teoría de la aproximación empiece a ser desarrollada casi por inercia. El primer resultado de entidad fue demostrado por Weierstrass en 1885 y hoy en día se conoce como el teorema de aproximación de Weierstrass:

Teorema 1 (teorema de aproximación de Weierstrass). *Toda función real continua definida en un intervalo compacto de \mathbb{R} puede ser aproximada mediante polinomios. En otras palabras, dados un intervalo compacto $I \subset \mathbb{R}$, una función $f \in C(I)$ y un $\epsilon > 0$, existe un polinomio p tal que $|f(t) - p(t)| < \epsilon$, para todo $t \in I$.*

Este teorema fue el punto de inicio para una fructífera teoría. Su importancia se ve reflejada en la cantidad de nuevas demostraciones que se dieron a conocer en los siguientes años utilizando técnicas muy variadas y firmadas por matemáticos como Picard, Fejér, Landau, de la Vallée Poussin, Runge, Phragmén, Lebesgue, Volterra y Bernstein. En 1937, Stone generalizó el teorema a funciones continuas en espacios compactos X y probó que toda subálgebra de $C(X)$ que contenga la función 1 y que separe puntos es densa en $C(X)$. Este resultado toma el nombre de teorema de Stone-Weierstrass.

En 1912, en el Congreso Internacional de Matemáticos de Cambridge, Bernstein planteó un problema a partir del resultado de Weierstrass cuando preguntó por la condición que un conjunto de números positivos $\{\lambda_n\}_{n \in \mathbb{N}}$ debía cumplir para que el conjunto de las combinaciones lineales finitas de $\{t^{\lambda_n}\}_{n \in \mathbb{N}}$ fuera denso en $C([0, 1])$. El propio Bernstein había sido capaz de dar algunos resultados parciales y conjeturó, con acierto, que la suma armónica $\sum_{n \in \mathbb{N}} 1/\lambda_n$ era decisiva. Tan solo dos años después, en 1914, Müntz confirmó la conjetura demostrando lo que hoy conocemos como el teorema de Müntz-Szász.

Teorema 2 (teorema de Müntz-Szász). *Sea $\{\lambda_n\}_{n \in \mathbb{N}}$ una sucesión estrictamente creciente de números positivos. Entonces, el conjunto de las combinaciones lineales finitas de las funciones $1, t^{\lambda_1}, t^{\lambda_2}, t^{\lambda_3}, \dots$ es denso en $C([0, 1])$ si y solo si $\sum_n 1/\lambda_n = +\infty$.*

En 1916, Szász publicó un artículo en el que completaba, mejoraba y simplificaba la demostración de Müntz. La prueba de Müntz utiliza técnicas de variable real y se basa en estimar la distancia entre una función continua cualquiera a ciertos subespacios finitos de polinomios viendo que esta distancia puede hacerse tan pequeña como se quiera. Nosotros veremos la prueba de Szász, la cual usa técnicas de variable compleja y ciertos argumentos del análisis funcional.

3. Preliminares

Antes de demostrar el teorema de Müntz-Szász debemos introducir una serie de conceptos y resultados previos que serán necesarios para una mejor comprensión de la sección 4. La referencia principal que hemos seguido es el libro de Rudin [4].

3.1. Introducción al análisis funcional

El análisis funcional es una rama de las matemáticas, en particular del análisis, que se basa fundamentalmente en el estudio de los espacios normados completos, también conocidos como los *espacios de Banach*. Esta especialidad tiene muchas aplicaciones en diversas áreas como, por ejemplo, en el análisis armónico y en las ecuaciones diferenciales.

Un **espacio normado** es un espacio vectorial sobre un cuerpo \mathbb{F} ($= \mathbb{R}$ o \mathbb{C}) dotado de una **norma**, es decir, una aplicación $\|\cdot\|_X : X \rightarrow [0, +\infty)$ que cumple las siguientes propiedades:

1. $\|x\|_X \geq 0 \forall x \in X$ y $\|x\|_X = 0$ si y solo si $x = 0$.
2. $\|\lambda x\|_X = |\lambda| \|x\|_X, \forall x \in X \forall \lambda \in \mathbb{F}$.
3. $\|x + y\|_X \leq \|x\|_X + \|y\|_X, \forall x, y \in X$.

Decimos que una sucesión $\{x_n\}_{n \in \mathbb{N}} \subset X$ **converge** en X (con la topología de la norma) si existe un elemento $x \in X$ tal que

$$\lim_{n \rightarrow \infty} \|x_n - x\|_X = 0.$$

Un espacio normado X es **completo** si las sucesiones de Cauchy son convergentes. Recordamos que una **sucesión de Cauchy** $\{x_n\}_{n \in \mathbb{N}}$ en X satisface que dado $\epsilon > 0$, existe $N \in \mathbb{N}$ tal que

$$\|x_n - x_m\|_X < \epsilon, \quad \forall n, m > N.$$

Un ejemplo básico de espacio de Banach es el espacio de las funciones continuas $C(I)$ con la norma

$$\|f\|_{C(I)} = \sup_{x \in I} |f(x)|,$$

donde $I = [a, b]$ es un intervalo compacto en \mathbb{R} . Es fácil comprobar que $\|\cdot\|_{C(I)}$ define una norma en $C(I)$. Claramente, el supremo se alcanza al ser I compacto y f continua. Además, la convergencia en $C(I)$ es la convergencia uniforme usual.

Como segundo ejemplo muy interesante vamos a introducir los **espacios de Lebesgue**. Estos espacios se definen a partir de la medida de Lebesgue. Recordemos que la medida exterior de Lebesgue de un subconjunto cualquiera A de \mathbb{R}^n se define como

$$m^*(A) = \inf \left\{ \sum_{j=1}^{\infty} l(Q_j)^n : Q_j \text{ } n\text{-cubos con } j \geq 1, A \subseteq \bigcup_{j=1}^{\infty} Q_j \right\},$$

donde $l(Q)$ denota la longitud del lado del cubo Q . Es decir, $m^*(A)$ es la medición más fina que podemos hacer cubriendo A con n -cubos. Asimismo, se dice que un subconjunto E de \mathbb{R}^n es **medible Lebesgue** si para todo subconjunto A de \mathbb{R}^n se verifica que

$$m^*(A) = m^*(A \cap E) + m^*(A \cap E^c).$$

Cabe destacar que la familia formada por todos los subconjuntos medibles Lebesgue forman una σ -álgebra, y la restricción de m^* a dicha σ -álgebra se conoce como la medida de Lebesgue, m . En este contexto, se dice que una función $f : \mathbb{R}^n \rightarrow \mathbb{C} \cong \mathbb{R}^2$ es **medible** (Lebesgue) si la preimagen de cualquier conjunto medible por f es medible.

Ya estamos en disposición de definir los espacios de Lebesgue $L^p(\Omega)$, con $1 \leq p \leq +\infty$, donde Ω es un subconjunto medible de \mathbb{R}^n . Distinguimos dos casos: si $1 \leq p < +\infty$, entonces

$$L^p(\Omega) = \{f \text{ medible} : \|f\|_{L^p(\Omega)} < +\infty\},$$

donde la norma viene dada por

$$\|f\|_{L^p(\Omega)} := \|f\|_p = \left(\int_{\Omega} |f(x)|^p dx \right)^{1/p}.$$

Por otro lado, si $p = +\infty$, entonces

$$L^\infty(\Omega) = \{f \text{ medible} : \|f\|_{L^\infty(\Omega)} < +\infty\},$$

donde la norma es la del **supremo esencial**, es decir,

$$\|f\|_{L^\infty(\Omega)} := \|f\|_\infty = \operatorname{ess\,sup}_{x \in \Omega} |f(x)| = \inf\{C > 0 : |f(x)| \leq C, \text{ a. e. } x \in \Omega\}.$$

Una observación importante es que los elementos de $L^p(\Omega)$ son, en realidad, clases de equivalencia $[f]$, donde $[f] = \{g : \Omega \rightarrow \mathbb{F} : f = g \text{ a. e. } x \in \Omega\}$, dado que si $E \subseteq \Omega$ es un conjunto medible, entonces

$$\int_E |f(x)|^p dx = \int_E |g(x)|^p dx.$$

Notamos también que si $\Omega = I$ y $f \in C(I)$, entonces el supremo esencial no es más que el supremo, es decir, $\|f\|_{C(I)} = \|f\|_{L^\infty(I)}$. Además, como I es compacto, tenemos que $C(I) \subset L^p(I)$, para todo $1 \leq p \leq +\infty$. En particular, $\|\cdot\|_{L^p(I)} \leq c \|\cdot\|_{C(I)}$, donde $c > 0$ es una constante que depende de la medida del intervalo. De hecho, todavía podemos decir más, y es que las funciones continuas son densas en los espacios de Lebesgue, excluyendo el caso $p = \infty$ [4, teorema 3.14].

Teorema 3. Sea $1 \leq p < +\infty$ y sea $f \in L^p(I)$, con $I = [a, b]$ compacto. Entonces, existe una sucesión de funciones continuas $\{f_n\}_{n \in \mathbb{N}} \subset C(I)$ tal que

$$\lim_{n \rightarrow +\infty} \|f_n - f\|_{L^p(I)} = 0.$$

Equivalentemente, $\overline{C(I)}^{L^p(I)} = L^p(I)$.

Este resultado es de gran utilidad ya que, como veremos en el corolario 15, en muchas ocasiones nos resultará más fácil trabajar con funciones continuas y a continuación, por densidad, obtener los resultados para las funciones en los espacios $L^p(I)$. El caso $p = \infty$ no es cierto dado que el límite uniforme de funciones continuas es continuo y, por tanto, cualquier función no continua que sea esencialmente acotada no se puede aproximar por funciones continuas con la norma del supremo esencial.

Los ejemplos de espacios de Banach que acabamos de ver son, en particular, espacios de funciones. Así, llamamos **funcional** a una aplicación que actúa sobre un espacio de funciones X y que toma valores sobre otro espacio normado Y , es decir, $L : X \rightarrow Y$. Decimos que el funcional L es **lineal** si

1. $L(x + x') = L(x) + L(x'), \forall x, x' \in X$, y
2. $L(\lambda x) = \lambda L(x), \forall x \in X$ y $\forall \lambda \in \mathbb{F}$,

y que el funcional L es **continuo** si para cualquier sucesión $\{x_n\}_{n=1}^\infty \subset X$ que converge a $x \in X$ se tiene que

$$\lim_{n \rightarrow +\infty} \|L(x_n) - L(x)\|_Y = 0.$$

Si L es lineal, como $L(x_n) - L(x) = L(x_n - x)$, está claro que L es continuo si y solo si lo es en $x = 0$. Además, no es difícil probar que esto equivale a que L esté **acotado**, es decir, a que exista una constante $c > 0$ tal que

$$\|L(x)\|_Y \leq c\|x\|_X, \quad \forall x \in X.$$

Se define la norma de L como la menor constante que satisface esta propiedad, es decir,

$$\|L\| = \sup_{\substack{x \in X \\ x \neq 0}} \frac{\|L(x)\|_Y}{\|x\|_X} = \sup_{\|x\|_X=1} \|L(x)\|_Y.$$

Por ejemplo, consideremos el funcional **multiplicación por un escalar**,

$$\begin{aligned} L_r : X &\longrightarrow X \\ x &\longmapsto L_r(x) = rx, \end{aligned}$$

para un cierto $r \in \mathbb{F}$. Claramente, L_r es lineal y continuo siendo su norma

$$\|L_r\| = \sup_{\|x\|_X=1} \|rx\|_X = |r|.$$

En los espacios de Banach, una gran parte del estudio involucra al **espacio dual**, es decir, el espacio de los funcionales lineales continuos sobre el cuerpo de escalares del espacio base X ($Y = \mathbb{F}$), y lo denotamos por

$$X^* = \{L : X \rightarrow \mathbb{F} \text{ tal que } L \text{ lineal y continuo}\}.$$

Por ejemplo, el dual de los espacios de Lebesgue $L^p(\Omega)$, con $1 \leq p < +\infty$, se identifica con $L^q(\Omega)$, donde $\frac{1}{p} + \frac{1}{q} = 1$ [4, teorema 6.16], a través del isomorfismo isométrico $T : L^q(\Omega) \rightarrow L^p(\Omega)^*$ definido como $T(g) = L_g$, donde

$$L_g(f) = \int_{\Omega} f(x)g(x) dx.$$

Por otro lado, el dual de las funciones continuas $C(I)$ no es un espacio de funciones, pero lo sabemos caracterizar gracias al teorema de representación de Riesz-Markov-Kakutani [4, teorema 6.19].

Teorema 4 (teorema de representación de Riesz-Markov-Kakutani). *El espacio de las medidas de Borel complejas regulares, $M(I)$, es el espacio dual de $C(I)$ vía*

$$\begin{aligned} M(I) &\longrightarrow C(I)^* \\ \mu &\longmapsto \left(\varphi \mapsto \langle \varphi, \mu \rangle = \int_0^1 \varphi d\mu \right) = \langle \cdot, \mu \rangle. \end{aligned}$$

Finalmente, para terminar esta introducción, presentamos un resultado que se obtiene como consecuencia de uno de los grandes teoremas del análisis funcional, conocido como el teorema de Hahn-Banach, el cual trata sobre la extensión de aplicaciones lineales continuas [4, teorema 5.16].

Teorema 5. *Sea M un subespacio lineal de un espacio lineal normado X , y sea $x_0 \in X$. Entonces, x_0 está en la clausura \overline{M} de M si y solo si no existe ningún funcional lineal acotado, T , definido en X tal que $T(x) = 0$ para todo $x \in M$ pero $T(x_0) \neq 0$.*

3.2. Algunos resultados del análisis complejo y de Fourier

En esta sección suponemos que el lector tiene conocimientos básicos de análisis complejo. En primer lugar, veamos unos resultados sobre funciones holomorfas. Denotamos por $\mathcal{H}(\Omega)$ al conjunto formado por las funciones holomorfas en un abierto Ω del plano complejo. Recordamos brevemente que una función holomorfa en el abierto Ω es una función derivable (respecto de la variable compleja) en todos los puntos de Ω , lo cual equivale a que f se pueda expresar localmente como serie de potencias (es decir, f es analítica en Ω). Asimismo, decimos que una función f es meromorfa en Ω si es holomorfa salvo en un conjunto de puntos aislados de Ω .

El siguiente teorema es un caso particular del teorema 15.23 del libro de Rudin [4] restringido a la familia H^∞ , la cual está formada por las funciones holomorfas en el disco unidad \mathbb{D} , que además son acotadas.

Teorema 6. Si $f \in H^\infty$ tiene ceros $\alpha_1, \alpha_2, \dots$ en \mathbb{D} y si

$$\sum_{n=1}^{\infty} 1 - |\alpha_n| = +\infty,$$

entonces $f(z) = 0$ para todo $z \in \mathbb{D}$.

La demostración del resultado que sigue se puede encontrar en el libro de Rudin [4, teorema 15.6].

Teorema 7. Sea Ω un dominio en el plano complejo. Supongamos que $f_n \in \mathcal{H}(\Omega)$ para $n = 1, 2, 3, \dots$, siendo f_n no idénticamente nula en ninguna componente de Ω , y

$$\sum_{n=1}^{\infty} |1 - f_n(z)|$$

converge uniformemente en subconjuntos compactos de Ω . Entonces, el producto

$$f(z) = \prod_{n=1}^{\infty} f_n(z)$$

converge uniformemente en subconjuntos compactos de Ω . Así, $f \in \mathcal{H}(\Omega)$.

El siguiente teorema es un resultado básico del análisis complejo [4, pág. 211], el cual afirma que una función holomorfa en un dominio Ω está determinada por sus valores en cualquier conjunto que tenga algún punto de acumulación en Ω . Con más detalle:

Teorema 8 (principio de identidad). Si f y g son dos funciones holomorfas en un dominio Ω y si $f(z) = g(z)$ para todo z en algún conjunto que tiene un punto de acumulación en Ω , entonces $f(z) = g(z)$ para todo $z \in \Omega$.

En la prueba del teorema de Müntz-Szász necesitaremos ver que ciertas funciones son holomorfas. Los siguientes resultados clásicos del análisis complejo nos serán de gran ayuda para este propósito.

Teorema 9 (teorema de Morera). Dada una función compleja continua, f , definida en un dominio Ω de \mathbb{C} , si

$$\oint_C f(z) dz = 0$$

para cada camino C cerrado y C^1 a trozos con soporte en Ω , entonces f es holomorfa en Ω .

Teorema 10 (teorema de Cauchy). Si f es holomorfa en un dominio simplemente conexo, entonces

$$\oint_C f(z) dz = 0$$

para cada camino cerrado rectificable C en el dominio.

Una vez comprobada la analiticidad, la fórmula de representación de Cauchy nos permitirá reescribir nuestra función de una manera específica.

Teorema 11 (fórmula de Cauchy). Si f es holomorfa dentro y en la frontera C de un dominio simplemente conexo, entonces, para cada z_0 dentro de C ,

$$f(z_0) = \frac{1}{2\pi i} \oint_C \frac{f(z)}{z - z_0} dz.$$

Para terminar con los preliminares, vamos a introducir muy brevemente dos tipos de funciones, que nos serán de gran utilidad en la demostración del teorema de Müntz-Szász. La primera es la **transformada de Fourier** de una función $g \in L^1(\mathbb{R}^n)$, que se define como

$$\widehat{g}(\xi) = \int_{\mathbb{R}^n} g(x) e^{-i x \cdot \xi} dx, \quad \xi \in \mathbb{R}^n.$$

Se satisface la siguiente propiedad básica.

Proposición 12. Si $g \in L^1(\mathbb{R}^n)$, entonces su transformada de Fourier \widehat{g} es acotada y continua.

Demostración. La acotación se deduce fácilmente de la definición, ya que

$$|\widehat{g}(\xi)| \leq \int_{\mathbb{R}^n} |g(x)| dx = \|g\|_{L^1(\mathbb{R}^n)}.$$

Por otra parte,

$$\lim_{\xi \rightarrow \xi_0} \widehat{g}(\xi) = \lim_{\xi \rightarrow \xi_0} \int_{\mathbb{R}^n} g(x) e^{-ix \cdot \xi} dx,$$

y como el valor absoluto del integrando esta acotado por $|g| \in L^1(\mathbb{R}^n)$, el teorema de la convergencia dominada nos permite meter el límite dentro de la integral y, por tanto,

$$\lim_{\xi \rightarrow \xi_0} \widehat{g}(\xi) = \int_{\mathbb{R}^n} g(x) e^{-ix \cdot \xi_0} dx = \widehat{g}(\xi_0). \quad \blacksquare$$

El segundo tipo de funciones son las *transformaciones de Möbius*. Una de las transformaciones más comunes viene dada por

$$\begin{aligned} \mathbb{D} &\longrightarrow \mathbb{H}_0 \\ z &\longmapsto \frac{1+z}{1-z}, \end{aligned}$$

donde $\mathbb{H}_\alpha = \{z \in \mathbb{C} : \operatorname{Re} z > \alpha\}$. Observemos que esta función está bien definida, ya que

$$\operatorname{Re} \left(\frac{1+z}{1-z} \right) = \frac{1-|z|^2}{|1-z|^2} > 0 \quad \forall z \in \mathbb{D}.$$

La transformación de Möbius inversa es

$$\begin{aligned} \mathbb{H}_0 &\longrightarrow \mathbb{D} \\ w &\longmapsto \frac{w-1}{w+1}. \end{aligned}$$

Para nuestro caso de interés consideraremos la transformación de Möbius

$$(1) \quad \begin{aligned} \mathbb{H}_{-1} &\longrightarrow \mathbb{D} \\ z &\longmapsto \frac{a-1-z}{a+1+z} = -\frac{\frac{z+1}{a} - 1}{\frac{z+1}{a} + 1}, \end{aligned}$$

para un $a > 0$ arbitrario.

4. Teorema de Müntz-Szász

Para demostrar el teorema 2, lo reformularemos de una manera más práctica y ligeramente más completa.

Teorema 13. Sean $0 < \lambda_1 < \lambda_2 < \lambda_3 < \dots$ y X la clausura en $C([0, 1])$ del conjunto de las combinaciones lineales finitas de las funciones

$$1, t^{\lambda_1}, t^{\lambda_2}, t^{\lambda_3}, \dots$$

- (a) Si $\sum_n 1/\lambda_n = +\infty$, entonces $X = C([0, 1])$.
- (b) Si $\sum_n 1/\lambda_n < +\infty$ y $\lambda \notin \{\lambda_n\}$, $\lambda \neq 0$, entonces $t^\lambda \notin X$.

Es inmediato ver que, en efecto, el teorema de Müntz-Szász (teorema 2) se deduce del teorema 13. Antes de enfrentarnos directamente a su demostración, probaremos la siguiente proposición que, como veremos, será de gran utilidad.

Proposición 14. Si $\sum 1/\lambda_n = \infty$, μ es una medida de Borel compleja en I y T es el funcional lineal acotado en $C(I)^* \cong M(I)$ asociado a μ tal que

$$(2) \quad T(t^{\lambda_n}) = \int_0^1 t^{\lambda_n} d\mu(t) = 0, \quad \forall n = 1, 2, \dots,$$

entonces

$$(3) \quad T(t^k) = \int_0^1 t^k d\mu(t) = 0, \quad k = 1, 2, \dots$$

Demostración. Para ver (3) usaremos algunas técnicas de análisis complejo, las cuales hemos presentado en la sección 3. Asumamos que (2) se cumple. Como los integrandos en (2) y (3) se anulan en $t = 0$, podemos asumir que la medida μ se concentra en $I = (0, 1]$.

Consideremos la función

$$(4) \quad f(z) = \int_I t^z d\mu(t) = \int_I e^{z \log t} d\mu(t),$$

que está bien definida en el semiplano complejo derecho \mathbb{H}_0 :

$$(5) \quad |f(z)| \leq \int_I |e^{z \log t}| d|\mu|(t) = \int_I e^{\operatorname{Re} z \log t} d|\mu|(t) = \int_I t^{\operatorname{Re} z} d|\mu|(t) \leq \|\mu\| < +\infty,$$

ya que, si $\operatorname{Re} z \geq 0$, entonces $|t^{\operatorname{Re} z}| \leq 1$ para todo $t \in I$.

Más aún, vamos a ver que f es holomorfa en \mathbb{H}_0 . Para ello nos serviremos del teorema de Morera (teorema 9). El primer paso es probar que f es continua. Como

$$f(z) - f(z_0) = \int_I t^z d\mu(t) - \int_I t^{z_0} d\mu(t) = \int_I (t^z - t^{z_0}) d\mu(t),$$

entonces

$$|f(z) - f(z_0)| \leq \int_I |t^z - t^{z_0}| d|\mu|(t).$$

Fijemos $\varepsilon > 0$. Como $(t, z) \mapsto t^z$ es una función continua en $[0, 1] \times \mathbb{H}_0$ (uniformemente en t , por ser $[0, 1]$ compacto), existe $\delta(\varepsilon) > 0$ tal que, si $|z - z_0| < \delta$, entonces $|t^z - t^{z_0}| < \varepsilon$, para todo $t \in I$. Así,

$$|f(z) - f(z_0)| \leq \varepsilon \int_I d|\mu|(t) = \varepsilon \|\mu\|,$$

que prueba la continuidad de f .

Ahora, sea γ un camino cerrado de clase C^1 en \mathbb{H}_0 . Entonces,

$$\oint_{\gamma} f(z) dz = \oint_{\gamma} \int_I t^z d\mu(t) dz = \int_I \oint_{\gamma} t^z dz d\mu(t) = 0.$$

La última igualdad se debe a que t^z es una función holomorfa, por lo que podemos aplicar el teorema de Cauchy (teorema 10). El intercambio en el orden de integración es legítimo por el teorema de Fubini, ya que

$$\oint_{\gamma} \int_I |t^z| d|\mu|(t) d|z| = \oint_{\gamma} \int_I t^{\operatorname{Re} z} d|\mu|(t) d|z| \leq \oint_{\gamma} \int_I d|\mu|(t) d|z| = \|\mu\| L(\gamma) < +\infty,$$

donde $L(\gamma)$ denota la longitud de la curva γ . Entonces, por el teorema de Morera (teorema 9), concluimos que f es una función holomorfa en \mathbb{H}_0 . Por otro lado, hemos probado en (5) que f está acotada en \mathbb{H}_0 .

Consideremos ahora la función

$$g(z) = f\left(\frac{1+z}{1-z}\right), \quad z \in \mathbb{D}.$$

Observamos que g es la composición entre una transformación de Möbius del disco al semiplano derecho (ver sección 3.2) y nuestra función f . Entonces,

- $g \in \mathcal{H}(\mathbb{D})$,
- g es acotada en \mathbb{D} (por ser f acotada).

Es decir, $g \in H^\infty$. Además, la hipótesis (2) dice que

$$f(\lambda_n) = T(t^{\lambda_n}) = 0, \quad \forall n = 1, 2, \dots,$$

luego $g(\alpha_n) = 0$, donde $\alpha_n = \frac{\lambda_n - 1}{\lambda_n + 1}$.

Afirmamos que

$$\sum_{n=1}^{\infty} 1/\lambda_n = +\infty \implies \sum_{n=1}^{\infty} 1 - |\alpha_n| = +\infty.$$

De hecho,

$$\sum_{n=1}^{\infty} 1 - \left| \frac{\lambda_n - 1}{\lambda_n + 1} \right| = \sum_{n=1}^{\infty} \frac{\lambda_n + 1 - |\lambda_n - 1|}{\lambda_n + 1}$$

y analizamos dos posibles casos.

- Si $0 < \lambda_n < 1$ para todo $n \in \mathbb{N}$, entonces $\lambda_n + 1 - |\lambda_n - 1| = \lambda_n + 1 + \lambda_n - 1 = 2\lambda_n$. Consecuentemente,

$$\sum_{n=1}^{\infty} 1 - |\alpha_n| = \sum_{n=1}^{\infty} \frac{2\lambda_n}{\lambda_n + 1} = +\infty,$$

ya que $\frac{2\lambda_n}{\lambda_n + 1} \rightarrow 0$, cuando $n \rightarrow +\infty$.

- Si a partir de un cierto $m \in \mathbb{N}$ se tiene que $\lambda_n \geq 1$ para todo $n \geq m$, entonces $\lambda_n + 1 - |\lambda_n - 1| = \lambda_n + 1 - \lambda_n + 1 = 2$. Por lo tanto,

$$\sum_{n=1}^{\infty} 1 - |\alpha_n| \geq \sum_{n=m}^{\infty} \frac{2}{\lambda_n + 1} = +\infty.$$

Aplicando el teorema 6, deducimos que $g(z) = 0$, $\forall z \in \mathbb{D}$. En particular, también

$$T(t^k) = \int_I t^k d\mu(t) = f(k) = g\left(\frac{k-1}{k+1}\right) = 0, \quad k = 1, 2, \dots$$

Esto concluye la prueba. ■

A continuación, trataremos de probar el teorema 13. Como veremos, la primera parte de la demostración se deduce directamente de la proposición 14, simplificando su estructura. Deberemos trabajar un poco más para conseguir el resultado de la segunda parte.

Demostración del teorema 13. Comencemos probando (a). Por el teorema de aproximación de Weierstrass (teorema 1), es suficiente probar que X contiene todas las funciones t^k , con $k = 1, 2, 3, \dots$. Supongamos, por el contrario, que existe $k_0 \in \mathbb{N}$ tal que $t^{k_0} \notin X$. Claramente, $t^{k_0} \in C([0, 1])$ y así, por el teorema 5, existe un funcional lineal y acotado $T : C([0, 1]) \rightarrow \mathbb{R}$ tal que

$$T(t^{k_0}) \neq 0 \quad \text{y} \quad T|_X \equiv 0.$$

Como T satisface las hipótesis del teorema de representación de Riesz-Markov-Kakutani (teorema 4), existe una medida de Borel compleja μ tal que

$$T(\varphi) = \int_0^1 \varphi(t) d\mu(t) \quad \forall \varphi \in C([0, 1]),$$

satisfaciendo además que

$$(i) \quad T(t^{k_0}) = \int_0^1 t^{k_0} d\mu(t) \neq 0,$$

$$(ii) \quad T(t^{\lambda_n}) = \int_0^1 t^{\lambda_n} d\mu(t) = 0 \quad \forall n = 1, 2, \dots$$

La afirmación (a) sigue inmediatamente de la proposición 14. En efecto, esta nos dice que $T(t^{k_0})$ es nulo y no nulo a la vez. Por lo tanto, la afirmación del comienzo debe ser falsa, $t^k \in X$ para todo $k \in \mathbb{N}$ y el primer apartado queda demostrado.

Procedemos ahora a demostrar (b). Asumamos que $\sum_n 1/\lambda_n < +\infty$. La prueba consiste en construir un funcional $T = \langle \cdot, \mu \rangle$, como en el apartado (a), tal que $T(t^{\lambda_n}) = 0$ para todo $n = 0, 1, 2, \dots$ (con $\lambda_0 = 0$), mientras que $T(t^\lambda) \neq 0$ para cada $\lambda > 0$ con $\lambda \notin \{\lambda_n\}_{n \in \mathbb{N}_0}$.

En vista del apartado (a), esto requiere encontrar una función holomorfa acotada en \mathbb{H}_{-1} , cuyos ceros sean precisamente estos λ_n . Dicha función debe poder expresarse como un producto infinito. Aplicando la maquinaria de las transformaciones de Möbius de \mathbb{H}_{-1} a \mathbb{D} , vista en (1), unos términos convenientes para dicho producto infinito son

$$\frac{\lambda_n - z}{2 + \lambda_n + z}.$$

Finalmente, la función que consideraremos es

$$f(z) = \frac{z}{(2+z)^3} \prod_{n=1}^{\infty} \frac{\lambda_n - z}{2 + \lambda_n + z}, \quad z \in \mathbb{C} \setminus \{-2 - \lambda_n\}_{n \in \mathbb{N}_0},$$

donde hemos añadido un término adicional, $\frac{1}{(2+z)^2}$, para garantizar una propiedad de integrabilidad que veremos más adelante.

Comencemos probando que f es meromorfa en \mathbb{C} , con polos en $\{-2 - \lambda_n\}_{n \in \mathbb{N}_0}$. En vista del teorema 7, es suficiente ver que la serie con términos

$$1 - \frac{\lambda_n - z}{2 + \lambda_n + z} = \frac{2z + 2}{2 + \lambda_n + z}$$

converge absolutamente y uniformemente sobre compactos en $\mathbb{C} \setminus \{-2 - \lambda_n\}_{n \in \mathbb{N}}$. Fijemos un subconjunto compacto K de $\mathbb{C} \setminus \{-2 - \lambda_n\}_{n \in \mathbb{N}}$. Por compacidad, existe $\alpha > 0$ tal que $K \subset \mathbb{H}_{-\alpha}$. Como $\sum_n 1/\lambda_n$ es una serie convergente de términos (decrecientes) positivos, ha de ser $\lambda_n \rightarrow +\infty$, cuando $n \rightarrow +\infty$. Por lo tanto, existe $N \in \mathbb{N}$ tal que $-2 - \lambda_n < -\alpha$, para todo $n > N$. Sea $C_K = \sup_{z \in K} |2z + 2|$. Ahora, para cada $z \in K$, si $n \leq N$, se tiene que

$$\left| \frac{2z + 2}{2 + \lambda_n + z} \right| \leq \frac{C_K}{\inf_{w \in K} |2 + \lambda_n + w|} \leq \frac{C_K}{\text{dist}(-2 - \lambda_n, K)},$$

y, si $n > N$, se tiene que

$$\left| \frac{2z + 2}{2 + \lambda_n + z} \right| \leq \frac{C_K}{\inf_{w \in \mathbb{H}_{-\alpha}} |2 + \lambda_n + w|} \leq \frac{C_K}{2 + \lambda_n - \alpha}.$$

Por lo tanto, la hipótesis de convergencia de la serie $\sum_n 1/\lambda_n < +\infty$ y el criterio M de Weierstrass nos permiten concluir que la serie converge absolutamente y uniformemente sobre K .

Veamos ahora que f está acotada por 1 en \mathbb{H}_{-1} . La parte del producto infinito está acotada por 1 ya que cada término del producto es una transformación de Möbius de \mathbb{H}_{-1} a \mathbb{D} . El término fuera del producto infinito está acotado también por 1 en \mathbb{H}_{-1} , ya que una parte es $\frac{z}{2+z}$, que es también una transformación de Möbius de \mathbb{H}_{-1} a \mathbb{D} , y el otro término, como $|2+z| \geq |\text{Re}(2+z)| \geq 1$ en \mathbb{H}_{-1} , satisface

$$\frac{1}{|2+z|^2} \leq 1, \quad \forall z \in \mathbb{H}_{-1}.$$

Veamos ahora que f está en L^1 cuando la restringimos a $\text{Re } z = -1$. Por la expresión de $|f|$, el producto infinito se puede acotar por 1 ya que permanece en el disco. Entonces,

$$\int_{\mathbb{R}} |f(-1 + ir)| \, dr \leq \int_{\mathbb{R}} \frac{|-1 + ir|}{|1 + ir|^3} \, dr = \int_{\mathbb{R}} \frac{dr}{1 + r^2} = \pi,$$

luego $f \in L^1(\{\text{Re } z = -1\})$.

El siguiente paso es considerar un z_0 fijo, con $\text{Re } z_0 > -1$, y aplicar la fórmula de Cauchy (11) para $f(z_0)$ a lo largo de la semicircunferencia con centro en -1 y radio $R > 1 + |z_0|$, tomado desde $-1 - iR$ a $-1 + R$ y hasta $-1 + iR$, y después continuado por el segmento de $-1 + iR$ a $-1 - iR$, como se muestra en la figura 1.

