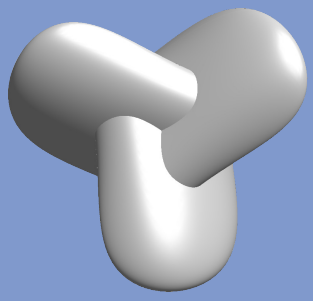


# Publicaciones Matemáticas del Uruguay

Proceedings del

## VI Coloquio Uruguayo de Matemática

*edited by*

Viviana Gubitosi
Ezequiel Maderna
Martín Reiris
Elisa Rocha

Volumen 17, Julio 2019

# Publicaciones Matemáticas del Uruguay

**Consejo Editor**

Diego Armentano
CMAT
diego@cmat.edu.uy

Jana Rodriguez Hertz
IMERL
jana@fing.edu.uy

Gonzalo Tornaría
CMAT / IMERL
tornaria@cmat.edu.uy

Armando Treibich
Université d'Artois / CURE
treibich@cmat.edu.uy

José L. Vieitez
Regional Norte
jvieitez@unorte.edu.uy

**Publicada por:**

CMAT — Facultad de Ciencias
IMERL — Facultad de Ingeniería

Universidad de la República

http://pmu.uy/

**ISSN: 0797-1443**

**Créditos:**

Diseño de tapa: J. Rodriguez Hertz
Editor LaTeX: I. Bustamante, G. Tornaría

©2019 PMU

Publicaciones Matemáticas del Uruguay

Volumen 17 Julio 2019

Prefacio ..... iii

*Minicursos*

Un paseo por la optimización
Marcelo Fiori ..... 1

Introducción a las Álgebras de Lie
Ana González ..... 25

La regularidad de los poliedros en la mirada moderna
Ángel Pereyra ..... 43

*Charlas Plenarias*

Holomorphic curves and celestial mechanics
Umberto L. Hryniewicz ..... 61

Solución de las ecuaciones de Hamilton-Jacobi como límite clásico de integrales con respecto al movimiento Browniano
José R. León ..... 81

Dinámica elemental de cubrimientos del anillo
Alvaro Rovella ..... 95

*Ponencias*

Complejidad en análisis numérico y el problema de valores propios
Diego Armentano ..... 109

Dendritaciones de superficies
Alfonso Artigue ..... 121

Ceros de polinomios y funciones aleatorias
Federico Dalmao ..... 127

Equivariant algebraic KK-Theory
Eugenia Ellis ..... 141

Amenabilidad, acciones de grupos, y equivalencia orbital
Eusebio Gardella ..... 153

Transformaciones en la enseñanza de la matemática para estudiantes ingresantes a la Facultad de Arquitectura, Diseño y Urbanismo
Omar Gil ..... 167

El número de oro, geometría, álgebra, aritmética
Gerardo González Sprinberg ..... 183

Teorías y modelos: una introducción a la lógica de primer orden
Alexandre Miquel ..... 195

Conjuntos minimales para homeomorfismos del círculo
Aldo Portela ..... 227

Wagner and Petersen are uniformly most-reliable graphs
Pablo Romero ..... 231

Una introducción a modelos matemáticos para composición automática de música tonal
Verónica Rumbo ..... 247

Semiprojectivity and semiinjectivity in different categories
Hannes Thiel ..... 263

Rigidez y geometricidad de acciones de grupos de superficies en el círculo
Maxime Wolff ..... 277

*Diplomados, ponencias*

Ley de Benford
María Caputi Zunini ..... 285

Criptografía sobre curvas elípticas
Horacio Castagna ..... 293

Las tareas enfocadas en similitudes y diferencias en el aprendizaje de las transformaciones lineales en la formación inicial de profesores de matemática
Ana Maldonado ..... 301

Un proyecto de intervención a partir de tareas de generalizar y particularizar: trabajo colaborativo entre investigador y formador de profesores
Victoria Mesa Pais ..... 307

# Prefacio

Este volumen colecciona varias de las diversas e interesantes contribuciones de investigadores nacionales e internacionales al VI Coloquio Uruguayo de Matemática, realizado en Montevideo entre el 20 y el 22 de Diciembre de 2017. Testifica al mismo tiempo el tipo y calidad de la actual investigación nacional (y sus vínculos internacionales) así como la magnitud de su expansión y diversificación a lo largo de los años.

Siguiendo la organización misma del evento, el volumen se divide en cuatro secciones: *Minicursos*, *Charlas Plenarias*, *Ponencias*, y *Diplomados, ponencias*. En ellas los autores exponen con estilos didácticos propios, desde una perspectiva panorámica o particular y en tópicos tradicionales o de frontera.

Entre investigadores nacionales e internacionales, docentes, estudiantes y público general, circularon más de trescientos participantes.

El Coloquio Uruguayo de Matemática es financiado por el PEDECIBA, el Instituto de Matemática y Estadística Rafael Laguardia (IMERL-FIng), y el Centro de Matemática del Uruguay (CMat-FCien).

Nuestro mayor agradecimiento a todos los que hicieron esto posible.

Viviana Gubitosi
Ezequiel Maderna
Martín Reiris
Elisa Rocha

Montevideo, Julio 2019.

# MINICURSOS

MINICURSO

# UN PASEO POR LA OPTIMIZACIÓN

MARCELO FIORI

Publicaciones Matemáticas del Uruguay
Volumen 17, Julio 2019, Páginas 1–21
S 0797-1443

# UN PASEO POR LA OPTIMIZACIÓN

MARCELO FIORI

## 1. Introducción

En general, un problema de optimización consiste en minimizar una función en cierto conjunto, y lo escribimos como:

$$\min_{x \in \mathcal{X}} f(x).$$

Muchas veces lo que importa, más que el valor mínimo, es dónde se alcanza el mismo. El problema de encontrar el o los puntos donde se minimiza la función, se expresa así:

$$\arg\min_{x \in \mathcal{X}} f(x).$$

La gran variabilidad de funciones $f$, conjuntos $\mathcal{X}$, y estructuras posibles, dan lugar a una gran cantidad de aplicaciones, preguntas interesantes, métodos, etc. Por ejemplo, $f$ puede ser una función convexa o no, diferenciable o no, $\mathcal{X}$ puede ser un subconjunto convexo de $\mathbb{R}^n$ o no, puede ser una variedad, puede ser un espacio de dimensión finita o no, etc.

Hay muchos ejemplos que ya son conocidos por todos los que pasaron por algun curso universitario de matemática. Por citar dos nada más: el problema de mínimos cuadrados y el de proyección de un punto a un conjunto dado (es decir, encontrar el punto del conjunto que minimiza la distancia al punto dado).

A lo largo de este mini curso intentaremos mostrar algunas aplicaciones o conexiones con diferentes áreas (geometría, álgebra, estadística, física, machine learning), algunos aspectos teóricos, algorítmicos, y algunas interpretaciones interesantes.

Una parte muy importante de la optimización es que podamos resolver el problema (sea lo que sea que esto quiere decir) en un tiempo *razonable* en una computadora. Durante mucho tiempo, el límite entre los problemas de optimización que se consideraban "realizables" y los que no, era la linealidad. Es decir, que la función $f$ fuera lineal, y el conjunto $\mathcal{X}$ fuera dado por igualdades y desigualdades de funciones lineales. Hoy se considera que este límite es la convexidad (que $f$ sea una función convexa, y $\mathcal{X}$ un conjunto convexo). Esto significa que hay mucho interés en algoritmos para problemas convexos (porque son los que se pueden resolver), pero también hay mucho interés en problemas no convexos (porque son los que "no se pueden resolver", y son por lo tanto un interesante desafío).

## 2. Motivación y ejemplos

Los siguientes ejemplos no son todos completamente independientes entre sí, y no están pensados para leer necesariamente en este orden, ni en este lugar. Es decir, se pueden saltear todos los ejemplos y pasar directamente a la sección 3, se puede

©2019 PMU

leer linealmente, o se puede ir alternando entre secciones de ejemplos y el resto de las secciones[1].

**2.1. El problema de Thomson (o el séptimo problema de Smale).** Consideremos $N$ electrones en la esfera $S^2$, llamemos $x_i$ al vector de $\mathbb{R}^3$ con las coordenadas del electrón número $i$, y consideremos todas las posiciones juntas en $x = (x_1, x_2, \dots, x_N)$. La energía de la interacción entre los electrones $i$ y $j$ (con cargas iguales $e_i = e_j = e$) es $U_{ij} = \frac{k_e e_i e_j}{\|x_i - x_j\|}$, donde $k_e$ es la constante de Coulomb. Entonces la energía potencial total del sistema, simplificando las unidades para eliminar las constantes, es la siguiente:

$$U(X) = \sum_{i<j} \frac{1}{\|x_i - x_j\|}.$$

El problema consiste en encontrar configuraciones de energía mínima. Es decir, encontrar las posiciones $x_i$ que minimicen la función $U$:

$$\arg \min_{s.t. \|x_i\|=1 \forall i} U(x),$$

donde la restricción $\|x_i\| = 1$ impone que los electrones se encuentren en la esfera.

Las configuraciones de mínima energía han sido estudiadas en profundidad solamente para algunos pocos valores de $N$. Por ejemplo, si tenemos dos electones, la solución es trivial, y consiste en posiciones diametralmente opuestas, que dan lugar a una energía de $\frac{1}{2}$. Para $N = 3$, los electrones se encuentran en los vértices de un triángulo equilátero en una circunferencia maximal. Para $N = 4$, los electrones forman un tetraedro regular. Sin embargo el problema todavía es de mucho interés para muchos valores de $N$, comportamiento asintótico, o algoritmos para hallar configuraciones mínimas.

Este problema es un caso particular del denominado séptimo problema de Smale [22], cuya motivación es encontrar buenos puntos de partida (raíces de polinomios) para algoritmos de homotopía que busquen todas las raíces de un polinomio dado en $\mathbb{C}$. Muy resumidamente, dado un polinomio en $\mathbb{C}$, un algoritmo de homotopía parte de un polinomio con raíces conocidas, y construye un camino desde el polinomio conocido al polinomio objetivo, intentando seguir el camino que recorren las raíces. Para esto, es necesario que las raíces del polinomio inicial estén tan alejadas entre sí como sea posible, y por eso se buscan configuraciones que minimicen $U(x)$.

Con las herramientas básicas descritas en este texto (específicamente en la sección 4.2), se puede programar muy fácilmente un algoritmo para encontrar soluciones aceptables para valores moderados de $N$.

**2.2. Problema de valores propios.** Uno de los problemas más fáciles de entender, pero a la vez más complejos, es el problema de encontrar algoritmos para hallar valores y vectores propios.

Sea $A$ una matriz simétrica real, a la que le queremos hallar sus valores y vectores propios. Es decir, buscamos una matriz $U$ ortogonal ($U \in \mathcal{O}(n)$) y una matriz $D$ diagonal tales que $A = UDU^T$. Si bien existen algoritmos muy utilizados como el método QR, que no están inspirados explícitamente en un problema de optimización, vamos a describir aquí otros dos enfoques que sí lo están.

[1]Un orden posible, al estilo Rayuela, puede ser 1, 2.1, 3, 2.4, 4, 2.2, 5, 2.3, 6, 2.5.

Ordenemos los valores propios de $A$: $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_n$, y sean $v_1, \ldots, v_n$ los vectores propios asociados. Comencemos definiendo el cociente de Rayleigh de $A$, que es la función $r_A : \mathbb{R}^n \setminus \{0\} \to \mathbb{R}$ dada por

$$r_A(x) = \frac{x^T Ax}{x^T x} = \frac{\langle Ax, x\rangle}{\langle x, x\rangle}.$$

Observar que $r_A$ es invariante por cambios de escala, o sea, $r_A(tx) = r_A(x)$, $\forall t \in \mathbb{R} \setminus \{0\}$, por lo que basta estudiar la función en $S^{n-1} = \{x \in \mathbb{R}^n : \|x\| = 1\}$. En este caso, podemos obviar la normalización del denominador, por lo que la función $r_A$ básicamente toma un vector de norma uno, y mide "cuán colineal" es $Ax$ con $x$ (y cuánto expande $A$ en esa dirección). El siguiente resultado captura esta idea:

**Teorema** (Courant-Fisher)**.** Se tiene que

$$\lambda_1 = \max_{\|x\|=1} r_A(x), \quad \lambda_n = \min_{\|x\|=1} r_A(x).$$

Además, los puntos críticos de $r_A$ en $S^{n-1}$ son los vectores propios $v_i$ de $A$.