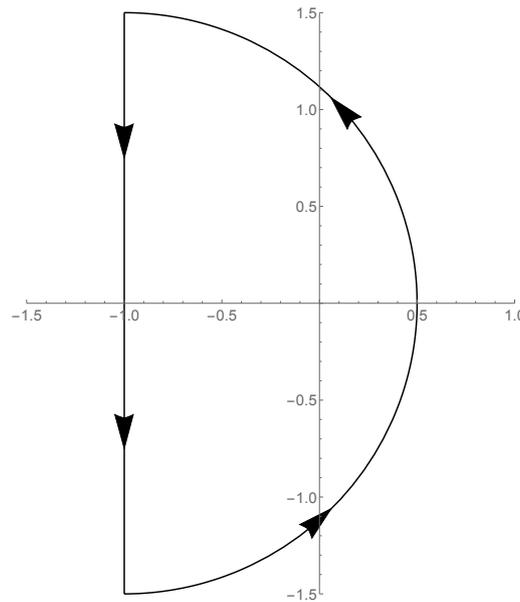


Figura 1: Camino en el que aplicamos la fórmula de Cauchy (en este dibujo, $R = 1,5$).

Llamando a esta curva C , una vez parametrizada, tenemos que

$$\begin{aligned} f(z_0) &= \frac{1}{2\pi i} \int_C \frac{f(z)}{z - z_0} \, dz \\ (6) \quad &= -\frac{1}{2\pi i} \int_{-R}^R \frac{f(-1 + is)}{-1 + is - z_0} i \, ds + \frac{1}{2\pi i} \int_{-\pi/2}^{\pi/2} \frac{f(-1 + R e^{i\theta})}{-1 + R e^{i\theta} - z_0} R i e^{i\theta} \, d\theta. \end{aligned}$$

Queremos ver que el segundo término de la integral a lo largo de la semicircunferencia, que denotamos por I_R , tiende a 0 cuando $R \rightarrow +\infty$. Utilizando de nuevo la cota dada por $|f(z)| \leq |z|/|2 + z|^3$, podemos escribir

$$\begin{aligned} |I_R| &\leq \frac{R}{2\pi} \int_{-\pi/2}^{\pi/2} \frac{|-1 + R e^{i\theta}|}{|1 + R e^{i\theta}|^3 | -1 + R e^{i\theta} - z_0|} \, d\theta \\ &\leq \frac{R}{2} \sup_{\theta \in (-\pi/2, \pi/2)} \frac{|-1 + R e^{i\theta}|}{|1 + R e^{i\theta}|^3 | -1 + R e^{i\theta} - z_0|}. \end{aligned}$$

La desigualdad triangular implica que

$$|-1 + R e^{i\theta}| \leq 1 + R \quad \text{y} \quad |1 + R e^{i\theta}| \geq R - 1.$$

Además, $|-1 + R e^{i\theta} - z_0| \geq R - |1 + z_0|$, luego

$$\frac{|-1 + R e^{i\theta}|}{|1 + R e^{i\theta}|^3 - |1 + R e^{i\theta} - z_0|} \leq \frac{R + 1}{(R - 1)^3(R - |1 + z_0|)}.$$

Como z_0 permanece fijo, el término $R - |1 + z_0|$ crece con R , por lo que puede ser acotado inferiormente por 1, si R es suficientemente grande. Entonces, a partir de dicho R ,

$$|I_R| \leq \frac{R}{2} \frac{R + 1}{(R - 1)^3} \rightarrow 0$$

cuando $R \rightarrow +\infty$. Aplicando esto y el teorema de la convergencia dominada sobre la integrabilidad de f en $\operatorname{Re} z = -1$, haciendo $R \rightarrow +\infty$ en (6) obtenemos

$$(7) \quad f(z_0) = \frac{1}{2\pi} \int_{\mathbb{R}} \frac{f(-1 + i s)}{1 + z_0 - i s} ds.$$

Recordando la identidad

$$(8) \quad \int_0^1 t^{z-is} dt = \frac{1}{z - is + 1} t^{z-is+1} \Big|_{t=0}^{t=1} = \frac{1}{z - is + 1} \quad (\operatorname{Re} z > -1),$$

observamos que (7) puede reescribirse como

$$(9) \quad f(z) = \int_0^1 t^z \left[\frac{1}{2\pi} \int_{-\infty}^{+\infty} f(-1 + i s) e^{-is \log t} ds \right] dt.$$

El intercambio en el orden de integración es legítimo: si el integrando en (9) es reemplazado por su valor absoluto, aparece una integral finita debido a que la restricción de f a la línea $\operatorname{Re} z = -1$ está en L^1 :

$$\begin{aligned} \frac{1}{2\pi} \int_0^1 \int_{-\infty}^{+\infty} |f(-1 + i s) e^{-is \log t} t^z| ds dt &= \frac{1}{2\pi} \int_0^1 t^{\operatorname{Re} z} \int_{-\infty}^{+\infty} |f(-1 + i s)| ds dt \\ &= \frac{M}{2\pi} \int_0^1 t^{\operatorname{Re} z} dt < +\infty \quad (\operatorname{Re} z > -1). \end{aligned}$$

Pongamos $g(s) = f(-1 + i s)$. Entonces, la integral interior en (9) es $\widehat{g}(\log t)$, donde \widehat{g} es la transformada de Fourier de g . Esta es una función continua y acotada en $(0, 1]$ (ver proposición 12). De este modo, tomando

$$d\mu(t) = \frac{1}{2\pi} \widehat{g}(\log t) dt$$

obtenemos una medida de Borel compleja que representa a f en la forma deseada (4), es decir,

$$f(z) = \int_I t^z d\mu(t),$$

que, además, por construcción se anula cuando $z = \lambda_n$. De manera equivalente, tenemos un funcional $T = \langle \cdot, \mu \rangle$ que se anula en t^{λ_n} y, por lo tanto, también lo hace en todas las combinaciones lineales de esas potencias. Pero no se anula en t^λ cuando $\lambda \neq \lambda_n$. En vista del teorema 5, t^λ no está en X cuando $\lambda \neq \lambda_n$ y el apartado (b) queda demostrado. ■

Si tratamos de extender el teorema de Müntz-Szász a los espacios de Lebesgue $L^p([0, 1])$, con $1 \leq p \leq +\infty$, nos damos cuenta de que el caso $p = \infty$ no es cierto. De hecho, acabamos de ver que

$$\overline{\operatorname{span}\{1, t^{\lambda_1}, t^{\lambda_2}, \dots\}}^{L^\infty([0,1])} = C([0, 1]) \subsetneq L^\infty([0, 1]).$$

Ahora bien, cuando $1 \leq p < +\infty$ sí que es posible obtener un resultado análogo de aproximación.

Corolario 15 (el teorema de Müntz-Szász para los espacios de Lebesgue). *Supongamos que*

$$0 < \lambda_1 < \lambda_2 < \lambda_3 < \dots$$

y sea X la clausura en $L^p([0, 1])$, con $1 \leq p < +\infty$, del conjunto de las combinaciones lineales finitas de las funciones

$$1, t^{\lambda_1}, t^{\lambda_2}, t^{\lambda_3}, \dots$$

(a) Si $\sum_n 1/\lambda_n = +\infty$, entonces $X = L^p([0, 1])$.

(b) Si $\sum_n 1/\lambda_n < +\infty$ y si $\lambda \notin \{\lambda_n\}$, $\lambda \neq 0$, entonces $t^\lambda \notin X$.

Demostración. (a) Recordamos que la convergencia uniforme es más fuerte que la convergencia en $L^p([0, 1])$, es decir, $\|\cdot\|_{L^p([0,1])} \leq \|\cdot\|_{L^\infty([0,1])}$. Llamamos $A = \text{span}\{1, t^{\lambda_1}, t^{\lambda_2}, \dots\} \subset L^p([0, 1])$. Trivialmente, $X = \overline{A}^{L^p([0,1])} \subset L^p([0, 1])$. La inclusión contraria se obtiene por el teorema de Müntz-Szász y la densidad de $C([0, 1])$ en $L^p([0, 1])$ (teorema 3):

$$L^p([0, 1]) = \overline{C([0, 1])}^{L^p([0,1])} = \overline{A}^{L^\infty([0,1])}^{L^p([0,1])} \subset \overline{A}^{L^p([0,1])}^{L^p([0,1])} = X.$$

(b) Si miramos la prueba que hemos dado anteriormente para el teorema de Müntz-Szász, notamos que $d\mu(t) = h(t) dt$, donde

$$h(t) = \frac{1}{2\pi} \widehat{g}(\log t), \quad t \in (0, 1].$$

Como \widehat{g} está acotada, $h \in L^\infty([0, 1]) \subset L^q([0, 1])$, con $1 < q \leq +\infty$, donde $\frac{1}{p} + \frac{1}{q} = 1$. Entonces, el funcional en $L^p([0, 1])^*$,

$$\begin{aligned} \langle \cdot, \bar{h} \rangle : L^p([0, 1]) &\longrightarrow \mathbb{R} \\ f &\longmapsto \langle f, \bar{h} \rangle = \int_0^1 f(t)h(t) dt, \end{aligned}$$

se anula en X , pero no en t^λ , para todo $\lambda > 0$, con $\lambda \notin \{\lambda_n\}_{n \in \mathbb{N}}$. Por tanto, por el teorema 5, X es un subespacio propio de $L^p([0, 1])$. ■

Referencias

- [1] LUXEMBURG, W.A.J. y KOREVAAR, J. «Entire functions and Müntz-Szász type approximation». En: *Trans. Amer. Math. Soc.* 157 (1971), págs. 23-37. <https://doi.org/10.2307/1995828>.
- [2] ORTIZ, Eduardo L. y PINKUS, Allan. «Herman Müntz: A Mathematicians Odyssey». En: *The Mathematical Intelligencer* 27 (2005), págs. 22-31. <https://doi.org/10.1007/BF02984810>.
- [3] PINKUS, Allan. «Density in Approximation Theory». En: *Surveys in Approximation Theory* 1 (2005), págs. 1-45. arXiv: math/0501328 [math.CA].
- [4] RUDIN, Walter. *Real and Complex Analysis*. Third Edition. McGraw-Hill, 1987. ISBN: 0070542341.
- [5] SZEGÖ, Gabor. «Otto Szász». En: *Bulletin of the American Mathematical Society* 60 (1954), págs. 261-263. <https://doi.org/10.1090/S0002-9904-1954-09794-X>.

TEMat

Influencia de los tamaños de clase en grupos finitos

✉ Víctor Manuel Ortiz Sotomayor
Universidad Politécnica de Valencia
vicorso@doctor.upv.es

Resumen: Presentamos una breve introducción sobre cómo los tamaños de las clases de conjugación de los elementos de un grupo finito afectan a su estructura. En particular, abordaremos dicho estudio en los grupos factorizados.

Palabras clave: grupos finitos, clases de conjugación, subgrupos de Sylow, grupos factorizados.

MSC2010: 20E45, 20D20, 20D40, 20D10.

Recibido: 28 de diciembre de 2016.

Aceptado: 14 de enero de 2017.

Agradecimientos: Este trabajo ha sido financiado por las «Ayudas para la contratación de personal investigador en formación de carácter predoctoral», otorgadas por la Generalitat Valenciana, España. Agradecer también a las profesoras Ana Martínez y María José Felipe (Universidad Politécnica de Valencia) por la constante orientación y motivación que en todo momento me han brindado durante el desarrollo de mi Trabajo Fin de Máster y mi actual tesis doctoral.

Referencia: ORTIZ SOTOMAYOR, Víctor Manuel. «Influencia de los tamaños de clase en grupos finitos». En: *TEMat*, 1 (2017), págs. 45-51. ISSN: 2530-9633. URL: <http://temat.anemat.com/articulo/2017-p45/>.

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

1. Introducción

En este trabajo solamente consideraremos grupos finitos. Además, supondremos que el lector conoce los conceptos básicos de un curso de Teoría de Grupos, por lo que algunas definiciones y pruebas serán omitidas y/o referidas. La notación que utilizaremos es la estándar en este contexto y está extraída principalmente del libro de Isaacs [9]. No obstante, para facilitar la lectura, recordamos a continuación alguna terminología específica.

Notación. Para un grupo finito G , sea x^G la **clase de conjugación** de un elemento $x \in G$, i.e., el conjunto formado por todos los elementos conjugados x en G . El tamaño de este conjunto lo denotaremos por $|x^G|$. Si p es un número primo, entonces p' es el conjunto de primos distintos de p . Además, diremos que $x \in G$ es un **p' -elemento** si su orden no es divisible por p , y diremos que es un **p -elemento** si su orden es una potencia de p . Finalmente, dado un número natural n , diremos que es **libre de cuadrados** si ningún número primo al cuadrado lo divide. ◀

La influencia que tienen los tamaños de las clases de conjugación sobre un grupo finito es un tema que ha sido extensamente estudiado durante los últimos 25 años. En concreto, se han obtenido resultados interesantes cuando se restringen las hipótesis a cierto subconjunto de elementos del grupo, como puedan ser los elementos de orden potencia de primo, los p' -elementos, etc. Un excelente resumen sobre esta materia puede encontrarse en el trabajo de Camina y Camina [3], y en el caso particular de que los tamaños de clase son libres de cuadrados también se puede consultar el *survey* de Felipe, Martínez Pastor y Ortiz Sotomayor [5]. Nuestro propósito es mostrar solamente una breve introducción al tema referido, mostrando ciertos resultados destacables con pruebas elementales e ilustrando algunos de ellos con ejemplos.

Una propiedad importante en la estructura de un grupo es el hecho de poseer algún factor directo. En este contexto, dado un p -subgrupo de Sylow P de un grupo finito G , diremos que G es **p -descomponible** si P es un factor directo de G , i.e., cuando $G = P \times H$ con H un p' -grupo (luego $H = O_{p'}(G)$ es el único p' -subgrupo de Hall de G). El siguiente teorema muestra una condición necesaria y suficiente para detectar dicha propiedad:

Teorema A ([2, lemma 3]). *Sean G un grupo finito y p un número primo divisor de $|G|$. Entonces, $|x^G|$ es una potencia de p para todo p -elemento $x \in G$ si y solo si G es p -descomponible.*

De hecho, ya en 1972 Camina [1] había probado otra condición equivalente para la p -descomponibilidad de un grupo G :

Teorema 1 ([1, lemma 1; corollary 1]). *Sean G un grupo finito y p un número primo divisor de $|G|$. Entonces:*

1. *Para todo p' -elemento $x \in G$, p no divide a $|x^G|$ si y solo si G es p -descomponible.*
2. *Para todo elemento $x \in G$, p no divide a $|x^G|$ si y solo si G es p -descomponible con p -subgrupo de Sylow abeliano.*

Observemos que la condición sobre los tamaños de clase en el punto 1 es dual a la del teorema A. Es más, en la sección 3 veremos que también existe una condición dual a la que aparece en punto 2 (ver sección 3, corolario 6). Cabe destacar que la segunda parte es una consecuencia inmediata de la primera (puede consultarse su prueba en el trabajo de Ortiz Sotomayor [12, teorema 2.2.1]).

Otra propiedad importante a reconocer en la estructura de un grupo es su resolubilidad. En este contexto, en 1990, Chillag y Herzog [4] probaron el siguiente resultado, mediante el uso de la clasificación de los grupos finitos simples (CGFS):

Teorema B ([4, proposition 5]). *Sea G un grupo finito. Supongamos que 4 no divide a ningún tamaño de clase de conjugación de G . Entonces, G es resoluble.*

Ejemplo 1. A diferencia de los resultados anteriores, es fácil encontrar ejemplos que muestran que el recíproco no es cierto, véase el grupo alternado de 4 letras. ◀

Es importante mencionar que en el artículo de Camina y Camina [2], los autores dan también una prueba evitando la CGFS. Destacar además que el teorema anterior ha sido extensamente generalizado (en el trabajo de Ortiz Sotomayor [12] se puede hallar gran parte del desarrollo).

Si nos centramos en el caso de los p -grupos, ya en 1951 Knoche [10] probó la siguiente condición equivalente entre sus tamaños de clase y el orden de su subgrupo derivado:

Teorema C ([10, Satz 2]). *Sea P un p -grupo. Entonces, para todo $x \in P$, p^2 no divide a $|x^P|$ si y solo si $|P'| \leq p$.*

En la próxima sección vamos a recopilar algunos resultados preliminares necesarios para probar, en la sección 3, los teoremas A, B (evitando la CGFS) y C. Finalmente destacamos que, como caso particular, en la última sección se combinará este estudio con los grupos factorizados, ya que es una línea de investigación novedosa y poco investigada.

2. Preliminares

Comenzamos esta sección con el siguiente resultado, el cual se le atribuye a Burnside:

Lema 2 ([8, corollary 4.16]). *Sea H un subgrupo de un grupo finito G . Si $G = \bigcup_{g \in G} H^g$, entonces $G = H$.*

Nota 1. De hecho, una propiedad básica es que un grupo no puede ser tampoco unión de dos subgrupos propios (véase el libro de Isaacs [8, problems 2.1]).

El siguiente lema es conocido como el «lema de Wielandt»:

Lema 3. *Sean G un grupo finito y p un primo divisor de $|G|$. Si $x \in G$ es un p -elemento con $|x^G|$ una potencia de p , entonces $x \in O_p(G)$.*

Demostración. Tenemos que, por el teorema de la órbita-estabilizador, $|x^G| = |G : C_G(x)|$ es una potencia de p . Además, escogemos un p -subgrupo de Sylow P de G tal que $x \in P$. Entonces $|G : P|$ es un p' -número y como ambos números son coprimos, podemos concluir que $G = P C_G(x)$ (ver [9, corollary X.12]).

Entonces tenemos que $\langle x^G \rangle = \langle x^{P C_G(x)} \rangle = \langle x^P \rangle \leq P$, luego es un p -subgrupo normal de G y, por tanto, $x \in \langle x^G \rangle \leq O_p(G)$. ■

Una propiedad que será importante en el teorema C es la que viene a continuación. Aunque aparece en el libro de Huppert [7, Hilfssatz 1.3 (a)], dicho libro está en alemán, luego incluimos aquí la prueba.

Lema 4. *Sean g y h elementos de un grupo finito G y n un número natural. Si $[[g, h], g] = 1$, entonces $[g^n, h] = [g, h]^n$.*

Demostración. Por inducción sobre n . El caso $n = 1$ es trivial. Supongamos cierto el caso $n - 1$ y probemos el caso n . Tenemos que $[g^n, h] = [g g^{n-1}, h] = [g, h]^{g^{n-1}} [g^{n-1}, h] = [g, h][g, h]^{n-1} = [g, h]^n$, donde la tercera igualdad se da debido a la inducción y a que $[[g, h], g] = 1$. ■

Finalmente, a la hora de trabajar con los tamaños de clase de subgrupos normales y cocientes de un grupo, es crucial el siguiente resultado sobre su divisibilidad:

Lema 5. *Sean N un subgrupo normal de un grupo G y $x \in G$. Entonces:*

1. $|x^N|$ divide a $|x^G|$, para todo $x \in N$.
2. $|(xN)^{G/N}|$ divide a $|x^G|$, para todo $x \in G$.

Demostración. Para el punto 1 tenemos que $|x^N| = |N : C_N(x)| = |N : N \cap C_G(x)| = |N C_G(x) : C_G(x)|$. Pero $N C_G(x) \leq G$, luego su orden divide al de G y, por tanto, el último término de la cadena de igualdades divide a $|G : C_G(x)|$.

Para el punto 2, claramente tenemos que $|(xN)^{G/N}| = |G/N : C_{G/N}(xN)|$ divide a $|G/N : C_G(x)N/N|$, pues $C_G(x)N/N \leq C_{G/N}(xN) \leq G/N$. Además, $|G/N : C_G(x)N/N| = |G : C_G(x)N|$ divide a su vez a $|G : C_G(x)|$, lo cual finaliza la prueba. ■

3. Pruebas de los teoremas A, B y C

Demostración del teorema A. Comencemos con la afirmación directa. Sea P un p -subgrupo de Sylow de G . Entonces, para cada $x \in P$, tenemos por el lema 3 que $x \in O_p(G)$, luego $P \leq O_p(G)$. La otra inclusión es trivial, por lo que $P = O_p(G)$. En consecuencia, $1 \trianglelefteq P \trianglelefteq G$, luego G es p' -separable y, por tanto, contiene algún p' -subgrupo de Hall H .

Tenemos que $G = PH$ con $P \trianglelefteq G$ y $P \cap H = 1$, luego basta con ver que $H \trianglelefteq G$. De nuevo, para cada $x \in P$, como $|x^G|$ no es divisible por ningún primo en p' , deducimos que existe algún $g_x \in G$ tal que $H^{g_x} \leq C_G(x)$. Como $G = PH$, tenemos que $g_x = ab$ con $a \in P$ y $b \in H$, y por tanto $H^{g_x} = H^a$, por lo que podemos asumir que $g_x \in P$. Luego $x \in P \cap C_G(H^{g_x}) = C_P(H^{g_x}) = C_P(H)^{g_x}$, para cada $x \in P$. Consecuentemente, $P \subseteq \bigcup_{g_x \in P} C_P(H)^{g_x} \subseteq P$, lo cual implica $P = \bigcup_{g_x \in P} C_P(H)^{g_x}$ y, por el lema 2, tenemos que $P = C_P(H)$. Por tanto, H es normal en $PH = G$ y entonces $G = P \times H$.

El recíproco es trivial pues P es el único p -subgrupo de Sylow de G y el producto de P con H es directo, luego cualquier p -elemento $x \in G$ va a cumplir que $H \leq C_G(x) \leq G$. Por tanto, $|x^G|$ divide a $|G : H| = |P|$. ■

Corolario 6 ([4, remark 3]). Sean G un grupo finito y p un número primo divisor de $|G|$. Entonces, $|x^G|$ es una potencia de p para todo elemento $x \in G$ si y solo si G es p -descomponible con p -complemento abeliano.

Demostración. Probaremos primero la implicación directa. Las hipótesis se cumplen en particular para todo p -elemento $x \in G$, luego podemos aplicar el teorema A y deducimos que el grupo es p -descomponible. Por tanto, tenemos que $G = P \times H$ con H un p' -subgrupo de Hall de G y P un p -subgrupo de Sylow de G . Veamos que H es abeliano: si $y \in H$, entonces por el lema 5, punto 1, tenemos que $|y^H|$ divide a $|y^G|$ (que es una potencia de p por hipótesis) y también divide trivialmente a $|H|$ (el cual es un p' -número), lo cual nos lleva a que $|y^H| = 1$ y, por tanto, $y \in Z(H)$ para todo $y \in H$. El recíproco es trivial, luego queda probado el resultado. ■

Pasamos ahora a probar el teorema B. Como veremos en su demostración, el siguiente resultado es clave. Aunque su prueba pueda parecer extensa, en su demostración solamente se usan conceptos elementales.

Lema 7 ([2, lemma 6]). Supongamos que G es un grupo simple y no abeliano de orden par. Entonces, existe un elemento de G con tamaño de clase divisible por 4.

Demostración. Sea $z \in G$ de orden 2, el cual existe por ser G de orden par. Tomemos $x \in G$ arbitrario y supongamos falso el resultado, luego podemos asumir que $|x^G| = |G : C_G(x)|$ es impar o igual a $2n$ con n impar. En el primer caso, existe un 2-subgrupo de Sylow P de G contenido en $C_G(x)$ y por tanto existe algún $g \in G$ tal que $z^g \in P \leq C_G(x)$. Veamos ahora qué pasa en el segundo caso.

Consideremos la siguiente acción de z sobre el conjunto de las $2n$ clases laterales de $C_G(x)$ en G , i.e., sobre $K := \{C = y_0 C, y_1 C, \dots, y_{2n-1} C\}$ con $C := C_G(x)$ e y_i representantes de las clases laterales, luego $|K| = 2n$. La acción viene dada por $z \cdot (y_j C) := (zy_j)C$. Deducimos que, o bien la acción es libre de puntos fijos, o bien $z^{y_j} \in C_G(x)$ para algún $0 \leq j \leq 2n - 1$. Veremos que si es libre de puntos fijos llegamos a una contradicción y, por tanto, siempre existirá algún $g \in G$ tal que $z^g \in C_G(x)$ o, equivalentemente, $x \in C_G(z)^g$. Luego

$$G \subseteq \bigcup_{g \in G} C_G(z)^g \subseteq G$$

y aplicando el lema 2 tenemos que $G = C_G(z)$, lo cual implica $z \in Z(G) = 1$, la contradicción final.

Observemos que la acción anterior viene asociada al homomorfismo

$$\begin{array}{rcl} \varphi: \langle z \rangle & \longrightarrow & \Sigma_{|K|} \\ z & \longmapsto & \varphi_z: K \longrightarrow K \\ & & y_j C \longmapsto (zy_j)C. \end{array}$$

Como $\text{Ker}(\varphi) \leq \langle z \rangle$ (el cual es isomorfo a un cíclico de orden 2), al ser la acción libre de puntos fijos deducimos que $\text{Ker}(\varphi) = 1$ y, por tanto, $\langle z \rangle$ es isomorfo a un subgrupo de $\Sigma_{|K|}$. Consecuentemente, al ser

z un elemento de orden 2, debe ser el producto de n transposiciones. Efectivamente, pues para cualquier $1 \leq j \leq 2n - 1$ tenemos que la acción consecutiva de z produce la cadena $y_j C \rightsquigarrow zy_j C \rightsquigarrow z^2 y_j C = y_j C$.

Por otro lado, G también actúa sobre K de la misma manera que lo hace $\langle z \rangle$. Además, el núcleo del homomorfismo asociado es un subgrupo normal de G , que es simple. Claramente no puede coincidir con G , pues en ese caso z fijaría a K , lo cual es una contradicción. Por tanto, G también es isomorfo a un subgrupo de $\Sigma_{|K|}$. Se sigue que $G \cap A_{|K|}$ es normal en G , luego $G \cap A_{|K|} = 1$ (si coincidiese con G , entonces $z \in G \leq A_{|K|}$, lo cual es una contradicción pues z es producto de n transposiciones, con n impar). Esto implica que G solamente contiene una única permutación par, la trivial. Pero en este caso se deduce que, dados $a, b \in G$ no triviales con $a \neq b$ (luego a y b van a ser permutaciones impares), necesariamente $ab = 1$ pues el producto es una permutación par y en G solamente hay una. Concluimos que dos elementos cualesquiera distintos entre sí en G son inversos, luego necesariamente G es un cíclico de orden 3, contradiciendo que no es abeliano. ■

Demostración del teorema B. Observemos primeramente que las hipótesis se heredan para subgrupos normales y cocientes de G : sea $N \trianglelefteq G$ y denotemos por $\bar{G} := G/N$. Entonces, por un lado, si $x \in N$ tenemos por el lema 5, punto 1, que $|x^N|$ divide a $|x^G|$, luego si 4 no divide a $|x^G|$ tampoco puede dividir a $|x^N|$; por otro lado, si $xN \in \bar{G}/N$ arbitrario, entonces por el lema 5, punto 2, tenemos que $|(xN)^{G/N}|$ divide a $|x^G|$, luego 4 tampoco divide en este caso.

Si el orden de G es impar, entonces es resoluble por el teorema de Feit-Thompson. Por tanto, podemos asumir que $|G|$ es par y que G no es abeliano. También podemos suponer por el lema anterior que G no es simple, luego que existe $1 < N < G$ normal en G . Entonces o bien N o G/N tienen orden par (en caso contrario, G tendría orden impar). Si ambos tienen orden par, como heredan las hipótesis, por inducción sobre $|G|$ tendríamos que ambos son resolubles y, por tanto, lo sería G también. Si solamente tiene alguno orden par, aplicamos inducción y el otro grupo sería resoluble por Feit-Thompson, lo cual finaliza la prueba. ■

Concluimos esta sección con la prueba del teorema C:

Demostración del teorema C. Empecemos por la implicación suficiente: si $|P'| = 1$, entonces P es abeliano y por tanto $|x^P| = 1$ para todo $x \in P$. Por tanto, podemos asumir que $|P'| = p$ y escoger un elemento $x \in P \setminus Z(P)$. Es fácil comprobar que la aplicación $\varphi: x^P \rightarrow [x, P] \leq P'$ definida por $\varphi(x^g) := [x, g] = x^{-1}x^g$ es biyectiva. Deducimos que $|x^P| = |[x, P]|$, el cual divide a $|P'| = p$.

Veamos la implicación directa: como $|P : C_P(x)| \leq p$ para todo $x \in P$, deducimos que $\Phi(P) \leq C_P(x)$ para todo $x \in P$, donde $\Phi(P)$ denota el subgrupo de Frattini de P . Por tanto $P' \leq \Phi(P) \leq Z(P)$. Entonces $P/Z(P)$ es isomorfo al cociente $(P/\Phi(P))/(Z(P)/\Phi(P))$, luego es elemental abeliano. Esto implica que $xZ(P)$ tiene orden p , luego $x^p \in Z(P)$ para todo $x \in P$. Aplicando el lema 4 (recordemos que $P' \leq Z(P)$), tenemos que $[x, y]^p = [x^p, y] = 1$ para todo $x, y \in P$, por tanto P' es abeliano y tiene todo elemento de orden p , luego es elemental abeliano. Veamos finalmente que solamente tiene un generador.

Sean $1 \neq [x, z]$ y $1 \neq [x', z']$ dos generadores de P' . Sea $a \in P \setminus (C_P(z) \cup C_P(z'))$, el cual existe por la nota 1. Además, $|P : C_P(z)| = p$ necesariamente, y como $C_P(z) < C_P(z)\langle x \rangle \leq P$, deducimos que $P = \langle x \rangle C_P(z)$. Por otro lado, como $a \in P = \langle x \rangle C_P(z)$, tenemos que $a = x^i t$ para algún número natural i y con $t \in C_P(z)$, luego

$$(1) \quad 1 \neq [a, z] = [x^i t, z] = [x^i, z]^t [t, z] = [x^i, z] = [x, z]^i,$$

donde la tercera y cuarta igualdades se dan debido a que $P' \leq Z(P)$. Por otro lado, también tenemos que $P = C_P(a)\langle z \rangle$, luego $z' = z^j k$ para algún número natural j y con $k \in C_P(a)$. Por tanto, tenemos que $1 \neq [a, z'] = [a, z^j k] = [a, z^j] \stackrel{(1)}{=} [x, z]^{i+j}$. Es más, también tenemos que $P = \langle x' \rangle C_P(z')$, luego $a = (x')^m w$ con m un número natural y $w \in C_P(z')$, con lo que $1 \neq [a, z'] = [(x')^m w, z'] = [x', z']^m$. Concluimos que $[x', z']^m = [x, z]^{i+j}$ y, por tanto, solamente hay un generador (que tiene orden p) en P' , luego $|P'| \leq p$. ■

4. El caso de los grupos factorizados

Paralelamente al estudio de la influencia de los tamaños de clase en la estructura de los grupos finitos, la investigación sobre grupos factorizados como producto de subgrupos ha ido tomando un interés creciente.

El caso en el que, además, los factores están conectados por ciertas condiciones de permutabilidad ha tenido un desarrollo notable. En esta amplia línea de investigación, quizás uno de los resultados más célebres podría ser el de Kegel y Wielandt, el cual afirma que el producto de dos grupos nilpotentes es resoluble; o también el teorema de Fitting, donde se prueba que si además los dos factores son normales, entonces el grupo también es nilpotente (en el libro de Isaacs [8, theorem 8.21] se da una prueba elemental del caso finito).

A la vista de estas dos perspectivas dentro de la Teoría de Grupos Finitos, la de los grupos factorizados y la de los tamaños de las clases de conjugación, uno se puede llegar a plantear de qué manera se podrían combinar ambas líneas. Quizás el primer artículo (hasta donde tenemos conocimiento) que realiza esta combinación podría ser el de Liu, Wang y Wei [11], donde se considera un grupo $G = AB$ factorizado como el producto de sus subgrupos (sub)normales A y B , y donde además se imponen condiciones del tipo « $|x^G|$ libre de cuadrados para todo $x \in A \cup B$ ». Queremos destacar la pérdida importante de información que supone este tipo de planteamientos, pues solamente se tiene información de los tamaños de clase de algunos elementos de los factores, los cuales, *a priori*, son bastantes menos que todos los del grupo. Cabe mencionar también que, por un lado, cuando se toma la factorización trivial $G = A = B$, normalmente se generalizan ciertos resultados conocidos para grupos arbitrarios no necesariamente factorizados; y por otro lado, cuando los factores son (sub)normales en el grupo, entonces las propiedades aritméticas de los tamaños de clase se heredan para los factores (ver lema 5), por lo que es más fácil dar información estructural sobre A y B . Un ejemplo sencillo de combinación de ambas técnicas podría ser el siguiente:

Teorema 8. *Sea $G = AB$ un grupo finito, el cual es producto de los subgrupos A y B . Supongamos que A es subnormal en G y que para todo $x \in A \cup B$, 4 no divide a $|x^G|$. Entonces G es resoluble.*

Demostración. Como A es subnormal en G , tenemos que $A \trianglelefteq A_1 \trianglelefteq A_2 \trianglelefteq \cdots \trianglelefteq A_n = G$ para ciertos subgrupos A_i de G , con $1 \leq i \leq n - 1$. Tomemos $x \in A$ arbitrario y, por el lema 5, punto 1, se sigue que $|x^A|$ divide a $|x^{A_1}|$. Razonando de esta forma con los demás términos de la serie anterior, podemos ver que $|x^{A_i}|$ divide a $|x^{A_{i+1}}|$ para todo i , luego $|x^A|$ divide a $|x^G|$. Como para todo $x \in A \cup B$ se cumple que 4 no divide a $|x^G|$, entonces para todo $x \in A$ tampoco puede dividir a $|x^A|$. Por el teorema B concluimos que A es resoluble.

Si $A = 1$ entonces $G = B$ y es resoluble por el teorema B. Luego podemos asumir que $A \neq 1$ y por tanto $F(A) \neq 1$, donde $F(A)$ denota el subgrupo de Fitting de A . Además, como $F(A) \leq F(G)$ por ser A subnormal en G , tenemos que $F(G) \neq 1$. Veamos que $G/F(G)$ es resoluble para finalizar la prueba. Tenemos que $G/F(G) = (AF(G)/F(G))(BF(G)/F(G))$ y $AF(G)/F(G)$ es subnormal en $G/F(G)$ por serlo A en G . Además, si tomamos $x \in F(G)$, luego 4 no divide a $|x^G|$ y, por el lema 5, punto 2, tampoco divide a $|(x F(G))^{G/F(G)}|$. Así, $G/F(G)$ hereda las hipótesis y, por inducción sobre $|G|$, es resoluble. ■

Queda claro que cuando $G = A = B$ se extiende el teorema B. También es evidente la ventaja que supone trabajar con productos de subgrupos que poseen algún factor (sub)normal, pues enseguida se pueden deducir propiedades estructurales sobre él.

Debemos destacar que no siempre es tan sencilla, o posible, una generalización a productos. Por ejemplo, Chillag y Herzog también probaron, entre otras cosas, que si un grupo finito G posee todos sus tamaños de clase libres de cuadrados, entonces $G/F(G)$ es cíclico de orden libre de cuadrados [4, theorem 1]. Supongamos que queremos extender este resultado a productos de factores normales. Entonces, nos encontramos con el siguiente ejemplo:

Ejemplo 2. Sea $G = A \times B$ el producto directo de $A = \Sigma_3$ un grupo simétrico de 3 letras y $B = D_{10}$ un grupo diédrico de orden 10. Entonces es fácil comprobar que $|x^G|$ es libre de cuadrados para todo $x \in A \cup B$, pero $F(G) = C_3 \times C_5$ y $G/F(G)$ es isomorfo a $C_2 \times C_2$ (el cuatro grupo de Klein), el cual ni es cíclico ni tiene orden libre de cuadrados. ◀

Esta combinación de ambas perspectivas de actualidad es el tema principal al cual el autor está dedicando su tesis doctoral. Finalmente, queremos mencionar que tanto en los artículos de Felipe, Martínez Pastor y Ortiz Sotomayor [5, 6] como en el trabajo de Ortiz Sotomayor [12] se puede consultar un primer avance al respecto, bastante más profundo que el aquí presentado.

Referencias

- [1] CAMINA, Alan Robert. «Arithmetical conditions on the conjugacy class numbers of a finite group». En: *J. London Math. Soc.* 5.2 (1972), págs. 127-132. <https://doi.org/10.1112/jlms/s2-5.1.127>.
- [2] CAMINA, Alan Robert y CAMINA, Rachel Deborah. «Implications of conjugacy class size». En: *J. Group Theory* 1.3 (1998), págs. 257-269. <https://doi.org/10.1515/jgth.1998.017>.
- [3] CAMINA, Alan Robert y CAMINA, Rachel Deborah. «The influence of conjugacy class sizes on the structure of finite groups: a survey». En: *Asian-Eur. J. Math.* 4 (2011), págs. 559-588. <https://doi.org/10.1142/S1793557111000459>.
- [4] CHILLAG, David y HERZOG, Marcel. «On the length of the conjugacy classes of finite groups». En: *J. Algebra* 131 (1990), págs. 110-125. [https://doi.org/10.1016/0021-8693\(90\)90168-N](https://doi.org/10.1016/0021-8693(90)90168-N).
- [5] FELIPE, María José; MARTÍNEZ PASTOR, Ana y ORTIZ SOTOMAYOR, Víctor Manuel. «On finite groups with square-free conjugacy class sizes». En: *International Journal of Group Theory* (2017). URL: http://ijgt.ui.ac.ir/article_21475.html.
- [6] FELIPE, María José; MARTÍNEZ PASTOR, Ana y ORTIZ SOTOMAYOR, Víctor Manuel. «Square-free class sizes in products of groups». En: *ArXiv e-prints* (2017). arXiv: 1703.08363 [math.GR].
- [7] HUPPERT, Bertram. *Endliche Gruppen I*. Berlin: Springer-Verlag, 1967. <https://doi.org/10.1007/978-3-642-64981-3>.
- [8] ISAACS, Irving Martin. *Algebra: A Graduate Course*. USA: Brooks/Cole Thompson Learning, 1994. ISBN: 0-534-19002-2.
- [9] ISAACS, Irving Martin. *Finite Group Theory*. USA: American Mathematical Society, 2008. ISBN: 978-0-8218-4344-4. <https://doi.org/10.4169/000298910x496804>.
- [10] KNOCHE, Hans Georg. «Über den Frobeniusschen Klassenbegriff in nilpotenten Gruppen». En: *Math. Z.* 55 (1951), págs. 71-83. <https://doi.org/10.1007/BF01212668>.
- [11] LIU, Xiaolei; WANG, Yanming y WEI, Huaquan. «Notes on the length of conjugacy classes of finite groups». En: *J. Pure Appl. Algebra* 196 (2005), págs. 111-117. <https://doi.org/10.1016/j.jpaa.2004.08.020>.
- [12] ORTIZ SOTOMAYOR, Víctor Manuel. *Clases de conjugación y grupos factorizados*. Trabajo Fin de Máster. Universidad Politécnica de Valencia, 2015. URL: <http://hdl.handle.net/10251/67883>.

TEMat

Este trabajo colaboró con una microcharla durante el XVII *Encuentro Nacional de Estudiantes de Matemáticas*, celebrado en Barcelona en julio de 2016.



Introducción a la lógica difusa y sus aplicaciones

✉ David Lobo Palacios
Universidad de Cádiz
david.lobouca.es

Resumen: La lógica clásica es una herramienta insustituible para razonar sobre aspectos del mundo real que son considerados como atemporales, objetivos, estables, etc. Sin embargo, la existencia de afirmaciones que no son totalmente ciertas ni totalmente falsas es evidente. De la imposibilidad de tratar este tipo de afirmaciones con herramientas de la lógica clásica surge la necesidad del desarrollo de las denominadas lógicas no clásicas, entre las cuales se encuentra la lógica difusa.

En este trabajo presentamos una introducción a la lógica difusa, así como algunas aplicaciones de esta lógica en problemas reales: el reconocimiento de textos manuscritos, la investigación espacial en Brasil, la programación lógica y el *big data*.

Palabras clave: lógica difusa, operador lógico, reconocimiento de textos manuscritos, investigación espacial, programación lógica, *big data*.

MSC2010: 03-94.

Recibido: 29 de enero de 2017.

Aceptado: 2 de mayo de 2017.

Referencia: LOBO PALACIOS, David. «Introducción a la lógica difusa y sus aplicaciones». En: *TEMat*, 1 (2017), págs. 53-67. ISSN: 2530-9633. URL: <http://temat.anemat.com/articulo/2017-p53/>.

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

1. Introducción

La principal característica del ser humano que le diferencia del resto de seres vivos es la capacidad de razonar. Dada una serie de premisas, el ser humano es capaz de obtener conclusiones. Por ejemplo, si suponemos cierto «el cielo es azul» y falso «Kepler era dentista», podemos deducir que afirmaciones como «el cielo es azul y Kepler era dentista» o «el cielo no es azul» son falsas. Así mismo, oraciones como «el cielo es azul o Kepler era dentista» o «que el cielo fuera azul no hizo que la población alemana del siglo XVI se hiciera dentista» son verdaderas. En realidad, estamos utilizando operadores lógicos para deducir la veracidad de afirmaciones que han sido construidas a partir de las anteriores. Todos estos operadores se encuentran inmersos en lo que denominamos *lógica clásica*, mediante la cual podemos trabajar con enunciados que son o bien ciertos, o bien falsos.

Sin embargo, ya en la era de Aristóteles se conocía la existencia de afirmaciones que no son totalmente ciertas ni totalmente falsas. La conocida como *paradoja del mentiroso* es un claro ejemplo de ello: «esta oración es falsa». Si suponemos que la oración anterior es verdadera, se deduce que es falsa. Si por el contrario suponemos que es falsa, obtenemos que es también verdadera. En cualquier caso, llegamos a una contradicción. Asimismo, para pronunciarse sobre el valor de verdad de «está lloviendo» hace falta conocer el lugar y el momento en el que se dice esta frase. Otras afirmaciones, como «el chico es listo», tienen una clara componente subjetiva. A un individuo puede parecerle que el chico es listo, pero a Stephen Hawking, no. Además, decir que oraciones como «mañana me tocará la Primitiva» son absolutamente ciertas o absolutamente falsas carece de sentido (por lo general, aunque puede darse el caso de que se jueguen todas las combinaciones posibles).

Con lo cual, cuando se desea representar el conocimiento y diseñar modos de razonamiento sobre aspectos dinámicos del mundo real, hay ocasiones en las que «la verdad de un enunciado no depende solo de la relación entre las palabras del lenguaje y los objetos del mundo, sino también del estado del mundo y el conocimiento de este estado» [22], por lo que se necesita de otras lógicas, las llamadas *lógicas no clásicas*.

La lógica difusa es una extensión de la lógica clásica a contextos en los que encontramos imprecisión o información incompleta. Las bases de la lógica difusa fueron presentadas en 1965 de la mano de Lofti Zadeh, natural de la actual República de Azerbaiyán y profesor de la Universidad de California, en Berkeley, con un artículo titulado «Fuzzy sets» [52]. En este artículo, Zadeh presenta unos conjuntos sin límites precisos, los cuales, según él, juegan un papel importante en el reconocimiento de formas, la interpretación de significados y, especialmente, en la abstracción, la esencia del proceso del razonamiento humano. Posteriormente, otros autores desarrollaron otros aspectos difusos de la matemática, con conceptos como medida difusa [8] o integral difusa [45].

El propósito de este trabajo es proporcionar al lector una noción general sobre la lógica difusa y despertar el interés del mismo en esta materia. Para ello, se mostrarán diferentes problemas reales en los que se han aplicado técnicas difusas. La diversidad de estos problemas permite imaginar el enorme rango de aplicaciones de la lógica difusa, así como la actualidad de dichas aplicaciones.

2. Conjuntos difusos y lógica difusa

La idea propuesta por Lofti Zadeh en 1965 sugiere que la pertenencia a un conjunto es la clave para decidir cuando nos enfrentamos a la incertidumbre. Nótese que la relación entre la lógica y la teoría de conjuntos es muy estrecha. Por ejemplo, en el caso clásico, decimos que «el chico es listo» es verdadero si y solo si el chico pertenece al conjunto de las personas listas.

Específicamente, la propuesta de Zadeh reside en asignar grados de verdad a las afirmaciones. O equivalentemente, asignar grados de pertenencia a un conjunto a cada elemento de un universo. Formalmente:

Definición 1. Sea A un conjunto en un universo U y $f: U \rightarrow [0, 1]$ la función de pertenencia al conjunto A . Si, para cada $x \in U$, $f(x)$ toma valores entre 0 y 1, decimos que A es un **conjunto difuso**. Por el contrario, si, para cada $x \in U$, $f(x)$ solo toma los valores 0 o 1, estamos ante un **conjunto clásico**. ◀

Observación 1. Cuando se define un conjunto clásico, se suele dar una serie de propiedades que han de satisfacer los elementos de un universo para pertenecer al conjunto. Sin embargo, un conjunto difuso se

define como una colección de pares de la forma $(x, f(x))$, donde x es un elemento del universo y $f(x)$ es su grado de pertenencia. ◀

Observación 2. En ocasiones se utilizan variantes más generales de la noción de conjunto difuso, considerando valores de pertenencia en una estructura algebraica L diferente al intervalo $[0, 1]$. Usualmente, se requiere que dicha estructura sea un conjunto parcialmente ordenado o un retículo, que no es más que un conjunto parcialmente ordenado tal que cada par de elementos del mismo tiene supremo e ínfimo en dicho conjunto. Este tipo de generalizaciones fueron introducidas por primera vez por Joseph Goguen [23]. ◀

En el siguiente ejemplo se muestra cómo podemos considerar un aspecto cotidiano como la altura de una persona en un ambiente difuso.

Ejemplo 1. Consideremos el universo U de los hombres vivos a 1 de enero de 1970 y sea $h: U \rightarrow \mathbb{R}^+$ la función que devuelve la altura (en metros) de cada elemento de U . Queremos definir el conjunto A de los hombres altos en U . Proponemos dos opciones para hacerlo:

- Como un conjunto clásico: $A = \{x \in U \mid h(x) \geq 1,80\}$.
- Como un conjunto difuso: $A = \{(x, f(x)) \mid x \in U\}$, siendo $f: U \rightarrow [0, 1]$ la función dada por

$$f(x) = \begin{cases} 0 & \text{si } h(x) \leq 1,7, \\ \frac{x - 1,7}{0,1} & \text{si } 1,7 < h(x) < 1,8, \\ 1 & \text{si } h(x) \geq 1,8. \end{cases}$$

Gráficamente, la función de pertenencia del conjunto A quedaría como en la figura 1. Se trata de una función de pertenencia de tipo trapezoidal.

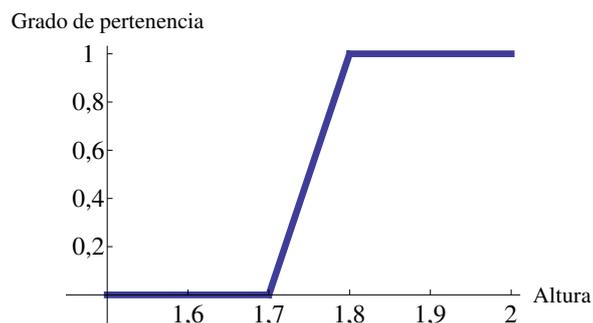


Figura 1: Función de pertenencia trapezoidal.

En el caso clásico, un hombre pertenece al conjunto de los altos si y solo si su altura es mayor o igual a 1,80 m. Ahora bien, un varón que mide 1,79 m podría sentirse molesto por no estar considerado como una persona alta, cuando alguien que mide 1,80 m sí que lo está. En efecto, esto podría ocurrir sea cual sea el listón que pongamos. No parece lógico imponer una altura exacta a partir de la cual decir que alguien es alto.

Cuando consideramos el problema en un ambiente difuso, con la función de pertenencia dada arriba, tenemos que a partir de una altura de 1,80 m una persona es alta (grado de pertenencia 1 a A), al igual que en el caso clásico. De igual manera, tenemos que si la altura de un hombre es menor de 1,70 m, entonces esta persona no es alta (grado de pertenencia 0 a A). Para valores entre 1,70 m y 1,80 m, hemos optado por realizar un ajuste lineal. Así, si alguien mide 1,79 m, entonces tiene un grado de pertenencia 0,9 al conjunto de los altos.

Naturalmente, podríamos tomar cualquier otro ajuste para definir la función de pertenencia. La elección de un ajuste u otro dependerá de la realidad que se desea modelizar. En la figura 2 vemos otra opción para la función de pertenencia, en este caso de tipo gamma.

En cualquier caso, la versión difusa de la definición del conjunto A se ajusta más a nuestra forma de pensar que la versión clásica. La razón principal es que el concepto «altura» no tiene una definición precisa, tiene un perfil difuso. ◀

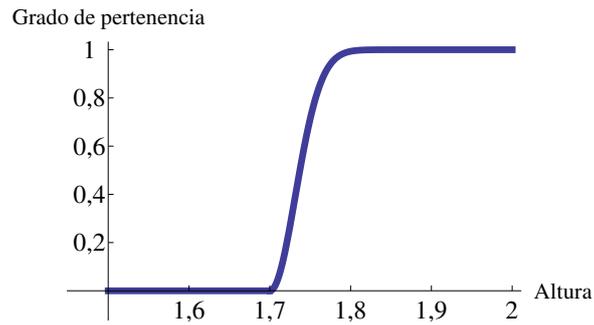


Figura 2: Función de pertenencia gamma.

El objetivo de la lógica difusa es proporcionar un marco de trabajo para representar el conocimiento y obtener reglas de inferencia en un ambiente de incertidumbre e imprecisión léxica. Es decir, queremos no solo asignar valores de verdad a las afirmaciones, sino también trabajar con ellas. Es por esto que necesitamos generalizar los operadores lógicos clásicos. Ahora bien, ¿cómo definimos un operador lógico difuso? Como veremos a continuación, existen infinitas maneras de hacerlo. De nuevo, en función de la realidad que queremos estudiar, tomaremos un operador difuso u otro.

2.1. T-normas y t-conormas

Los operadores lógicos clásicos, aquellos que hemos utilizado (a veces inconscientemente) durante toda nuestra vida, se pueden resumir en conjunción, disyunción, implicación y negación.

En las siguientes líneas, nos centraremos en cómo generalizar la conjunción clásica al caso difuso. Este operador es utilizado a diario mediante la conjunción copulativa «y» del lenguaje natural. Denotado por \wedge , queda resumido en el cuadro 1:

Cuadro 1: Conjunción clásica.

p	q	$p \wedge q$
0	0	0
1	0	0
0	1	0
1	1	1

Es decir, $p \wedge q$ (p y q) es cierto si y solo si tanto p como q son ciertos.

Nuestro propósito es definir un operador de $[0, 1] \times [0, 1]$ en $[0, 1]$ que extienda la conjunción clásica. En consecuencia, cualquier conjunción difusa debería satisfacer, al menos, la tabla anterior. Ahora bien, existen infinidad de operadores binarios que cumplen dicha tabla. Algunos ejemplos son los siguientes:

- $p \wedge q = p * q$
- $p \wedge q = \text{mín}(p, q)$
- $p \wedge q = \text{máx}(0, p + q - 1)$

Tomaremos aquella conjunción que mejor se adapte al contexto que queremos estudiar. En ocasiones, exigiremos propiedades adicionales a la conjunción, tales como conmutatividad, asociatividad y monotonía.

En efecto, la generalización de conjunción clásica que ha recibido más atención en los últimos años cumple las propiedades mencionadas. Se trata de la norma triangular o t-norma, la cual definimos a continuación.

Definición 2. Una **norma triangular** (o **t-norma**) es una operación binaria T en el intervalo unidad $[0, 1]$, es decir, una función $T: [0, 1] \times [0, 1] \rightarrow [0, 1]$, tal que para todo $x, y, z \in [0, 1]$ se satisfacen las siguientes propiedades:

1. $T(x, y) = T(y, x)$ (conmutatividad);
2. $T(x, T(y, z)) = T(T(x, y), z)$ (asociatividad);
3. si $y \leq z$, entonces $T(x, y) \leq T(x, z)$ (monotonía);
4. $T(x, 1) = x$ (condición de frontera).

Huelga decir que la t-norma no es la única generalización de la conjunción clásica en el ámbito difuso. Por ejemplo, la semicópula [3] es un operador menos restrictivo que la t-norma y también generaliza la conjunción clásica.

Observemos que los ejemplos de conjunción difusa mencionados anteriormente son, en realidad, t-normas. De hecho, se trata de las t-normas más utilizadas en la bibliografía.

Ejemplo 2. Las t-normas producto, Gödel y Lukasiewicz son los ejemplos más usuales de este tipo de operador. También es interesante la t-norma drástica, por ser la más pequeña de todas.

$$\begin{array}{ll}
 T_P(x, y) = x \cdot y & \text{(t-norma producto)} \\
 T_G(x, y) = \min(x, y) & \text{(t-norma Gödel)} \\
 T_L(x, y) = \max(0, x + y - 1) & \text{(t-norma Lukasiewicz)} \\
 T_D(x, y) = \begin{cases} 0 & \text{si } x, y \in [0, 1), \\ \min(x, y) & \text{en otro caso} \end{cases} & \text{(t-norma drástica)}
 \end{array}$$

Además, la t-norma Gödel es la mayor de todas. Esto es, cualquier t-norma toma valores entre la t-norma drástica y la t-norma Gödel. Formalmente:

Proposición 1. Sea T una t-norma. Entonces, para todo $x, y \in [0, 1]$, se verifica que

$$(1) \quad T_D(x, y) \leq T(x, y) \leq T_G(x, y).$$

De forma similar, las conormas triangulares o t-conormas generalizan la disyunción clásica (denotada por \vee), resumida en el cuadro 2:

Cuadro 2: Disyunción clásica.

p	q	$p \vee q$
0	0	0
1	0	1
0	1	1
1	1	1

En la siguiente definición, podemos observar que las t-normas y t-conormas difieren solamente con respecto a las condiciones de frontera.

Definición 3. Una **conorma triangular** (o **t-conorma**) es una operación binaria S en el intervalo unidad $[0, 1]$, es decir, una función $S: [0, 1] \times [0, 1] \rightarrow [0, 1]$, tal que para todo $x, y, z \in [0, 1]$ se satisfacen las siguientes propiedades:

1. $S(x, y) = S(y, x)$ (conmutatividad);
2. $S(x, S(y, z)) = S(S(x, y), z)$ (asociatividad);
3. si $y \leq z$, entonces $S(x, y) \leq S(x, z)$ (monotonía);
4. $S(x, 0) = x$ (condición de frontera).

Nótese que la única diferencia entre una t-norma y una t-conorma es que el 1 es el elemento neutro de una t-norma, mientras que en el caso de una t-conorma el 0 es el elemento neutro.

Ejemplo 3. Un ejemplo clásico de t-conorma es el operador máximo, conocido usualmente en la literatura como t-conorma Gödel.

$$S_G(x, y) = \max(x, y) \quad (\text{t-conorma Gödel}) \quad \blacktriangleleft$$

El resultado que sigue muestra cómo obtener una t-conorma a partir de una t-norma y viceversa [25].

Proposición 2. Una función $S: [0, 1] \times [0, 1] \rightarrow [0, 1]$ es una t-conorma si y solo si existe una t-norma T tal que para todo $x, y \in [0, 1]$ se cumple que

$$(2) \quad S(x, y) = 1 - T(1 - x, 1 - y).$$

Un ejercicio interesante para familiarizarse con los operadores lógicos difusos es hallar las t-conormas asociadas a las t-normas producto, Gödel, Lukasiewicz y drástica. Naturalmente, la t-conorma obtenida a partir de la t-norma Gödel es la definida en el ejemplo 3.

2.2. Implicaciones difusas y negaciones

Para finalizar esta sección, introduciremos las implicaciones difusas y discutiremos el papel que juega la negación en un paradigma de incertidumbre e información incompleta.

Las implicaciones difusas generalizan la implicación de la lógica clásica conservando las propiedades de monotonía. Siguiendo la nomenclatura usada en programación lógica (una rama de la lógica de la cual hablaremos más adelante), escribiremos « \leftarrow » en lugar de « \rightarrow ».

Definición 4. Decimos que un operador $\leftarrow: [0, 1] \times [0, 1] \rightarrow [0, 1]$ es una **implicación difusa** si para cada $y, y_1, y_2, z, z_1, z_2 \in [0, 1]$ se cumplen las siguientes propiedades:

1. si $z_1 \leq z_2$, entonces $z_1 \leftarrow y \leq z_2 \leftarrow y$. Es decir, \leftarrow es creciente en el consecuente (argumento de la izquierda);
2. si $y_1 \leq y_2$, entonces $z \leftarrow y_2 \leq z \leftarrow y_1$. Esto es, \leftarrow es decreciente en el antecedente (argumento de la derecha). ◀

Como se observa, la definición usual de implicación es muy general. Dependiendo del entorno de aplicación se exigen propiedades adicionales, como es el caso de la implicación residuada, introducida por primera vez por Pavelka [38].

En efecto, si nos paramos a pensar en el papel de la implicación en lógica clásica, resulta un elemento esencial en la regla del *modus ponens*, que es el principal método de deducción. A fin de generalizar también esta regla, se define una nueva implicación que está relacionada con el concepto de operador residual.

Definición 5. Dada una t-norma $T: [0, 1] \times [0, 1] \rightarrow [0, 1]$, se dice que \leftarrow es una **implicación residuada de T** si existe una implicación difusa $\leftarrow: [0, 1] \times [0, 1] \rightarrow [0, 1]$ tal que

$$(3) \quad T(x, y) \leq z \text{ si y solo si } x \leq z \leftarrow y$$

para todo $x, y, z \in [0, 1]$. Entonces se escribe como \leftarrow_T , y al par (T, \leftarrow_T) se le llama **par residuo** o **par adjunto**. La propiedad dada por la ecuación (3) se conoce como **propiedad de adjunción**. ◀

El siguiente resultado se deduce directamente de la propiedad de adjunción.

Proposición 3. Sea (T, \leftarrow_T) un par adjunto. Entonces, para cada $y, z \in [0, 1]$ se tiene que $y \leq z$ si y solo si $z \leftarrow y = 1$.

A continuación se muestran las implicaciones adjuntas de las t-normas presentadas en el ejemplo 2.

Ejemplo 4. Las implicaciones residuadas correspondientes a las t-normas producto, Gödel y Lukasiewicz se definen, para todo $y, z \in [0, 1]$, como

$$\begin{aligned} z \leftarrow_P y &= \begin{cases} 1 & \text{si } y \leq z, \\ z/y & \text{en otro caso;} \end{cases} \\ z \leftarrow_G y &= \begin{cases} 1 & \text{si } y \leq z, \\ z & \text{en otro caso;} \end{cases} \\ z \leftarrow_L y &= \min(1, 1 - y + z). \end{aligned}$$

El caso del operador de negación en un ambiente lógico no clásico es especialmente peculiar. Un pensamiento común entre los iniciados a las lógicas no clásicas es que no es necesario considerar un operador de negación, puesto que en su lugar podría tomarse un mayor número de conjuntos. Por ejemplo, si A es el conjunto clásico de los científicos y x es un elemento del universo, en lugar de decir que $n(x)$ (la negación de x) pertenece a A , podríamos definir el conjunto B de los no científicos y decir que x pertenece a B .

Sin embargo, por suerte o por desgracia, el ser humano no utiliza siempre la misma negación. En el lenguaje natural se utilizan diferentes tipos de negación en función del sentido que se le da a dicho operador. Pensemos, por ejemplo, en la proposición «Dios existe». A día de hoy, parece imposible demostrar que dicha afirmación es cierta o es falsa. En principio, no es posible asignar un valor de verdad a dicho enunciado. Sin embargo, las personas religiosas (en todas sus vertientes) dirían que dicha afirmación es cierta, mientras que las personas ateas dirían que es falsa. ¿Cómo es posible? La razón es que la comunidad religiosa utiliza información explícita, en este caso sus creencias, para pronunciarse sobre el valor de verdad de la afirmación. Cuando no conoce el valor de verdad de una afirmación, tiene en cuenta algunos aspectos externos a dicha afirmación. Por el contrario, podemos suponer que el grupo de ateos niega una proposición cuando no conoce el valor de verdad de la misma.

Nuestro objetivo es desarrollar una teoría lógica que modelice el uso que hace el ser humano de la negación. Por consiguiente, podemos definir diferentes tipos de negación y diferentes operadores de negación difusos, lo cual aporta a las lógicas no clásicas un grado más de flexibilidad. Algunos de los operadores de negación más importantes son los definidos a partir de implicaciones residuadas de una t-norma [9, 19, 42]; las negaciones ordinarias, estudiadas por Trillas, Esteva y Domingo [16-18, 50], y los pares de negaciones débiles, introducidos por Georgescu y Popescu [21].

En general, un operador de negación difuso viene definido como sigue.

Definición 6. Un operador unario sobre el intervalo unidad $n: [0, 1] \rightarrow [0, 1]$ se dice que es una **negación** si satisface las siguientes propiedades:

1. Si $x \leq y$, entonces $n(y) \leq n(x)$.
2. $n(0) = 1$, $n(1) = 0$.

Si para todo $x \in [0, 1]$ se satisface $x \leq n(n(x))$, decimos que el operador de negación n es una **negación ordinaria**, mientras que si se verifica la condición $n(n(x)) \leq x$, se trata de una **negación débil**. Decimos que n es una **negación fuerte** (o **involutiva**) si $x = n(n(x))$, para todo $x \in [0, 1]$.

Ejemplo 5. El operador negación más usado es el que se define como $n(x) = 1 - x$, para todo $x \in [0, 1]$, el cual es una negación fuerte.

No es difícil imaginar que, con la generalidad de la teoría que acabamos de desarrollar, las aplicaciones de esta sean muy numerosas y variadas. En la siguiente sección mostramos algunas aplicaciones interesantes de la lógica difusa en problemas reales.

3. Aplicaciones de la lógica difusa

En nuestro día a día tenemos que trabajar con elementos o ideas que no tienen una definición precisa. Hasta el momento, cuando a la hora de resolver un problema encontramos que puede aparecer cierto

error, nuestra atención se centra en controlar dicho error. Es decir, asegurar que el error cometido no exceda de ciertos límites.

La lógica difusa surge como una potente herramienta para trabajar con problemas del perfil mencionado. Además, no solo permite disminuir el error cometido (y acotarlo), sino que considera el problema completo desde una nueva perspectiva más general. De esta forma, se aumenta el nivel de abstracción del problema, lo cual conlleva, por lo general, una mejor modelización del mismo.

3.1. Clasificación y reconocimiento de manuscritos

La primera aplicación que mostraremos fue presentada por el profesor László T. Kóczy [26] en el 7.º *Simposio Europeo sobre Inteligencia Computacional y Matemáticas* (ESCIM 2015), celebrado en Cádiz en 2015.

Hoy día es relativamente sencillo encontrar aplicaciones capaces de reconocer textos impresos prácticamente sin cometer ningún error. Sin embargo, cuando se trata de textos manuscritos, la búsqueda se convierte en toda una odisea. No es difícil imaginar el motivo por el que no existen (de momento) aplicaciones de este tipo que aseguren, y efectivamente tengan, una alta precisión. A saber, el umbral de aceptación del usuario en cuanto a la precisión de un sistema de reconocimiento de manuscritos se sitúa en el 97 % [28].

Lo cierto es que reconocer un texto escrito a mano puede resultar bastante complejo, ya que cada persona tiene su propia caligrafía. Por ejemplo, el par de caracteres «cl» puede confundirse fácilmente (dependiendo de la persona que lo escriba) con la letra «d». Así, en castellano, la palabra «clara» podría ser reconocida erróneamente como la palabra «dara», y viceversa.

Usualmente, para alcanzar el nivel de precisión deseado, los métodos de reconocimiento de manuscritos aplican transformaciones geométricas muy complejas; hasta el punto de que algunos métodos recientes no disponen de la suficiente capacidad de cómputo para dar un resultado en un tiempo razonable. Es decir, podríamos obtener un resultado muy bueno, pero dentro de 800 años, lo cual no resulta útil a efectos prácticos. La lógica difusa aparece como una alternativa a estos métodos, proporcionando una solución aceptable para el reconocimiento de manuscritos.

El punto de partida de este trabajo es la consideración de cada carácter escrito a mano como una señal, es decir, una secuencia de movimientos. Para hacerlo, se toma un sistema de coordenadas cartesianas de dimensión 2 en el plano de la hoja donde se encuentra el texto. Posteriormente, se considera cada carácter como una lista cronológicamente ordenada de pares de coordenadas. De esta forma, es posible utilizar métodos de clasificación y procesamiento de señales para reconocer caracteres manuscritos.

En particular, en la presentación de Kóczy se mostró una familia de reconocedores unitrazo y multitrazo (también influye la cantidad de trazos con la que se escribe un carácter) denominada *reconocedores basados en el difuso* o *FUBAR*, por sus siglas en inglés [46-48]. Tras optimizar el algoritmo FUBAR multitrazo mediante el uso de métodos estadísticos y metaheurísticos, se alcanzó una precisión del 99,49 % [47]. Este porcentaje de acierto es significativamente superior al obtenido por métodos similares anteriores, como el 86,03 % del algoritmo Graffiti 2 obtenido por Költringer y Grechenig [27], o el 96,7 % de la mejor versión del algoritmo \$N\$ [1].

Una explicación más exhaustiva de todo lo expuesto anteriormente puede encontrarse en el artículo de Tormási y Kóczy [49].

3.2. Investigación espacial en Brasil

La siguiente aplicación de la lógica difusa fue también presentada en el congreso ESCIM 2015, esta vez por parte de Sandra Sandri, investigadora del *Instituto Nacional de Pesquisas Espaciais* (INPE) en Brasil.

Brasil es uno de los países más grandes del mundo, y uno de los más importantes desde el punto de vista biológico. Además de contener el 60 % de la superficie de selva amazónica de la Tierra [36], el 46 % de la matriz energética de este país está formada por energías renovables [44]. Se trata del mayor productor en agricultura tropical y es líder mundial en tecnología de biocombustibles.

Por todo ello, estudiar los efectos de fenómenos como la meteorología, la deforestación y el cambio climático en el territorio brasileño resulta de vital importancia. Estas son algunas de las tareas de las que se encarga el Instituto Nacional de Pesquisas Espaciais, el centro líder en innovación espacial en Brasil. Para tal propósito, el INPE recoge datos de satélites y de diversos sistemas terrestres y, mediante un proceso de investigación y desarrollo tecnológico, proporciona resultados o productos beneficiosos para la sociedad. Cabe destacar su política de acceso libre a los datos, ya que tanto las imágenes como los productos tecnológicos obtenidos por el instituto se encuentran disponibles en internet.

El motor fundamental de los estudios realizados por el INPE es la lógica difusa. A continuación mencionamos algunas aplicaciones recientes de la inteligencia computacional realizadas por el INPE, y en las cuales subyace la lógica difusa.

Para comenzar, destacamos el uso de sistemas neurodifusos para extraer conocimiento de sistemas caóticos. Debido a que numerosos fenómenos del mundo real tienen algún tipo de régimen caótico, estudios como este tienen una gran relevancia. Un sistema neurodifuso [29] no es más que una combinación de dos técnicas: redes neuronales artificiales y sistemas difusos basados en reglas. Mediante herramientas de aprendizaje de redes neuronales se generan reglas de inferencia de un sistema difuso. De esta forma, se maximizan las ventajas de cada técnica y se minimizan sus inconvenientes. Actualmente, estos sistemas están siendo utilizados para generar automáticamente reglas que permitan predecir cambios de régimen en sistemas caóticos, tales como el atractor de Lorenz y el problema de las tres ondas [43]. El objetivo final es aplicar esta metodología para generar reglas de cambio de régimen en el clima de la región nordeste de Brasil.

Por otro lado, se están empleando autómatas celulares difusos en aplicaciones georreferenciadas. La idea de autómatas celulares fue propuesta por John Von Neumann [37] y consistía en un sistema lógico autorreplicante. En pocas palabras, un autómata celular es un modelo de un sistema dinámico que evoluciona en pasos discretos. El interés de estos modelos es que permite estudiar un gran número de sistemas dinámicos, tales como incendios, inundaciones, fragmentación urbana, deforestación, epidemias o avalanchas.

La principal aportación del INPE en esta línea ha sido la reproducción del círculo SIR (susceptible-infectado-recuperado) para el dengue [12], lo cual puede ser utilizado para comprobar hipótesis sobre controles de calidad.