De esta manera, formulamos el problema de hallar valores propios como un problema de optimización (por lo menos los valores propios más grande y más chico, aunque se puede extender esta idea fácilmente).

Veamos otra forma de atacar el problema. Como tenemos (o buscamos tener) $A = UDU^T$, podemos pensar equivalentemente que buscamos una matriz ortogonal $U$ tal que $U^T AU$ sea diagonal. Una manera de expresar esto, es intentando minimizar los elementos fuera de la diagonal de $U^T AU$. Si denotamos como $diag(M)$ a la matriz que tiene solamente los elementos en la diagonal de $M$, entonces buscamos

$$min_{U \in \mathcal{O}(n)} \|diag(U^T AU) - U^T AU\|_F^2.$$

Esta función suma los cuadrados de las entradas fuera de la diagonal de $U^T AU$. Como $A$ es simétrica, sabemos que tiene que haber una matriz $U$ que haga que $U^T AU$ sea diagonal, y por lo tanto el mínimo de la función objetivo será cero, y se alcanzará en una matriz ortogonal que diagonaliza $A$.

En realidad, se puede ver que si consideramos una matriz diagonal $N$ cualquiera, el problema

$$min_{U \in \mathcal{O}(n)} \|N - U^T AU\|_F^2$$

también tiene como soluciones las matrices $U$ que diagonalizan $A$. Más aún, si $N$ tiene los elementos ordenados (por ejemplo $N = diag(1, 2, 3, \ldots, n)$), entonces el punto crítico correspondiente a los valores propios ordenados de menor a mayor, es un atractor.

Tenemos entonces otras dos formulaciones en términos de problemas optimización. Sobre los métodos para resolverlos, se comenta brevemente en algunos pasajes de estas notas.

**2.3. Problemas de isomorfismos de grafos.** Los que siguen son otros problemas simples de enunciar, pero de los que todavía queda mucho por entender. Hace muy poco, László Babai revolucionó a la comunidad con un resultado sobre el problema de isomorfismo de grafos [4], pero aún quedan muchas preguntas planteadas. En esta sección describiremos los problemas, formularemos un problema de optimización equivalente, y comentaremos algunos resultados recientes.

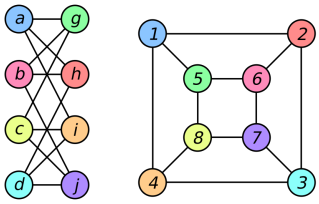
**Definición.** Un grafo $G = (V, E)$ consiste en un conjunto $V = \{v_1, v_2, \ldots, v_n\}$ de vértices y un conjunto de aristas $E$, que son pares no ordenados de $V$.

Intuitivamente, un grafo es un conjunto de puntos unidos por aristas, por lo que hay muchas formas de "dibujarlo". En particular, cualquier re-ordenación de los nodos, da lugar al mismo dibujo. El concepto de isomorfismo captura este concepto:

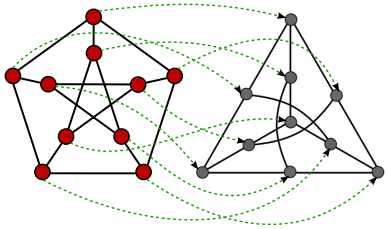
**Definición.** Decimos que dos grafos $G_1 = (V_1, E_1)$ y $G_2 = (V_2, E_2)$ son isomorfos sii existe una biyección $\varphi : V_1 \to V_2$ que preserve la estructura. Esto es, que $(v_i, v_j) \in E_1 \iff (\varphi(v_i), \varphi(v_j)) \in E_2$.

Un isomorfismo de un grafo en sí mismo se denomina automorfismo, y el conjunto de automorfismos de un grafo forma un grupo con la composición.

**Ejemplo.** *Los grafos de la siguiente figura son isomorfos, siendo $\varphi$ la que asigna a un vértice de un grafo, el vértice del mismo color en el otro grafo.*




**Ejemplo.** *Los grafos de la siguiente figura son isomorfos, con la biyección representada con las flechas punteadas.*

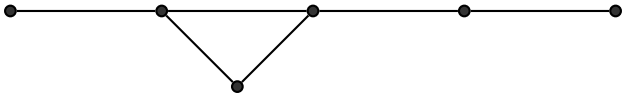


El problema de isomorfismo de grafos (GIP por sus siglas en inglés) consiste en, dados dos grafos, decidir si son isomorfos o no. Computacionalmente, es un problema difícil, cuya complejidad se encuentra estrictamente entre P y NP (a no ser que estas clases colapsen).

Otro problema relacionado es el de automorfismo de grafos, que consiste determinar si el grupo de automorfismos de un grafo es trivial o no.

Observar que los dos problemas presentados arriba busca respuestas del tipo *Sí o no*, pero no busca encontrar elementos (isomorfismos o automorfismos) explícitamente.

**Ejemplo.** *El grafo de la figura tiene grupo de automorfismo trivial.*



Dado un grafo $G = (V, E)$, definimos su matriz de adyacencia como $A \in \mathrm{Sym}(n)$

$$A_{ij} = \begin{cases} 1 & \text{si } (v_i, v_j) \in E \\ 0 & \text{si no} \end{cases}$$

*Comentario*
Si $A$ y $B$ son matrices de adyacencia de dos grafos isomorfos, entones $A = PBP^T$ para alguna matrix de permutación $P \in \mathcal{P}(n)$.

En términos de las matrices de adyacencia, el problema de isomorfismo de grafos es: ¿$\exists P \in \mathcal{P}(n) / A = PBP^T$?

Por otro lado, el denominado Graph Matching Problem (GMP) consiste en encontrar el isomorfismo entre dos grafos (o el "más cercano" en algún sentido, si no son isomorfos). Esto es:

$$P^* = \arg \min_{P \in \mathcal{P}(n)} \|A - PBP^T\|_F^2,$$

o lo que es lo mismo (multiplicando por $P$ a derecha, que no modifica la norma),

$$P^* = \arg \min_{P \in \mathcal{P}(n)} \|AP - PB\|_F^2, \tag{1}$$

El problema es que $\mathcal{P}(n)$ es un conjunto discreto con $n!$ elementos, por lo que computacionalmente no es viable probar todos los elementos del conjunto (que son todas las formas de re-ordenar los nodos) para encontrar la solución.

Un camino posible consiste en relajar el dominio, pasando del conjunto de permutaciones $\mathcal{P}(n)$ a su envolvente convexa, que resulta ser (teorema de Birkhoff-von Neumann) el conjunto de matrices de entradas no negativas cuyas filas y columnas suman uno. Estas matrices se denominan *doblemente estocásticas*, y denotaremos a este conjunto como $\mathcal{D}$. Es decir: $\mathcal{D}(n) = \{M \in \mathbb{R}^{n \times n} : M\mathbf{1} = \mathbf{1}, M^T\mathbf{1} = \mathbf{1}, M_{ij} \geqslant 0 \, \forall i, j\}$.

El problema resultante es entonces

$$\hat{P} = \arg \min_{P \in \mathcal{D}(n)} \|AP - PB\|_F^2, \tag{2}$$

que es ahora un problema convexo, y por lo tanto lo podemos resolver (por ejemplo, con técnicas de 4.2). Por supuesto que la solución $\hat{P}$ de este problema no necesariamente será solución de (1), ya que agrandamos el conjunto. Una opción es resolver (2), y luego buscar la matriz de permutación más cercana a $\hat{P}$ (que se puede hacer en tiempo polinomial), pero en principio no hay garantías de que esta matriz de permutación corresponda a la solución del problema original. Hay también otros enfoques posibles [14, 17].

Una pregunta interesante es bajo qué condiciones la solución del problema (2) que *podemos* resolver, coincide con la solución del problema que *queremos* resolver. Algunos resultados parciales se pueden encontrar en [2, 16], y resultados asintóticos en [19].

En particular, algunos de estos resultados, que vienen de estudiar en qué casos dos problemas de optimización son equivalentes, dan lugar a resultados puramente algebraicos, como por ejemplo que si el espectro de la matriz de adyacencia de un grafo tiene ciertas propiedades, entonces el grupo de automorfismos es trivial.

**2.4. Machine Learning.** Uno de los campos que está más de moda, donde la optimización es fundamental, es el aprendizaje automático (Machine Learning, o Inteligencia Artificial, son también denominaciones de conceptos muy similares). Dicho muy breve e incompletamente, consiste en hacer algoritmos para poder tomar decisiones de manera automática, generalmente basados en datos.

Ejemplos de aplicaciones de estos métodos se ven todos los días: sistemas de recomendación (por ejemplo los de Netflix o Spotify), etiquetado automático de imágenes, transcripción automática de música, traducción entre idiomas en tiempo real, o el AlphaGo (que logró un nivel superior al humano en el ancestral juego chino Go), son algunos de los miles de ejemplos.

Quizás el problema más sencillo de comprender es el de clasificación. Supongamos que tenemos elementos que queremos separar en dos clases, por ejemplo peras de manzanas[2], y para ello contamos con medidas de los elementos (digamos, ancho y alto). A partir de ver solamente las medidas, queremos decidir si se trata de una pera o de una manzana. Para ello, en general se cuenta con cierta cantidad de datos de ejemplo, para poder entrenar al clasificador.

Es decir, tenemos dos conjuntos de puntos $(x_1, x_2, \ldots, x_N)$ y $(y_1, y_2, \ldots, y_M)$, distribuidos en $\mathbb{R}^n$, y buscamos una función $f : \mathbb{R}^n \to \mathbb{R}$ tal que $f$ sea negativa para una clase, y positiva para la otra:

$$f(x_i) > 0, \quad i = 1, \ldots, N, \qquad\qquad f(y_i) < 0, \quad i = 1, \ldots, M.$$

Un primer paso, que da lugar luego a uno de los clasificadores más usados (SVM), consiste en intentar separar linealmente los conjuntos. Es decir, buscamos un hiperplano que deje al conjunto de los $x_i$ de un lado, y al conjunto de los $y_i$ del otro. Por supuesto que esto no siempre es posible, y para esto están las modificaciones y variaciones que no mencionaremos en esta oportunidad.

Si los conjuntos son linealmente separables, en general hay muchos hiperplanos que cumplen la función. Pero el problema de clasificación consiste en aprender el hiperplano, para luego clasificar nuevas instancias que llegarán. Por lo tanto no todos los hiperplanos son igualmente buenos.

Buscamos entonces el hiperplano $a^T x + b$ que ofrezca el mayor margen de separación posible, para que cuando llegue una nueva instancia, sea más probable clasificarla correctamente. Una manera de plantear esto es mediante la variable auxiliar $t$, que mide justamente el margen (en realidad el margen es $2t$, ver figura):

$$\begin{array}{ll} \text{maximizar} & t \\ \text{s.t.} & a^T x_i + b \geqslant t, i = 1, \ldots, N \\ & a^T y_i + b \leqslant -t, i = 1, \ldots, N \\ & \|a\|_2 \leqslant 1 \end{array}$$

Es decir, maximizar el margen de separación, con las restricciones de separar los conjuntos de puntos. La restricción sobre la norma de $a$ es porque sino se podría escalar todo por una constante y obtener márgenes arbitrariamente grandes. Este problema es convexo, y por lo tanto se puede resolver eficientemente. Se puede ver una discusión más amplia sobre esto en [8].

[2]Puede ser también correo spam de no-spam, usos fraudulentos de tarjetas de crédito de usos normales, imágenes, etc.

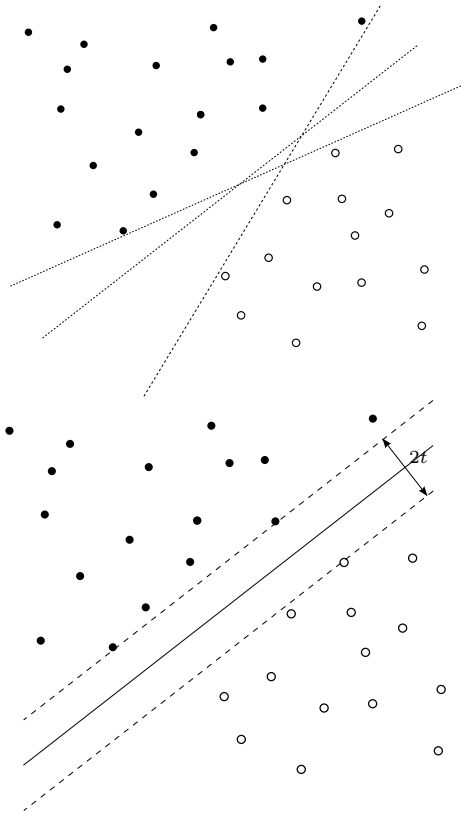



FIGURA 1. Conjuntos de puntos linealmente separables. A a izquierda, distintos hiperplanos que separan los conjuntos. A la derecha, el hiperplano con mayor margen.