Antes de concluir este apartado, comentaremos el trabajo desarrollado por el INPE en cuanto al índice de vegetación en el país. En este sentido, el marco de estudio establecido ha sido la aplicación de perceptrones multicapa difusos para la clasificación de series temporales LUCC [39, 40]. Naturalmente, las series temporales que interesan son aquellas relacionadas con el índice de vegetación, como las de las zonas de bosque, de pasto, de cultivo y de deforestación. Otras aplicaciones interesantes de la lógica difusa por parte del INPE han sido relativas a la reparación de satélites y a la obtención de patrones de clorofila del mar.

3.3. Programación lógica

En esta sección introduciremos al lector en lo que podríamos denominar un área de aplicación de la lógica difusa, ya que debido a su extensión se ha convertido en una rama de la lógica difusa en lugar de en una mera aplicación de esta.

El objetivo principal de la programación lógica, como el de tantas otras ramas de la matemática, es entender y modelizar sistemas que cambian en el tiempo, es decir, sistemas dinámicos. A grandes rasgos, un programa lógico podría definirse como un conjunto de reglas, o implicaciones, deducidas de un sistema dinámico. Cada regla contiene el conjunto de premisas que se han de dar para obtener un consecuente [30].

Por ejemplo, supongamos que tenemos un sistema dinámico, digamos una máquina, y deseamos tener en cuenta diversos aspectos de la misma, tales como alta temperatura, comportamiento ruidoso, alto consumo de combustible, etc. Estos aspectos reciben el nombre de *símbolos proposicionales del sistema*. Tras poner en funcionamiento la máquina, observamos que, por ejemplo, cuando la temperatura de la máquina es alta, esta se comporta ruidosamente, y un alto consumo de combustible implica que la temperatura sea alta. Al conjunto de deducciones realizadas lo llamamos programa lógico.

En el caso clásico, consideramos que cada símbolo proposicional solo toma los valores *cierto* o *falso*, y las deducciones que realizamos son totalmente ciertas o totalmente falsas. Sin embargo, como ya hemos visto, en ocasiones los símbolos proposicionales no son ni totalmente ciertos ni totalmente falsos. De la misma forma, podemos tener que la mayoría de las veces que la máquina tiene cierto grado de temperatura elevado, la máquina se comporta ruidosamente (en mayor o menor medida), mientras que otras veces la máquina no hace ningún ruido a pesar de la alta temperatura. Con lo cual, cuando consideramos un programa lógico difuso, el proceso de definición del programa se complica considerablemente, puesto que no solo hemos de deducir reglas de inferencia en el sistema, sino que hemos de dotar de un grado de verdad a dichas reglas [51].

Llegados a este punto, dejaremos de lado la rigurosidad del lenguaje a fin de hacer el texto más adecuado para los no iniciados en la materia. Una introducción formal a la programación lógica difusa requiere la definición de diversas estructuras algebraicas y conceptos que no son absolutamente necesarios para tener una idea general del área, lo cual es el objetivo principal de la presente sección. Aquellos que deseen acceder a más información pueden dirigirse a la bibliografía o contactar con el autor.

El siguiente ejemplo muestra un programa lógico difuso que contiene la información de una máquina similar a la mencionada anteriormente.

Ejemplo 6. Consideremos el retículo $\langle [0, 1], \leq \rangle$ y el par adjunto producto $(*, \leftarrow)$.

Sea el conjunto de símbolos proposicionales $\Pi_{\mathbb{P}} = \{p, q, s, t, u\}$ y tomemos el siguiente programa lógico residuado \mathbb{P} (denominado así por tener como operadores un par adjunto, o residuado) valuado en $[0, 1]$ y consistente en cinco reglas y un hecho (una regla que no tiene ningún símbolo proposicional en el antecedente):

$$\begin{array}{ll} r_1 : \langle p \leftarrow q * \neg s ; & 0,5 \rangle, & r_2 : \langle q \leftarrow \neg t ; & 0,2 \rangle, \\ r_3 : \langle t \leftarrow \neg q ; & 0,6 \rangle, & r_4 : \langle s \leftarrow t * u * \neg p ; & 0,4 \rangle, \\ r_5 : \langle p \leftarrow q * t ; & 0,6 \rangle, & r_6 : \langle u \leftarrow ; & 0,4 \rangle, \end{array}$$

siendo

$p \equiv$ alto consumo de combustible,
 $q \equiv$ bajo nivel de aceite,
 $s \equiv$ comportamiento ruidoso,
 $t \equiv$ sobrecalentamiento y
 $u \equiv$ bajo nivel de agua.

Así, por ejemplo, de la regla r_1 se deduce que en el 50 % de ocasiones en las que el nivel de aceite es bajo y la máquina no se comporta ruidosamente, el consumo de combustible es alto. A partir del resto de reglas se pueden obtener otros comportamientos de la máquina. ◀

Una vez definido un programa lógico, tenemos modelizado el sistema dinámico. El siguiente paso será utilizar dicha modelización para obtener información sobre el sistema. Una herramienta muy interesante para este fin es la técnica del punto fijo.

Nótese que el conjunto de reglas de un programa lógico determina, o, mejor dicho, simula, el comportamiento de un sistema dinámico. Pero el programa es independiente del estado del sistema en un momento determinado, es decir, de los valores de verdad que toman los símbolos proposicionales del sistema en un momento determinado. La pregunta que nos hacemos es «¿podemos encontrar estados del sistema de forma que, aunque el sistema esté en funcionamiento, es decir, que las reglas del programa estén actuando sobre los símbolos proposicionales, siempre tengamos el mismo estado?». Estos estados serán denominados *modelos* del programa lógico.

La respuesta es depende, depende de la estructura algebraica (recordemos la observación 2) y de los operadores lógicos que intervienen en la definición del programa lógico. En particular, con objeto de definir la semántica de un programa lógico, se buscan condiciones bajo las cuales exista un único modelo mínimo, en lugar de un conjunto de modelos minimales. Lo cierto es que, cuando consideramos un programa lógico monótono, es decir, sin ningún operador de negación en el antecedente de las reglas, podemos asegurar la existencia y la unicidad del modelo mínimo [5]. Lo anterior también es cierto para programas lógicos residuados [14] y multiadjuntos [33] (en este caso se consideran varios pares adjuntos en la definición del programa lógico).

Ejemplo 7. El ejemplo 6 es un programa lógico residuado sin negaciones. Por lo tanto, podemos asegurar que tiene un modelo mínimo.

En efecto, realizando los cálculos pertinentes se obtiene que dicho modelo es:

Alto consumo de combustible: 0,0416

Bajo nivel de aceite: 0,0909

Comportamiento ruidoso: 0,0836

Sobrecalentamiento: 0,5454

Bajo nivel de agua: 0,4

Desafortunadamente, al incluir un operador de negación en la estructura algebraica sobre la que se define un programa lógico, ya sea clásico, difuso o de otro tipo, no podemos asegurar la existencia de modelos del programa (por ejemplo, en el artículo de Madrid y Ojeda-Aciego [31] se prueba lo anterior para el caso residuado). La búsqueda de condiciones que aseguren la existencia y la unicidad de modelos en un entorno de programación lógica con negaciones es un tema de investigación muy activo en la actualidad.

En el caso clásico, Constantini [13] da una condición sintáctica para asegurar la existencia de modelos. En cuanto a la programación lógica no clásica, debido a la pluralidad de teorías no clásicas existente, existen numerosos resultados acerca de la existencia de modelos.

En particular, el estudio de la existencia y unicidad de modelos de un programa lógico residuado con negaciones definido en $[0, 1]$ ha sido realizado previamente por Madrid y Ojeda-Aciego [32]. Recientemente, Cornejo, Lobo y Medina [10] han generalizado los resultados obtenidos en el trabajo anteriormente mencionado. Además, en una continuación de dicho trabajo, se ha desarrollado un entorno de programación multiadjunta con negaciones y se han obtenido condiciones para la existencia de modelos definidos en un programa lógico multiadjunto con negaciones [11].

3.4. *Big data*

La siguiente y última aplicación de la lógica difusa que mostraremos se refiere al *big data*, uno de los marcos de trabajo que más repercusión, tanto científica como mediática, ha tenido en los últimos tiempos.

Hoy en día, seis mil millones de fuentes digitales generan varios zettabytes (10^{21} bytes) de datos al año, y se espera que el número de fuentes aumente a cincuenta mil millones en 2020 [35]. Actualmente se dispone de herramientas para almacenar dicha información, pero desafortunadamente no hay forma de manejarla, y de nada vale guardar información que no se puede utilizar. Ante la incapacidad de manejar tal volumen de datos, se analizan y se extraen patrones ocultos en los datos que nos permiten resumir o deducir la información total contenida en ellos.

Sin embargo, la tecnología requerida para procesar grandes cantidades de datos suele ir acompañada de un alto coste computacional. En consecuencia, se requieren estrategias más eficientes que los métodos tradicionales. Esta es la principal motivación por la que algunas de las mayores compañías del panorama actual, tales como IBM [2], Oracle [41], Microsoft [24] y Google [6, 7, 15, 34], están invirtiendo ingentes cantidades de dinero con el objetivo de optimizar las herramientas de control de flujo de información.

En el mundo del *big data* resulta esencial establecer una jerarquía en los datos de los cuales se desea extraer información. Una estrategia habitual a la hora de elaborar esta clasificación consiste en definir una serie de características, presentes en algunos de los datos, y deducir qué características hacen que un dato pertenezca a una clase o a otra. Actualmente se estudian técnicas de clasificación de datos que consideren dichas características con un perfil difuso [4, 20].

Por ejemplo, supongamos que nuestro objetivo es clasificar un conjunto de textos (en ciertas clases que hemos definido previamente). Las características de los textos en las que nos basaremos para clasificarlos podrían ser palabras clave o grupos de palabras. De esta forma, la aparición de ciertas palabras en un texto nos indica a qué clase pertenece dicho texto.

El proceso de inferencia a partir del conjunto de palabras clave se lleva a cabo mediante algún algoritmo de clasificación, como la máquina de vectores de soporte (SVM) o el clasificador bayesiano ingenuo (*naive-Bayes*). El porcentaje de acierto en la clasificación de los textos depende tanto del algoritmo usado

como del conjunto de textos. Así mismo, observemos que una selección adecuada de palabras clave resulta esencial para obtener un buen resultado en la clasificación.

Ahora bien, ¿qué quiere decir que una palabra clave *aparezca* en un texto? Si una palabra clave se encuentra un número elevado de veces en un texto, debería tener más relevancia para su clasificación que otra que solo está presente una vez. Aún más, la aparición de una palabra clave debería tener más peso si lo hace en el título o en el encabezado del texto que si lo hace en el cuerpo del mismo, por ejemplo. Podemos cuantificar todos estos aspectos haciendo uso de técnicas difusas en lugar de los algoritmos clásicos de clasificación. Se trata de una idea que aún está en desarrollo, pero que presenta una interesante línea de trabajo, dada la necesidad y la importancia de manejar grandes cantidades de información en la sociedad actual.

4. Conclusiones

En el lenguaje cotidiano, solemos utilizar conceptos o ideas que no tienen una definición precisa. En este trabajo hemos presentado una lógica no clásica que nos permite escribir y trabajar con esta imprecisión matemáticamente. Se han introducido los principales operadores usados en este paradigma lógico y se ha discutido sobre la importancia de los operadores de negación en un ambiente no clásico. Con el objetivo de evidenciar la versatilidad de la teoría lógica presentada, se han mostrado pinceladas acerca de cuatro aplicaciones de la lógica difusa en problemas reales: la clasificación y reconocimiento de manuscritos, la investigación espacial, la programación lógica y el *big data*.

La lógica difusa es una teoría lógica muy moderna, de apenas medio siglo de vida. Pese a ello, estamos viviendo una revolución a pasos agigantados en la inteligencia computacional. El actual rango de aplicaciones es muy amplio, y se encuentra en constante crecimiento. La razón es bien simple: habitamos un mundo en el que, paradójicamente, lo único exacto son las matemáticas. Todo lo demás es impreciso, difuso.

Aprovechamos estas líneas para agradecer al lector el interés mostrado en el presente trabajo, así como le animamos a profundizar en la temática, para lo cual recomendamos emplear las fuentes mencionadas durante el trabajo.

Referencias

- [1] ANTHONY, Lisa y WOBROCK, Jacob O. «A Lightweight Multistroke Recognizer for User Interface Prototypes». En: *Proceedings of Graphics Interface 2010*. GI '10. Ottawa, Ontario, Canada: Canadian Information Processing Society, 2010, págs. 245-252. ISBN: 978-1-56881-712-5. URL: <http://dl.acm.org/citation.cfm?id=1839214.1839258>.
- [2] BARRANCO FRAGOSO, R. *Qué es Big Data*. <https://www.ibm.com/developerworks/ssa/local/im/ques-big-data/>. (Visitado 18-06-2012).
- [3] BASSAN, Bruno y SPIZZICHINO, Fabio. «Relations among univariate aging, bivariate aging and dependence for exchangeable lifetimes». En: *Journal of Multivariate Analysis* 93.2 (abr. de 2005), págs. 313-339. <https://doi.org/10.1016/j.jmva.2004.04.002>.
- [4] BING, L. y CHAN, K. C. C. «A Fuzzy Logic Approach for Opinion Mining on Large Scale Twitter Data». En: *2014 IEEE/ACM 7th International Conference on Utility and Cloud Computing*. Dic. de 2014, págs. 652-657. <https://doi.org/10.1109/UCC.2014.105>.
- [5] BLAIR, H. A.; BROWN, A. L. y SUBRAHMANYAN, V. S. «Monotone Logic Programming». En: *Intensional Logics for Programming*. Ed. por Cerro, L. Farinas del y Penttonen, M. Oxford: Clarendon Press, 1992, págs. 1-22.
- [6] BURROWS, Mike. «The Chubby Lock Service for Loosely-coupled Distributed Systems». En: *Proceedings of the 7th Symposium on Operating Systems Design and Implementation*. OSDI '06. Seattle, Washington: USENIX Association, 2006, págs. 335-350. ISBN: 1-931971-47-1. URL: <http://dl.acm.org/citation.cfm?id=1298455.1298487>.

- [7] CHANG, Fay; DEAN, Jeffrey; GHEMAWAT, Sanjay; HSIEH, Wilson C.; WALLACH, Deborah A.; BURROWS, Mike; CHANDRA, Tushar; FIKES, Andrew y GRUBER, Robert E. «Bigtable: A Distributed Storage System for Structured Data». En: *Proceedings of the 7th USENIX Symposium on Operating Systems Design and Implementation - Volume 7*. OSDI '06. Seattle, WA: USENIX Association, 2006, págs. 15-15. URL: <http://dl.acm.org/citation.cfm?id=1267308.1267323>.
- [8] CHOQUET, Gustave. «Theory of capacities». eng. En: *Annales de l'institut Fourier* 5 (1954), págs. 131-295. URL: <http://eudml.org/doc/73714>.
- [9] CINTULA, Petr; KLEMENT, Erich Peter; MESIAR, Radko y NAVARA, Mirko. «Residuated logics based on strict triangular norms with an involutive negation». En: *Mathematical Logic Quarterly* 52.3 (2006), págs. 269-282. ISSN: 1521-3870. <https://doi.org/10.1002/maLq.200510032>.
- [10] CORNEJO, M.E.; LOBO, D. y MEDINA, J. «Stable models in normal residuated logic programs». En: *7th European Symposium on Computational Intelligence and Mathematics (ESCIM 2015)*. Ed. por Kacprzyk, J.; Kóczy, László T y Medina, J. Cádiz, Spain, 2015, págs. 150-155.
- [11] CORNEJO, M.E.; LOBO, D. y MEDINA, J. «Towards Multi-adjoint Logic Programming with Negations». En: *8th European Symposium on Computational Intelligence and Mathematics (ESCIM 2016)*. Ed. por Kóczy, László T y Medina, J. Sofia, Bulgaria, 2016, págs. 24-29.
- [12] COSTA, Walley da; MEDEIROS, Lílíam y SANDRI, Sandra. «A Fuzzy Cellular Automata for SIR Compartmental Models». En: *Fuzzy Logic and Applications: 10th International Workshop, WILF 2013, Genoa, Italy, November 19-22, 2013. Proceedings*. Ed. por Masulli, Francesco; Pasi, Gabriella y Yager, Ronald. Cham: Springer International Publishing, 2013, págs. 234-247. ISBN: 978-3-319-03200-9. https://doi.org/10.1007/978-3-319-03200-9_24.
- [13] COSTANTINI, S. «On the existence of stable models of non-stratified logic programs». En: *Journal of Theory and Practice of Logic Programming* (2006), 6(1-2):169-212.
- [14] DAMÁSIO, Carlos Viegas y PEREIRA, Luís Moniz. «Monotonic and Residuated Logic Programs». En: *Symbolic and Quantitative Approaches to Reasoning with Uncertainty: 6th European Conference, ECSQARU 2001 Toulouse, France, September 19-21, 2001 Proceedings*. Ed. por Benferhat, Salem y Besnard, Philippe. Berlin, Heidelberg: Springer Berlin Heidelberg, 2001, págs. 748-759. ISBN: 978-3-540-44652-1. https://doi.org/10.1007/3-540-44652-4_66.
- [15] DEAN, Jeffrey y GHEMAWAT, Sanjay. «MapReduce: Simplified Data Processing on Large Clusters». En: *Commun. ACM* 51.1 (ene. de 2008), págs. 107-113. ISSN: 0001-0782. <https://doi.org/10.1145/1327452.1327492>.
- [16] ESTEVA, F. «Negaciones en retículos completos». En: *Stochastica* I (1975), págs. 49-66.
- [17] ESTEVA, F y DOMINGO, X. «Sobre funciones de negación en $[0,1]$ ». En: *Stochastica* IV (1980), págs. 141-166.
- [18] ESTEVA, F; TRILLAS, E. y DOMINGO, X. «Weak and Strong Negation Functions in Fuzzy Set Theory». En: *Proc. XI Int. Symposium on Multivalued Logic*. 1981, págs. 23-26.
- [19] ESTEVA, Francesc; GODO, Lluís; HÁJEK, Petr y NAVARA, Mirko. «Residuated fuzzy logics with an involutive negation». English. En: *Archive for Mathematical Logic* 39.2 (2000), págs. 103-124. ISSN: 0933-5846. <https://doi.org/10.1007/s001530050006>.
- [20] FERNÁNDEZ, A.; CARMONA, C.J.; JESUS, M.J. del y HERRERA, F. «A View on Fuzzy Systems for Big Data: Progress and Opportunities». En: *International Journal of Computational Intelligence Systems* 9.1 (2016). TIN2014-57251-P, P11-TIC-7765, UJA2014/06/15, págs. 69-80.
- [21] GEORGESCU, G. y POPESCU, A. «Non-commutative fuzzy structures and pairs of weak negations». En: *Fuzzy Sets and Systems* 143 (2004), págs. 129-155.
- [22] GOCHET, Paul; GRÉGOIRE, Eric; GRIBOMONT, Pascal; LOUIS, Georges; SANCHEZ, Eduardo; SNYERS, Dominique y WODON, Pierre. *From Standard Logic to Logic Programming: Introducing a Logic Based Approach to Artificial Intelligence*. Ed. por Thayse, André. New York, NY, USA: John Wiley & Sons, Inc., 1988. ISBN: 0-471-91838-5.
- [23] GOGUEN, J.A. «L-fuzzy sets». En: *Journal of Mathematical Analysis and Applications* 18.1 (1967), págs. 145-174. ISSN: 0022-247X. [https://doi.org/http://dx.doi.org/10.1016/0022-247X\(67\)90189-8](https://doi.org/http://dx.doi.org/10.1016/0022-247X(67)90189-8).

- [24] GRANZEN, Achim. «CIO considerations for Big Data». En: Microsoft Services, 2012.
- [25] KLEMENT; MESIAR y PAP. *Triangular Norms*. Trends in Logic. Kluwer Academic Publishers, 2000.
- [26] KÓCZY, László T. *Classification and recognition of movement sequences*. 7th European Symposium on Computational Intelligence and Mathematics, Cádiz. 7 de oct. de 2015. URL: <http://escim2015.uca.es/program/keynote-speakers/>.
- [27] KÖLTRINGER, Thomas y GRECHENIG, Thomas. «Comparing the Immediate Usability of Graffiti 2 and Virtual Keyboard». En: *CHI '04 Extended Abstracts on Human Factors in Computing Systems*. CHI EA '04. Vienna, Austria: ACM, 2004, págs. 1175-1178. ISBN: 1-58113-703-6. <https://doi.org/10.1145/985921.986017>.
- [28] LALOMIA, Mary. «User Acceptance of Handwritten Recognition Accuracy». En: *Conference Companion on Human Factors in Computing Systems*. CHI '94. Boston, Massachusetts, USA: ACM, 1994, págs. 107-108. ISBN: 0-89791-651-4. <https://doi.org/10.1145/259963.260086>.
- [29] LIN, Chin-Teng y LEE, C. S. George. *Neural Fuzzy Systems: A Neuro-fuzzy Synergism to Intelligent Systems*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1996. ISBN: 0-13-235169-2.
- [30] LLOYD, J. W. *Foundations of Logic Programming*. New York, NY, USA: Springer-Verlag New York, Inc., 1984. ISBN: 0-387-13299-6.
- [31] MADRID, N. y OJEDA-ACIEGO, M. «On coherence and consistence in fuzzy answer set semantics for residuated logic programs». En: *Lect. Notes in Computer Science* (2009), 5571:60-67.
- [32] MADRID, N. y OJEDA-ACIEGO, M. «On the existence and unicity of stable models in normal residuated logic programs». En: *International Journal of Computer Mathematics* 89.3 (2012), págs. 310-324.
- [33] MEDINA, J.; OJEDA-ACIEGO, M. y VOJTÁŠ, P. «Multi-adjoint logic programming with continuous semantics». En: *Logic Programming and Non-Monotonic Reasoning, LPNMR'01*. Lecture Notes in Artificial Intelligence 2173. 2001, págs. 351-364.
- [34] MELNIK, Sergey; GUBAREV, Andrey; LONG, Jing Jing; ROMER, Geoffrey; SHIVAKUMAR, Shiva; TOLTON, Matt y VASSILAKIS, Theo. «Dremel: Interactive Analysis of Web-scale Datasets». En: *Proc. VLDB Endow.* 3.1-2 (sep. de 2010), págs. 330-339. ISSN: 2150-8097. <https://doi.org/10.14778/1920841.1920886>.
- [35] MORABITO, V. *Big Data and Analytics: Strategic and Organizational Impacts*. Springer International Publishing, 2015. ISBN: 9783319106656. URL: <https://books.google.es/books?id=9lx0BgAAQBAJ>.
- [36] MUNDO, Redacción BBC. *Amazonía: la deforestación amenaza la mitad de las especies de árboles, tipo @ONLINE*. 2015. URL: http://www.bbc.com/mundo/noticias/2015/11/151121_amazonia_arboles_extincion_am.
- [37] NEUMANN, John Von. *Theory of Self-Reproducing Automata*. Ed. por Burks, Arthur W. Champaign, IL, USA: University of Illinois Press, 1966.
- [38] PAVELKA, J. «On fuzzy logic I, II, III». En: *Zeitschr. f. Math. Logik und Grundl. der Math.* 25 (1979).
- [39] PIMENTEL, Toni; RAMOS, Fernando M. y SANDRI, Sandra. «Using Fuzzy Multilayer Perceptrons for the Classification of Time Series». En: *Fuzzy Logic and Applications: 10th International Workshop, WILF 2013, Genoa, Italy, November 19-22, 2013. Proceedings*. Ed. por Masulli, Francesco; Pasi, Gabriella y Yager, Ronald. Cham: Springer International Publishing, 2013, págs. 60-67. ISBN: 978-3-319-03200-9. https://doi.org/10.1007/978-3-319-03200-9_7.
- [40] PIMENTEL, Toni; RAMOS, Fernando M. y SANDRI, Sandra. «Classification of Land Use and Land Cover in the Brazilian Amazon Using Fuzzy Multilayer Perceptrons». En: *Int. J. Nat. Comput. Res.* 5.1 (ene. de 2015), págs. 57-71. ISSN: 1947-928X. <https://doi.org/10.4018/ijncr.2015010104>.
- [41] PLUNKETT, Tom; MACDONALD, Brian; NELSON, Bruce; HORNICK, Mark; SUN, Helen; MOHIUDDIN, Khader; HARDING, Debra; MISHRA, Gokula; STACKOWIAK, Robert; LAKER, Keith y SEGLEAU, David. *Oracle Big Data Handbook*. 1st. McGraw-Hill Osborne Media, 2013. ISBN: 0071827269, 9780071827263.
- [42] SAN-MIN, Wang. «Logics for residuated pseudo-uninorms and their residua». En: *Fuzzy Sets and Systems* 218.0 (2013). Theme: Logic and Algebra, págs. 24-31. ISSN: 0165-0114. <https://doi.org/http://dx.doi.org/10.1016/j.fss.2012.11.018>.

- [43] SANTOS, Pettras Leonardo Bueno dos; CAMPOS VELHO, Haroldo Fraga de; CINTRA, Rosangela y SANDRI, Sandra. «Chaotic Systems Predictability Using Neuro-Fuzzy Systems and Neural Networks with Bred Vectors». En: *Recent Developments and New Directions in Soft Computing*. Ed. por Zadeh, Lotfi A.; Abbasov, Ali M.; Yager, Ronald R.; Shahbazova, Shahnaz N. y Reformat, Marek Z. Cham: Springer International Publishing, 2014, págs. 295-313. ISBN: 978-3-319-06323-2. https://doi.org/10.1007/978-3-319-06323-2_19.
- [44] SANTOS-VARELA, A. *El mercado de las energías renovables en Brasil, tipo @ONLINE*. 2013. URL: http://espanha-brasil.org/img/documentos/16_document.pdf (visitado 01-10-2013).
- [45] SUGENO, M. *Theory of fuzzy integrals and its applications*. Tesis doct. Tokyo Institute of Technology, 1974.
- [46] TORMÁSI, A. y KÓCZY, L. T. «Fuzzy-based multi-stroke character recognizer». En: *2013 Federated Conference on Computer Science and Information Systems*. Sep. de 2013, págs. 671-674.
- [47] TORMÁSI, A. y KÓCZY, L. T. «Improving the Accuracy of a Fuzzy-Based Single-Stroke Character Recognizer by Antecedent Weighting». En: *Recent Developments and New Directions in Soft Computing*. Ed. por Zadeh, Lotfi A.; Abbasov, Ali M.; Yager, Ronald R.; Shahbazova, Shahnaz N. y Reformat, Marek Z. Cham: Springer International Publishing, 2014, págs. 165-179. ISBN: 978-3-319-06323-2. https://doi.org/10.1007/978-3-319-06323-2_11.
- [48] TORMÁSI, Alex y BOTZHEIM, János. «Single-Stroke Character Recognition with Fuzzy Method». En: *New Concepts and Applications in Soft Computing*. Ed. por Balas, Valentina Emilia; Fodor, János y Várkonyi-Kóczy, Annamária R. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, págs. 27-46. ISBN: 978-3-642-28959-0. https://doi.org/10.1007/978-3-642-28959-0_2.
- [49] TORMÁSI, Alex y KÓCZY, László T. «Identification of the initial rule-base of a multi-stroke fuzzy-based character recognition method with meta-heuristic techniques». En: *Czasopismo Techniczne* (2015).
- [50] TRILLAS, E. «Sobre negaciones en la teoría de conjuntos difusos». En: *Stochastica* III (1979), págs. 47-60.
- [51] VOJTÁŠ, P. «Fuzzy logic programming». En: *Fuzzy sets and systems* 124.3 (2001), págs. 361-370.
- [52] ZADEH, L.A. «Fuzzy sets». En: *Information and Control* 8 (1965), págs. 338-353.

TEMat

Aproximación avariciosa en espacios de Banach

✉ Pablo M. Berná
Universidad de Murcia
pmb11991@gmail.com

Resumen: En este artículo presentamos una pequeña interacción entre teoría de aproximación y análisis funcional. Queremos estudiar ciertos tipos de convergencia del llamado algoritmo de aproximación *greedy* y, para ello, veremos qué condiciones han de cumplir las bases de los espacios de Banach con las cuales trabajamos.

Palabras clave: quasi greedy bases, greedy bases, bases incondicionales.

MSC2010: 46B15, 41A65.

Recibido: 28 de diciembre de 2016.

Aceptado: 13 de febrero de 2017.

Agradecimientos: Quiero agradecer su tiempo, apoyo y dedicación a los profesores Óscar Blasco (Universitat de València), con el cual empecé a desarrollar mi investigación durante el Trabajo de Fin de Máster, y Gustavo Garrigós (Universidad de Murcia), mi tutor actual de doctorado.

Referencia: BERNÁ, Pablo M. «Aproximación avariciosa en espacios de Banach». En: *TEMat*, 1 (2017), págs. 69-78. ISSN: 2530-9633. URL: <http://temat.anemat.com/articulo/2017-p69/>.

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

1. Introducción

En este artículo vamos a tratar un problema que se remonta a Fourier [8, 9] y Taylor [12]: cómo representar una función de forma aproximada. Esta cuestión se ha vuelto muy interesante y son muchos los investigadores que han trabajado para dar solución a este problema en distintos ámbitos. El primer paso que se ha de hacer para representar una función es elegir un sistema de representación. Un sistema de representación en un espacio vectorial topológico \mathbb{H} es una colección de elementos $\{e_j\}_{j=1}^{\infty}$ de tal forma que para cualquier $f \in \mathbb{H}$ existe una sucesión de escalares $\{a_j\}_{j=1}^{\infty}$ tal que

$$f = \sum_{j=1}^{\infty} a_j e_j,$$

convergiendo en \mathbb{H} . Tradicionalmente, un sistema de este tipo tiene características como minimalidad, ortogonalidad, etc. Algunos de los sistemas clásicos son el sistema trigonométrico, el sistema algebraico, el sistema *wavelet*, etc. Se pueden consultar en el artículo de Korobeinik [11] más propiedades y ejemplos sobre sistemas de representación.

En general, podemos hablar de estructuras genéricas como un sistema biortogonal completo, pero, en esta ocasión, supondremos que estamos trabajando con una base de Schauder $\mathcal{B} = \{e_n\}_{n=1}^{\infty}$ en un espacio de Banach real $(\mathbb{X}, \|\cdot\|)$ con los funcionales biortogonales asociados $\{e_n^*\}_{n=1}^{\infty}$, donde $e_n^* \in \mathbb{X}^* \forall n \in \mathbb{N}$.

A modo de recordatorio, si \mathbb{X} es un espacio de Banach, \mathbb{X}^* es el espacio dual, que se define como el conjunto de las aplicaciones lineales y continuas que van de \mathbb{X} a \mathbb{K} ($= \mathbb{R}$ o \mathbb{C}). Los funcionales biortogonales $\{e_n^*\}_n$ asociados a una base $\{e_n\}_n$ son elementos del dual \mathbb{X}^* tales que $e_n^*(e_n) = 1$ y $e_n^*(e_j) = 0$ con $j \neq n$. Una base de Schauder $\mathcal{B} = \{e_n\}_{n=1}^{\infty}$ es aquella en la que para cada $x \in \mathbb{X}$ existe una única sucesión de coeficientes $\{a_n\}_{n=1}^{\infty}$, con $a_n \in \mathbb{R} \forall n \in \mathbb{N}$, tal que

$$x = \sum_{n=1}^{\infty} a_n e_n.$$

Además, podemos dar los coeficientes a_n usando los funcionales biortogonales, donde

$$e_n^*(x) = \sum_{j=1}^{\infty} a_j e_n^*(e_j) = a_n.$$

Una vez que tenemos fijada la base, buscamos un algoritmo de aproximación, que no es más que una colección de aplicaciones $(T_m)_{m=1}^{\infty}$, con $T_m : \mathbb{X} \rightarrow \mathbb{X}$, tal que $T_m(x)$ es una combinación lineal de m elementos. Un ejemplo clásico de este tipo de algoritmos es el algoritmo lineal de las sumas parciales $\{S_m\}_{m=1}^{\infty}$, donde

$$S_m \left(\sum_{n=1}^{\infty} a_n e_n \right) = \sum_{n=1}^m a_n e_n.$$

En este artículo vamos a presentar un algoritmo introducido por Konyagin y Temlyakov [10], el llamado algoritmo *greedy* (o algoritmo avaricioso). Para definir el algoritmo, tomamos $x = \sum_{n=1}^{\infty} e_n^*(x) e_n$ y consideramos el reordenamiento

$$\sum_{n=1}^{\infty} e_{\rho(n)}^*(x) e_{\rho(n)},$$

donde ρ es el **ordenamiento greedy natural**, esto es, $\rho : \mathbb{N} \rightarrow \mathbb{N}$, $\text{supp}(x) = \{j : e_j^*(x) \neq 0\} \subset \rho(\mathbb{N})$ y $(|e_{\rho(n)}^*(x)|)_{n=1}^{\infty}$ es no-creciente y con la condición de que si se tiene algún par de índices $i < j$ con $|e_{\rho(i)}^*(x)| = |e_{\rho(j)}^*(x)|$, entonces $\rho(i) < \rho(j)$. Esto último quiere decir que si tenemos coeficientes que en valor absoluto coinciden, entonces al aplicar ρ mantienen el orden que tenían en la serie original. Definimos entonces la suma *greedy* de orden m como

$$\mathcal{G}_m(x) := S_m \left(\sum_{n=1}^{\infty} e_{\rho(n)}^*(x) e_{\rho(n)} \right) = \sum_{n=1}^m e_{\rho(n)}^*(x) e_{\rho(n)}.$$

En otras palabras, la suma *greedy* de orden m está formada por m coeficientes que, en valor absoluto, son los de mayor tamaño. Esto último es lo que justifica el término «avaricioso». El algoritmo *greedy* es la colección $\{\mathcal{G}_m\}_{m=1}^{\infty}$.