*2.4.1. Matrix completion.* El problema de *completado de matrices* consiste, como lo indica su nombre, en completar una matriz conociendo solamente algunas de sus entradas. Por supuesto que este problema, así enunciado, no tiene solución (o mejor dicho, tiene infinitas soluciones), a no ser que asumamos alguna estructura sobre la matriz. Generalmente se denomina como *matrix completion* al problema de encontrar la matriz de menor rango, que complete las entradas conocidas. Este modelo responde a varios problemas importantes, de los cuales mencionaremos un par de ejemplos.

El primero es quizás el que hizo famoso el problema de matrix completion, y es el denominado problema de Netflix, que cobró relevancia en medios de prensa especializados porque ofrecía un premio de un millón de dólares. Se pretendía mejorar el sistema de recomendación de Netflix. La matriz sintetiza la información sobre los gustos de los usuarios. Cada fila de la matriz corresponde a un usuario, y cada columna a una película. En la entrada $(i, j)$, si el usuario $i$ vio la película $j$, se encuentra el puntaje que le asignó. Es por lo tanto una matriz que tiene muy pocos elementos conocidos (pues cada usuario ve relativamente pocas películas), pero de la que es deseable inferir el resto de las entradas, porque de esta manera, el proveedor del servicio puede saber si recomendar o no determinada película a determinado

usuario. ¿Por qué esta matriz debería ser de bajo rango? Porque se cree que son algunos pocos factores los que contribuyen al gusto de cada persona. Desde otro punto de vista, también se puede pensar que cada usuario es (aproximadamente) combinación lineal de otros pocos usuarios.

Para el otro ejemplo, supongamos que tenemos una red de sensores inalámbricos, distribuidos aleatoriamente en alguna región, y cada sensor puede estimar la distancia a sus vecinos más cercanos, basándose en la potencia de la señal. Con estos datos, formamos una matriz (incompleta) de distancias, y nos gustaría reconstruir la geometría de la red de sensores (o sea, el resto de la matriz). Si los sensores están ubicados en un plano, entonces esta matriz tiene rango dos (y rango tres si se encuentran en el espacio tridimensional).

Volviendo a la formulación del problema, si denotamos por $\Omega$ al conjunto de índices para los cuales se conocen las entradas de la matriz, y por $M_{ij}$ a dichas entradas, entonces el problema es:

$$\begin{array}{ll} \min\limits_{X\in\mathbb{R}^{n\times n}} & rank(X) \\ \text{s.t.} & X_{ij} = M_{ij} \quad \forall (i,j) \in \Omega. \end{array} \tag{3}$$

Lamentablemente este problema es NP-hard (y no solo eso, sino que además todos los algoritmos conocidos requieren tiempo doblemente exponencial, tanto en teoría como en práctica), por lo tanto es necesario buscar alternativas.

La más exitosa surge de observar que el rango de una matriz coincide con la cantidad de valores singulares no nulos. Por lo tanto, el problema (3) es equivalente a minimizar la cantidad de valores singulares no nulos. En vez de hacer esto, una relajación de este problema consiste en considerar la denominada norma nuclear, que es la suma de los valores singulares:

$$\|X\|_* = \sum_{i=1}^{n} \sigma_i(X)$$

de donde el problema relajado es:

$$\begin{array}{ll} \min\limits_{X\in\mathbb{R}^{n\times n}} & \|X\|_* \\ \text{s.t.} & X_{ij} = M_{ij} \quad \forall (i,j) \in \Omega. \end{array} \tag{4}$$

Este problema tiene la ventaja de ser convexo, y por lo tanto se puede resolver eficientemente. La relajación del rango a la norma nuclear está fuertemente relacionada con el contenido de la próxima sección. La pregunta, como en varios de los ejemplos aquí presentados donde aparecen relajaciones de problemas, es cuándo la solución del problema (4) coincide con la solución de (3). Una discusión detallada del problema, y resultados en el sentido de la equivalencia de los problemas, se puede encontrar por ejemplo en [10, 12].

*2.4.2. Sparse modeling.* Consideremos una matriz de rango completo $A \in \mathbb{R}^{n\times m}$ con $n < m$, y el sistema de ecuaciones indeterminado $Ax = b$. Este sistema tiene infinitas soluciones, y para determinadas aplicaciones, a uno le gustaría elegir alguna de ellas, con cierto criterio, que sea particularmente útil para la aplicación. Una opción, por ejemplo, es elegir la solución de menor norma:

$$\begin{array}{ll} \min\limits_{x\in\mathbb{R}^{m}} & \|x\|_2^2 \\ \text{s.t.} & Ax = b. \end{array} \tag{5}$$

Este problema tiene solución en forma cerrada, que es la conocida $x = A^T(AA^T)^{-1}b$. Sin embargo, supongamos que queremos buscar la solución más "esparsa", la que tiene menor cantidad de elementos no nulos. Esto puede estar justificado de alguna manera, por el deseo de encontrar la solución "más simple", y de hecho en la última década ha habido una cantidad increíble de resultados teóricos y aplicaciones de este problema. Haciendo un poco de abuso de notación (ya que no es formalmente una norma), llamaremos $\|x\|_0$ a la cantidad de elementos no nulos del vector $x$. Se puede pensar $\|x\|_0$ como el límite de $\|x\|_p^p$ cuando $p \to 0$. El problema entonces lo podemos formular de la siguiente manera:

$$\begin{aligned} \min_{x\in\mathbb{R}^m} \quad & \|x\|_0 \\ \text{s.t.} \quad & Ax = b. \end{aligned} \tag{6}$$

Este problema, sin embargo, resulta ser NP-hard. Una manera de resolver una versión aproximada de (6) (ciertamente no la única), es considerar la norma uno, que es la norma convexa más cercana a $\|x\|_0$. El problema resultante, que es ahora convexo, y por lo tanto se puede resolver, es

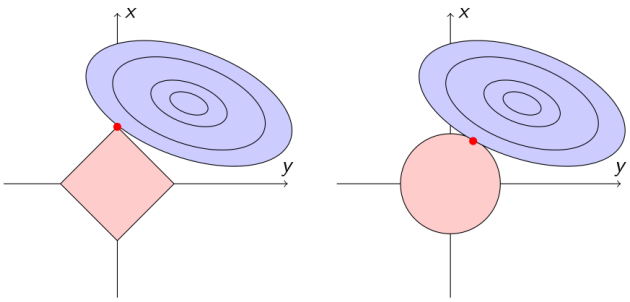
$$\begin{aligned} \min_{x\in\mathbb{R}^m} \quad & \|x\|_1 \\ \text{s.t.} \quad & Ax = b. \end{aligned} \tag{7}$$

Observar que la función $\|x\|_1$ es convexa pero no diferenciable, por lo cual algoritmos que puedan operar en estas condiciones son de particular interés.

La pregunta, como en varios otros pasajes de estas notas, es cuándo la solución del problema (7), que podemos resolver, coincide con la solución del problema (6), que es el que queremos resolver originalmente. Hay varios resultados en este sentido [9, 11]. Daremos aquí simplemente una intuición geométrica. Para esto, consideremos el problema similar:

$$\begin{aligned} \min_{x\in\mathbb{R}^m} \quad & \|Ax - b\|_2^2 \\ \text{s.t.} \quad & \|x\|_1 \leqslant \alpha. \end{aligned} \tag{8}$$

En la siguiente figura, se encuentran las curvas de nivel de $\|Ax-b\|_2^2$, y la restricción, que es simplemente que $x$ pertenezca a la bola de radio $\alpha$ con la norma uno. Al costado, se observa la misma situación pero utilizando la norma dos para la restricción. Se puede obsevar que en el caso de la norma uno, la solución se encuentra en un vértice (y esta propiedad es genérica), por lo que varias de sus coordenadas son cero.




*Comentario*
Relacionando estos conceptos con el problema de la sección 2.4.1, observar que la

norma nuclear es al rango, lo que la norma uno es a la pseudo-norma cero. Mientras una cuenta elementos no nulos (en el caso del rango, valores singulares no nulos), la otra suma sus valores absolutos.

La cantidad de aplicaciones de estas técnicas es tan grande que no podríamos listarlas aquí. Mostramos solamente una de ellas, para ejemplificar la fortaleza de la formulación, y referimos al lector interesado a otros textos para profundizar [9, 20].

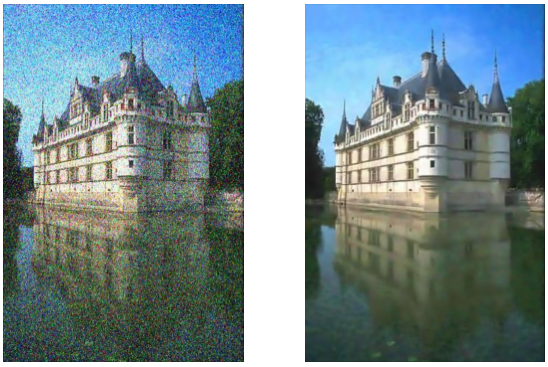


FIGURA 2. Denoising de imágenes utilizando técnicas de sparse modeling. A la izquierda la imagen de entrada, a la derecha el resultado de eliminar el ruido mediante el problema de optimización descrito, y una buena elección de la matriz $A$.

**2.5. Maximum Likelihood.** Consideremos una familia de distribuciones de probabilidad en $\mathbb{R}^m$, indexadas por un vector $x \in \mathbb{R}^n$, con densidades $p_x(\cdot)$ (un modelo paramétrico). Cuando las consideramos como función del parámetro $x$, con un $y \in \mathbb{R}^m$ fijo, a la función $p_x(y)$ se le suele llamar función de verosimilitud (likelihood en inglés), y en general es conveniente trabajar con el logaritmo de la misma (log-likelihood): $l(x) = \log p_x(y)$.

Un problema usual es estimar el valor del parámetro $x$, basados en una observación $y$. Uno de los métodos más usados es el llamado de Máxima Verosimilitud (o Maximum Likelihood), que consiste en estimar el parámetro $x$ como:

$$\hat{x}_{ML} = \arg\max_x p_x(y) = \arg\max_x l(x).$$

Intuitivamente, se selecciona el valor del parámetro $x$ que hace los datos más probables.

Cuando por ejemplo la función $l(x)$ es cóncava, el problema resulta convexo. Veamos algunos ejemplos particulares.

*2.5.1. Modelos lineales con ruido.* Supongamos que tenemos un modelo lineal:

$$y_i = a_i^T x + v_i, \qquad i = 1, \dots m,$$

donde $x \in \mathbb{R}^n$ es el vector de parámetros que queremos estimar, $y_i$ son las medidas con las que contamos, y los $v_i$ representan el ruido o errores de medida, que asumiremos i.i.d. con densidad $p(\cdot)$ en $\mathbb{R}$. La función de verosimilitud $p_x(y)$ es entonces:

$$p_x(y) = \prod_{i=1}^{m} p(y_i - a_i^T x),$$

de donde la función de log-likelihood resulta

$$l(x) = \sum_{i=1}^{m} \log p(y_i - a_i^T x).$$

Si asumimos que el ruido es Gaussiano, $p(z) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{z^2}{2\sigma^2}}$, entonces resulta

$$l(x) = -\frac{m}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\|Ax - y\|_2^2,$$

donde $A$ es la matriz formada por los $a_i$. El estimador de máxima verosimilitud entonces coincide con el problema de mínimos cuadrados, o dicho de otra forma, cuando hacemos mínimos cuadrados estamos asumiendo que el ruido es Gaussiano.

Si el ruido fuera Laplaciano, $p(z) = \frac{1}{2a} e^{-\frac{|z|}{a}}$, con $a > 0$, entonces se puede ver que el estimador de ML consiste en:

$$\hat{x} = \arg\min_x \|Ax - y\|_1,$$

que también es un problema convexo.

*2.5.2. Inversa de covarianza para variables Gaussianas.* Supongamos que $y = (y_1, \ldots, y_n)$ es una variable aleatoria Gaussiana $y \sim \mathcal{N}(0, \Sigma)$, y sea $\Theta = \Sigma^{-1}$, que se denomina matriz de concentración o matriz de precisión, y también resulta definida positiva, al igual que $\Sigma$. El *problema de selección de covarianza* [13] consiste en estimar el patrón de ceros de la matriz de precisión. Este problema es de particular interés para analizar dependencias entre variables, dado que dos coordenadas $y_i$ e $y_j$ son condicionalmente independientes dadas las otras coordenadas, si y solo si $\Theta(i,j) = 0$. Esta propiedad, además, motiva la representación de la estructura de dependencia condicional en términos de un grafo (con aristas entre los nodos $i$ y $j$ cuando $\Theta(i,j) \neq 0$).

Empezemos de todas maneras con $\Sigma$ como parámetro a estimar. La densidad de $y$ es:

$$p_\Sigma(y) = \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}} e^{-\frac{y^T \Sigma^{-1} y}{2}}.$$

Dadas $N$ muestras independientes $Y_1, \ldots, Y_N \in \mathbb{R}^n$, queremos estimar $\Sigma$. El logaritmo de la función de verosimilitud es entonces

$$\begin{aligned} l(\Sigma) &= \log p_\Sigma(Y_1, \ldots, Y_N) \\ &= -\frac{Nn}{2}\log(2\pi) - \frac{N}{2}\log\det\Sigma - \frac{1}{2}\sum_{i=1}^{N} Y_i^T \Sigma^{-1} Y_i \\ &= -\frac{Nn}{2}\log(2\pi) - \frac{N}{2}\log\det\Sigma - \frac{N}{2}\operatorname{tr}(\Sigma^{-1} S) \end{aligned}$$

donde $S = \frac{1}{N}\sum_{i=1}^{N} Y_i Y_i^T$ es la matriz empírica de covarianza.

Así planteado, el problema de optimizar $l(\Sigma)$ no es convexo, pero sin embargo se puede trabajar un poco para dejar el problema más amigable. Pasemos entonces, ahora sí, a usar la inversa de $\Sigma$ como parámetro. Resulta

$$l(\Theta) = -\frac{Nn}{2}\log(2\pi) + \frac{N}{2}\log\det\Theta - \frac{N}{2}\operatorname{tr}(\Theta S),$$

por lo que el estimador de máxima verosimilitud es

$$\max_{\Theta \succ 0} \log\det\Theta - \operatorname{tr}(\Theta S).$$

Ahora la función objetivo es cóncava, por lo que el problema resulta convexo (pues queremos maximizar, y no minimizar). Se puede ver que el gradiente de la función objetivo es $\Theta^{-1} - S$, por lo que (utilizando conceptos básicos, que son recordados en 3) cuando no tenemos más información o restricciones sobre $\Sigma$, el estimador de máxima verosimilitud de la matriz de covarianza es simplemente $\hat{\Sigma}_{ML} = S$, la covarianza empírica.

Volviendo al problema de selección de covarianza, muchas veces se tiene como información que los elementos no nulos de $\Theta$ son relativamente pocos, por lo que, a la hora de estimar la inversa de covarianza, se desea penalizar la cantidad de entradas no nulas. Una manera de hacer esto (como fue mencionado en 2.4.2), consiste en agregar una penalización de norma $\ell_1$, quedando el problema:

$$\max_{\Theta>0} \log\det\Theta - \operatorname{tr}(S\Theta) - \lambda\|\Theta\|_1.$$

Este problema sigue siendo convexo, y tiene muchísimas aplicaciones, generalizaciones, e implementaciones eficientes [5, 15].

## 3. (Algunas) condiciones de optimalidad y subdiferenciales

El objetivo más ambicioso en optimización es conseguir el mínimo global de la función. Esto es posible para algunos casos, pero para muchos otros casos no lo es. Entonces uno se conforma con encontrar un mínimo local, o incluso algún punto que verifique alguna condición relacionada con esto. Veamos algunos ejemplos.

Desde los primeros cursos de cálculo vemos condiciones necesarias de optimalidad. Por ejemplo, para funciones derivables de una variable, si la función tiene un mínimo relativo en $x_0$, entonces su derivada se anula en ese punto. Por supuesto que esto es solamente una condición necesaria (el contraejemplo más usado es $x^3$ en el origen), pero muchas veces buscar puntos donde se anule la derivada es lo mejor que podemos hacer.

Esta condición de optimalidad se generaliza de manera directa a funciones diferenciables de $\mathbb{R}^n$ a $\mathbb{R}$, donde resulta que el gradiente debe anularse, pero también tiene generalizaciones para funciones no diferenciables, y para puntos que no son interiores al dominio. Tenemos por ejemplo el siguiente resultado inmediato:

**Proposición 1.** Sea $f : \mathcal{X} \to \mathbb{R}$ diferenciable, donde $\mathcal{X}$ es un subconjunto convexo de $\mathbb{R}^n$.

a) Si $f$ tiene un mínimo local en $x^*$, entonces $\langle \nabla f(x^*), x - x^* \rangle \geqslant 0$ para todo $x \in \mathcal{X}$.
b) Si además $f$ es convexa, entonces la condición anterior es suficiente.

La interpretación de la primera condición es clara: dado que el conjunto $\mathcal{X}$ es convexo, las direcciones $(x - x^*)$ son las factibles, es decir, las direcciones hacia las cuales uno se puede mover desde $x^*$, manteniéndose dentro del conjunto $\mathcal{X}$. El producto interno $\langle \nabla f(x^*), x - x^* \rangle$ no es otra cosa que la derivada direccional de $f$ en la dirección $(x - x^*)$, y por lo tanto la condición se puede leer así: en todas las direcciones que uno se pueda mover desde $x^*$, la función $f$ *crece*, y por lo tanto $x^*$ tiene que ser un mínimo local.

Si llamamos $P_{\mathcal{X}} : \mathbb{R}^n \to \mathcal{X}$ a la proyección sobre $\mathcal{X}$ ($P_{\mathcal{X}}(z) = \arg\min_{x\in\mathcal{X}} \|z - x\|$), entonces la condición a) es equivalente a que $x^* = P_{\mathcal{X}}(x^* - \alpha\nabla f(x^*))$ para todo $\alpha \geqslant 0$. Esto es: si estando en un mínimo local $x^*$, nos movemos cualquier distancia $\alpha$ en la dirección de máximo descenso, y proyectamos de nuevo al conjunto $\mathcal{X}$, entonces

volvemos a caer en $x^*$. Esta interpretación está muy relacionada con la forma de algunos algoritmos iterativos que comentaremos más adelante.

Para poder extender estas condiciones a funciones no diferenciables (y darles una interpretación geométrica), necesitamos una extensión del concepto de diferencial. Haremos esto para funciones convexas:

**Definición.** Dada una función $f : \mathbb{R}^n \to \mathbb{R}$ convexa y $x \in \mathbb{R}^n$, decimos que $d \in \mathbb{R}^n$ es un subgradiente de $f$ en $x$ sii $f(z) \geqslant f(x) + \langle z - x, d\rangle$ para todo $z \in \mathbb{R}^n$.

Esto es una generalización del concepto de plano tangente en un punto: $d$ es un subgradiente si el plano que define queda completamente por debajo de la función. En la figura 3 se puede ver una interpretación geométrica en el caso unidimensional.

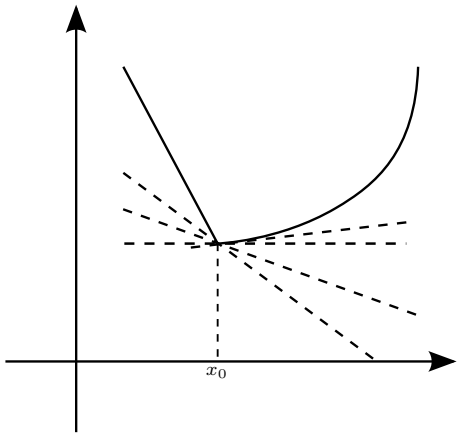



FIGURA 3. Función no diferenciable en $x_0$, con rectas que dejan el gráfico de la función completamente por arriba (subgradientes).

Cuando $f$ es diferenciable en un punto, entonces el único subgradiente posible es el gradiente tradicional. En los puntos de no diferenciabilidad, puede haber más de un subradiente,como se ve en la figura. Al conjunto de todos los subgradientes de $f$ en un punto $x$ se denomina subdiferencial, y se denota como $\partial f(x)$. Para la función $f(x) = |x|$, tenemos por ejemplo $\partial f(1) = \{1\}$ y $\partial f(0) = [-1, 1]$. Es decir, todas las rectas por el origen con pendiente entre $-1$ y $1$, dejan al gráfico de $f$ por arriba.

Ahora podemos dar una condición de optimalidad más general:

**Proposición 2.** Sea $f : \mathcal{X} \to \mathbb{R}$ convexa, donde $\mathcal{X}$ es un subconjunto convexo de $\mathbb{R}^n$. Entonces $x^*$ minimiza $f$ sii existe un subgradiente $d \in \partial f(x^*)$ tal que $\langle d, z - x^*\rangle \geqslant 0$ para todo $z \in \mathcal{X}$.

Observar que si $f$ es diferenciable, obtenemos la condición anterior, y que si $x^*$ está en el interior de $\mathcal{X}$, la condición queda $0 \in \partial f(x^*)$, que generaliza la condición de anular el gradiente.

Veamos una interpretación geométrica de las condiciones de optimalidad. Para esto, definamos antes el cono normal:

**Definición.** Dado un conjunto convexo $\mathcal{X}$ y un punto $x \in \mathcal{X}$, llamamos cono normal a $\mathcal{X}$ por $x$ al conjunto

$$N_{\mathcal{X}}(x) = \{g \in \mathbb{R}^n : \langle g, (z - x)\rangle \leqslant 0, \forall z \in \mathcal{X}\}.$$

Es decir, son las direcciones $g$ que forman un ángulo de al menos $\pi/2$ con las direcciones factibles $z - x$. Dicho de otra manera, son los vectores $g$ que definen un hiperplano que soporta al conjunto $\mathcal{X}$ en $x$.

Esto permite generalizar el concepto que el gradiente es normal a las curvas de nivel. Si $f$ es una función convexa (no necesariamente diferenciable), fijamos $x_0 \in \mathbb{R}^n$, y consideramos el conjunto de nivel $C = \{z \in \mathbb{R}^n : f(x) \leqslant f(x_0)\}$. Entones es fácil ver que la condición para que $g \in \partial f(x_0)$, es equivalente a que $g$ pertenezca al cono normal al conjunto de nivel por $x_0$: $g \in N_C(x_0)$. Si el borde de $C$ es diferenciable, entones el cono normal es simplemente la semirrecta perpendicular al borde, por lo que en este caso, recuperamos lo sabido del gradiente y las curvas de nivel.

Volvamos ahora a las condiciones de optimalidad. La condición dada en la proposición 2 (que exista un subgradiente $d$ tal que $\langle d, z - x^* \rangle \geqslant 0$ para todo $z \in \mathcal{X}$), se traduce en que exista un subgradiente $d \in \partial f(x^*)$ tal que $-d \in N_{\mathcal{X}}(x^*)$. Veamos un par de ejemplos.

Consideremos primero una función $f$ diferenciable, pero un conjunto $\mathcal{X}$ convexo, con borde no diferenciable. Como $f$ es diferenciable (y por lo tanto el único subgradiente es el gradiente), la condición de optimalidad es $-\nabla f(x^*) \in N_{\mathcal{X}}(x^*)$. Este es el caso que se puede ver en la figura 4.

Si ahora $f$ es no diferenciable, pero asumimos que el mínimo se alcanza en un punto $x^*$ donde el borde del conjunto $\mathcal{X}$ es diferenciable, entonces el cono normal a $\mathcal{X}$ será simplemente una semirrecta, a la cual deberá pertenecer el opuesto de algún subgradiente de $f$. Este es el caso de la figura 5. Se puede pensar, como ejercicio, cómo es geométricamente la condición de optimalidad cuando la función es no diferenciable y el mínimo se alcanza en un punto de no diferenciabilidad del borde de $\mathcal{X}$.