Ejemplo 1. Sea $x = (1, 0, -1, 1/2, 2, 0, 1/3, 1/4, 1/5, 1/6, 1/1, 1/8 \dots) \in c_0$, donde c_0 se define como $c_0 := \{x = (x_n)_n \in \mathbb{K}^{\mathbb{N}} : \lim_{n \rightarrow \infty} |x(n)| = 0\}$. Reordenando el elemento x con ρ nos quedaría así: $x_\rho = (2, 1, -1, 1/2, 1/3, 1/4, 1/5, 1/6, \dots)$. Entonces, unos ejemplos de sumas *greedy* de distinto orden serían

$$\mathcal{G}_2(x) = 2e_5 + e_1, \quad \mathcal{G}_4(x) = 2e_5 + e_1 - e_3 + \frac{1}{2}e_4. \quad \blacktriangleleft$$

Pues bien, una vez que tenemos definido el algoritmo podemos hacernos varias preguntas sobre la convergencia, como, por ejemplo, ¿cuándo converge el algoritmo? ¿Cuándo el algoritmo produce la mejor aproximación? Para dar respuesta a estas preguntas, vamos a estudiar qué condiciones debe satisfacer la base de Schauder con la que estamos trabajando.

2. Bases *quasi greedy*

La primera pregunta a la que vamos a dar respuesta es a cuándo el algoritmo *greedy* converge. Para ello, Konyagin y Temlyakov [10] introdujeron el término de base *quasi greedy*.

Definición 1. Diremos que \mathcal{B} es una base *quasi greedy* si existe una constante $C \geq 1$ tal que

$$(1) \quad \|\mathcal{G}_m(x)\| \leq C\|x\|, \quad \forall x \in \mathbf{X}, m \in \mathbb{N}.$$

Denotamos por C_{qg} a la constante más pequeña que satisface la desigualdad (1). ◀

A priori, parece que esta definición no tiene nada que ver con la convergencia del algoritmo, pero en el año 2000 Wojtaszczyk [14] probó el siguiente teorema:

Teorema 1. Una base \mathcal{B} es *quasi greedy* si y solo si

$$\lim_{m \rightarrow \infty} \|x - \mathcal{G}_m(x)\| = 0, \quad \forall x \in \mathbf{X}.$$

Por tanto, tenemos caracterizada la convergencia en norma del algoritmo a través de la acotación uniforme de las sumas *greedy*. Veamos un par de ejemplos:

Ejemplo 2. Sea \mathbf{X} el espacio (real) de Banach formado por todas las sucesiones $(a_n)_{n \in \mathbb{N}}$ con

$$\|a\| := \sup_{M \geq 1} \left| \sum_{n=1}^M a_n \right| < \infty.$$

Vamos a ver que la base canónica no es *quasi greedy*, es decir, tenemos que comprobar que

$$\sup_{a, m} \frac{\|\mathcal{G}_m(a)\|}{\|a\|} = \infty.$$

Para ello, dado $m \in \mathbb{N}$, tomamos el elemento $b = (-1, 2 - \frac{\varepsilon}{m}, -2, 2 - \frac{\varepsilon}{m}, -2, \dots, 2 - \frac{\varepsilon}{m}, -2, 0, \dots)$, que tiene norma $1 + \varepsilon$, con $\varepsilon > 0$. Los cardinales de los conjuntos $\{j : b_j = 2 - \varepsilon/m\}$ y $\{j : b_j = -2\}$ son m . Entonces,

$$\sup_{a, m'} \frac{\|\mathcal{G}_{m'}(a)\|}{\|a\|} \geq \frac{\|\mathcal{G}_m(b)\|}{1 + \varepsilon} = \frac{\|(-2, -2, \dots, -2)\|}{1 + \varepsilon} = \frac{2m}{1 + \varepsilon}.$$

Haciendo ahora $\varepsilon \rightarrow 0$ y $m \rightarrow \infty$, obtenemos el resultado. ◀

Ejemplo 3. La base canónica del espacio $\ell^1 \oplus c_0$, $\mathcal{B} = \{e_n\}_n$ con $e_i = (0, \dots, 0, 1, 0, \dots)$, estando el 1 en la posición i -ésima, es *quasi greedy*. Esto es debido a que esta base es incondicional (luego veremos la relación entre estos dos conceptos). ◀

3. Bases greedy

En teoría de aproximación es útil saber desde un punto de vista teórico cuándo un algoritmo es eficiente, en otras palabras, cuándo el algoritmo produce una muy buena aproximación. En nuestro caso, queremos estudiar cuándo el error *greedy* de orden m , $\|x - \mathcal{G}_m(x)\|$, es comparable con el error óptimo de aproximar x usando cualquier combinación de m elementos que puede generarse con los elementos de la base, el cual es denotado por $\sigma_m(x)$. Definimos formalmente $\sigma_m(x)$ como

$$\sigma_m(x) := \inf \left\{ \left\| x - \sum_{j \in A} a_j e_j \right\| : a_j \in \mathbb{R}, A \subset \mathbb{N}, |A| = m \right\}.$$

Obviamente, siempre tenemos que $\sigma_m(x) \leq \|x - \mathcal{G}_m(x)\|$ ya que la suma *greedy* de orden m es una combinación de m elementos particular. Por tanto, estamos interesados en saber cuándo $\|x - \mathcal{G}_m(x)\|$ está acotado superiormente (salvo por una constante) por $\sigma_m(x)$.

Definición 2. Decimos que \mathcal{B} es una base **greedy** si existe una constante $C \geq 1$ tal que

$$(2) \quad \|x - \mathcal{G}_m(x)\| \leq C\sigma_m(x), \quad \forall x \in \mathbb{X}, m \in \mathbb{N}.$$

Denotamos por C_g a la constante más pequeña que satisface la desigualdad (2). ◀

Nota 1. Una relación directa es que si la base es *greedy* entonces es *quasi greedy*. Esto se da ya que $\sigma_m(x) \rightarrow 0$ cuando $m \rightarrow \infty$ en cualquier espacio de Banach, por lo que $\|x - \mathcal{G}_m(x)\| \rightarrow 0$ cuando m crece a infinito y por el teorema 1, la base es *quasi greedy*. ◀

Una lista de ejemplos de bases *greedy* es la siguiente:

1. Cualquier base ortonormal en un espacio de Hilbert \mathbb{H} es *greedy* con constante $C_g = 1$, es decir,

$$\|x - \mathcal{G}_m(x)\| = \sigma_m(x), \quad \forall x \in \mathbb{H}, \forall m \in \mathbb{N}.$$

2. La base canónica $\mathcal{B} = \{e_n\}_{n=1}^\infty$, donde $e_i = (0, \dots, 0, 1, 0, \dots)$ con el 1 en la posición i -ésima, es *greedy* en el espacio ℓ^p , $1 \leq p < \infty$, con constante $C_g = 1$.
3. El sistema de Haar es una base *greedy* en los espacios $L_p([0, 1])$ con $1 < p < \infty$. Este resultado fue probado por Temlyakov [13].

Ahora bien, si queremos saber si una base es o no *greedy* por definición es algo complicado. Por ello, Konyagin y Temlyakov dieron una caracterización basada en dos nociones: bases incondicionales y democráticas.

Tradicionalmente, la definición clásica de base incondicional es aquella en la que cualquier reordenada de la serie que define cualquier elemento del espacio es convergente. En este caso, vamos a usar una definición equivalente que involucra al operador proyección P_A .

Definición 3. Una base \mathcal{B} es **suppression unconditional** (o **incondicional de tipo supresivo**) si existe una constante $K \geq 1$ tal que

$$(3) \quad \|P_A(x)\| \leq K\|x\|, \quad \forall x \in \mathbb{X}, \forall A \subset \mathbb{N},$$

donde $P_A(\sum_{n=1}^\infty a_n e_n) = \sum_{n \in A} a_n e_n$.

Denotamos por K_s a la constante más pequeña que satisface la desigualdad (3). ◀

Nota 2. La equivalencia entre base incondicional de tipo supresivo y base incondicional puede encontrarse en el trabajo de Facenda Aguirre [7]. ◀

Nota 3. Cuando la base no es incondicional de tipo supresivo diremos que la base es condicional. ◀

Nota 4. Las sumas *greedy* son en realidad unas proyecciones sobre unos conjuntos particulares, es decir, $\mathcal{G}_m(x) = P_B(x)$, con $|B| = m$ y $B = \{\rho(1), \dots, \rho(m)\}$. Por tanto, si \mathcal{B} es incondicional de tipo supresivo, es decir, $\|P_A(x)\| \leq C\|x\|$ para todo $x \in \mathbb{X}$ y $A \subset \mathbb{N}$, entonces la base es *quasi greedy*. Luego, en el ejemplo 3, la base canónica en el espacio $\ell^1 \oplus c_0$ es *quasi greedy* ya que la base es incondicional de tipo supresivo. ◀

Definición 4. Decimos que una base \mathcal{B} es **democrática** si para todo par de conjuntos finitos A y B de mismo cardinal, existe una constante $C \geq 1$ tal que

$$(4) \quad \frac{1}{C} \|\mathbf{1}_B\| \leq \|\mathbf{1}_A\| \leq C \|\mathbf{1}_B\|,$$

donde $\mathbf{1}_S = \sum_{n \in S} e_n$.

Denotamos por C_d la constante más pequeña que satisface la desigualdad (4). ◀

Con estas dos nociones, Konyagin y Temlyakov [10] probaron el siguiente teorema:

Teorema 2. Una base \mathcal{B} es greedy si y solo si \mathcal{B} es incondicional de tipo supresivo y democrática. Además, si \mathcal{B} es greedy con constante C_g , entonces \mathcal{B} es incondicional de tipo supresivo y democrática con constantes $K_s \leq C_g$ y $C_d \leq C_g$. Recíprocamente, si \mathcal{B} es incondicional de tipo supresivo y democrática con constantes K_s y C_d , respectivamente, entonces \mathcal{B} es greedy con constante $C_g \leq K_s + K_s^3 C_d$.

Demostración. En esta prueba vamos a ver solo que si \mathcal{B} es greedy, entonces \mathcal{B} es democrática e incondicional de tipo supresivo. La prueba completa se encuentra en el artículo de Konyagin y Temlyakov [10]. Supongamos pues que la base \mathcal{B} es greedy.

- Para ver que es democrática, tomamos dos conjuntos arbitrarios A y B de mismo cardinal, y definimos el vector $x = (1 + \varepsilon)\mathbf{1}_{B \setminus A} + \mathbf{1}_A$, con $\varepsilon > 0$. Entonces, si denotamos por $m = |A \setminus B| = |B \setminus A|$, se tiene que $\mathcal{G}_m(x) = (1 + \varepsilon)\mathbf{1}_{B \setminus A}$, y usando que la base es greedy se obtiene que

$$\|\mathbf{1}_A\| = \|x - \mathcal{G}_m(x)\| \leq C_g \sigma_m(x) \leq C_g \|x - \mathbf{1}_{A \setminus B}\| \leq C_g \|\mathbf{1}_B\| + C_g \varepsilon \|\mathbf{1}_{B \setminus A}\|.$$

Haciendo tender ε a 0, obtenemos que la base es democrática con constante $C_d \leq C_g$.

- Para ver que es incondicional de tipo supresivo, cogemos un elemento $x \in \mathbb{X}$ de soporte B finito, un conjunto $A \subset B$ y $\alpha \in \mathbb{R}$ que satisface la condición

$$\alpha > \sup_{n \in A} |e_n^*(x)| + \sup_{n \in B \setminus A} |e_n^*(x)|.$$

Ahora, definimos el elemento $y = x + \alpha \mathbf{1}_{B \setminus A} = \sum_{n \in B \setminus A} (\alpha + e_n^*(x)) + P_A(x)$. Gracias a la condición que imponemos a α , se puede afirmar que $\mathcal{G}_m(y) = \sum_{n \in B \setminus A} (\alpha + e_n^*(x))$, con $m = |B \setminus A|$, y por tanto, usando que la base es greedy obtenemos que

$$\|P_A(x)\| = \|y - \mathcal{G}_m(y)\| \leq C_g \sigma_m(y) \leq C_g \|y - \alpha \mathbf{1}_{B \setminus A}\| = C_g \|x\|.$$

Usando ahora un argumento sencillo de densidad, se deduce que $\|P_A(x)\| \leq C_g \|x\|$, para todo elemento $x \in \mathbb{X}$ y todo conjunto $A \subset \mathbb{N}$, luego \mathcal{B} es incondicional de tipo supresivo con constante $K_s \leq C_g$. ■

Veamos un ejemplo sencillo de base democrática e incondicional de tipo supresivo (y por tanto greedy).

Ejemplo 4. Sean $\mathbb{X} = \ell^p$ con $1 \leq p < \infty$ y $\mathcal{B} = \{e_n\}_n$ la base canónica. Sean $A \subset \mathbb{N}$ y $x \in \mathbb{X}$. Entonces,

$$\|P_A(x)\|_p = \left(\sum_{n \in A} |e_n^*(x)|^p \right)^{1/p} \leq \left(\sum_{n=1}^{\infty} |e_n^*(x)|^p \right)^{1/p} = \|x\|_p.$$

Por lo tanto, la base es incondicional de tipo supresivo con constante $K_s = 1$. Veamos que la base es democrática. Sean $A, B \subset \mathbb{N}$ de cardinal m , entonces

$$\left\| \sum_{n \in A} e_n \right\|_p = \left(\sum_{n \in A} 1 \right)^{1/p} = m^{1/p} = \left(\sum_{n \in B} 1 \right)^{1/p} = \left\| \sum_{n \in B} e_n \right\|_p,$$

por lo que la base es democrática con constante $C_d = 1$. Usando el teorema 2, la base es greedy. ◀

En general, *quasi greedy* no implica que la base sea incondicional de tipo supresivo (ver el ejemplo 6 de la sección 5), aunque, recientemente, Albiac y Ansorena [1] han probado el siguiente teorema:

Teorema 3. Una base \mathcal{B} es incondicional de tipo supresivo con $K_s = 1$ si y solo si \mathcal{B} es quasi greedy con constante $C_{qg} = 1$, es decir,

$$\|P_A(x)\| \leq \|x\|, \forall x \in \mathbb{X}, \forall A \subset \mathbb{N} \iff \|\mathcal{G}_m(x)\| \leq \|x\|, \forall x \in \mathbb{X}, \forall m \in \mathbb{N}.$$

4. El problema de la constante 1

Desde el punto de vista de la teoría de aproximación, cuando una base es *greedy*, es decir, que para todo elemento $x \in \mathbb{X}$ y todo natural m ,

$$\sigma_m(x) \leq \|x - \mathcal{G}_m(x)\| \leq C_g \sigma_m(x),$$

se dice que el algoritmo produce la «casi» mejor aproximación. Decimos «casi» debido a que tenemos una constante C_g en la definición. En el caso de que $C_g = 1$, se dice que el algoritmo produce la mejor aproximación ya que tendríamos la igualdad $\|x - \mathcal{G}_m(x)\| = \sigma_m(x)$. La pregunta es «¿cuándo se consigue la constante $C_g = 1$?».

Gracias al teorema de caracterización de las bases *greedy* de Konyagin y Temlyakov, si tenemos que la base es *greedy* con $C_g = 1$, entonces tenemos que la base es democrática e incondicional con constantes $K_s = C_d = 1$, lo cual hace que la situación sea perfecta. En cambio, en el recíproco solo se puede afirmar que si la base es democrática e incondicional con constantes $C_d = K_s = 1$, entonces la base es *greedy* con constante $C_g \leq 2$. Por lo que ¿cómo se puede recuperar que $C_g = 1$? Pues bien, esta respuesta fue dada por Albiac y Wojtaszczyk [3] en 2006.

Para dar solución a esta cuestión, estos dos autores introdujeron la llamada propiedad (A). La definición formal de esta propiedad es la siguiente: definimos $M(x)$ como el conjunto de índices asociados a los coeficientes de x de mayor tamaño en valor absoluto y Π_G el conjunto de permutaciones que mueven los coeficientes $e_j^*(x)$ con $j \in M(x)$ a posiciones donde tenemos ceros, es decir, las permutaciones mueven esos coeficientes fuera del soporte de x . Entonces, decimos que \mathcal{B} satisface la **propiedad (A)** si

$$\|x\| = \left\| \sum_{n \in \text{supp}(x)} \varepsilon_{\pi(n)} e_n^*(x) e_{\pi(n)} \right\|$$

para toda permutación $\pi \in \Pi_G$ y signos ε_k tales que $\varepsilon_{\pi(n)} = 1$ si $\pi(n) = n$ y $\varepsilon_{\pi(n)} = \pm 1$ si $\pi(n) \neq n$.

Nota 5. Los coeficientes $e_j^*(x)$ con $j \notin M(x)$ quedan en la misma posición. Además, en el caso en el que el $\text{supp}(x)$ sea \mathbb{N} , la única permutación posible es la identidad. ◀

Ejemplo 5. Un ejemplo de permutación y cambio de signo sería el siguiente: supongamos que tenemos un elemento x en un espacio de Banach de tal forma que su sucesión de coeficientes es $(3, 3, 3, 0, -1, 0, 0, 2, \dots)$. En este caso, $M(x) = \{1, 2, 3\}$, escogemos la permutación π de la forma $\pi(1) = 1, \pi(2) = 4, \pi(3) = 6$ y el cambio de signo ε de la forma $\varepsilon_{\pi(2)} = -1$ y $\varepsilon_{\pi(3)} = 1$. Con lo cual, obtenemos el siguiente vector: $(3, 0, 0, -3, -1, 3, 0, 2, \dots)$. ◀

Usando pues esta nueva propiedad, Albiac y Wojtaszczyk [3] probaron el siguiente teorema:

Teorema 4. Una base es greedy con constante $C_g = 1$ si y solo si la base es incondicional de tipo supresivo con constante $K_s = 1$ y satisface la propiedad (A).

Vemos entonces que para poder tener la mejor aproximación es necesario reemplazar la propiedad de democracia por esta propiedad más general. Posteriormente, en el año 2014, Dilworth *et al.* [6] reformularon la propiedad (A) de una forma diferente para dar un teorema con constantes generales. La idea de reformular la propiedad (A) es que, de una cierta forma, esa propiedad involucra una cierta simetría.

Definición 5. Decimos que una base \mathcal{B} es **simétrica para coeficientes grandes** si existe una constante $C \geq 1$ tal que

$$(5) \quad \left\| x + t \sum_{n \in A} \varepsilon_n e_n \right\| \leq C \left\| x + t \sum_{n \in B} \varepsilon'_n e_n \right\|$$

para todo $x \in \mathbb{X}$ y todo par de conjuntos A y B tales que $|A| = |B| < \infty, A \cap B = \emptyset, \text{supp}(x) \cap (A \cup B) = \emptyset$ y para cualesquiera colecciones de signos $\varepsilon_n, \varepsilon'_n \in \{\pm 1\}$ y $t = \sup_{n \in \text{supp}(x)} |e_n^*(x)|$. ◀

Denotamos por C_a a la constante más pequeña que satisface la desigualdad (5).

Nota 6. Cuando $C_a = 1$ recuperamos la definición de Albiac y Wojtaszczyk. ◀

Nota 7. Si \mathcal{B} es simétrica para coeficientes grandes, entonces es democrática. Esto es sencillo de probar: sean C y D dos conjuntos arbitrarios de mismo cardinal. Tomando en (5) $x = \mathbf{1}_{C \cap D}$, $A = C \setminus D$, $B = D \setminus C$ y signos $\varepsilon_k = \varepsilon'_k = 1$, obtenemos el resultado. ◀

Gracias a esta reformulación, Dilworth *et al.* [6] probaron el siguiente teorema:

Teorema 5. Si \mathcal{B} es base greedy con constante C_g , entonces es incondicional de tipo supresivo con constante $K_s \leq C_g$ y simétrica para coeficientes grandes con constante $C_a \leq C_g$. Recíprocamente, si la base \mathcal{B} es incondicional de tipo supresivo con constante K_s y simétrica para coeficientes grandes con constante C_a , entonces la base es greedy con constante $C_g \leq K_s^2 C_a$.

Nota 8. Este teorema nos permite recuperar la constante 1 de la base greedy, ya que si tenemos que $K_s = C_a = 1$, entonces $1 \leq C_g \leq 1^2 \cdot 1 = 1$. ◀

5. Bases almost greedy

Hasta ahora hemos introducido las nociones de base greedy y quasi greedy. Si comparamos estas bases desde el punto de vista de la teoría de aproximación, se podría decir que son puntos extremos, es decir, en las bases quasi greedy solo sabemos que el algoritmo converge, pero en las bases greedy el algoritmo no solo converge, sino que produce la «casi» mejor aproximación. ¿Existirá un punto intermedio entre ambas bases? La respuesta es «sí».

En este caso, vamos estudiar cuando $\|x - \mathcal{G}_m(x)\|$ es comparable al error óptimo de aproximación usando proyecciones. Este error es denotado por $\tilde{\sigma}_m(x)$, donde

$$\tilde{\sigma}_m(x) := \inf\{\|x - P_A(x)\| : |A| = m\}.$$

Dado que la suma greedy es una proyección, como hemos comentado en la sección 3 (ver nota 4), siempre se tiene que $\tilde{\sigma}_m(x) \leq \|x - \mathcal{G}_m(x)\|$. Por ello, al igual que pasaba en las bases greedy, queremos averiguar cuándo $\|x - \mathcal{G}_m(x)\|$ está acotado superiormente por $\tilde{\sigma}_m(x)$.

Definición 6. Decimos que \mathcal{B} es una base almost greedy si existe una constante $C \geq 1$ tal que

$$\|x - \mathcal{G}_m(x)\| \leq C \tilde{\sigma}_m(x), \forall x \in \mathbb{X}, \forall m \in \mathbb{N}. \quad \blacktriangleleft$$

Nota 9. Si la base es greedy entonces la base es almost greedy. Esta implicación es trivial porque $\sigma_m(x) \leq \tilde{\sigma}_m(x)$. ◀

Ahora bien, saber por definición si una base es o no almost greedy no es tarea fácil. Por ello, en el año 2003, Dilworth *et al.* [5] probaron el siguiente teorema:

Teorema 6. Una base \mathcal{B} es almost greedy si y solo si \mathcal{B} es democrática y quasi greedy.

Ejemplo 6. Veamos un ejemplo de una base que es almost greedy pero no es greedy. En el espacio $c_{00} := \{x = (x_n)_n \in \mathbb{K}^{\mathbb{N}} : x(n) = 0 \text{ si } n > n_0 \text{ para cierto } n_0 \in \mathbb{N}\}$, definimos la siguiente norma de una sucesión:

$$\|(a_n)_n\| := \max \left\{ \left(\sum_{n=1}^{\infty} |a_n|^2 \right)^{1/2}, \sup_N \left| \sum_{n=1}^N \frac{a_n}{\sqrt{n}} \right| \right\}.$$

Tomamos ahora el espacio de Banach \mathbb{X} que surge de completar c_{00} en c_0 bajo esta norma. Entonces, en este espacio, la base canónica es una base quasi greedy y condicional (ver [2, pág. 287]). Veamos que es democrática: sea A un conjunto de cardinal m . Entonces

$$\left(\sum_{n \in A} 1 \right)^{1/2} = m^{1/2},$$

mientras que

$$\sum_{n \in A} \frac{1}{\sqrt{n}} \leq \sum_{n=1}^m \frac{1}{\sqrt{n}} \leq \int_0^m \frac{dx}{\sqrt{x}} = 2m^{1/2}.$$

Por tanto,

$$m^{1/2} \leq \|\mathbf{1}_A\| \leq 2m^{1/2},$$

lo cual implica que la base es democrática con constante $C_d \leq 2$. Usando entonces el teorema 6, la base canónica en este espacio es *almost greedy* y, usando el teorema 2, no es *greedy* debido a que la base es condicional. ◀

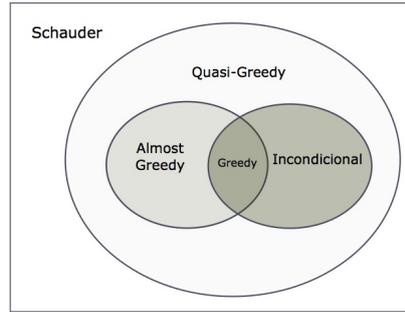


Figura 1: Las relaciones existentes entre las distintas clases de bases que se han visto a lo largo del artículo.

6. Una nueva caracterización de las bases *greedy*

En esta sección vamos a ver una nueva caracterización de las bases *greedy* usando un nuevo funcional. Este funcional es el error óptimo de aproximar $x \in \mathbf{X}$ usando polinomios de coeficientes constantes. Dicho funcional, el cual es denotado por $\mathcal{D}_m^*(x)$, es definido como

$$\mathcal{D}_m^*(x) := \inf \left\{ \left\| x - \sum_{n \in A} \alpha \varepsilon_n e_n \right\| : \alpha \in \mathbb{R}, A \subset \mathbb{N}, |A| = m, \varepsilon_n \in \{\pm 1\} \right\}.$$

Obviamente, tenemos que $\sigma_m(x) \leq \mathcal{D}_m^*(x)$ para todo elemento del espacio x y todo natural m . La motivación del uso de este funcional reside justamente en la prueba de la implicación que hemos hecho del teorema 2: para poder demostrar la democracia y la incondicionalidad hubiese bastado usar este nuevo funcional en vez de $\sigma_m(x)$, es decir, si tuviéramos una base en la cual se cumpla que

$$\|x - \mathcal{G}_m(x)\| \leq C \mathcal{D}_m^*(x), \forall x \in \mathbf{X}, \forall m \in \mathbb{N},$$

entonces la base sería democrática e incondicional de tipo supresivo. En el artículo de Blasco y Berná [4] podemos encontrar la demostración del siguiente teorema:

Teorema 7. Sea \mathcal{B} una base de un espacio de Banach \mathbf{X} . Son equivalentes:

- \mathcal{B} es *greedy*, es decir, existe una constante $C_1 \geq 1$ tal que

$$\|x - \mathcal{G}_m(x)\| \leq C_1 \sigma_m(x), \forall x \in \mathbf{X}, m \in \mathbb{N}.$$

- Existe una constante $C_2 \geq 1$ tal que

$$\|x - \mathcal{G}_m(x)\| \leq C_2 \mathcal{D}_m^*(x), \forall x \in \mathbf{X}, m \in \mathbb{N}.$$

- \mathcal{B} es incondicional de tipo supresivo y simétrica para coeficientes grandes.
- \mathcal{B} es incondicional de tipo supresivo y democrática.

Por tanto, este teorema nos dice que se pueden caracterizar las bases *greedy* usando el funcional $\mathcal{D}_m^*(x)$, y esto es llamativo ya que, *a priori*, el funcional $\mathcal{D}_m^*(x)$ es más fácil de manejar que el funcional $\sigma_m(x)$ ya que solo involucra una constante en la definición y, además, este funcional y $\sigma_m(x)$ tienen comportamiento totalmente opuesto, al menos en espacios de Hilbert, tal y como mostramos en este teorema (el cual está probado por Blasco y Berná [4]):

Teorema 8. Sean \mathbb{H} un espacio de Hilbert y \mathcal{B} una base ortonormal. Entonces,

$$\lim_{m \rightarrow \infty} \mathcal{D}_m^*(x) = \|x\|, \quad \forall x \in \mathbb{H}.$$

El comportamiento opuesto de este nuevo funcional y $\sigma_m(x)$ se deduce de la nota 1.

Para demostrar este último teorema, se hace uso de la siguiente fórmula que puede encontrarse en el artículo de Blasco y Berná [4]:

Proposición 9. Sean \mathbb{H} un espacio de Hilbert y \mathcal{B} una base ortonormal. Entonces,

$$\mathcal{D}_m^*(x) = \sqrt{\|x\|^2 - \frac{1}{m} \sup \{ \langle x, \mathbf{1}_{\varepsilon A} \rangle^2 : |A| = m, \{\varepsilon_n\}_n \in \{\pm 1\} \}},$$

donde $\mathbf{1}_{\varepsilon A} = \sum_{n \in A} \varepsilon_n e_n$.

Por otro lado, a la vista del teorema 8, es natural que nos preguntemos si esa propiedad del límite se mantiene en cualquier espacio de Banach, pero, desgraciadamente, no se tiene respuesta actualmente, es un problema abierto. Solo se dispone del siguiente resultado parcial que se puede encontrar en el artículo de Blasco y Berná [4]:

Proposición 10. Sean $\mathbb{X} = \ell^p$, $1 < p < \infty$, $\mathcal{B} = (e_n)_n$ la base canónica y $B \subset \mathbb{N}$ de cardinal N . Entonces,

$$\mathcal{D}_m^*(\mathbf{1}_B) = \begin{cases} (N - m)^{1/p} & \text{si } m \leq N, \\ N^{1/p} \left(1 + \left(\frac{m}{N} - 1\right)^{-1/(p-1)}\right)^{-1/p'} & \text{si } m \geq N, \end{cases}$$

donde $p' = \frac{p}{p-1}$. Por tanto, $\lim_{m \rightarrow \infty} \mathcal{D}_m^*(\mathbf{1}_B) = \|\mathbf{1}_B\|$.

Nota 10. En la proposición anterior, si $p = 1$ también se obtiene el mismo resultado, pero tenemos una expresión distinta para el funcional $\mathcal{D}_m^*(\mathbf{1}_B)$:

$$\mathcal{D}_m^*(\mathbf{1}_B) = \begin{cases} N - m & \text{si } m \leq N, \\ m - N & \text{si } N \leq m \leq 2N, \\ N & \text{si } m \geq 2N. \end{cases}$$

Referencias

- [1] ALBIAC, Fernando y ANSOARENA, José Luis. «Characterization of 1-quasi-greedy bases». En: *Journal of Approximation Theory* 201 (2016), págs. 7-12. <https://doi.org/10.1016/j.jat.2015.08.006>.
- [2] ALBIAC, Fernando y KALTON, Nigel. *Topics in Banach Space Theory*. Graduate Texts in Mathematics. Second Edition, Springer New York, 2016. ISBN: 978-3-319-31557-7. <https://doi.org/10.1007/978-3-319-31557-7>.
- [3] ALBIAC, Fernando y WOJTASZCZYK, Przemysław. «Characterization of 1-greedy bases». En: *Journal of Approximation Theory* 138 (2006), págs. 65-86. <https://doi.org/10.1016/j.jat.2005.09.017>.
- [4] BLASCO, Óscar y BERNÁ, Pablo M. «Characterization of greedy bases in Banach spaces». En: *Journal of Approximation Theory* 215 (2017), págs. 28-39. <https://doi.org/10.1016/j.jat.2016.11.008>.
- [5] DILWORTH, Stephen; KALTON, Nigel; KUTZAROVA, Denka y TEMPLYAKOV, Vladimir N. «The thresholding greedy algorithm, greedy bases, and duality». En: *Constructive approximation* 19 (2003), págs. 575-597. <https://doi.org/10.1007/s00365-002-0525-y>.

- [6] DILWORTH, Stephen; KUTZAROVA, Denka; ODELL, Edward; SCHLUMPRECHT, Thomas y ZSÁK, Andras. «Renorming spaces with greedy bases». En: *Journal of Approximation Theory* 188 (2014), págs. 39-56. <https://doi.org/10.1016/j.jat.2014.09.001>.
- [7] FACENDA AGUIRRE, José A. *Geometría de Espacios de Banach*. Universidad de Sevilla, Secretariado de Publicaciones, 1998.
- [8] FOURIER, Joseph. *Mémoire sur la propagation de la chaleur dans les corps solides*. 1807.
- [9] FOURIER, Joseph. *Théorie analytique de la chaleur*. 1822. <https://doi.org/10.1017/CB09780511693229>.
- [10] KONYAGIN, Sergey V. y TEMLYAKOV, Vladimir N. «A remark on greedy approximation in Banach spaces». En: *East J. Approx* 5 (1999), págs. 365-379.
- [11] KOROBĚIŇNIK, Yurii F. «On absolutely representing systems in spaces of infinitely differentiable functions». En: *Studia Mathematica* 139.2 (2000), págs. 175-188.
- [12] TAYLOR, Brook. *Methodus Incrementorum Directa et Inversa*. 1715. <https://doi.org/10.3931/e-rara-4090>.
- [13] TEMLYAKOV, Vladimir N. «The best m-term approximation and greedy algorithms». En: *Advances in Computational Mathematics* 8 (1998), págs. 249-265. <https://doi.org/10.1023/A:1018900431309>.
- [14] WOJTASZCZYK, Przemysław. «Greedy algorithm for general biorthogonal systems». En: *Journal of Approximation Theory* 107.2 (2000), págs. 293-314. <https://doi.org/10.1006/jath.2000.3512>.

TEMat

Problemas directos e inversos en combinatoria aditiva: las desigualdades de Plünnecke

✉ Alberto Espuny Díaz
University of Birmingham
axe673@bham.ac.uk

Resumen: La combinatoria aditiva es una rama de las matemáticas que ha experimentado un gran crecimiento en el último medio siglo. En este artículo nos centramos en presentar la teoría de los problemas directos e inversos relacionados con los conjuntos suma, una pequeña parte de todos los resultados en este contexto. En particular, nos centramos en presentar y demostrar unas desigualdades tradicionales de esta área, conocidas como desigualdades de Plünnecke-Ruzsa, así como una de sus aplicaciones más directas en un problema inverso.

Palabras clave: combinatoria aditiva, conjuntos suma, desigualdades de Plünnecke.

MSC2010: 11P70, 11B13.

Recibido: 22 de enero de 2017.

Aceptado: 23 de abril de 2017.

Agradecimientos: Me gustaría agradecer a los profesores Oriol Serra y Juanjo Rué por introducirme al mundo de la combinatoria aditiva, así como por toda la ayuda que me brindaron durante la realización de este trabajo.

Referencia: ESPUNY DÍAZ, Alberto. «Problemas directos e inversos en combinatoria aditiva: las desigualdades de Plünnecke». En: *TEMat*, 1 (2017), págs. 79-89. ISSN: 2530-9633. URL: <http://temat.anemat.com/articulo/2017-p79/>.

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

1. Introducción

La combinatoria aditiva es una rama de las matemáticas que se desarrolló especialmente a partir de mediados del siglo pasado. Aunque algunos resultados aislados se conocían desde tiempo antes, esta teoría tuvo un gran impulso a partir del intento de Schnirelmann de resolver la conjetura de Goldbach. Aunque los resultados dados por esta teoría fueron rápidamente mejorados con otros métodos, quedó un interés por los resultados propios de la teoría que hizo que se convirtiese en un área de investigación muy activa, en la que aún hoy quedan muchos problemas abiertos.