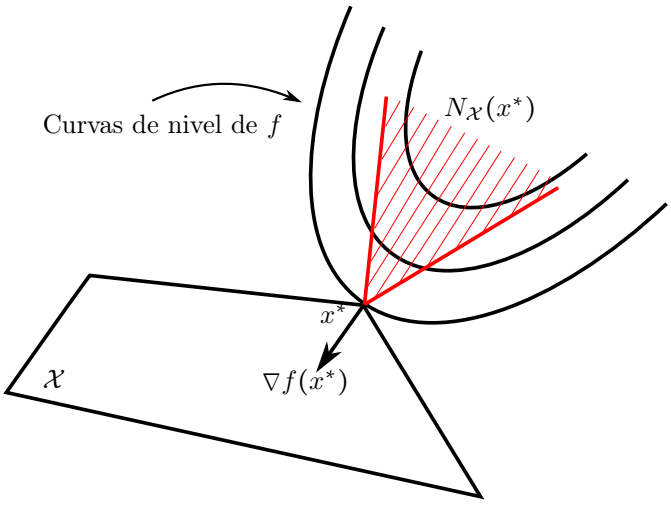



FIGURA 4. Curvas de nivel de una función diferenciable, y el mínimo se alcanza en un vértice de $\mathcal{X}$. En rojo, el cono normal a $\mathcal{X}$ por $x^*$, y se puede observar que el opuesto del gradiente pertenece al cono normal.

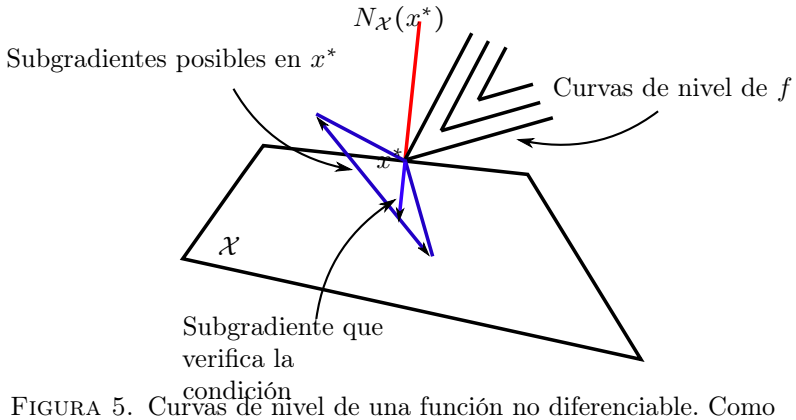



FIGURA 5. Curvas de nivel de una función no diferenciable. Como el mínimo se alcanza en un punto de diferenciabilidad del borde de $\mathcal{X}$, el cono normal es una semirrecta (en rojo). En azul, los posibles subgradientes de $f$ en $x^*$ (observar que forman un ángulo de $\pi/2$ con la curva de nivel por $x^*$), y el subgradiente cuyo opuesto pertenece al cono normal.

## 4. Métodos iterativos

**4.1. Optimización sin restricciones.** Hay una gran cantidad de métodos para resolver problemas de optimización, algunos bastante generales, y otros que dependen de la estructura del problema. La gran mayoría de los métodos son iterativos, y muchos de ellos son además de la siguiente forma: se comienza desde un punto $x_0$, y luego se construye una sucesión mediante $x_{k+1} = T(x_k)$, donde $T$ es un operador que depende obviamente del problema. Este operador $T$ puede incluir por ejemplo información de derivadas de distintos órdenes de $f$, o no. De esto depende, entre otras cosas, la complejidad (la cantidad de cálculos que hay que hacer para pasar de $x_k$ a $x_{k+1}$), pero también la velocidad de convergencia, es decir, cuántas iteraciones $k$ se necesitan para llegar a estar "cerca" de la solución.

El primer ejemplo es siempre el descenso por gradiente, o métodos relacionados. Consideremos primero que queremos minimizar una función diferenciable $f : \mathbb{R}^n \to \mathbb{R}$ (es decir, el conjunto $\mathcal{X}$ es todo el espacio, lo que usualmente se denomina como optimización sin restricciones). Lo que buscamos son puntos estacionarios, donde el gradiente de $f$ se anule. Entonces, si en determinado momento estamos parados en un punto $x_k$, y su gradiente es no nulo, significa que (al menos) en la dirección $-\nabla f(x_k)$, la función es decreciente. Más aún, es la dirección de máximo decrecimiento de $f$, localmente. Por lo tanto tiene sentido moverse un poco en esa dirección, para calcular un nuevo punto $x_{k+1}$, que será mejor que $x_k$, en el sentido que su valor funcional será menor. Esto es:

$$x_{k+1} = x_k - \alpha \nabla f(x_k),$$

donde $\alpha$ es el "poco" que nos queremos mover. Este valor de $\alpha$ hay que elegirlo en cada iteración $k$, de manera de asegurarse (al menos) que $f(x_{k+1}) < f(x_k)$.[3] Una

[3]Tenemos garantizado, puesto que la derivada de $f$ en la dirección $-\nabla f(x_k)$ es negativa, que existe un entorno $(0, \varepsilon)$ tal que para todo $\alpha$ en ese entorno, se tiene $f(x_{k+1}) < f(x_k)$.

opción es elegir el "mejor" valor de $\alpha$, en el sentido que sea el que lleve a $f(x_{k+1})$ a ser el más chico posible (recordemos que la dirección ya está elegida)[4].

Surgen varias preguntas. Por ejemplo, con alguna elección "razonable" del paso $\alpha$ en cada iteración, la sucesión generada $x_k$, ¿converge a un punto estacionario? ¿Hay alguna dirección mejor que la de máximo descenso?

La respuesta a ambas preguntas es positiva.

En el caso de la convergencia, pidiendo algunas hipótesis más, se puede ver que hay valores de $\alpha$ fijos, constantes para toda iteración (es decir, que no hay que elegirlos en cada $k$), que garantizan convergencia de $x_k$ a un punto estacionario.

Un algoritmo aceptable entonces para hallar puntos estacionarios (y si la función es convexa, la solución del problema de optimización), sería algo así:

**Input:** punto inicial $x_0$, paso $\bar{\alpha}$, tolerancia $tol$
**Output:** punto $x_k$ con $\|\nabla f(x_k)\| < tol$
**while** $\|\nabla f(x_k)\| \geqslant tol$ **do**
  Calcular $\nabla f(x_k)$;
  Hacer $x_{k+1} = x_k - \bar{\alpha}\nabla f(x_k)$
**end**

**Algoritmo 1:** Descenso por gradiente

Sobre la segunda pregunta. ¿Por qué habría mejores direcciones que la dirección de máximo decenso?. Localmente, mirar hacia $-\nabla f$ es lo mejor que podemos hacer. Sin embargo, cuando las curvas de nivel son "estiradas", se produce un zigzagueo no deseado:

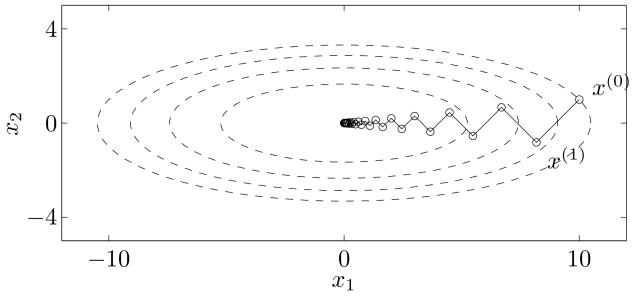



FIGURA 6. Descenso por gradiente para la función $f(x_1, x_2) = x_1^2 + 10x_2^2$. En punteado las curvas de nivel. Figura tomada de [8].

La información de esta geometría está en las derivadas segundas, por lo tanto debemos incorporar estas derivadas para lograr mejores trayectorias. Esta mejora viene de la mano de un mayor costo computacional: hay que calcular derivadas primeras y derivadas segundas, y eventualmente invertir una matriz.

El método de segundo orden más clásico es el de Newton, que consiste en corregir parcialmente la dirección del gradiente, usando información de segundo orden. Específicamente, el método consiste en iterar:

$$x_{k+1} = x_k - \alpha \left(\nabla^2 f(x_k)\right)^{-1} \nabla f(x_k),$$

[4]Observar que esto implica resolver otro problema de optimización, esta vez en $\mathbb{R}$. Es decir, hallar el valor de $\alpha$ que minimice $f(x_k - \alpha\nabla f(x_k))$. Hay soluciones eficientes que no requieren resolver esta minimización, como por ejemplo el método de Armijo (ver en [6]).

donde $\nabla^2 f(x_k)$ es la matriz Hessiana. Esta iteración surge de considerar la aproximación de Taylor de segundo orden alrededor de $x_k$, y hallar el mínimo de esa función.

El método de Newton requiere calcular muchas derivadas más, pero da lugar a convergencias más rápidas. Existen métodos intermedios (considerando la diagonal de la Hessiana solamente, o aproximaciones similares), que manejan un compromiso entre la convergenia y la cantidad de cálculos.

De la misma manera, hay mejoras al método de Newton, algunos que tienen en cuenta no solo $x_k$ sino también el punto anterior $x_{k-1}$, haciendo uso de la historia de la iteración, o también considerando derivadas de orden superior. En cualquier caso, a no ser que los métodos sean el objeto principal de estudio, o se requiera una convergencia ridículamente rápida, en general, para minimizar funciones diferenciables en $\mathbb{R}^n$, basta con el método de descenso por gradiente o (algún) método de Newton.

**4.2. Optimización con restricciones.** Cuando se tienen algunas restricciones en la optimización (es decir, que el conjunto $\mathcal{X}$ no es todo el espacio), hay que hacer modificaciones a los métodos presentados. Describiremos aquí solamente uno de ellos.

Supongamos entonces que tenemos una función $f : \mathcal{X} \to \mathbb{R}$, con $\mathcal{X} \subset \mathbb{R}^n$ convexo, y partimos de cierto punto $x_0 \in \mathcal{X}$ como antes. El objetivo es construir una sucesión $x_k$ que converja a un mínimo local (o al menos a un punto que satisfaga la condición de optimalidad).

En cualquier momento, estando en $x_k$, podemos calcular una dirección de descenso (la dirección opuesta al gradiente por ejemplo), e intentar movenos en esa dirección un paso $\alpha$. El problema es que este movimiento nos puede llevar fuera del conjunto factible $\mathcal{X}$. Para asegurarnos que esto no pasa, y aprovechando la convexidad de $\mathcal{X}$, la dirección en la que nos moveremos estará dada por $\bar{x} - x_k$ para algún $\bar{x} \in \mathcal{X}$. Es decir, estando en $x_k$, nos moveremos en dirección a $\bar{x}$.

Estos métodos se pueden escribir entonces de forma general como:

$$x_{k+1} = x_k + \alpha(\bar{x} - x_k),$$

donde resta definir la elección de $\bar{x} \in \mathcal{X}$, y el paso $\alpha$.

Una opción, que da lugar al *Método de gradiente proyectado*, es tomar $\bar{x} = P_{\mathcal{X}}(x_k - s\nabla f(x_k))$. Es decir, nos movemos en la dirección opuesta al gradiente, y si caemos fuera de $\mathcal{X}$, lo proyectamos, y usamos este punto para la dirección. En particular si tomamos $\alpha = 1$ (que es una posible elección dentro de la familia del método de gradiente proyectado), la iteración resulta:

$$x_{k+1} = P_{\mathcal{X}}\left(x_k - s_k \nabla f(x_k)\right).$$

Como primer comentario sobre la efectividad de este método, basta decir que tenemos resultados de convergencia similares al descenso por gradiente clásico.

De todas maneras, hagamos un chequeo sobre el comportamiento del método. ¿Cuándo se estanca el método? Mirándolo en la iteración

$$x_{k+1} = x_k + \alpha(\bar{x} - x_k),$$

es claro que el algoritmo se estanca cuando $\bar{x} = x_k$. Esto, en este caso particular que estamos viendo aquí, significa que

$$x_k = P_{\mathcal{X}}\left(x_k - s\nabla f(x_k)\right),$$

que, si recordamos, es exactamente la condición de estacionareidad, equivalente a la condición de optimalidad que vimos al principio.

## 5. Algoritmos iterativos como discretizaciones de flujos

En la sección anterior vimos algoritmos iterativos, presentados directamente en su naturaleza discreta. Básicamente definimos las sucesiones expresando, a partir de la posición en determinado momento $x_k$, hacia dónde y cuánto nos movemos, para luego calcular el siguiente iterado $x_{k+1}$. Sin embargo, algunos de estos métodos pueden ser interpretados como discretizaciones de ecuaciones diferenciales convenientes.

Volvamos a considerar el problema de optimización sin restricciones (supongamos por ahora que $f$ es diferenciable y su gradiente es Lipschitz):

$$\arg\min_{x\in\mathbb{R}^n} f(x)$$

donde recordemos que buscamos un punto donde se anule el gradiente.

Es natural entonces pensar en el flujo gradiente:

$$\begin{cases} \dot{x}(t) = -\nabla f\left(x(t)\right) & \forall\, t > 0 \\ x(0) = x_0 \end{cases}$$

Claramente los puntos críticos de esta ecuación diferencial son los que buscamos, y además es fácil ver que si $x(t)$ es una solución, entonces $\frac{df(x(t))}{dt} = -\|\nabla f(x(t))\|^2 < 0$, y por lo tanto $f(x(t))$ es decreciente con $t$. Si además tenemos la hipótesis de convexidad, entonces $x(t)$ converge al mínimo global.

Si discretizamos la ecuación, aproximando la derivada por el cociente incremental en iterados sucesivos, tenemos:

$$\frac{x_{k+1} - x_k}{h} = -\nabla f(x_k).$$

Esta discretización, que corresponde al método de Euler (o de Euler explícito, o *forward Euler*), da lugar justamente a la iteración estudiada arriba como el método de descenso por gradiente. En efecto, resulta:

$$x_{k+1} = x_k - h\nabla f(x_k).$$

Hay otro método de discretización, que en general presenta mejores condiciones de estabilidad, llamado el método de Euler implícito (o *backward Euler*). La diferencia está en el punto donde se evalúa la derivada:

$$\frac{x_{k+1} - x_k}{h} = -\nabla f(x_{k+1}).$$

Si bien ahora no podemos despejar $x_{k+1}$ para poder entender directamente la iteración correspondiente, resulta que es exactamente la condición de optimalidad de la iteración $x_{k+1} = prox_{hf}(x_k)$, donde $prox$ es el denominado operador proximal, que es la base de varios métodos modernos de optimización, sobretodo para funciones no diferenciables. No comentaremos detalles de esto aquí, pero referimos a [6, 7] para encontrar una profundización del tema.

Volviendo al flujo gradiente, una propiedad interesante sobre la unicidad de la solución y la convergencia al mínimo global, es que vale también en casos bastante

más generales. Por ejemplo, si $f$ es una función convexa pero no necesariamente diferenciable, podemos estudiar la siguiente inclusión diferencial:

$$\begin{cases} \dot{x}(t) \in -\partial f\,(x(t)) & \text{para casi todo } t > 0 \\ x(0) = x_0 \end{cases}$$

donde $\partial f$ denota el subdiferencial de $f$ como fue definido anteriormente.

Se puede probar fácilmente la siguiente

**Proposición 3.** Sean $x_1$ y $x_2$ dos soluciones de la inclusión diferencial, entonces tenemos que para todo $t$ se cumple $|x_1(t) - x_2(t)| \leqslant |x_1(0) - x_2(0)|$. En particular, esto implica la unicidad de la solución.

La discretización de esta inclusión diferencial con el método *forward Euler* da lugar al método proximal para funciones no diferenciables.

Sobre flujos gradiente en espacios más generales, se puede leer en [21].

## 6. Optimización en variedades

En la sección 4.2 vimos un ejemplo para optimizar una función $f$ sobre un conjunto $\mathcal{X} \subset \mathbb{R}^n$, pero que no utilizaba la estructura o la geometría de $\mathcal{X}$ (más que para la proyección sobre el conjunto). En muchos casos, el conjunto $\mathcal{X}$ tiene una estructura diferenciable que puede ayudarnos para diseñar métodos de optimización, que por algún motivo pueden ser más convenientes.

Hay muchos ejemplos de esto, pero en particular los problemas de las secciones 2.2 y 2.3 responden a esta situación.

En efecto, en el problema de isomorfismo de grafos (o de automorfismo de grafos) buscamos una matriz de permutación $P$ que conjugue $A$ con $B$, es decir que $A = PBP^T$. Ahora, las matrices de permutación son en particular ortogonales[5], y por lo tanto es interesante buscar matrices $Q$ ortogonales tales que $A = QBQ^T$, o equivalentemente, estudiar el siguiente problema:

$$\arg \min_{Q \in \mathcal{O}(n)} \|A - QBQ^T\|_F^2,$$

donde $\mathcal{O}(n)$ es el grupo ortogonal. Por supuesto que esto no soluciona el problema de isomorfismo de grafos (pues el resultado no necesariamente es una matriz de permutación), pero de todas maneras es útil.

En el caso del problema de valores y vectores propios sucede algo similar. Dada una matriz $A$ a diagonalizar, se desea encontrar una matriz ortogonal que resuelva

$$min_{U \in \mathcal{O}(n)} \|N - U^T AU\|_F^2$$

donde $N$ es una matriz diagonal, por ejemplo $N = diag(1, 2, 3, \ldots, n)$.

Como fue mencionado arriba, una opción es considerar la función a minimzar $f : \mathbb{R}^{n \times n} \to \mathbb{R}$, y un conjunto $\mathcal{X} = \{U \in \mathbb{R}^{n \times n} : UU^T = Id\}$, y utilizar un método de optimización con restricciones como por ejemplo: $U_{k+1} = P_{\mathcal{X}}\,(U_k - \alpha \nabla f(U_k))$.

Otra opción es considerar directamente $f : \mathcal{O}(n) \to \mathbb{R}$. La función $f$ (cualquiera de las presentadas como ejemplo) es diferenciable, y como $\mathcal{O}(n)$ tiene una estructura de variedad Riemanniana, podemos definir el flujo gradiente asociado a $f$ sobre $\mathcal{O}(n)$:

$$\dot{U}(t) = -\operatorname{grad} f(U(t)), \tag{9}$$

[5]De hecho, son la intersección delas matrices ortogonales con las de entradas no negativas.

donde grad $f$ es el gradiente de $f$ definido por la métrica en $\mathcal{O}(n)$[6]. Se tiene por lo tanto que si $U(t)$ es solución de (9), entonces $U(t) \in \mathcal{O}(n)$ para todo $t$.

Se puede estudiar entonces la dinámica de (9), y obtener resultados como por ejemplo (para el problema de valores propios) que el flujo es completo, que converge a puntos de equilibrio, y que hay un abierto denso de condiciones iniciales tales que el flujo converge a la solución deseada. Ver por ejemplo [3].

Ahora, esto define un flujo continuo, que hay que discretizar para poder implementarlo. Como antes, hay que tener ciertos cuidados a la hora de discretizar. Por ejemplo, cuando en determinado momento se está en un punto $U_k$, y se desea mover una cantidad $\alpha$ en la dirección $-\operatorname{grad} f(U_k)$, que es un elemento del espacio tangente, en general se caerá fuera de la variedad, por lo que hay que proyectar de nuevo a la misma. Una forma eficiente de hacer esto es mediante las llamadas retracciones.

Se puede encontrar un estudio detallado de esta teoría en [1, 18], y en las notas de curso [3].

## Referencias


[1] P.A. Absil, R. Mahony, and R. Sepulchre, *Optimization algorithms on matrix manifolds*, Princeton University Press, Princeton, NJ, USA, 2007.

[2] Y. Aflalo, A. Bronstein, and R. Kimmel, *Graph matching: relax or not?*, arXiv:1401.7623 (2014).

[3] Diego Armentano, *Notas de curso: Optimización en variedades*, 2017. `http://www.cmat.edu.uy/~diego/documents/notas-opt-var-2017.pdf`.

[4] László Babai, *Graph isomorphism in quasipolynomial time*, arXiv preprint arXiv:1512.03547 (2016).

[5] O. Banerjee, L. El Ghaoui, A. D'Aspremont, and G. Natsoulis, *Convex optimization techniques for fitting sparse Gaussian graphical models*, Proceedings of the 23rd international conference on Machine learning - ICML '06 (2006), 89–96.

[6] Dimitri P Bertsekas, *Nonlinear programming*, Athena scientific Belmont, 1999.

[7] ———, *Convex optimization algorithms*, Athena Scientific Belmont, 2015.

[8] Stephen Boyd and Lieven Vandenberghe, *Convex optimization*, Cambridge university press, 2004.

[9] Alfred M Bruckstein, David L Donoho, and Michael Elad, *From sparse solutions of systems of equations to sparse modeling of signals and images*, SIAM review **51** (2009), no. 1, 34–81.

[10] Emmanuel J Candès and Benjamin Recht, *Exact matrix completion via convex optimization*, Foundations of Computational mathematics **9** (2009), no. 6, 717.

[11] Emmanuel J Candès, Justin Romberg, and Terence Tao, *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information*, IEEE Transactions on information theory **52** (2006), no. 2, 489–509.

[12] Emmanuel J Candès and Terence Tao, *The power of convex relaxation: Near-optimal matrix completion*, IEEE Transactions on Information Theory **56** (2010), no. 5, 2053–2080.

[13] A. Dempster, *Covariance selection*, Biometrics **28** (1972), no. 1, 157–175.

[14] Nadav Dym, Haggai Maron, and Yaron Lipman, *Ds++: A flexible, scalable and provably tight relaxation for matching problems*, ACM Trans. Graph. **36** (November 2017), no. 6, 184:1–184:14.

[15] Marcelo Fiori, Pablo Musé, Ahamd Hariri, and Guillermo Sapiro, *Multimodal graphical models via group lasso*, Signal Processing with Adaptive Sparse Structured Representations (2013).


[6] Es decir: sea $(M, \langle\cdot,\cdot,\rangle_x)$ una variedad Riemanniana, y $f : M \to \mathbb{R}$ una función diferenciable. Definimos el *gradiente* de $f$ como el vector tangente $\operatorname{grad} f(x) \in T_x M$, $x \in M$, que satisface

$$Df(x)v_x = \langle \operatorname{grad} f(x), v_x \rangle, \quad \forall\, v_x \in T_x M.$$

[16] Marcelo Fiori and Guillermo Sapiro, *On spectral properties for graph matching and graph isomorphism problems*, Information and Inference: A Journal of the IMA **4** (2015), no. 1, 63–76.

[17] Marcelo Fiori, Pablo Sprechmann, J.T. Vogelstein, Pablo Musé, and Guillermo Sapiro, *Robust multimodal graph matching: Sparse coding meets graph matching*, Advances in Neural Information Processing Systems 26 (2013), 127–135.

[18] Uwe Helmke and John B Moore, *Optimization and dynamical systems*, Springer Science & Business Media, 2012.

[19] V. Lyzinski, D. Fishkind, M. Fiori, J.T. Vogelstein, C.E. Priebe, and G. Sapiro, *Graph matching: Relax at your own risk*, arXiv preprint arXiv:1405.3133 (2014).

[20] Julien Mairal, Francis Bach, and Jean Ponce, *Sparse modeling for image and vision processing*, Foundations and Trends® in Computer Graphics and Vision **8** (2014), no. 2-3, 85–283.

[21] Filippo Santambrogio, {*Euclidean, metric, and Wasserstein*} *gradient flows: an overview*, Bulletin of Mathematical Sciences **7** (2017), no. 1, 87–154.

[22] Steve Smale, *Mathematical problems for the next century*, Mathematical Intelligencer **20** (1998), 7–15.

IMERL, Facultad de Ingeniería, Universidad de la República. Montevideo, Uruguay.
*Email address*: mfiori@fing.edu.uy

MINICURSO

# INTRODUCCIÓN A LAS ÁLGEBRAS DE LIE

ANA GONZÁLEZ

Publicaciones Matemáticas del Uruguay
Volumen 17, Julio 2019, Páginas 25–39
S 0797-1443

# INTRODUCCIÓN A LAS ÁLGEBRAS DE LIE

ANA GONZÁLEZ

## Introducción

Estas notas corresponden al mini-curso "Introducción a las álgebras de Lie" del 6to Coloquio Uruguayo de Matemática. El curso consiste en una rápida introducción a las álgebras de Lie, presentando los conceptos necesarios para realizar la clasificación de las mismas en dimensiones bajas. La determinación de listas completas de álgebras de Lie es un problema cuya complejidad aumenta a medida que se incrementa la dimensión del álgebra. Por ejemplo, la clasificación de las álgebras de Lie nilpotentes es un problema abierto que ha sido tratado por diversos autores.