La combinatoria aditiva se puede entender como una parte de la combinatoria aritmética, que se especializa en conocer propiedades de conjuntos contenidos en grupos, con los que se puede operar, y de entender la interacción entre las operaciones. La combinatoria aditiva sería, entonces, aquella parte en la que se consideran solo sumas y restas, y el objetivo sería entender la estructura aditiva de los conjuntos. En esta área se combinan técnicas de teoría de números, análisis armónico, combinatoria, geometría y teoría ergódica para dar lugar a resultados muy diversos. A menudo se puede utilizar como ejemplo para mostrar cómo la interacción de técnicas muy diversas puede dar lugar a resultados cada vez más generales e interesantes. Para un lector más interesado en el tema que quiera conocer todas las ramas de la combinatoria aditiva recomendamos el libro de Tao y Vu [15]; nosotros nos vamos a concentrar en problemas de *conjuntos suma*, para los que las desigualdades de Plünnecke van a resultar esenciales. Para un estudio en mayor profundidad de las desigualdades de Plünnecke recomendamos la monografía de Espuny Díaz [1] o los capítulos de Ruzsa [13]; también se pueden leer las notas de Petridis [7].

Lo primero que necesitamos es explicar qué es un conjunto suma. En general, vamos a tratar con conjuntos finitos en cualquier grupo conmutativo, pero el lector puede pensar en conjuntos de números enteros para tener una primera idea.

Definición 1. Sea $(G, +)$ un grupo abeliano. Sea $A \subseteq G$ un conjunto no vacío cualquiera. El **conjunto suma** de A se define como

$$A + A = \{a + b : a, b \in A\}.$$

Si tenemos un segundo conjunto $B \subseteq G$, el conjunto suma de ambos se define como

$$A + B = \{a + b : a \in A, b \in B\}.$$

Se puede definir el inverso de un conjunto como «el conjunto de los elementos inversos», es decir,

$$-A = \{-a : a \in A\},$$

ya que los elementos inversos existen por ser G un grupo. Eso permite definir también el **conjunto diferencia**

$$A - A = \{a - b : a, b \in A\}.$$

En general, se define el conjunto suma iterado de manera inductiva, como

$$mA = A + (m - 1)A = \{a_1 + a_2 + \dots + a_m : a_i \in A\},$$

para $m \in \mathbb{N}$, y se pueden definir similarmente conjuntos de sumas y restas iteradas para $k, l \in \mathbb{N}$ como

$$kA - lB = \{a_1 + \dots + a_k - b_1 - \dots - b_l : a_i \in A, b_j \in B\}. \quad \blacktriangleleft$$

Nótese que estas definiciones no tienen sentido si uno de los conjuntos es vacío. Por lo tanto, aunque a veces no se indique explícitamente, todos los resultados de este artículo suponen que los conjuntos con que se trabaja tienen al menos un elemento.

La combinatoria aditiva se enfrenta principalmente a dos tipos de problemas: los problemas directos y los inversos. Los primeros son los que primero se plantearía uno: conociendo algo sobre la estructura del conjunto que tengo (y sobre el grupo en el que está, al que llamaremos el grupo ambiente), ¿qué podemos decir sobre la estructura de los conjuntos suma? En particular, ¿qué podemos decir sobre su tamaño? Y los problemas inversos funcionan exactamente al revés: dada cierta información sobre los conjuntos

suma (por ejemplo, su tamaño respecto al conjunto inicial), ¿qué podemos decir sobre la estructura del conjunto?

Una de las condiciones con las que vamos a trabajar es conocer el número de sumas de pares de elementos. Así, por ejemplo, uno de los problemas que nos interesan es, sabiendo cuántas sumas hay en $A + A$, ¿qué podemos decir sobre el número de diferencias? ¿Y sobre el número de triples sumas? Esto nos permitirá obtener información sobre la estructura de los conjuntos.

2. Cotas triviales

Se pueden dar ciertas cotas triviales sobre el tamaño de los conjuntos suma y diferencia que se aplican a cualquier conjunto. Si denotamos el tamaño de un conjunto A como $|A|$, refiriéndonos al número de elementos que tiene, para el conjunto suma básico tenemos las siguientes cotas.

Lema 1. *Sea A un conjunto finito no vacío en un grupo abeliano. Entonces,*

$$|A| \leq |A + A| \leq \binom{|A| + 1}{2}.$$

Demostración. La cota inferior es consecuencia del hecho de que $a + A \subseteq A + A$ para cualquier $a \in A$, y el conjunto $a + A$ es una traslación de A , que se puede entender como una permutación en el grupo ambiente, y por tanto tiene el mismo tamaño que A .

La cota superior es consecuencia de considerar el máximo número posible de sumas. El número de parejas de elementos que se pueden tomar es $\binom{|A|}{2}$ (consideramos parejas sin ordenar, ya que estamos en un grupo conmutativo); además, hay que considerar las parejas de la forma (a, a) , $a \in A$, de las que hay $|A|$ en total. La suma de las dos cantidades da el resultado. ■

Similarmente, se pueden establecer cotas triviales para las diferencias y las sumas iteradas.

Ejercicio 1. Encontrar las cotas triviales para el tamaño del conjunto diferencia $A - A$, donde A es un conjunto finito no vacío en un grupo abeliano. ◀

Ejercicio 2. Dados dos conjuntos finitos no vacíos A y B en un grupo abeliano G , demostrar que $|A + B| \leq |A||B|$. ◀

Ejercicio 3. Encontrar cotas triviales para el tamaño de conjuntos suma iterados nA y $A_1 + A_2 + \dots + A_n$ dados conjuntos finitos no vacíos A, A_1, \dots, A_n en un grupo abeliano. ◀

Nótese que, en general, podemos tener igualdad en las cotas triviales, para lo que basta con construir conjuntos adecuados. Entonces, para que el problema de encontrar cotas resulte interesante, debemos imponer restricciones sobre los conjuntos que consideramos.

3. Problemas directos e inversos

Para ilustrar mejor el tipo de problemas a los que nos enfrentamos, vamos a considerar un ejemplo: trabajamos con conjuntos finitos de números enteros.

Proposición 2. *Si $A \subseteq \mathbb{Z}$ es tal que $|A| = n > 0$, entonces $|A + A| \geq 2n - 1$.*

Demostración. Usando el orden normal de los números enteros, etiquetamos los elementos de A de manera que $A = \{a_1, a_2, \dots, a_n\}$ con $a_1 < a_2 < \dots < a_n$. Usando este orden, resulta evidente que

$$a_1 + a_1 < a_1 + a_2 < a_1 + a_3 < \dots < a_1 + a_n < a_2 + a_n < \dots < a_n + a_n,$$

y está claro que todos estos elementos están en $A + A$. Como en esa lista hay $2n - 1$ términos y todos ellos son diferentes, entonces $|A + A| \geq 2n - 1$. ■

Así, este resultado es un ejemplo en el que usamos la estructura del grupo ambiente (los enteros) para obtener una cota sobre el tamaño de los conjuntos suma que es mejor que la cota trivial. Se trata de una solución a un problema directo.

Proposición 3. Dado $A \subseteq \mathbb{Z}$ con $|A| = n > 0$, $|A + A| = 2n - 1$ si y solo si A es una progresión aritmética.

Demostración. Si A es una progresión aritmética el resultado es trivial, así que vamos a demostrar la otra implicación. Como hemos visto en la demostración de la proposición 2, podemos ordenar los elementos del conjunto suma como

$$a_1 + a_1 < a_1 + a_2 < a_1 + a_3 < a_1 + a_4 < \dots < a_1 + a_n < a_2 + a_n < \dots < a_n + a_n,$$

pero este no es el único orden posible. Por ejemplo, podemos considerar el orden dado por

$$a_1 + a_1 < a_2 + a_1 < a_2 + a_2 < a_2 + a_3 < \dots < a_2 + a_n < a_3 + a_n < \dots < a_n + a_n.$$

También podemos considerar

$$a_1 + a_1 < a_2 + a_1 < a_3 + a_1 < a_4 + a_1 < \dots < a_n + a_1 < a_n + a_2 < \dots < a_n + a_n$$

(en general, podemos pensar que tenemos elementos en un cuadrado de lado n , y hay tantas formas de ordenar de menor a mayor $2n - 1$ sumas como caminos desde la esquina inferior izquierda hasta la esquina opuesta moviéndonos solo hacia arriba o hacia la derecha). Como cada lista tiene $2n - 1$ elementos diferentes, los elementos de cada lista están en $A + A$ y estamos suponiendo que $|A + A| = 2n - 1$, esto quiere decir que todas las listas tienen que ser iguales. Comparando las dos primeras listas que hemos dado, tenemos que $a_1 + a_i = a_2 + a_{i-1}$ para $i \in \{2, 3, \dots, n\}$. Eso quiere decir que $a_i - a_{i-1} = a_2 - a_1$ para todo i , es decir, la diferencia entre cada dos elementos del conjunto es siempre la misma. Y esta es la caracterización de una progresión aritmética. ■

Este es un ejemplo perfecto de un problema inverso. Partiendo de una condición sobre el tamaño de los conjuntos suma, hemos llegado a una propiedad estructural muy fuerte sobre los conjuntos que consideramos. Se pueden obtener otros resultados similares e interesantes en muchos casos.

Ejercicio 4. Sean $A, B \subseteq \mathbb{Z}$ dos conjuntos finitos no vacíos de enteros, con $|A| = n$, $|B| = m$. Demostrar que $|A + B| \geq n + m - 1$, con igualdad si y solo si A y B son progresiones aritméticas con la misma diferencia. ◀

Ejercicio 5. Dado un conjunto no vacío A en un grupo abeliano, demostrar que $|A + A| = |A|$ si y solo si A es una clase lateral de un subgrupo del grupo. ◀

Ejercicio 6 (teorema de Cauchy-Davenport (ver [13, pág. 142])). Sea p un número primo y sean A y B conjuntos no vacíos en $\mathbb{Z}/p\mathbb{Z}$. Demostrar que $|A + B| \geq \min\{p, |A| + |B| - 1\}$. ¿Cuándo se da la igualdad? ◀

En general, la idea que se puede obtener de las proposiciones 2 y 3 y los ejercicios 4, 5 y 6 es que, cuanto más pequeña sea la suma, mayor es la estructura de los conjuntos. El siguiente paso es preguntarse qué ocurre (en el caso de los enteros) si no tenemos igualdad. El siguiente resultado, cuya demostración se puede encontrar en el libro de Nathanson [5], da una primera respuesta a esa pregunta.

Teorema 4 (Freïman). Sea $A \subseteq \mathbb{Z}$ un conjunto finito tal que $|A| \geq 3$ y $|A + A| = 2n - 1 + b \leq 3n - 4$, con n y b enteros. Entonces, A está contenido en una progresión aritmética de longitud $n + b$.

Vemos que se pierde un poco de estructura en el conjunto, pero aún tenemos una condición muy fuerte. Sin embargo, algo así deja de ser cierto si $|A + A| = 3n - 3$: ya se pueden encontrar ejemplos en los que el conjunto no está contenido en una progresión aritmética.

Estos resultados se pueden seguir generalizando. Uno puede definir progresiones d -dimensionales (que quiere decir que se trabaja con d diferencias, no que estemos en un espacio d -dimensional), y entonces se puede demostrar que si el conjunto suma es «pequeño» (es decir, lineal en el tamaño del conjunto original), entonces el conjunto tiene que estar contenido en una progresión aritmética generalizada, con dimensión y tamaño que dependen del factor lineal del conjunto suma (este también es un teorema de Freïman [3, theorem 2]).

Todos los resultados (que no ejercicios) de esta sección se aplican solo al caso particular en que el grupo ambiente es el de los números enteros. Obviamente, también funcionan en algunos otros grupos (en los racionales o los reales tenemos los mismos resultados), pero no funcionan en todos los grupos conmutativos. Sin embargo, sí se pueden generalizar a cualquier grupo abeliano: el resultado más general es el que presentamos a continuación.

Teorema 5 (Green-Ruzsa [4]). *Sea G un grupo abeliano y sea $A \subseteq G$ un conjunto finito tal que $|A + A| \leq \alpha|A|$ para algún $\alpha \in \mathbb{Q}$. Entonces, A está contenido en un conjunto de la forma $H + P$, donde H es un subgrupo de G y P es una progresión aritmética generalizada, tal que la dimensión de P es como máximo d y $|H||P| \leq \alpha'|A|$, donde d y α' son funciones solo de α .*

En este artículo no vamos a demostrar este teorema, pero sí una versión más débil que ya permite dar conclusiones muy interesantes. En cualquier caso, es un buen ejemplo del tipo de resultados inversos que se persiguen en combinatoria aditiva.

4. La desigualdad triangular de Ruzsa

En esta sección presentamos una de las desigualdades que más aplicación tiene para conseguir nuevas desigualdades de tamaños de conjuntos suma, y mostramos algunas de sus aplicaciones. El resultado se puede formular como sigue.

Proposición 6 (desigualdad triangular de Ruzsa [11]). *Sean X, Y y Z tres conjuntos finitos no vacíos en un grupo abeliano. Entonces,*

$$|X||Y - Z| \leq |X - Y||X - Z|.$$

Demostración. Fijemos un elemento $a = y - z \in Y - Z$. Este elemento se puede escribir de $|X|$ formas diferentes como $(x - z) - (x - y)$ (una forma para cada $x \in X$). Pero esto quiere decir que hay por lo menos $|X|$ formas de escribirlo como la diferencia de un elemento de $X - Z$ y un elemento de $X - Y$, de modo que

$$|Y - Z| \leq \frac{|(X - Y) - (X - Z)|}{|X|} \leq \frac{|X - Y||X - Z|}{|X|},$$

donde la última desigualdad es consecuencia de las cotas triviales (ejercicio 2). ■

La desigualdad triangular de Ruzsa se puede utilizar para obtener muchos otros resultados. También se pueden obtener resultados similares. Aquí presentamos algunos de ellos, y dejamos otros como ejercicios.

Corolario 7. *Sean A, B y C tres conjuntos finitos no vacíos en un grupo abeliano. Entonces,*

$$|A||B - C| \leq |A + B||A + C|.$$

Demostración. Basta con aplicar la proposición 6 tomando $X = A, Y = -B$ y $Z = -C$, y tener en cuenta que $|B - C| = |C - B|$ por la conmutatividad del grupo ambiente. ■

Corolario 8. *Sea A un conjunto finito en un grupo abeliano tal que $|A + A| \leq \alpha|A|$. Entonces,*

$$|A - A| \leq \alpha^2|A|.$$

Demostración. Aplicamos la proposición 6 tomando $X = A, Y = Z = -A$. Como $| -A + A| = |A - A|$, tenemos que

$$|A||A - A| \leq |A + A|^2,$$

y el resultado es consecuencia de aplicar la condición del enunciado. ■

Corolario 9. *Sea A un conjunto finito en un grupo abeliano tal que $|3A| \leq \alpha|A|$. Entonces,*

$$|2A - 2A| \leq \alpha^2|A|.$$

Demostración. Aplicamos la proposición 6 tomando $X = A, Y = Z = -2A$. ■

Ejercicio 7. Sea A un conjunto finito en un grupo abeliano tal que $|3A| \leq \alpha|A|$. Demostrar que

$$|4A| \leq \alpha^3|A|. \quad \blacktriangleleft$$

Ejercicio 8. Sea A un conjunto finito en un grupo abeliano tal que $|3A| \leq \alpha|A|$. Demostrar que

$$|nA| \leq \alpha^{2n-5}|A|, \quad |mA - A| \leq \alpha^{2m-2}|A| \quad \text{y} \quad |mA - lA| \leq \alpha^{2(l+m-3)}|A|$$

para todo $n \geq 3$ y $m, l \geq 2$. \blacktriangleleft

En general, observamos que la desigualdad triangular de Ruzsa nos permite obtener muchas desigualdades más generales, y la idea siempre es que si tenemos una cota (lineal, por ejemplo) para el tamaño de un determinado conjunto suma, entonces podemos conseguir cotas (también lineales) con parámetros que son una función (exponencial) de la constante conocida para distintos conjuntos suma.

Esta desigualdad se puede aplicar en casos muy diversos, pero hay algunas cosas que podemos preguntarnos y no puede responder. Por ejemplo, ¿hay un recíproco del corolario 8? Es decir, dada una cota sobre $|A - A|$, ¿podemos obtener cotas sobre $|A + A|$? Esta pregunta no se puede resolver con la desigualdad triangular de Ruzsa. De forma similar, la desigualdad triangular nunca nos permite obtener cotas para sumas de más de dos conjuntos si las cotas que conocemos son solo sobre la suma de dos. Así, aún no podemos responder otra de las preguntas principales de este artículo: dada una cota sobre el tamaño del conjunto suma, ¿qué cota podemos dar sobre el tamaño del conjunto suma de tres conjuntos? Para poder dar solución a estas dudas recurrimos a las desigualdades de Plünnecke.

5. Las desigualdades de Plünnecke

Las desigualdades de Plünnecke permiten dar cotas para la suma iterada de un conjunto una vez se conoce una cota para una suma de menos elementos. Así, si se conoce una cota para el tamaño del conjunto suma de un conjunto, se pueden dar cotas para el conjunto suma iterado tres, cuatro o, en general, k veces. Este resultado fue descubierto por Plünnecke a finales de los años sesenta [8], pero pasó desapercibido hasta que fue redescubierto por Ruzsa a finales de los ochenta [9, 10]. El teorema es en realidad más fuerte que estas cotas, y se puede escribir como sigue.

Teorema 10 (desigualdades de Plünnecke). *Sean A y B conjuntos finitos en un grupo abeliano tales que $|A + B| \leq \alpha|A|$. Entonces, existe un conjunto $X \subseteq A$ tal que para todo $k \geq 1$ se cumple que*

$$|X + kB| \leq \alpha^k|X|.$$

Esto permite obtener cotas para conjuntos suma de manera directa usando cotas triviales.

Corolario 11. *Sea A un conjunto finito en un grupo abeliano tal que $|A + A| \leq \alpha|A|$. Entonces, para todo $k \geq 1$,*

$$|kA| \leq \alpha^k|A|.$$

Demostración. Tomamos $B = A$ en el teorema 10. Así, para cada $k \geq 1$ podemos escribir

$$|kA| \leq |X + kA| \leq \alpha^k|X| \leq \alpha^k|A|,$$

para algún $X \subseteq A$. \blacksquare

La demostración original de las desigualdades de Plünnecke aplica técnicas de teoría de grafos. Plünnecke construyó lo que él llamó *grafos conmutativos*, que crecían de acuerdo a las propiedades conmutativas de los grupos ambiente, y dando cotas sobre el crecimiento en estos grafos logró demostrar su resultado¹. Esta es una demostración larga y laboriosa, para la que se deben tener en cuenta muchos detalles, y para escribirla hacen falta conocimientos básicos de teoría de grafos. Sin embargo, recientemente Petridis presentó una nueva demostración elemental de este resultado [6], que es la que presentamos aquí. Toda su demostración se basa en el siguiente lema.

¹En realidad, el resultado demostrado por Plünnecke, aunque equivalente para casi todas las aplicaciones, es ligeramente diferente al teorema 10, dando un resultado más débil en lo que se refiere al conjunto X para el que las cotas se cumplen pero permitiendo cotas ligeramente mejores si la hipótesis es sobre un conjunto de la forma $A + jB$ para algún $j > 1$.

Lema 12 (Petridis [6, proposition 2.1]). Sean X y B dos conjuntos finitos en un grupo tales que

$$\lambda := \frac{|X+B|}{|X|} \leq \frac{|Z+B|}{|Z|}$$

para todo $Z \subseteq X$. Entonces, para cualquier conjunto C finito en el mismo grupo ambiente,

$$|C+X+B| \leq \lambda|C+X|.$$

Demostración. La demostración la vamos a hacer por inducción sobre el tamaño del conjunto C . Para preparar lo que vamos a necesitar, démosle un orden a los elementos de C (nos vale cualquier orden), $C = \{c_1, c_2, \dots, c_r\}$. Definimos entonces los conjuntos X_1, \dots, X_r como $X_1 := X$ y $X_i := \{x \in X : c_i + x \notin \{c_1, \dots, c_{i-1}\} + X\}$ para todo $2 \leq i \leq r$ (es decir, cada X_i es el conjunto de elementos de X que al sumarles c_i dan lugar a sumas que no habíamos conseguido con los anteriores elementos de C). Esta definición de los conjuntos X_i nos permite escribir $C+X$ como una unión disjunta de conjuntos,

$$C+X = \bigsqcup_{i=1}^r (c_i + X_i).$$

De hecho, para cada subconjunto de C con los primeros j elementos de nuestro orden podemos escribir la misma igualdad como unión disjunta,

$$\{c_1, \dots, c_j\} + X = \bigsqcup_{i=1}^j (c_i + X) = \bigsqcup_{i=1}^j (c_i + X_i),$$

de modo que al considerar cardinales tenemos que

$$(1) \quad |\{c_1, \dots, c_j\} + X| = \sum_{i=1}^j |c_i + X_i| = \sum_{i=1}^j |X_i|.$$

El caso base de la inducción es fácil: cuando C tiene solo un elemento, $C = \{c\}$, tenemos que

$$|C+X+B| = |c+X+B| = |X+B| = \lambda|X| = \lambda|c+X| = \lambda|C+X|,$$

donde la tercera igualdad es consecuencia de la definición de λ . Para demostrar el caso general, supongamos que ya conocemos la desigualdad hasta conjuntos de tamaño $r-1$, y demostrémosla para conjuntos de tamaño r . Para ello, definimos el complemento de X_r en X , $X_r^c = X \setminus X_r$. Por definición de los conjuntos X_i tenemos que $c_r + X_r^c \subseteq \{c_1, \dots, c_{r-1}\} + X$, de modo que $c_r + X_r^c + B \subseteq \{c_1, \dots, c_{r-1}\} + X + B$ y

$$C+X+B \subseteq (\{c_1, \dots, c_{r-1}\} + X + B) \cup ((c_r + X + B) \setminus (c_r + X_r^c + B)).$$

Podemos observar que $(c_r + X + B) \setminus (c_r + X_r^c + B) = c_r + ((X+B) \setminus (X_r^c + B))$, ya que se trata de una traslación de los conjuntos. De este modo, tomando cardinales obtenemos las cotas

$$\begin{aligned} |C+X+B| &\leq |(\{c_1, \dots, c_{r-1}\} + X + B) \cup ((c_r + X + B) \setminus (c_r + X_r^c + B))| \\ &\leq |\{c_1, \dots, c_{r-1}\} + X + B| + |(c_r + X + B) \setminus (c_r + X_r^c + B)| \\ &= |\{c_1, \dots, c_{r-1}\} + X + B| + |X+B| - |X_r^c + B|, \end{aligned}$$

ya que $X_r^c + B \subseteq X + B$. En la última expresión, podemos acotar el término de la izquierda por la hipótesis de inducción, de modo que

$$|\{c_1, \dots, c_{r-1}\} + X + B| \leq \lambda |\{c_1, \dots, c_{r-1}\} + X| = \lambda \sum_{i=1}^{r-1} |X_i|,$$

donde la igualdad viene de la expresión (1). Para acotar el término de la derecha, la hipótesis del enunciado establece que $|X_r^c + B| \geq \lambda|X_r^c|$ y que $|X+B| = \lambda|X|$, de modo que

$$|X+B| - |X_r^c + B| \leq \lambda(|X| - |X_r^c|) = \lambda|X_r|.$$

Juntando ambas expresiones y aplicando (1) otra vez, concluimos que

$$|C + X + B| \leq \lambda \sum_{i=1}^r |X_i| = \lambda |C + X|. \quad \blacksquare$$

Demostración del teorema 10. Sea $X \subseteq A$ un subconjunto que minimice el cociente $\frac{|Z + B|}{|Z|}$, para $Z \subseteq A$, y sea λ el valor de este mínimo. En particular, $\lambda \leq \alpha$. La demostración se hace por inducción sobre k . El caso base vendría dado para $k = 1$, cuando tenemos que $|X + B| = \lambda |X| \leq \alpha |X|$.

Supongamos que el resultado está demostrado para $k - 1$. Entonces tenemos que

$$|X + kB| = |(k - 1)B + X + B| \leq \lambda |(k - 1)B + X| \leq \alpha |(k - 1)B + X| \leq \alpha^k |X|,$$

donde la primera desigualdad es consecuencia del lema 12 y la tercera, de la hipótesis de inducción. \blacksquare

Ahora ya hemos conseguido contestar la pregunta de cómo acotar el número de triples sumas dada una cota sobre el número de dobles sumas (las desigualdades de Plünnecke nos dicen que si $|A + A| \leq \alpha |A|$ entonces $|A + A + A| \leq \alpha^3 |A|$). También tenemos respuesta a la pregunta de cómo acotar $|A + A|$ si lo que conocemos es una cota sobre $|A - A|$: basta con tomar $B = -A$ en el enunciado del teorema 10 y aplicar cotas triviales de la misma forma que en el corolario 11 para obtener que $|2A| = |2B| \leq \alpha^2 |A|$. Más en general, nos podemos preguntar cómo acotar el tamaño de sumas y diferencias iteradas de conjuntos dadas cotas sobre conjuntos suma. Una combinación de algunos resultados previos da respuesta a esta pregunta.

Teorema 13 (desigualdades de Plünnecke-Ruzsa [10]). Sean A y B dos conjuntos finitos en un grupo abeliano tales que $|A + B| \leq \alpha |A|$. Entonces, para $l, m \geq 1$ enteros cualesquiera,

$$|lB - mB| \leq \alpha^{l+m} |A|.$$

Demostración. El teorema 10 nos garantiza la existencia de un conjunto $T \subseteq A$ tal que $|T + kB| \leq \alpha^k |T|$, para todo $k \geq 1$. Ahora basta con aplicar la proposición 6 tomando $X = T$, $Y = -mB$ y $Z = -lB$. Entonces,

$$|T||lB - mB| \leq |T + lB||T + mB| \leq \alpha^l |T| \alpha^m |T| \leq \alpha^{l+m} |T||A|.$$

El resultado se obtiene dividiendo por $|T|$. \blacksquare

El nuevo método de Petridis, además de permitir demostrar los teoremas 10 y 13 de manera sencilla, tiene aplicación en otros muchos problemas. Por ejemplo, permite demostrar la siguiente desigualdad, análoga a la desigualdad triangular de Ruzsa pero que no se deduce de esta.

Proposición 14 ([7, lemma 1.7]). Sean A, B y C tres conjuntos no vacíos en un grupo abeliano. Entonces,

$$|A||B + C| \leq |A + B||A + C|.$$

Demostración. Sea X el conjunto que minimiza el cociente $\frac{|Z + B|}{|Z|}$ para todo $Z \subseteq A$. Por el lema 12, tenemos que

$$|C + B| \leq |C + X + B| \leq |C + X| \frac{|X + B|}{|X|} \leq |C + A| \frac{|A + B|}{|A|},$$

por definición de X y por las cotas triviales. El enunciado se obtiene multiplicando ambos lados de la desigualdad por $|A|$ y reordenando los términos (ya que el grupo ambiente es conmutativo). \blacksquare

Ejercicio 9. Mejorar las cotas del ejercicio 8 en vista de la proposición 14: si $|3A| \leq \alpha |A|$, entonces

$$|nA| \leq \alpha^{n-2} |A|, \quad |mA - A| \leq \alpha^m |A| \quad \text{y} \quad |mA - lA| \leq \alpha^{l+m-2} |A|$$

para todo $n \geq 3$ y $m, l \geq 2$. \blacktriangleleft

Además de esto, el lema 12 también tiene muchas aplicaciones para el caso en que se consideran conjuntos en grupos no abelianos. El estudio de este caso es más complejo que el del caso conmutativo, pero muchas de las técnicas y resultados expuestos aquí se pueden extender. Además, un lector cuidadoso podrá darse cuenta de que en la demostración del lema 12 no hemos utilizado en ningún momento la conmutatividad del grupo ambiente, lo cual quiere decir que el mismo resultado es cierto en grupos no conmutativos, y utilizando esto se pueden obtener otros resultados en este contexto.

Ejercicio 10. Encontrar las versiones no conmutativas de las proposiciones 6 y 14, así como las cotas del ejercicio 8 (en el caso no abeliano queremos cotas sobre $|\varepsilon_1 A + \varepsilon_2 A + \dots + \varepsilon_n A|$ para $n \geq 3$, con $\varepsilon_i \in \{-1, 1\}$ para todo i). ◀

Las desigualdades de Plünnecke, sin embargo, dejan de ser ciertas en el caso no conmutativo. El motivo de esto es que se pueden construir conjuntos en grupos no abelianos tales que su conjunto suma es «pequeño» (lineal en el tamaño del conjunto), pero cuyo conjunto triple suma es «grande» (cuadrático). Se puede encontrar un ejemplo detallado de este fenómeno en la monografía de Espuny Díaz [1, example 5.2].

De este modo, el estudio de las propiedades de los conjuntos suma se vuelve más complejo y queda fuera de los objetivos de este artículo. Se puede leer una discusión sobre los problemas del caso no conmutativo en el libro de Ruzsa [13]. Para sortear estos problemas, se han buscado otras condiciones que sí permitan establecer cotas del estilo de las desigualdades de Plünnecke. Se pueden encontrar resultados positivos en los artículos de Tao [14], Petridis [6] y Espuny Díaz [2].

6. Aplicación: un problema inverso

Para cerrar este artículo, vamos a demostrar un resultado inverso muy importante. Se trata de un caso particular del teorema 5 que se puede demostrar utilizando las desigualdades de Plünnecke-Ruzsa. Este caso particular no cubre todas las posibilidades que hay en los grupos abelianos, pero sí sirve para ver, de nuevo, el tipo de resultados que se buscan en combinatoria aditiva, y para ver lo útiles que resultan las desigualdades de Plünnecke. Su demostración se debe a Ruzsa [12], quien combinó todas las técnicas anteriores de manera elegante y sencilla.

Teorema 15 (Freiman-Ruzsa). *Sea G un grupo abeliano tal que todos sus elementos tienen orden acotado. Sea r una cota superior para el orden de todos los elementos. Sea $A \subseteq G$ un conjunto finito tal que $|A + A| \leq \alpha|A|$. Entonces, A está contenido en un subgrupo H de G tal que*

$$|H| \leq \alpha^2 r^{\alpha^4} |A|.$$

Demostración. Definimos un conjunto $X \subseteq 2A - A$ que sea maximal sujeto a la condición de que para cada $x \in X$, los conjuntos $x + A$ sean disjuntos; este conjunto existe porque $2A - A$ es finito. Eso quiere decir que podemos escribir

$$X + A = \bigcup_{x \in X} (x + A),$$

siendo la unión disjunta, de modo que $|X + A| = |X||A|$. Por otra parte, ya que $X \subseteq 2A - A$, entonces $X + A \subseteq 3A - A$. Aplicando el teorema 13, tenemos que

$$|X||A| = |X + A| \leq |3A - A| \leq \alpha^4 |A|,$$

de modo que $|X| \leq \alpha^4$.

Sea $t \in 2A - A$ un elemento cualquiera. Por la maximalidad de X , tenemos que $(t + A) \cap (X + A) \neq \emptyset$: de ser esta intersección vacía, habríamos encontrado un elemento t que podríamos añadir a X sin romper la condición, de modo que X no sería maximal y llegaríamos a una contradicción. Esto quiere decir que $t \in X + A - A$. Como podemos hacer lo mismo para cualquier elemento t , tenemos que $2A - A \subseteq X + A - A$. Ahora podemos demostrar por inducción que $jA - A \subseteq (j - 1)X + A - A$. En efecto, tenemos que

$$jA - A = (j - 1)A - A + A \subseteq (j - 2)X + A - A + A = (j - 2)X + 2A - A \subseteq (j - 2)X + X + A - A = (j - 1)X + A - A,$$

donde las inclusiones vienen dadas por la hipótesis de inducción y el caso base, respectivamente.

Sea H el subgrupo de G generado por A , y sea I el subgrupo generado por X . Dado que el orden de todos los elementos de G está acotado, tenemos que $H = \bigcup_{j \geq 2} (jA - A)$, ya que $H = \bigcup_{j \geq 1} (jA)$, y sumar y restar A no puede hacer que nos salgamos del subgrupo. También podemos escribir $I = \bigcup_{j \geq 1} (jX)$. Así, teniendo en cuenta el resultado que acabamos de demostrar por inducción,

$$H = \bigcup_{j \geq 2} (jA - A) \subseteq \bigcup_{j \geq 2} ((j-1)X + A - A) = \bigcup_{j \geq 1} (jX) + A - A = I + A - A.$$

Así, basta con acotar el tamaño de $I + A - A$ para obtener una cota para el tamaño del subgrupo H en el que A está contenido. Podemos acotar $|I|$ considerando que cualquier elemento g de I se tiene que poder escribir como $n_1 x_1 + n_2 x_2 + \dots + n_{|X|} x_{|X|}$, donde x_i son los elementos de X escritos en un orden arbitrario y $0 \leq n_i < r$. El tamaño de I es como mucho el número de estas expresiones, que es $r^{|X|}$. Finalmente, usamos la cota que hemos obtenido para el tamaño de X , una de las desigualdades triviales (ejercicio 2) y el corolario 8 (que también se puede obtener usando el teorema 13) para obtener la cota

$$|H| \leq |I + A - A| \leq |I||A - A| \leq \alpha^2 r^{\alpha^4} |A|. \quad \blacksquare$$

Nota. Se pueden encontrar soluciones a los ejercicios 4, 5 y 8 (y, por tanto, al ejercicio 7), así como a la primera parte del ejercicio 10, en la monografía de Espuny Díaz [1, proposition 1.2, proposition 1.1, corollary 4.7, theorem 5.18 y lemma 6.6, respectivamente], algunas de ellas escritas de manera más general que en este artículo pero pudiéndose obtener estas de manera inmediata. La solución del ejercicio 9 sigue las mismas líneas que la del ejercicio 8. Finalmente, la solución de la segunda parte del ejercicio 10 se puede encontrar en un artículo de Espuny Díaz [2], en la demostración del teorema 1.7; también sigue las líneas del ejercicio 8. ◀

Referencias

- [1] ESPUNY DÍAZ, Alberto. *Classical and modern approaches for Plünnecke-type inequalities*. Trabajo Final de Grado. Universitat Politècnica de Catalunya, 2015. URL: <http://hdl.handle.net/2117/77022>.
- [2] ESPUNY DÍAZ, Alberto. «Explicit bounds for growth of sets in non-abelian groups». En: *Reports@SCM* 3.1 (2017). Preprint, págs. 17-26. ISSN: 2385-4227. URL: <http://revistes.iec.cat/index.php/reports/article/view/142624>.
- [3] FREĪMAN, Gregory A. *Foundations of a structural theory of set addition*. Traducido del ruso, Translations of Mathematical Monographs, Vol 37. American Mathematical Society, Providence, R. I., 1973, págs. vii+108.
- [4] GREEN, Ben y RUZSA, Imre Z. «FreĪman's theorem in an arbitrary abelian group». En: *J. Lond. Math. Soc. (2)* 75.1 (2007), págs. 163-175. ISSN: 0024-6107. <https://doi.org/10.1112/jlms/jdl021>.
- [5] NATHANSON, Melvyn B. *Additive number theory*. Vol. 165. Graduate Texts in Mathematics. Inverse problems and the geometry of sumsets. Springer-Verlag, New York, 1996, págs. xiv+293. ISBN: 0-387-94655-1. <https://doi.org/10.1007/978-1-4757-3845-2>.
- [6] PETRIDIS, Giorgis. «New proofs of Plünnecke-type estimates for product sets in groups». En: *Combinatorica* 32.6 (2012), págs. 721-733. ISSN: 0209-9683. <https://doi.org/10.1007/s00493-012-2818-5>.
- [7] PETRIDIS, Giorgis. *Introduction to the Theory of Set Addition*. Notas para el Block Course Towards the Polynomial Freiman-Ruzsa Conjecture. Oct. de 2014. URL: https://mat-web.upc.edu/people/juan.jose.rue/BlockCourse-FUBerlin-Week1/Berlin_notes.pdf.
- [8] PLÜNNECKE, Helmut. «Eine zahlentheoretische Anwendung der Graphentheorie». En: *J. Reine Angew. Math.* 243 (1970), págs. 171-183. ISSN: 0075-4102.
- [9] RUZSA, Imre Z. «Addendum to: An application of graph theory to additive number theory». En: *Sci. Ser. A Math. Sci. (N.S.)* 4 (1990/1991), págs. 93-94.

-
- [10] RUZSA, Imre Z. «An application of graph theory to additive number theory». En: *Sci. Ser. A Math. Sci. (N.S.)* 3 (1989), págs. 97-109. ISSN: 0716-8446.
- [11] RUZSA, Imre Z. «Sums of finite sets». En: *Number theory (New York, 1991–1995)*. Springer, New York, 1996, págs. 281-293. https://doi.org/10.1007/978-1-4612-2418-1_21.
- [12] RUZSA, Imre Z. «An analog of Freiman's theorem in groups». En: *Astérisque* 258 (1999). Structure theory of set addition, págs. xv, 323-326. ISSN: 0303-1179.
- [13] RUZSA, Imre Z. «Sumsets and structure». En: *Combinatorial number theory and additive group theory*. Adv. Courses Math. CRM Barcelona. Birkhäuser Verlag, Basel, 2009, págs. 87-210. <https://doi.org/10.1007/978-3-7643-8962-8>.
- [14] TAO, Terence. «Product set estimates for non-commutative groups». En: *Combinatorica* 28.5 (2008), págs. 547-594. ISSN: 0209-9683. <https://doi.org/10.1007/s00493-008-2271-7>.
- [15] TAO, Terence y VU, Van H. *Additive combinatorics*. Vol. 105. Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, 2006, págs. xviii+512. ISBN: 978-0-521-85386-6; 0-521-85386-9. <https://doi.org/10.1017/CB09780511755149>.

TEMat

El efecto de Talbot: de la óptica a la ecuación de Schrödinger

✉ Daniel Eceizabarrena
BCAM - Basque Center for Applied
Mathematics
deceizabarrena@bcamath.org

Resumen: El objetivo de este artículo es dar a conocer un bello efecto óptico que se denomina efecto de Talbot. Primero, describiremos el fenómeno y comentaremos su descubrimiento a mediados del siglo XIX. A continuación, analizaremos las razones del fenómeno y lo justificaremos mediante algunos cálculos apoyados en las leyes físicas que gobiernan la propagación de ondas, así como por medios analíticos. Por último, veremos que este efecto se reproduce en algunas soluciones de la ecuación de Schrödinger, que muestran el mismo comportamiento que la luz.

Palabras clave: efecto de Talbot, óptica, ondas, Helmholtz, Schrödinger, sumas de Gauss.

MSC2010: 78A99.

Recibido: 1 de marzo de 2017.

Aceptado: 1 de mayo de 2017.

Agradecimientos: En primer lugar, quiero mostrar mi agradecimiento a Luis Vega por haberme presentado este tema y por su ayuda en la clarificación de varias cuestiones. Asimismo, me gustaría nombrar a Sandeep Kumar, con quien comencé a entender el proceso físico que nos ocupa; a Aitor Eceizabarrena, quien muy acertadamente me asistió en la traducción de fragmentos de artículos, y a Markel Epelde, cuyas sugerencias y comentarios me han sido de gran ayuda.

Este trabajo ha contado con la financiación del Ministerio de Educación, Cultura y Deporte para la Formación de Profesorado Universitario FPU15/03078.

Referencia: ECEIZABARRENA, Daniel. «El efecto de Talbot: de la óptica a la ecuación de Schrödinger». En: *TEMat*, 1 (2017), págs. 91-106. ISSN: 2530-9633. URL: <http://temat.anemat.com/articulo/2017-p91/>.

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

1. Introducción y orígenes

El efecto de Talbot es un fenómeno óptico microscópico que consiste en la repetición de un patrón que la luz forma cuando atraviesa una rejilla. La causa reside en la *difracción* que sufre la luz en su calidad de onda y que se genera cuando esta pasa cerca de un obstáculo o una apertura. La difracción es un efecto inherente a toda onda, aunque la apreciamos de manera más clara y visible cuando experimentamos con ondas de agua y de luz. En el caso del agua, si una ola se dirige hacia una pared con una pequeña apertura vertical, las olas resultantes que atraviesan la apertura se propagarán de manera circular en vez de recta. En el caso de la luz, la difracción hace que los diferentes colores se separen, generando efectos llamativos como cuando miramos al reflejo de un CD.

Consideremos que, en un plano bidimensional con ejes x y z , tenemos una pared en $z = 0$ en la que hemos realizado un agujero. Hagamos que una onda plana, cuyos frentes de onda forman líneas rectas, se dirija perpendicularmente hacia la pared. Cuando alcance la pared, el único punto por el que la onda pasará será el agujero, que se comportará como una fuente de una onda esférica (en este caso bidimensional, una onda circular) con la misma longitud de onda y velocidad que la onda original. Esta situación se muestra en la figura 1 y es el ejemplo más básico de difracción.

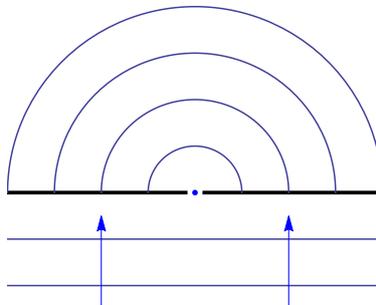


Figura 1: El caso más básico de la difracción de las ondas. Intuitivamente, podemos pensar en una ola dirigiéndose hacia un muro con una apertura, donde se generan nuevas olas circulares.

Si en lugar de hacer un solo agujero en la pared la perforamos de tal manera que haya agujeros separados de manera equidistante formando una rejilla, cada uno de los agujeros se comportará como una fuente de ondas esféricas. Además, en este caso deberemos tener en cuenta la interferencia entre ondas de diferentes fuentes. Esta es la situación en la que ocurre el efecto de Talbot, cuando las ondas que enviamos hacia la rejilla son de luz, como describimos a continuación. Imaginemos que la distancia entre los agujeros es muy pequeña (más adelante concretaremos el significado de *muy pequeño*) y que colocamos una lente frente a la rejilla, en el lado contrario a la fuente de luz, de tal manera que la lente capte la luz que sale de la rejilla y amplifique la imagen recibida para que la podamos ver con los ojos. Según movemos la lente adelante o atrás, la imagen recibida varía como consecuencia de la interferencia. Pero existe una distancia particular, conocida como **distancia de Talbot** y que denotaremos como z_T , desde la que la lente reproduce la propia rejilla con una precisión asombrosa. El observador en z_T es, por lo tanto, capaz de ver la rejilla. Este efecto, en el que la interferencia de las diferentes ondas de luz difractadas genera una copia de la rejilla, es lo que se conoce como **efecto de Talbot** (véase la figura 2).

El efecto de Talbot recibe su nombre del científico británico William Henry Fox Talbot (1800-1877), quien contaba entre sus intereses disciplinas tan diferentes como la fotografía, la arqueología o la política. Por supuesto, fue el tiempo que le dedicó a la primera el que generó el tema que nos ocupa. En la cuarta entrega de la serie de artículos denominados *Facts Relating to Optical Science* [10] que publicó en 1836, Talbot describió los resultados de dirigir la luz hacia dos prismas pegados entre sí y dio una pista de la razón por la que los fenómenos de difracción en los campos cercano y lejano reciben los nombres de difracción de Fresnel y de Fraunhofer respectivamente. Talbot escribió que fue Fresnel quien observó por primera vez luz difractada a través de un microscopio y Fraunhofer quien hizo lo propio a través del telescopio. Anteriormente, Newton y Grimaldi habían realizado sus observaciones, menos precisas, directamente sobre una hoja de papel que utilizaban como pantalla receptora de la luz. Pero en lo que nos concierne, lo más interesante que Talbot reflejó en su artículo es el siguiente experimento. En sus palabras,

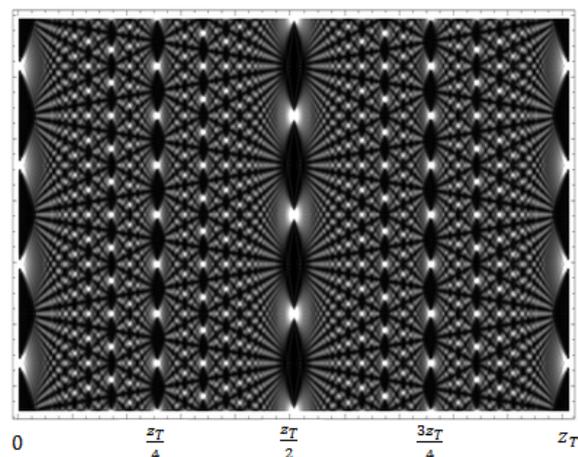


Figura 2: La visualización del efecto de Talbot, en lo que se denomina una **alfombra de Talbot**. La imagen muestra la luz que recibe un observador o una pantalla situada a diferentes distancias de la rejilla. El extremo izquierdo representa la rejilla mencionada, que recibe la luz desde el lado izquierdo. Vemos que en el extremo derecho esta rejilla se repite. En distancias intermedias, se forman diferentes patrones de luz. Hay una versión interactiva de la imagen en la dirección web de Quantum Interactive [8]. Versión modificada de una figura de Ben Goodman con licencia CC BY-SA 3.0 que se distribuye bajo la misma licencia que la original, disponible en https://commons.wikimedia.org/wiki/File:Optical_Talbot_Carpet.png.

«a unos tres o seis metros del punto radiante, coloqué en la trayectoria del rayo de luz una rejilla vertical equidistante fabricada por Fraunhofer». Talbot describe una rejilla equidistante como «una plancha de cristal cubierta con pan de oro con varios centenares de cortes paralelos, de tal manera que la luz sea transmitida en intervalos idénticos». Después, describe lo que observó al mirar hacia ella:

Entonces observé la luz que había pasado a través de la rejilla mediante una lente de considerable poder de aumento. La imagen era muy curiosa, formada por una alternancia regular de líneas o bandas rojas y verdes, en dirección paralela a la de los cortes de la rejilla. Alejando un poco la lente, las bandas cambiaban de color gradualmente, y se convertían alternativamente en azules y amarillas. Cuando alejaba la lente un poco más, las bandas recuperaban el color rojo y verde. Y este cambio se repetía un número indefinido de veces, a medida que la distancia entre la rejilla y la lente aumentaba. En todos los casos, las bandas presentaban dos colores complementarios.

Era muy curioso observar que aunque la lente estaba lo suficientemente lejos de la rejilla como para no enfocarla, la apariencia de las bandas era perfectamente distinguible y nítida.

Sin embargo, esto solo ocurre cuando el punto radiante tiene un diámetro aparente *muy pequeño*, en cuyo caso la lente puede ser alejada de la rejilla hasta unos 50 centímetros, sin afectar demasiado a la belleza y nitidez de las bandas de color. Por lo que si la fuente de luz fuera un mero punto matemático, parece posible que la distancia pudiera ser incrementada sin límite...

Las observaciones de Talbot, pese a ser meramente cualitativas, fueron la primera mención de lo que hoy llamamos el efecto de Talbot, descrito gráficamente en la figura 2.

Al parecer, los experimentos de Talbot pasaron desapercibidos durante varios años. No fue sino hasta 1881 cuando John William Strutt, tercer barón de Rayleigh (más conocido como Lord Rayleigh) trató de nuevo el fenómeno en su artículo «On Copying Diffraction-gratings, and on some Phenomena connected therewith» [6]. Tras reproducir el experimento de Talbot, llevó a cabo un exhaustivo análisis de los resultados, tratando de darle una explicación teórica. Sin embargo, se podría deducir de la lectura del artículo que la intención de Rayleigh no era esa. Más bien, como sugiere el propio nombre del artículo, pretendía en un principio obtener copias de rejillas de difracción originales. Aunque lo parezca, no es para nada

un objetivo trivial, dado el gran número de hendiduras en cada rejilla. A lo largo del artículo, Rayleigh asegura que «en verano de 1879 se me presentó la oportunidad de llevar mis hipótesis a la práctica gracias a la amabilidad del señor Rutherford, quien me obsequió con una maravillosa rejilla de cristal con unas 12 000 hendiduras, a 6800 por centímetro». Dicho sea de paso que esta frase nos muestra la magnitud microscópica del experimento: en efecto, la distancia entre las hendiduras mencionada anteriormente es *mu*y pequeña. Mientras Rayleigh, en el mismo artículo [6], se enfrenta a numerosos problemas técnicos para obtener una copia satisfactoria, escribe:

El verano pasado descubrí accidentalmente que Fox Talbot había hecho unas observaciones parecidas varios años atrás, y su detenido estudio me llevó a variar ligeramente el enfoque de mis procedimientos.

Inmediatamente después reproduce el experimento de Talbot, con la salvedad de hacer sus observaciones a través de algunos cristales de color rojo y verde frente a sus ojos. Esto le permite observar el experimento como si la luz emitida fuera monocromática roja o verde, y así realizar un análisis para una sola longitud de onda, la propia de cada color. Tras observar el mismo fenómeno que Talbot, registra las distancias, tanto para la luz roja como para la verde, en las que la rejilla se reproduce más y menos nítidamente. Después, deduce tras varios cálculos que, tal y como Talbot conjeturó, existe una distancia z_T en la que la rejilla es reproducida con altísima precisión. Incluso es capaz de formular una expresión explícita para ella,

$$(1) \quad z_T = \frac{\lambda}{1 - \sqrt{1 - \frac{\lambda^2}{d^2}}},$$

donde λ es la longitud de onda y d es la separación de las hendiduras de la rejilla. Cuando la longitud de onda es considerablemente menor que la distancia entre las hendiduras, es decir, cuando $\lambda \ll d$, es aceptable truncar la expansión en serie de Taylor de la raíz cuadrada en (1), $\sqrt{1 - x^2} = 1 - x^2/2 - x^4/8 + \dots$, a partir de la cuarta potencia. Esto nos permite simplificar la expresión (1), consiguiendo

$$(2) \quad z_T = \frac{2d^2}{\lambda}.$$

Gracias a estos cálculos, es posible copiar una rejilla colocando un material fotosensible a distancia z_T , logrando así el objetivo que Rayleigh perseguía. También, como curiosidad, la fórmula (2) permite deducir fácilmente la longitud de onda de diferentes luces monocromáticas midiendo la distancia de Talbot de manera experimental.

Los cálculos de Rayleigh son completamente analíticos, aunque hay momentos en los que se vale de algunas aproximaciones, y se basan en la expresión matemática explícita del comportamiento de las ondas planas cuando atraviesan una rejilla, que Rayleigh trata en su libro *The Theory of Sound* [7].

El artículo está organizado de la siguiente forma. En la sección 2 deduciremos la aproximación de la distancia de Talbot (2) por medio de argumentos físicos bastante intuitivos que permitirán hacernos una idea clara de lo que Rayleigh desarrolla. A continuación, en la sección 3 trataremos de plasmar las ideas principales del mencionado razonamiento de Rayleigh. Finalmente, en la sección 4 veremos que el efecto de Talbot no es un fenómeno exclusivo de la luz, sino que también se reproduce en ciertas soluciones de la ecuación de Schrödinger cuando tomamos como dato inicial un modelo de la rejilla mencionada.

2. Una primera aproximación a la distancia de Talbot

El objetivo de esta sección es deducir la existencia de la distancia de Talbot (en su versión aproximada (2)) mediante argumentos físicos, que son muy visuales y permiten comprender mejor las razones por las que el fenómeno ocurre. Seguiremos los argumentos presentados en el blog *Skulls in the Stars* [9].

El efecto de Talbot es, como mencionamos al comienzo, un fenómeno de difracción y lo explicaremos por medio de lo que se conoce como el **principio de Huygens**. Para ello, necesitamos definir ciertos conceptos. La **fase** de una onda en un punto es la fracción de un ciclo completo en el que la onda se

encuentra con respecto a un punto de referencia. Un **frente de onda** es el conjunto de puntos adyacentes que tienen la misma fase. Por ejemplo, los círculos y las líneas debajo de la pared en la figura 1 representan frentes de onda. Según el principio de Huygens, todo punto en un frente de onda de una onda particular se comporta como fuente de una onda esférica con las mismas propiedades que la onda original (con la misma longitud de onda y velocidad, entre otras). Esta ley lleva el nombre del físico holandés Christiaan Huygens (1629-1695), quien la describió por primera vez en su libro *Traité de la lumière* [5]. Este principio es especialmente útil cuando consideramos la situación de la rejilla descrita anteriormente. En efecto, si una onda plana se dirige perpendicularmente hacia una rejilla, cada una de las hendiduras se comportará cual fuente de una onda esférica, como se muestra en la figura 3.

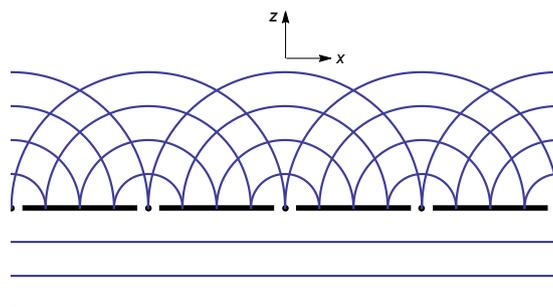


Figura 3: El principio de Huygens: una onda plana pasando a través de una rejilla genera ondas esféricas.

Por lo tanto, la superposición de cada una de estas ondas esféricas determinará la onda resultante del paso por la rejilla. En consecuencia, analizaremos la interacción entre las nuevas ondas, intentando detectar los nuevos frentes de onda que se generan. Antes de entrar en los razonamientos, enumeramos algunas definiciones y propiedades razonables que supondremos ciertas:

- La **longitud de onda** (es decir, la longitud de un ciclo completo de la onda) de la luz determina su color, y viceversa. Por lo tanto, como hizo Lord Rayleigh, trabajaremos con una luz monocromática, de tal manera que la longitud de onda sea constante. La denotaremos como λ .
- La fase de la onda viene dada por $2\pi y/\lambda$, donde y es la distancia recorrida por la onda respecto a algún punto de referencia. Se mide en radianes (módulo 2π). Definimos el **número de onda** como $k = 2\pi/\lambda$, por lo que la expresión para la fase es ky .
- La rejilla que Talbot describe en su artículo es «una plancha de cristal con varios centenares de hendiduras paralelas». Por lo tanto, podemos reducir el problema a un contexto bidimensional, en el plano cuyo vector normal viene dado por la dirección de las hendiduras. Así, la rejilla se convierte en una rejilla unidimensional, como la que se muestra en la figura 3, donde x es la dirección de la rejilla y z es la dirección perpendicular. Situaremos la rejilla en $z = 0$.
- Las hendiduras de la rejilla estarán separadas de manera equidistante y la distancia entre dos sucesivas aperturas será d . Además, a pesar de que la separación d de la rejilla de Rutherford era del orden de 10^{-6} metros, la longitud de onda de la luz visible es del orden de 10^{-7} metros, y la separación en la de Talbot sería bastante mayor. Por eso, podemos considerar $\lambda \ll d$.

Analícemos, pues, la interferencia entre las ondas esféricas generadas en cada una de las hendiduras. Como hemos dicho, queremos determinar frentes de onda que se generan tras el paso por la rejilla. Para una mejor comprensión, en la figura 4 mostramos algunos de estos frentes. En la imagen, la separación entre dos frentes consecutivos generados en una misma apertura es precisamente λ .

Para empezar, es fácil distinguir una línea recta paralela a la rejilla que es tangente a todos los frentes con radio $j\lambda$, para cada $j \in \mathbb{N}$ (en la figura 4a). Tomemos ahora un punto de hendidura $(S, 0)$, el frente de radio λ generado en la siguiente hendidura $(S + d, 0)$, el de radio 2λ creado en $(S + 2d, 0)$, y así sucesivamente. Es fácil ver mediante argumentos geométricos básicos que la recta que parte de $(S, 0)$ y que es tangente al frente de radio λ mencionado también es tangente a todos los demás frentes (figura 4b). En esa imagen, las líneas rojas son tangentes a todas las circunferencias consideradas en cada caso. De la misma manera,

si tomamos el frente con centro en $(S + d, 0)$ y radio 2λ , el centrado en $(S + 2d, 0)$ con radio 4λ y los que siguen, podemos dibujar una nueva recta tangente a cada uno de los frentes señalados, más inclinada que la anterior (figura 4c).

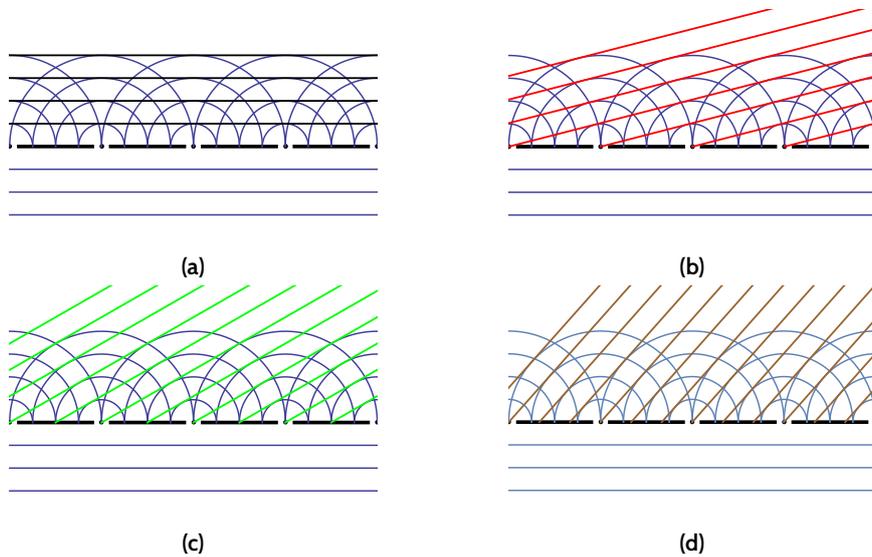


Figura 4: La creación de diferentes frentes de onda.

En cada imagen, cada línea define un nuevo frente de onda de una onda plana, que viaja en una dirección particular. Estas direcciones están determinadas por el ángulo θ entre el frente y la rejilla. Además, la superposición de todas estas ondas planas será la onda resultante del paso por la rejilla.

Si continuamos dibujando frentes de onda con este método, al ser los frentes circulares cada vez más grandes, llegará un momento en el que la apertura quedará dentro del círculo que consideremos, por lo que no existirá la tangente de la que estamos hablando. Por lo tanto, la cantidad de direcciones para los nuevos frentes es finita y, además, los ángulos correspondientes cumplirán $\theta \leq \pi/2$.

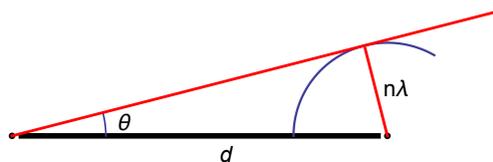


Figura 5: La determinación del ángulo θ correspondiente a un frente de onda.

En la figura 4 podemos distinguir fácilmente la situación que mostramos en la figura 5. Cada dirección de los frentes de onda corresponde a un valor $n \in \mathbb{N} \cup \{0\}$. Por ejemplo, la figura 4a corresponde a $n = 0$. Asimismo, las figuras 4b, 4c y 4d corresponden a los valores $n = 1$, $n = 2$ y $n = 3$, respectivamente. Además, el triángulo dibujado en la figura 5 es rectángulo para todo n , por lo que

$$\sin \theta = \frac{n\lambda}{d}, \quad n = 0, 1, 2, \dots$$

Como hemos dicho, θ no puede ser arbitrario. Un ángulo $\theta = \pi/2$ tendría un seno igual a 1, que correspondería al valor $n^* = d/\lambda$. Entonces, n debe ser menor que n^* , y como es entero, vemos que podemos escoger valores

$$(3) \quad n = 0, 1, 2, \dots, \left\lfloor \frac{d}{\lambda} \right\rfloor.$$

Estos valores representan los frentes que se desplazan hacia la izquierda o hacia delante. Pero también debemos tener en cuenta los frentes que se dirigen hacia la derecha, dado que de la misma manera que en la figura 4 hemos tomado centros $(S, 0), (S + d, 0), (S + 2d, 0), \dots$ para construir los diferentes frentes, podemos tomar centros en $(S, 0), (S - d, 0), (S - 2d, 0), \dots$. Por lo tanto, el número de direcciones para los frentes de onda que debemos considerar es $2 \lfloor d/\lambda \rfloor + 1$. Cada $n \in \{-\lfloor d/\lambda \rfloor, \dots, -1, 0, 1, \dots, \lfloor d/\lambda \rfloor\}$ corresponderá a una de estas direcciones.

Los realizados hasta ahora son cálculos preliminares que necesitamos para deducir la distancia de Talbot, que se mide en el eje z . Para ello, vamos a descomponer la onda en partes perpendiculares y paralelas a la rejilla, y nos centraremos en la fase de la onda en la dirección perpendicular z . Si encontramos una distancia en la que la fase de todos los frentes de onda sea la misma, entonces la luz recibida allí será idéntica a la emitida por la rejilla, dado que en la rejilla, la cual tomamos como referencia, todos los frentes de onda están en fase cero. Como consecuencia, se producirá el efecto de Talbot. ¿Pero qué significa la fase en la dirección z si hemos visto que los frentes de onda están generalmente inclinados? Fijemos una dirección en la figura 5 de tal manera que el frente de ondas no sea paralelo a la rejilla (es decir, un $n \neq 0$). Si fijamos un x en la rejilla, lo que tenemos en la dirección z es una onda que tendrá una longitud de onda, λ_z , diferente a la original. Del mismo modo, si fijamos z , una onda viaja en la dirección x con una longitud de onda λ_x , tal y como se muestra en la figura 6.

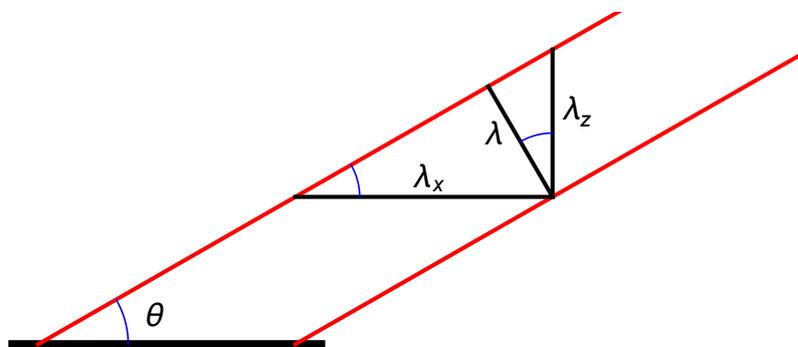


Figura 6: La descomposición de la longitud de onda λ correspondiente a una dirección particular de los frentes con $n \neq 0$.

La rejilla es equidistante con separación d . Por lo tanto, se deduce inmediatamente que para cada z fijo, la onda que viaja en la dirección x tendrá un periodo espacial d , dado que dos puntos separados por una distancia d están en la misma situación con respecto a la rejilla. Por otro lado, el periodo espacial de la onda es precisamente λ_x (que por definición es el mínimo periodo), por lo que necesariamente

$$(4) \quad m\lambda_x = d$$

para algún $m \in \mathbb{N}$. Además, tal y como observamos en la figura 4, $m = |n|$.

Por otro lado, también es inmediato ver que $\lambda = \lambda_x \sin \theta$ y $\lambda = \lambda_z \cos \theta$, dado que los ángulos marcados en la figura 6 son iguales a θ . De ahí deducimos que

$$(5) \quad \frac{1}{\lambda^2} = \frac{1}{\lambda_x^2} + \frac{1}{\lambda_z^2}.$$

Como decíamos, queremos trabajar con la fase en z , considerando como referencia la propia rejilla en $z = 0$. Calculemos el número de onda en esa dirección. Por (4) y (5) tenemos

$$k_z = \frac{2\pi}{\lambda_z} = 2\pi \sqrt{\frac{1}{\lambda^2} - \frac{1}{\lambda_x^2}} = 2\pi \sqrt{\frac{1}{\lambda^2} - \frac{n^2}{d^2}} = \frac{2\pi \sqrt{1 - n^2 \lambda^2 / d^2}}{\lambda},$$

y por lo tanto, la fase es

$$(6) \quad zk_z = \frac{2\pi z}{\lambda} \sqrt{1 - \frac{n^2 \lambda^2}{d^2}}.$$

Es importante observar que cuando $n = 0$, tenemos $\lambda = \lambda_z$ y $\lambda_x = 0$, por lo que $k_z = k$ y $zk_z = 2\pi z/\lambda$, que es precisamente lo que nos diría la ecuación (6). Por lo tanto, la fórmula (6) es válida para todo valor de n .

En este momento, necesitamos hacer algunas aproximaciones. En efecto, conocemos la serie de Taylor de la raíz cuadrada que se muestra en (6), dado que

$$\sqrt{1-x^2} = \sum_{j=0}^{\infty} \binom{1/2}{j} (-1)^j x^{2j} = 1 - \frac{x^2}{2} - \frac{x^4}{8} - \dots, \quad \text{cuando } x < 1.$$

Para valores de n pequeños, $n\lambda/d \ll 1$, dado que $\lambda \ll d$, y por ello, tomando los dos primeros términos de la serie, podemos aproximar la fase por medio de

$$\frac{2\pi z}{\lambda} \left(1 - \frac{n^2 \lambda^2}{2d^2} \right) = \frac{2\pi z}{\lambda} - 2\pi \frac{z\lambda n^2}{2d^2}.$$

En principio, para valores de n grandes, tenemos $n\lambda/d \sim 1$ y la aproximación realizada es demasiado imprecisa. Pero volviendo a la figura 4, los valores grandes de n corresponden a frentes muy inclinados que, en comparación con los más horizontales, no contribuirán mucho (dado que en una rejilla de longitud finita los frentes más inclinados tienen poco recorrido en la dirección z , que es la que nos interesa) y los podemos despreciar. Por ello, podemos dar por válida la aproximación anterior. Nótese que solo el segundo término depende de n . Además, si consideramos el punto z para el que

$$(7) \quad \frac{z\lambda}{2d^2} = 1,$$

entonces la dependencia en n desaparece porque logramos un múltiplo de 2π . En consecuencia, si se satisface (7), todos los frentes de onda tienen la misma fase con respecto a $z = 0$. Como hemos dicho, en $z = 0$ y en tiempo $t = 0$ todos los frentes están en fase 0. La fase, por supuesto, cambia con el tiempo, pero siempre será la misma para todos los frentes. Por lo tanto, lo que vemos es que esta situación se reproduce cuando (7) se cumple. En otras palabras, una pantalla situada en el punto z determinado por (7) recibe el mismo patrón de luz que el que se ha emitido desde $z = 0$. O si miráramos desde esa distancia hacia la rejilla, veríamos precisamente las hendiduras de esta.

Así pues, la distancia dada por la condición (7), que es

$$z = z_T = \frac{2d^2}{\lambda},$$

es la distancia de Talbot, distancia en la que el fenómeno descrito al principio ocurre. Coincide, precisamente, con la expresión aproximada (2) dada por Lord Rayleigh.

Es interesante observar que la elección de cualquier múltiplo de la distancia de Talbot lleva a la misma conclusión. Por lo tanto, la rejilla se reproduce una y otra vez cada vez que alejamos el punto de observación z_T unidades de espacio. Se podría decir que la luz se comporta como una copia de la rejilla original en cada punto $z = l z_T$ con $l \in \mathbb{N}$.

3. El razonamiento de Lord Rayleigh

En esta sección desarrollaremos unos cálculos con un mayor nivel de corrección. Nos basaremos en la argumentación de Lord Rayleigh en su artículo [6] y en su libro *The Theory of Sound* [7].

Nuestro punto de partida es la conocida **ecuación de ondas**, que podemos expresar como

$$(8) \quad \frac{\partial^2 u}{\partial t^2} = c^2 \Delta u, \quad \mathbf{x} \in \mathbb{R}^n, \quad t > 0,$$

donde $u = u(\mathbf{x}, t)$ representa el desplazamiento de la onda en el punto \mathbf{x} y en el momento t respecto a un punto de referencia, c representa la velocidad con la que la onda viaja y Δ es el operador laplaciano.