La teoría de Lie tiene múltiples aplicaciones a otras ciencias, tales como Física, Ingeniería, Economía y Finanzas. Por ejemplo, en Física, los grupos y las álgebras de Lie son muy utilizados como herramientas en el estudio de las simetrías, no sólo de las clásicas en el espacio-tiempo, sino también en la teoría de super-cuerdas.

A continuación detallaremos como será organizado el curso. Comenzaremos construyendo la categoría de álgebras de Lie: (álgebras de Lie y los morfismos entre ellas). Luego veremos una serie de ejemplos que nos permitirán familiarizarnos con estos conceptos. Continuaremos con costrucciones algebraicas que serán de utilidad para la clasificación en dimensiones bajas y finalizaremos presentando la lista completa de álgebras de Lie de dimensión 1, 2 y 3. Para culminar daremos un breve recorrido por algunos problemas en economía y finanzas que usan teoría de Lie y desarrollaremos uno de estos ejemplos para familiarizar al lector con estas aplicaciones.

El objetivo de este curso es motivacional. Por lo cual será muy incompleto y se omitirán algunas de las demostraciones. Para cubrir esto se incluyen varias referencias, que ayudan a profundizar las ideas que acá se presentan.

## 1. Conceptos básicos

En esta sección desarrollaremos los conceptos básicos de la teoría de álgebras de Lie. Conceptos y los resultados necesarios para comprender la clasificación que se realiza en la sección 2 y el ejemplo desarrollado en la sección 3.

**Definición 1.** *Sea* $\Bbbk$ *un cuerpo. Decimos que* $L$ *es un* álgebra de Lie *sobre* $\Bbbk$ *si es un* $\Bbbk$*-espacio vectorial y admite un producto (bilineal), llamado corchete de Lie*

$$\begin{array}{rccc} [\,,\,]: & L\times L & \to & L \\ & (x,y) & \mapsto & [x,y] \end{array}$$

*Que satisface las siguientes propiedades*

1. $[x,x]=0$, $\forall x\in L$.

©2019 PMU

2. $[x,[y,z]]+[y,[z,x]]+[z,[x,y]]=0$, $\forall x,y\in L$. *Esta propiedad se conoce como* condición de Jacobi.

*Un álgebra de Lie se dice* abeliana *si* $\forall x,y\in L$ *se tiene que el corchete de Lie*

$$[x,y]=0.$$

*Observación* 2.

$$0=[x+y,x+y]=[x,x]+[x,y]+[y,x]+[y,y]=[x,y]+[y,x]\Rightarrow[x,y]=-[y,x].$$

Entonces, la propiedad (1) puede sustituirse por la siguiente propiedad:
(1') $[x,y]=-[y,x]$, $\forall x,y\in L$.

*Observación* 3. Supongamos que el álgebra es de dimensión finita y consideremos $B=\{e_1,\dots,e_n\}$ una base de la misma (como espacio vectorial sobre $\Bbbk$), por lo cual todo elemento $x\in L$ puede escribirse como

$$x=\sum_{i=1}^{n}\xi_i e_i,\quad \xi_i\in\Bbbk.$$

Para probar que un espacio vectorial es un álgebra de Lie basta verificar que los elementos de la base satisfacen las condiciones de Lie.
Veamos sólo la condición (1). La condición (2) se prueba de forma similar pero los cálculos son más extensos.

$$[x,x]=\left[\sum_{i=1}^{n}\xi_i e_i,\sum_{j=1}^{n}\xi_j e_j\right]=\sum_{i,j=1}^{n}\xi_i\xi_j\big[e_i,e_j\big]=0$$
$$\Rightarrow[x,x]=0,\ \forall x\in L.$$

**Definición 4** (**Estructura constante**). *Si* $L$ *es un álgebra de Lie sobre* $\Bbbk$ *con base* $\mathcal{B}=\{e_1,\dots,e_n\}$ *entonces* $[\,,\,]$ *queda completamente determinado si conocemos los productos* $[e_i,e_j]$.
*Definamos escalares* $a_{ij}^k\in\Bbbk$ *tales que*

$$[e_i,e_j]=\sum_{k=1}^{n}a_{ij}^k e_k.$$

*Estos escalares se llaman* estructura constante *de* $L$ *con respecto a la base* $\mathcal{B}$. *Diferentes bases pueden dar, en general, diferentes estructuras constantes.*

Por las propiedades (1) y (1') es suficiente conocer la estructura constante $a_{ij}^k$, $1\le i<j\le n$ para tener determinada la estructura de Lie sobre $L$.

**Ejemplo 1.1.** *Consideremos el cuerpo de los reales* $\mathbb{R}$ *y el espacio vectorial* $\mathbb{R}^3$ *sobre* $\mathbb{R}$. *Observemos que el producto vectorial*

$$\begin{array}{rccc}\wedge: & \mathbb{R}^3\times\mathbb{R}^3 & \to & \mathbb{R}^3\\ & (x,y) & \mapsto & x\wedge y\end{array}$$

*define una estructura de Lie sobre* $\mathbb{R}^3$.

*Recordemos que* $(x_1,x_2,x_3)\wedge(y_1,y_2,y_3)=(x_2y_3-x_3y_2,x_3y_1-x_1y_3,x_1y_2-x_2y_1)$. *Sabemos que* $\wedge$ *es una función bilineal. Veamos que verifica las condiciones de Lie.*

(a) $x\wedge x=(x_2x_3-x_3x_2,x_3x_1-x_1x_3,x_1x_2-x_2x_1)=(0,0,0)$, *para todo* $x\in\mathbb{R}^3$.

(b) *Condición de Jacobi:*

$$[x,[y,z]] = x \wedge (y \wedge z) = \big(x_2y_1z_2 - x_2y_2z_1 - x_3y_3z_1 + x_3y_1z_3, x_3y_2z_3 - x_3y_3z_2 - x_1y_1z_2 + x_1y_2z_1, x_1y_3z_1 - x_1y_1z_3 - x_2y_2z_3 + x_2y_3z_2\big)$$

$$[y,[z,x]] = y \wedge (z \wedge x) = \big(y_2z_1x_2 - y_2z_2x_1 - y_3z_3x_1 + y_3z_1x_3, y_3z_2x_3 - y_3z_3x_2 - y_1z_1x_2 + y_1z_2x_1, y_1z_3x_1 - y_1z_1x_3 - y_2z_2x_3 + y_2z_3x_2\big)$$

$$[z,[x,y]] = z \wedge (x \wedge y) = \big(z_2x_1y_2 - z_2x_2y_1 - z_3x_3y_1 + z_3x_1y_3, z_3x_2y_3 - z_3x_3y_2 - z_1x_1y_2 + z_1x_2y_1, z_1x_3y_1 - z_1x_1y_3 - z_2x_2y_3 + z_2x_3y_2\big)$$

*Sumando obtenemos que*

$$[x,[y,z]] + [y,[z,x]] + [z,[x,y]] = 0.$$

**Ejemplo 1.2.** *Cualquier espacio vectorial $V$ tiene un corchete de Lie dado por*

$$[x,y] = 0, \quad \forall x, y \in V.$$

*Esta es una estructura de Lie abeliana sobre $V$.*
*En particular el cuerpo puede considerarse como un álgebra de Lie de dimensión 1 con este corchete.*

**Ejemplo 1.3.** *Sea $V$ un espacio vectorial de dimensión finita sobre $\Bbbk$. Denotamos por $\mathrm{gl}(V)$ el conjunto de todas las transformaciones lineales de $V$ en $V$. Este es de nuevo un espacio vectorial sobre $\Bbbk$ y podemos definir un corchete que le de estructura de álgebra de Lie de la siguiente forma:*

$$\begin{array}{rccc} [\,,\,]: & \mathrm{gl}(V)\times\mathrm{gl}(V) & \to & \mathrm{gl}(V) \\ & (T,S) & \mapsto & T\circ S - S\circ T \end{array}$$

*Es claro que la composición de transformaciones lineales es una operación bilineal. Veamos las condiciones de Lie.*

$$[T,T] = TT - TT = 0,$$

$$[T,[S,U]] + [S,[U,T]] + [U,[T,S]]$$

$$=$$

$$T(SU-US)-(SU-US)T+S(UT-TU)-(UT-TU)S+U(TS-ST)-(TS-ST)U$$

$$=$$

$$TSU-TUS-SUT+UST+SUT-STU-UTS+TUS+UTS-UST-TSU+STU$$

$$= 0.$$

**Ejemplo 1.4.** *Consideremos el espacio vectorial $\mathrm{gl}(n,\Bbbk)$ de las matrices cuadradas sobre $\Bbbk$. En este espacio vectorial también podemos definir un corchete de Lie dado por:*

$$\begin{array}{rccc} [\,,\,]: & \mathrm{gl}(n,\Bbbk)\times\mathrm{gl}(n,\Bbbk) & \to & \mathrm{gl}(n,\Bbbk). \\ & (M,N) & \mapsto & MN - NM \end{array}$$

*Una forma de probar que es un álgebra de Lie es considerar la base canónica de las matrices $\{E_{ij}, 1 \le i,j \le n\}$ y verificar (1) y (2) sobre estas matrices. Para ello basta observar que*

$$\big[E_{ij}, E_{kl}\big] = \delta_{jk}E_{il} - \delta_{il}E_{kj}.$$

**Ejemplo 1.5.** *Sea* $\mathrm{sl}(n,\Bbbk)$ *el subespacio de* $\mathrm{gl}(n,\Bbbk)$ *de todas las matrices de traza cero. Para matrices cuadradas arbitrarias* $M, N$*, la matriz* $MN - NM$ *tiene traza* $0$*, así* $[M,N] = MN - NM$*, define una estructura de Lie sobre* $\mathrm{sl}(n,\Bbbk)$*. Esta álgebra de Lie es conocida como* álgebra lineal especial.

**Ejemplo 1.6.** *Sea* $\mathrm{b}(n,\Bbbk)$ *el conjunto de las matrices triangulares superiores en* $\mathrm{gl}(n,\Bbbk)$*. Esta es un álgebra de Lie con el mismo corchete de Lie de* $\mathrm{gl}(n,\Bbbk)$*. Similarmente,* $\mathrm{n}(n,\Bbbk)$ *el conjunto de matrices estríctamente triangulares superiores en* $\mathrm{gl}(n,\Bbbk)$ *es un álgebra de Lie con el mismo corchete que* $\mathrm{gl}(n,\Bbbk)$*.*

### 1.1. Elementos de la Teoría de álgebras de Lie.

Los dos ejemplos presentados anteriormente sugieren la existencia de la noción de subálgebra de Lie. Veamos a continuación la definición.

**Definición 5.** *Sea* $L$ *un álgebra de Lie, decimos que* $S$ *es una* subálgebra de Lie *de* $L$ *si es un subespacio vectorial* $S \subset L$ *tal que*

$$[x,y] \in S, \quad \forall x, y \in S.$$

**Definición 6.** *Un* ideal $I$ *de un álgebra de Lie* $L$ *es un subespacio de* $L$ *tal que:*

$$[x,y] \in I, \quad \forall x \in L, y \in I.$$

*Observación* 7. Por la condición (1'), $[x,y] = -[y,x]$, así que no necesitamos distinguir entre ideales a izquierda o a derecha. Por ejemplo $\mathrm{sl}(n,\Bbbk)$ es un ideal de $\mathrm{gl}(n,\Bbbk)$ y $\mathrm{n}(n,\Bbbk)$ es un ideal de $\mathrm{b}(n,\Bbbk)$.

Un ideal siempre es una subálgebra. Pero, una subálgebra no necesariamente es un ideal. Por ejemplo $\mathrm{b}(n,\Bbbk)$ es una subálgebra de $\mathrm{gl}(n,\Bbbk)$, pero si $n > 2$, no es un ideal. Para ver esto notar que $E_{11} \in \mathrm{b}(n,\Bbbk)$ y $E_{21} \in \mathrm{gl}(n,\Bbbk)$ y

$$[E_{21}, E_{11}] = E_{21} \notin \mathrm{b}(n,\Bbbk).$$

Un álgebra de Lie es por sí misma un ideal de $L$, porque $\forall x, y \in L$ se tiene que $[x,y] \in L$. Así como también $\{0\}$ es un ideal de $L$, a estos ideales se los llama *ideales triviales* de $L$.