Usualmente se añade al problema una condición inicial correspondiente al tiempo $t = 0$, y viene dada por una función $u(x, 0) = u_0(x)$ para $x \in \mathbb{R}^n$. Tiene sentido trabajar con esta ecuación, dado que estamos tratando un fenómeno de la luz, que no deja de ser una onda.

Como venimos haciendo, trabajaremos con dos dimensiones espaciales que llamaremos (x, z) y consideramos la rejilla equidistante en $z = 0$. El dominio a analizar es, pues, el semiplano superior $z > 0$. Lo primero que queremos ver es que no será necesario estudiar la situación en todo el semiplano, sino que nos podremos reducir a un dominio más pequeño. Observamos primero, como ya hemos mencionado anteriormente en la sección 2, que dado que el origen de las ondas es la rejilla equidistante con separación d , la solución que buscamos debe ser periódica en x con periodo d . Esto se cumple dado que la ecuación de ondas (8) es invariante por traslaciones; es decir, si $u(x, t)$ es solución, $u(x + d, t)$ también lo será. Tienen dato inicial $u(x, 0)$ y $u(x + d, 0)$ respectivamente. Nuestra rejilla juega el papel de dato inicial, por lo que al ser equidistante, podemos suponer que el dato inicial será periódico de periodo d . En consecuencia, los datos mencionados son idénticos. La unicidad de la solución para un dato inicial dado muestra entonces que $u(x, t) = u(x + d, t)$ y por lo tanto la periodicidad de nuestra solución. Gracias a ello, bastará con analizar la solución en la franja $0 \leq x \leq d$.

Pero podemos decir más aún. Por ejemplo, si trazamos una línea vertical en una apertura de la rejilla, observamos en la figura 7a que la velocidad transversal de la onda es nula: las componentes horizontales de la velocidad se anulan entre sí, resultando una velocidad vertical. Lo mismo ocurre en los puntos medios entre las aperturas (figura 7b). De estas propiedades podemos deducir dos hechos:

- La velocidad transversal (es decir, en la dirección x) en las líneas $x = 0$, $x = d/2$ y $x = d$ es nula, por lo que

$$(9) \quad \frac{\partial u}{\partial x}(0, z, t) = \frac{\partial u}{\partial x}(d/2, z, t) = \frac{\partial u}{\partial x}(d, z, t) = 0, \quad \forall z > 0, \forall t > 0.$$

- La solución es par con respecto a la «pared» marcada en $x = d/2$. Y es que si tomamos un $x_0 > 0$, la situación del punto $(d/2 - x_0, z)$ respecto a $(0, 0)$ y a $(d/2, 0)$ es exactamente la misma que la situación de $(d/2 + x_0, z)$ respecto a $(d, 0)$ y $(d/2, 0)$ respectivamente. Se da, por tanto, una simetría respecto a la línea $x = d/2$.

Gracias a estos argumentos, bastará con resolver el problema para la franja $0 \leq x \leq d/2$. De esta solución parcial podremos recuperar la solución completa extendiéndola primero de manera par a $0 \leq x \leq d$ y después de manera periódica a toda la recta. Esto se puede comprobar analíticamente gracias a las condiciones sobre $\partial u / \partial x$ (9) y a la invarianza de la ecuación ondas (8) respecto a traslaciones y reflexiones pares.

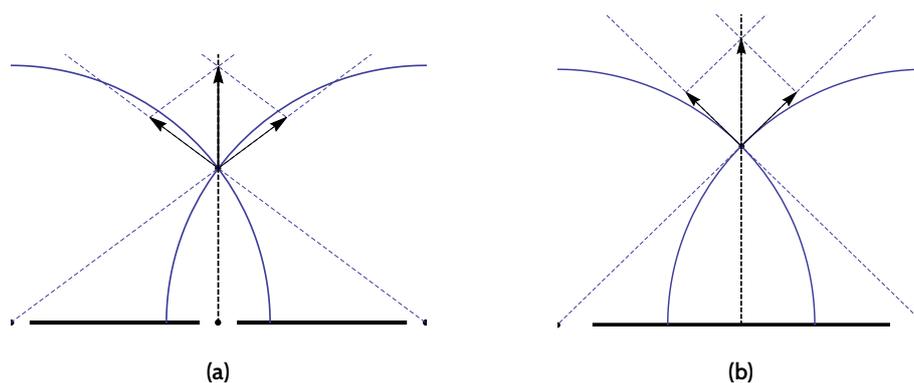


Figura 7: Las «paredes» del experimento de Talbot.

Por lo tanto, hemos conseguido reducirnos al problema

$$(10) \quad \begin{cases} \frac{\partial^2 u}{\partial t^2} = c^2 \Delta u, & 0 \leq x \leq d/2, \quad z > 0, \quad t > 0, \\ \frac{\partial u}{\partial x}(0, z, t) = \frac{\partial u}{\partial x}(d/2, z, t) = 0, & z > 0, \quad t > 0. \end{cases}$$

Lo resolveremos mediante el método de separación de variables. Si escribimos $u(x, z, t) = v(x, z)T(t)$, logramos

$$(11) \quad v(x, z)T''(t) = c^2 \Delta v(x, z)T(t) \iff \frac{T''(t)}{c^2 T(t)} = \frac{\Delta v(x, z)}{v(x, z)} = -\mu,$$

donde $\mu \in \mathbb{R}$ es una constante, dado que una igualdad entre funciones en variables distintas implica que ambas son constantes. Para la variable temporal tenemos una ecuación de segundo grado que depende del signo de μ . En efecto, si $\mu < 0$, las soluciones son exponenciales, y si $\mu = 0$ son lineales. No nos interesan, dado que una onda no se comporta de esa manera. En cambio, si $\mu > 0$, entonces la solución oscila. Más concretamente, tenemos

$$T(t) = Ae^{ic\sqrt{\mu}t} + Be^{-ic\sqrt{\mu}t}$$

para ciertas constantes A y B . Si llamamos $\mu = k^2$ (veremos más adelante que k será precisamente el número de onda), entonces podemos tratar la solución como

$$(12) \quad T(t) = Ae^{ickt} + Be^{-ickt} := e^{\pm ickt}, \quad \text{o} \quad T(t) = A \cos(ckt) + B \sin(ckt).$$

Para las variables espaciales, en las ecuaciones (11) vemos que

$$(13) \quad \Delta u(x, z) + k^2 u(x, z) = 0.$$

A la ecuación (13) se la llama **ecuación de Helmholtz**, y es una versión estacionaria de la ecuación de ondas. Volvamos a separar variables escribiendo $v(x, z) = X(x)Z(z)$, de tal manera que después de dividirla por u , la ecuación (13) se convierte en

$$(14) \quad -\frac{X''(x)}{X(x)} = \frac{Z''(z)}{Z(z)} + k^2 = \eta,$$

donde de nuevo η es una constante, porque volvemos a tener una igualdad entre funciones de variables distintas. Las condiciones sobre la $\partial u / \partial x$ se convierten en $X'(0) = X'(d/2) = 0$.

Observamos primero que la ecuación para X es la misma que para T (salvo sustitución de $c^2\mu$ por η), por lo que también depende del signo de η . Si $\eta < 0$, las soluciones son exponenciales, y es fácil ver que las condiciones de contorno mencionadas fuerzan a que la solución sea nula, solución que no nos interesa. Si $\eta = 0$, entonces la solución es de la forma $X(x) = Ax + B$ para ciertas constantes A y B . En este caso, las condiciones de contorno fuerzan que $A = 0$, por lo que la solución es constante. En cambio, si $\eta > 0$, siguiendo los mismos pasos que para lograr las expresiones (12) y llamando $\eta = j^2$, tenemos $X(x) = A \cos(jx) + B \sin(jx)$ con A y B constantes. La condición $X'(0) = 0$ implica que $B = 0$. Como consecuencia, $0 = X'(d/2) = -jA \sin(jd/2)$, por lo que se debe cumplir que $jd/2 = p\pi$ con $p \in \mathbb{Z}$. El caso $\eta = 0$ se incluye en $p = 0$. Así pues, tenemos soluciones distintas para cada valor de $p \in \mathbb{Z}$, que denotamos por

$$(15) \quad X_p(x) = A_p \cos\left(\frac{2\pi px}{d}\right),$$

donde A_p es una constante para cada valor de p .

Centrémonos ahora en la ecuación (14) para Z , que vuelve a ser igual a la de T una vez sustituida la constante $c^2\mu$ por $k^2 - \eta = k^2 - j^2$. Así pues, las soluciones que nos interesan son aquellas con $k^2 - j^2 \geq 0$. Utilizando la caracterización de j en función de p , definimos

$$(16) \quad v_p^2 = k^2 - \frac{4\pi^2 p^2}{d^2} \geq 0.$$

Con esta notación, al igual que en la expresión (12), tenemos $Z_p(z) = e^{\pm iv_p z}$. Si juntamos esto con las expresiones (12) y (15) que hemos obtenido para X y T , logramos un conjunto de soluciones

$$(17) \quad u_p(x, z, t) = B_p e^{\pm iv_p z} e^{\pm ickt} \cos\left(\frac{2\pi px}{d}\right)$$

para los valores de p determinados por la condición (16), donde B_p es constante para cada p .

Una primera simplificación consiste en darse cuenta de que v_p depende de p^2 y de que el coseno es una función par. Por ello, $u_p = u_{-p}$, y nos podemos limitar a $p \in \mathbb{N} \cup \{0\}$.

Los signos \pm de la expresión (17) se pueden determinar. Y es que si tomáramos la exponencial con $ckt + v_p z$, entonces será constante en rectas de la forma $ckt + v_p z = c$. Aquí, cuando t crece, z decrece, así que esta expresión determina una onda que viaja hacia atrás, al igual que la expresión con $-ckt - v_p z$. Pero nuestra onda, que viaja en z , va hacia adelante dado que estamos considerando una pared en $z = 0$. Por ello, nos quedamos con las formas $\pm(ckt - v_p z)$.

Al tener el conjunto de soluciones u_p para $p = 0, 1, 2, \dots$, la linealidad de la ecuación nos permite trabajar con la suma de todas ellas como solución. Así, escribiéndola en su forma real, logramos

$$(18) \quad u(x, z, t) = \sum_{p \leq kd/2\pi} (C_p \cos(ckt - v_p z) + D_p \sin(ckt - v_p z)) \cos\left(\frac{2\pi p x}{d}\right),$$

donde la suma recorre los números enteros no negativos y C_p y D_p son constantes para cada p .

La expresión (18) deja al descubierto las razones de nombrar a k así. Se ve que ck es la frecuencia de la onda. Dicho de otro modo, si T es el periodo, entonces $k = 2\pi/(cT)$. Por definición, $cT = \lambda$ es la longitud de onda, por lo que k es el número de onda. Adecuando con esta notación la cota de la suma, logramos

$$(19) \quad u(x, z, t) = \sum_{p \leq d/\lambda} (C_p \cos(ckt - v_p z) + D_p \sin(ckt - v_p z)) \cos\left(\frac{2\pi p x}{d}\right).$$

Mientras que la longitud de onda determina el color de la luz, es la amplitud la que marca su intensidad. Por lo tanto, estamos interesados en calcular la amplitud de la onda, que será el máximo de nuestra función (19). Fijando un punto espacial (x, z) , tratémosla como función del tiempo $u(t)$. Un simple vistazo a su forma nos dice que no será fácil maximizarla igualando su derivada a cero. Pero de la misma manera que en la expresión (11), se ve que $u''(t) + c^2 k^2 u(t) = 0$. Si multiplicamos la ecuación por $u'(t)$, formando derivadas logramos $[(u')^2]' + c^2 k^2 (u^2)' = 0$, por lo que

$$(20) \quad [u'(t)]^2 + c^2 k^2 u(t)^2 = C$$

para alguna constante C y para todo $t \in \mathbb{R}$. Si tomamos un punto t_0 que maximice la función, es decir, con $u'(t_0) = 0$, entonces en (20) vemos que el valor de la función es $u(t_0)^2 = C/c^2 k^2$, que será precisamente el cuadrado del máximo. Pero C viene determinada en la propia igualdad (20), por lo que llamando A a la amplitud, logramos

$$(21) \quad A^2 = u(t_0)^2 = u(t)^2 + \frac{1}{c^2 k^2} [u'(t)]^2.$$

Tras varios cálculos utilizando la fórmula (21), la expresión (19) y alguna identidad trigonométrica básica para lograr la amplitud de nuestra onda, veremos que

$$A^2 = C_0^2 + D_0^2 + 2[(C_0 C_1 + D_0 D_1) \cos(kz - v_1 z) + (C_0 D_1 - C_1 D_0) \sin(kz - v_1 z)] \cos\left(\frac{2\pi x}{d}\right) + (C_1^2 + D_1^2) \cos^2\left(\frac{2\pi x}{d}\right) + \dots,$$

donde solo hemos desarrollado los términos que conciernen exclusivamente a los índices $p = 0$ y $p = 1$. En este momento, Lord Rayleigh considera que es suficiente quedarse con estos términos como aproximación. Dicha aproximación resulta ser periódica en z con periodo Z determinado por $(k - v_1)Z = 2\pi$. Resolviendo para Z logramos

$$Z = \frac{2\pi}{k - v_1} = \frac{2\pi}{\frac{2\pi}{\lambda} - \sqrt{\left(\frac{2\pi}{\lambda}\right)^2 - \frac{4\pi^2}{d^2}}} = \frac{\lambda}{1 - \sqrt{1 - \frac{\lambda^2}{d^2}}},$$

que es precisamente la expresión (1) que dimos al principio.

Los cálculos realizados muestran la existencia de la distancia de Talbot, y además nos ofrecen una expresión explícita para ella. Sin embargo, si uno mira la alfombra de Talbot de la figura 2, se dará cuenta de que el fenómeno es más complejo de lo que hemos deducido. Obsérvese que la distancia de Talbot, z_T , es precisamente la anchura de la imagen, y es en el extremo derecho donde la rejilla original de la parte izquierda se repite. Pero en puntos intermedios, hay muchos patrones fácilmente distinguibles. Por ejemplo, en $z = z_T/2$, la rejilla se repite salvo por el hecho de que está trasladada por una distancia de $d/2$. En $z = z_T/4$, la rejilla se ha duplicado y ahora tiene una separación de $d/2$. Y en un tercio de la distancia de Talbot, la rejilla se ha triplicado, pasando a tener un tercio de la separación original.

Esta imagen muestra que el efecto de Talbot no solo consiste en la repetición de la rejilla en la distancia z_T , sino que genera infinidad de imágenes autosemejantes de la rejilla original, en diferentes distancias y con diferentes separaciones. Estas observaciones nos podrían hacer pensar en fractales. De hecho, en 1996 (más de 100 años después de los avances de Lord Rayleigh y más de 150 años después que Talbot), M.V. Berry y S. Klein mostraron en su artículo «Integer, fractional and fractal Talbot effects» [2] que la alfombra de Talbot posee una estructura fractal.

4. El efecto de Talbot en la ecuación de Schrödinger

En las secciones anteriores hemos deducido el efecto de Talbot mediante distintos métodos en el caso de un rayo de luz que atraviesa una rejilla. En lo que sigue veremos que la ecuación de Schrödinger, que modeliza el comportamiento de un sistema cuántico que presenta un comportamiento ondulatorio, también reproduce el efecto de Talbot, como muestran F. de la Hoz y L. Vega en su artículo [4].

El problema del valor inicial para la **ecuación libre de Schrödinger** es

$$(22) \quad \begin{cases} u_t(\mathbf{x}, t) = i\Delta u(\mathbf{x}, t), & \text{para } (\mathbf{x}, t) \in \mathbb{R}^n \times \mathbb{R}, \\ u(\mathbf{x}, 0) = u_0(\mathbf{x}), & \text{para } \mathbf{x} \in \mathbb{R}^n, \end{cases}$$

donde u_t representa la derivada en tiempo de la función u . Aquí, i es la unidad imaginaria, Δ el operador laplaciano y la función u representa la función de onda del sistema cuántico en cuestión.

En esta sección trabajaremos en el caso unidimensional $n = 1$ y con la conocida **invarianza galileana**. Si la denotamos como G , esta invarianza viene dada por

$$(23) \quad Gu(x, t) = e^{i\lambda x - i\lambda^2 t} u(x - 2\lambda t, t), \quad \forall \lambda \in \mathbb{R}.$$

La invarianza galileana es una simetría de la ecuación de Schrödinger; es decir, si u es una solución de la ecuación, entonces Gu también lo es. Esto se puede comprobar fácilmente a mano. En particular, evaluando la expresión (23) en $t = 0$, la invarianza galileana para el dato inicial es

$$(24) \quad Gu_0(x) = e^{i\lambda x} u_0(x).$$

Una interesante utilidad de las simetrías consiste en considerar datos iniciales invariantes. En efecto, si consideramos un dato u_0 para el que $Gu_0 = u_0$, entonces dado que u y Gu resuelven el sistema (22), la unicidad de soluciones de la ecuación implica que $Gu = u$, obteniendo una ecuación para la solución. En el caso de la invarianza galileana, uno de los casos más sencillos es considerar la distribución delta de Dirac como dato inicial, que genera la solución fundamental de la ecuación de Schrödinger, la cual denotaremos como $U(x, t)$ (la delta de Dirac se puede representar intuitivamente como una *función* que vale ∞ en $x = 0$ y 0 en todo otro punto y que cumple $\int_{\mathbb{R}} \delta = 1$. Se puede encontrar información accesible sobre la teoría de distribuciones y soluciones fundamentales en los capítulos 2 y 3 del libro de Vladimirov [11] y también en el capítulo 6 del libro de Cerdà [3]). Así, como $G\delta = \delta$, las soluciones generadas por ambos datos son iguales y, por lo tanto,

$$GU(x, t) = U(x, t), \quad (x, t) \in \mathbb{R} \times \mathbb{R}.$$

En otras palabras,

$$U(x, t) = e^{i\lambda x - i\lambda^2 t} U(x - 2\lambda t, t), \quad \forall \lambda \in \mathbb{R}.$$

Recordemos que en el experimento de Talbot tenemos una rejilla periódica con hendiduras separadas por una distancia d . Si pensáramos en esta rejilla como el dato inicial de la ecuación de Schrödinger, podríamos identificar cada una de las aperturas con una delta de Dirac, de tal manera que tendríamos

$$(25) \quad u_0(x) = \sum_{k \in \mathbb{Z}} \delta(x - kd).$$

Este dato representa deltas separadas por una distancia d y, por lo tanto, es periódico de periodo d . Además, vemos que para todo $m \in \mathbb{Z}$,

$$e^{2\pi i m \frac{x}{d}} u_0(x) = u_0(x),$$

lo cual indica que (25) no varía mediante una cierta invarianza galileana. Esta igualdad se cumple dado que en puntos $x \neq ld$ con $l \in \mathbb{Z}$ la función es cero, y en puntos $x = ld$ el valor de la exponencial es 1. Por lo tanto, si u es la solución correspondiente, por (23) y (24) con $\lambda = 2\pi m/d$ tenemos que

$$(26) \quad u(x, t) = e^{2\pi i \frac{m}{d} x - 4\pi^2 i \frac{m^2}{d^2} t} u(x - 4\pi m t/d, t)$$

para todo $m \in \mathbb{Z}$.

Asimismo, es importante observar que, al igual que vimos en la sección 3, gracias a que la ecuación de Schrödinger es invariante por translaciones y a la unicidad de soluciones, el hecho de que el dato inicial sea periódico implica que la solución también lo será en su variable espacial x . Por lo tanto, gracias a esta periodicidad, podemos trabajar con series de Fourier. Consideremos el j -ésimo coeficiente de u , que es

$$(27) \quad \widehat{u}(j, t) = \frac{1}{d} \int_0^d u(x, t) e^{-2\pi i \frac{j}{d} x} dx.$$

Utilizando (26), cambiando de variables $x \rightarrow x + 4\pi m t/d$ y por la periodicidad, podemos decir que

$$\begin{aligned} \widehat{u}(j, t) &= \frac{1}{d} \int_0^d e^{2\pi i \frac{m}{d} x - 4\pi^2 i \frac{m^2}{d^2} t} u(x - 4\pi m t/d, t) e^{-2\pi i \frac{j}{d} x} dx \\ &= \frac{1}{d} \int_0^d e^{2\pi i \frac{m-j}{d} (x+4\pi m t/d) - 4\pi^2 i \frac{m^2}{d^2} t} u(x, t) dx \\ &= \frac{1}{d} e^{8\pi^2 i m \frac{m-j}{d^2} t - 4\pi^2 i \frac{m^2}{d^2} t} \int_0^d e^{2\pi i \frac{m-j}{d} x} u(x, t) dx \\ &= e^{8\pi^2 i m \frac{m-j}{d^2} t - 4\pi^2 i \frac{m^2}{d^2} t} \widehat{u}(j - m, t). \end{aligned}$$

Al ser esta ecuación válida para cualesquiera $j, m \in \mathbb{Z}$, tomemos $j = m$ para obtener

$$\widehat{u}(j, t) = e^{-4\pi^2 i \frac{j^2}{d^2} t} \widehat{u}(0, t)$$

y así poder escribir la serie de Fourier de u como

$$(28) \quad u(x, t) = \sum_{k \in \mathbb{Z}} \widehat{u}(k, t) e^{2\pi i \frac{k}{d} x} = \widehat{u}(0, t) \sum_{k \in \mathbb{Z}} e^{2\pi i (\frac{k}{d} x - 2\pi \frac{k^2}{d^2} t)}.$$

Esta es la expresión que nos permitirá detectar el efecto de Talbot. Primero observamos que la última suma en (28) cumple la ecuación del problema (22), lo cual implica que $\widehat{u}(0, t)$ debe ser constante como función de t . En particular, $\widehat{u}(0, t) = \widehat{u}(0, 0)$. Es bien sabido, y fácil de ver en (27), que el coeficiente 0 de Fourier de la solución u no es más que la integral de su dato inicial multiplicada por el inverso de su periodo, por lo que

$$\widehat{u}(0, t) = \widehat{u}(0, 0) = \frac{1}{d} \int_0^d u(x, 0) dx = \frac{1}{d}.$$

Esto se cumple porque en un periodo integramos una sola delta de Dirac, cuya integral es 1. Por otro lado, podemos deducir una identidad importante si miramos el dato inicial en (28). Y es que evaluando $t = 0$ obtenemos

$$(29) \quad \sum_{k \in \mathbb{Z}} \delta(x - kd) = u(x, 0) = \frac{1}{d} \sum_{k \in \mathbb{Z}} e^{2\pi i \frac{k}{d} x}.$$

Haremos uso de esta identidad más adelante.

La propia expresión (28) muestra que u es periódica en t de periodo $d^2/2\pi$, y por lo tanto volvemos a ver el dato inicial. Es precisamente el efecto de Talbot que hemos descrito en las secciones anteriores. Pero en este caso podemos obtener aún más información. Evaluemos la solución (28) en los tiempos

$$t_{p,q} = \frac{d^2}{2\pi} \frac{p}{q}, \quad \text{con } p \in \mathbb{Z}, \quad q \in \mathbb{N}, \quad (p, q) = 1,$$

que son múltiplos racionales del periodo $d^2/2\pi$. Entonces,

$$u(x, t_{p,q}) = \frac{1}{d} \sum_{k \in \mathbb{Z}} e^{2\pi i \left(\frac{k}{d} x - 2\pi \frac{k^2}{d^2} \frac{d^2}{2\pi} \frac{p}{q} \right)} = \frac{1}{d} \sum_{k \in \mathbb{Z}} e^{2\pi i \left(\frac{k}{d} x - k^2 \frac{p}{q} \right)}.$$

Fijemos el número q . Sabemos que todo $k \in \mathbb{Z}$ se puede escribir como $k = \kappa q + m$ donde $\kappa \in \mathbb{Z}, 0 \leq m < q$. Por lo tanto, reescribimos la suma anterior renombrando a κ de nuevo como k para obtener

$$\begin{aligned} u(x, t_{p,q}) &= \frac{1}{d} \sum_{k \in \mathbb{Z}} \sum_{m=0}^{q-1} e^{2\pi i \left(\frac{kq+m}{d} x - (kq+m)^2 \frac{p}{q} \right)} = \frac{1}{d} \sum_{k \in \mathbb{Z}} \sum_{m=0}^{q-1} e^{2\pi i \left(\frac{kq+m}{d} x - 2\pi i m^2 \frac{p}{q} \right)} \\ &= \frac{1}{d} \sum_{m=0}^{q-1} e^{2\pi i \frac{m}{d} x - 2\pi i m^2 \frac{p}{q}} \sum_{k \in \mathbb{Z}} e^{2\pi i q \frac{k}{d} x}. \end{aligned}$$

Utilicemos (29) para escribir

$$u(x, t_{p,q}) = \frac{1}{q} \sum_{m=0}^{q-1} e^{2\pi i \frac{m}{d} x - 2\pi i m^2 \frac{p}{q}} \sum_{k \in \mathbb{Z}} \delta(x - kd/q).$$

De la misma manera que antes, separamos la suma en los enteros (la suma en k) módulo q para escribir

$$u(x, t_{p,q}) = \frac{1}{q} \sum_{m=0}^{q-1} \sum_{k \in \mathbb{Z}} \sum_{r=0}^{q-1} e^{2\pi i \frac{m}{d} x - 2\pi i m^2 \frac{p}{q}} \delta \left(x - d \frac{kq+r}{q} \right).$$

Vemos que la suma de las deltas determina la forma de los puntos x en los que tendremos una contribución no nula. En otras palabras, si $x \neq d \frac{lq+s}{q}$ para l y s enteros, la suma de deltas será nula y el valor de x en la exponencial no tiene ningún efecto en este caso. Por otro lado, si $x = d \frac{lq+s}{q}$ con l y s enteros, solo sobrevive la delta que corresponde a l y a s , y por lo tanto podemos sustituir x por $d \frac{lq+s}{q}$ en la exponencial. Así, obtenemos

$$\begin{aligned} u(x, t_{p,q}) &= \frac{1}{q} \sum_{m=0}^{q-1} \sum_{k \in \mathbb{Z}} \sum_{r=0}^{q-1} e^{2\pi i \frac{m}{d} d \frac{kq+r}{q} - 2\pi i m^2 \frac{p}{q}} \delta \left(x - d \frac{kq+r}{q} \right) \\ &= \frac{1}{q} \sum_{m=0}^{q-1} \sum_{k \in \mathbb{Z}} \sum_{r=0}^{q-1} e^{2\pi i m \frac{r}{q} - 2\pi i m^2 \frac{p}{q}} \delta \left(x - d \frac{kq+r}{q} \right) \\ &= \frac{1}{q} \sum_{m=0}^{q-1} \sum_{k \in \mathbb{Z}} \sum_{r=0}^{q-1} e^{2\pi i \frac{r(m-pm^2)}{q}} \delta \left(x - d \frac{kq+r}{q} \right). \end{aligned}$$

Observamos que hemos generado lo que se conoce como una **suma cuadrática generalizada de Gauss**, que denotamos por

$$(30) \quad G(a, b, c) = \sum_{m=0}^{c-1} e^{2\pi i \frac{am^2+bm}{c}}$$

para enteros a, b, c . Con esta notación, obtenemos

$$(31) \quad u(x, t_{p,q}) = \frac{1}{q} \sum_{k \in \mathbb{Z}} \sum_{r=0}^{q-1} G(-p, r, q) \delta \left(x - dk - d \frac{r}{q} \right).$$

La expresión (31) es precisamente a lo que llamamos **efecto de Talbot**. Vamos a ver que la solución u tiene el mismo comportamiento que el rayo de luz que atraviesa una rejilla de difracción, pero para eso necesitamos conocer el valor de la suma de Gauss. Afortunadamente, las sumas cuadráticas de Gauss son objetos ampliamente estudiados. El lector puede consultar el primer capítulo del libro *Gauss and Jacobi sums* [1] y el apéndice del artículo de Hoz y Vega [4]. Sabemos decir, por ejemplo, que cuando $(p, q) = 1$ como en nuestro caso,

$$(32) \quad G(-p, r, q) = \begin{cases} \sqrt{q} e^{i\theta_r}, & \text{si } q \text{ es impar,} \\ \sqrt{2q} e^{i\theta_r}, & \text{si } q \text{ es par y } q/2 \equiv r \pmod{2}, \\ 0, & \text{si } q \text{ es par y } q/2 \not\equiv r \pmod{2}. \end{cases}$$

Aquí, el ángulo θ_r depende de r y también de p y de q , pero al ser p y q los valores que determinan el tiempo $t_{p,q}$, podremos considerar que son valores fijos cada vez que determinamos un tiempo. Analicemos algunos casos particulares para convencernos de que realmente tenemos el efecto de Talbot ante nuestros ojos.

- Sea $q = 1$, tal que $t_{p,1} = pd^2/2\pi$. Entonces, en la expresión de u (31), se tiene $r = 0$ y

$$u(x, t_{p,1}) = G(-p, 0, 1) \sum_{k \in \mathbb{Z}} \delta(x - kd) = \sum_{k \in \mathbb{Z}} \delta(x - kd),$$

dado que gracias a la definición (30) es fácil ver que $G(-p, 0, 1) = 1$. Por eso, vemos que en los tiempos $t_{p,1}$ recuperamos el dato inicial. La solución es, por lo tanto, periódica en tiempo con periodo $t_{1,1}$, que es el periodo mínimo. Podríamos decir, por lo tanto, que t_{11} es la **distancia de Talbot**,

$$(33) \quad z_T = t_{1,1} = \frac{d^2}{2\pi}.$$

- Sea $q = 2$, que significa que $t_{1,2}$ es la mitad de la distancia de Talbot. Esperamos, pues, recuperar el dato inicial trasladado por medio periodo. Para verlo, escribamos

$$u(x, t_{1,2}) = \frac{1}{2} \sum_{r=0}^1 G(-1, r, 2) \sum_{k \in \mathbb{Z}} \delta \left(x - dk - d \frac{r}{2} \right).$$

De acuerdo con la fórmula (32), o calculándola directamente, la suma de Gauss es nula cuando $r = 0$, y por lo tanto tenemos

$$u(x, t_{1,2}) = \sum_{k \in \mathbb{Z}} \delta(x - kd - d/2).$$

Es un tren de deltas centradas en cada $d(k + 1/2)$, que son los puntos medios entre las deltas iniciales. Es precisamente el fenómeno que esperábamos desde que vimos la alfombra en la figura 2.

- Sea $q = 4$, que corresponde a un cuarto de la distancia de Talbot. Esperamos que la solución sea el dato inicial duplicado. Si miramos a la fórmula (31), vemos que solo sobreviven los sumandos con $r = 0, 2$, y

$$u(x, t_{1,4}) = \frac{1}{\sqrt{2}} \left[e^{i\theta_0} \sum_{k \in \mathbb{Z}} \delta(x + kd) + e^{i\theta_2} \sum_{k \in \mathbb{Z}} \delta(x + kd + k/2) \right].$$

Esto muestra que hay deltas en los mismos puntos que el dato inicial (aunque en este caso vienen multiplicadas por ciertas constantes complejas unimodulares), así como en los puntos medios.

- En general, podemos comprobar que para diferentes valores de q la solución de la ecuación de Schrödinger se comporta como la alfombra de Talbot que vimos en la figura 2, en todos los múltiplos racionales de la distancia de Talbot z_T determinada en (33).

En definitiva, hemos visto que el mismo efecto que observamos y explicamos en la sección 1 y que deducimos en las secciones 2 y 3 se refleja en la solución de la ecuación de Schrödinger con el dato inicial periódico, que proponemos como paralelismo al *dato inicial* que de alguna manera representa la rejilla en el caso de la luz.

Referencias

- [1] BERNDT, Bruce C.; EVANS, Ronald J. y WILLIAMS, Kenneth S. *Gauss and Jacobi sums*. Canadian Mathematical Society Series of Monographs and Advanced Texts. John Wiley & Sons, Inc., New York, 1998. ISBN: 0-471-12807-4.
- [2] BERRY, Michael V. y KLEIN, S. «Integer, fractional and fractal Talbot effects». En: *J. Modern Opt.* 43.10 (1996), págs. 2139-2164. ISSN: 0950-0340. <https://doi.org/10.1080/095003496154761>.
- [3] CERDÀ, Joan. *Linear functional analysis*. Vol. 116. Graduate Studies in Mathematics. American Mathematical Society, Providence, RI; Real Sociedad Matemática Española, Madrid, 2010. ISBN: 978-0-8218-5115-9. <https://doi.org/10.1090/gsm/116>.
- [4] HOZ, Francisco de la y VEGA, Luis. «Vortex filament equation for a regular polygon». En: *Nonlinearity* 27.12 (2014), págs. 3031-3057. ISSN: 0951-7715. <https://doi.org/10.1088/0951-7715/27/12/3031>.
- [5] HUYGENS, Christiaan. *Traité de la lumière*. Leiden: Pieter van der Aa, 1690.
- [6] LORD RAYLEIGH, John William Strutt. «On Copying Diffraction-gratings, and on some Phenomena connected therewith». En: *Philos. Mag.* (Mar. de 1881).
- [7] LORD RAYLEIGH, John William Strutt. *The Theory of Sound*. Dover, 1894.
- [8] QUANTUM NANOPHYSICS GROUP, Universidad de Viena. *Quantum Interactive. The Talbot-Effect*. URL: <http://interactive.quantumnano.at/advanced/quantum-experiments/talbot-effect/> (visitado 10-12-2016).
- [9] SKYSKULL, Dr. *Skulls in the Stars. Rolling out the (optical) carpet*. URL: <https://skullsinthestars.com/2010/03/04/rolling-out-the-optical-carpet-the-talbot-effect/> (visitado 26-04-2017).
- [10] TALBOT, Henry Fox. «Facts Relating to Optical Science. No. IV». En: *Philos. Mag.* (Dic. de 1836).
- [11] VLADIMIROV, Vasily S. *Equations of mathematical physics*. Vol. 3. Marcel Dekker, Inc., New York, 1971.

TEMat, volumen 1. Julio de 2017.

e-ISSN: 2530-9633

© 2017 Asociación Nacional de Estudiantes de Matemáticas.

© 2017 los autores de los artículos.

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