Un ejemplo importante de un ideal no-trivial es el centro de $L$, el cual definimos como sigue:

**Definición 8.** *Se define el* centro *de un álgebra de Lie como:*

$$Z(L) = \{x \in L : \ [x,y] = 0 \, \forall y \in L\}.$$

**Teorema 9.** *Sea* $L$ *un álgebra de Lie y* $Z(L)$ *su centro. Entonces*

$$L = Z(L) \Leftrightarrow L \text{ es abeliana.}$$

*Demostración.* Si $L = Z(L) \Rightarrow [x,y] = 0, \forall x, y \in L$ lo cual prueba que $L$ es abeliana. Recíprocamente, si $L$ es abeliana entonces $[x,y] = 0$ para todo $x, y \in L$, por lo tanto $x \in Z(L)$ para todo $x \in L$. Es decir, $L = Z(L)$. □

**Definición 10 (Homomorfismos entre álgebras de Lie).** *Sean* $L_1$ *y* $L_2$ *dos álgebras de Lie sobre el mismo cuerpo* $\Bbbk$*, se dice que una función* $\varphi : L_1 \to L_2$ *es un* homomorfismo de Lie *si* $\varphi$ *es una transformación lineal tal que*

$$\varphi\big([x,y]\big) = \big[\varphi(x), \varphi(y)\big], \quad \forall x, y \in L_1.$$

*Decimos que* $\varphi$ *es un* isomorfismo *si* $\varphi$ *es además biyectiva.*

Un homomorfismo importante en la categoría de álgebras de Lie es el siguiente.

**Definición 11** (**Homomorfismo adjunto**). *Si $L$ es un álgebra de Lie el* homomorfismo adjunto *se define como:*

$$\begin{array}{rccll} \mathrm{ad}: & L & \to & \mathrm{gl}(L) & \\ & x & \mapsto & \mathrm{ad}(x): & L \to L \\ & & & & y \mapsto [x,y] \end{array}$$

Observar que ad es una transformación lineal. Es un homomorfismo de Lie ya que:

$$\mathrm{ad}\big([x,y]\big) = \big[\mathrm{ad}(x), \mathrm{ad}(y)\big] = \mathrm{ad}(x)\circ\mathrm{ad}(y) - \mathrm{ad}(y)\circ\mathrm{ad}(x).$$

$$\begin{array}{rcl} \big(\mathrm{ad}\big([x,y]\big)\big)(z) & = & [[x,y],z] = -[z,[x,y]] \\ & = & [x,[y,z]] + [y,[z,x]] \\ & = & \mathrm{ad}(x)([y,z]) + \mathrm{ad}(y)([z,x]) \\ & = & \mathrm{ad}(x)\big(\mathrm{ad}(y)(z)\big) + \mathrm{ad}(y)\big(-\mathrm{ad}(x)(z)\big) \\ & = & \big[\mathrm{ad}(x), \mathrm{ad}(y)\big](z). \end{array}$$

**Teorema 12.** *Sean $L_1$ y $L_2$ dos álgebras de Lie abelianas. $L_1$ y $L_2$ son isomorfas si y sólo si tienen la misma dimensión.*

*Demostración.* Si son isomorfas, en particular son isomorfas como espacios vectoriales, entonces tienen la misma dimensión.
Recíprocamente, si tienen la misma dimensión son isomorfas como espacios vectoriales. Además, como son abelianas $[x,y]=0$, $\forall x,y \in L_1$ y $[x',y']=0$, $\forall x',y' \in L_2$, entonces el isomorfismo preserva las estructuras de Lie. Es decir, es un isomorfismo de álgebras de Lie. □

**Definición 13** (**Álgebra derivada**). *Sea $L$ un álgebra de Lie, definimos el* álgebra derivada *de $L$, como la subálgebra $L' = [L,L]$. Es decir:*

$$L' = \mathrm{span}\big\{[x,y] : x,y \in L\big\} = \Big\{t = \sum c_i[x_i,y_i] : \ x_i, y_i \in L, c_i \in \Bbbk\Big\}.$$

**Ejemplo 1.7.** *Sea $L$ el álgebra de matrices de la forma*

$$L = \left\{ \begin{pmatrix} 0 & a \\ 0 & 0 \end{pmatrix} : a \in \mathbb{C} \right\}$$

*Si $M = \begin{pmatrix} 0 & a \\ 0 & 0 \end{pmatrix}$ y $N = \begin{pmatrix} 0 & b \\ 0 & 0 \end{pmatrix}$ entonces $[M,N] = MN - NM = 0$. Por lo tanto $L' = \{0\}$.*

**Definición 14** (**Álgebra de Lie cociente**). *Si $I$ es un ideal de un álgebra de Lie $L$, entonces es en particular un subespacio de $L$, así el espacio cociente $L/I$ se puede dotar de una estructura de álgebra de Lie de manera natural definiendo*

$$[w+I, z+I] = [w,z] + I, \quad \forall w,z \in L.$$

**Teorema 15** (**Teoremas de isomorfismos**). (a) *Sea $\varphi : L_1 \to L_2$ un homomorfismo de álgebras de Lie. Entonces, $N(\varphi)$ es un ideal de $L_1$, $Im(\varphi)$ es una subálgebra de $L_2$ y*

$$L_1/N(\varphi) \cong Im(\varphi).$$

(b) *Si $I$ y $J$ son ideales de un álgebra de Lie $L$ tales que $I \subset J$. Entonces $J/I$ es un ideal de $L/I$ y*

$$(L/I)/(J/I) \cong L/J.$$

## 1.2. Álgebras de Lie resolubles.

**Definición 16** (**Serie de derivadas**). *Se define la* serie de derivadas *de $L$ como la serie*

$$L \supseteq L^{(1)} \supseteq L^{(2)} \supseteq \cdots$$

*donde*

$$L^{(1)} = L' \text{ y } L^{(k)} = \left[L^{(k-1)}, L^{(k-1)}\right] \; k \geq 2.$$

**Ejemplo 1.8** (**Álgebra de Heisenberg**). *El álgebra de Heisenberg $\mathcal{H}_n$ es el álgebra de Lie real de dimensión $2n+1$ que tiene por base los elementos:*

$$\left\{P_1, \ldots, P_n, Q_1, \ldots, Q_n, C\right\}$$

*y el corchete de Lie definido por*

$$\left[P_i, P_j\right] = \left[Q_i, Q_j\right] = \left[P_i, C\right] = \left[Q_i, C\right] = [C, C] = 0, \quad \left[P_i, Q_j\right] = C.$$

*Estudiemos el caso particular de $n = 1$: consideremos la base $\left\{P, Q, C\right\}$ dada por*

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, Q = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, C = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

*Entonces una forma de definir la forma matricial de los elementos del álgebra es:*

$$pP + qQ + cC = \begin{pmatrix} 0 & p & c \\ 0 & 0 & q \\ 0 & 0 & 0 \end{pmatrix}.$$

*Tomando el corchete de Lie $[P, Q] = PQ - QP$, resulta evidente que $[P, Q] = C$ y $[P, C] = [Q, C] = 0$.*
*Observar que*

$$\left[\begin{pmatrix} 0 & p & c \\ 0 & 0 & q \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & p' & c' \\ 0 & 0 & q' \\ 0 & 0 & 0 \end{pmatrix}\right] = \begin{pmatrix} 0 & 0 & pq' - qp' \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

*Entonces* $\mathcal{H}_1' = \left\{ \begin{pmatrix} 0 & 0 & \alpha \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} : \alpha \in \mathbb{C} \right\}$.

*Ahora* $\left[\begin{pmatrix} 0 & 0 & \alpha \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & \beta \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}\right] = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$. *Entonces* $\mathcal{H}_1^{(2)} = 0$. *Por lo cual el álgebra de Heisenberg es resoluble.*

Por otro lado, si $L = \mathrm{sl}(2, \mathbb{C})$, tenemos que $L = L'$ y por lo tanto $L^{(n)} = L$ para todo $n \geq 1$, por lo cual $\mathrm{sl}(2, \mathbb{C})$ no es resoluble.

**Teorema 17.** *Sean $L_1$ y $L_2$ dos álgebras de Lie y $\varphi : L_1 \to L_2$ isomorfismo entre ellas. Entonces*

1. $\varphi(L_1') = L_2'$.
2. $\varphi\big(Z(L_1)\big) = Z(L_2)$.

La demostración de este resultado queda como ejercicio para el lector.

**Ejercicio 1:** Sean $L_1$ y $L_2$ dos álgebras de Lie. Consideremos $L = L_1 \oplus L_2 = \{(x_1, x_2) : x_i \in L_i\}$ la suma directa de los espacios vectoriales subyacentes. Si definimos $[(x_1, x_2), (y_1, y_2)] = ([x_1, y_1], [x_2, y_2])$, probar que $L$ es un álgebra de Lie (la cual se llama *suma directa* de $L_1$ y $L_2$).
Probar además que si $L = L_1 \oplus L_2$ entonces $Z(L) = Z(L_1) \oplus Z(L_2)$ y $L' = L_1' \oplus L_2'$.

## 2. Álgebras de Lie de dimensión baja

En esta sección estudiaremos las álgebras de Lie de dimensiones 1, 2 y 3.

### 2.1. Álgebras de Lie abelianas.

Las álgebras de Lie abelianas son fáciles de comprender. Para todo número natural $n$ hay un álgebra de Lie abeliana de dimensión $n$, donde para cualquier par de elementos el corchete de Lie es cero.
Aplicando el Teorema 12 sabemos que esta es única a menos de isomorfismos.

Para entender entonces cuantas álgebras de Lie, esencialmente diferentes (no isomorfas) hay nos concentraremos en las no abelianas.
Sabemos que álgebras de Lie de diferentes dimensiones no pueden ser isomorfas, ya que si lo fueran en particular serían isomorfas como espacios vectoriales, lo que implica que deben tener la misma dimensión. Más aún, si $L$ es no abeliana, entonces su álgebra derivada $L'$ es no nula y su centro $Z(L)$ es un ideal propio. Además, por el Teorema 17, álgebras derivadas y centros se preservan por isomorfismos, así que es razonable usar la dimensión de $L$ y las propiedades de $L'$ y $Z(L)$ para poder clasificar las álgebras de Lie.
A continuación estudiaremos las álgebras de Lie de dimensión 1, 2 y 3.

### 2.2. Álgebras de Lie de dimensión 1 y 2.

Toda álgebra de Lie de dimensión 1 es abeliana:

Supongamos que $L$ es un álgebra de Lie no abeliana de dimensión 2 sobre $\Bbbk$. El álgebra derivada $L'$ debe tener como mucho dimensión 1, porque si $\{x, y\}$ es una base de $L$ entonces $L'$ esta generado por $[x, y]$. Por otro lado, el álgebra derivada no puede ser cero porque $L$ no es abeliana. Entonces $L'$ tiene exactamente dimensión 1. Consideremos un elemento no nulo $x \in L'$ y completemos a una base $\{x, z\}$ de $L$ tal que $[x, z] \in L'$, este elemento es no nulo (porque $L$ es no abeliana) $\Rightarrow \exists \alpha \in \Bbbk$ tal que $[x, z] = \alpha x$. Podemos cambiar $z$ por $y = \alpha^{-1} z$ de donde $[x, y] = x$.
Acabamos de probar que si un álgebra de Lie de dimensión 2 no abeliana existe, esta admite una base de la forma $\{x, y\}$ tal que $[x, y] = x$.

**Ejercicio 2:** Sea $L$ un espacio vectorial de dimensión 2 con base $\{x, y\}$. Definimos una función bilineal $[\,,\,]$ en $L$ tal que $[u, u] = 0$, para todo $u \in L$. Probar que la identidad de Jacobi se verifica y por lo tanto $L$ es un álgebra de Lie.

**Teorema 18.** *Para un cuerpo cualquiera $\Bbbk$, a menos de isomorfismos existe una única álgebra de Lie no abeliana de dimensión 2. Esta álgebra admite una base $\{x, y\}$ tal que $[x, y] = x$. El centro de esta álgebra es* 0.

### 2.3. Álgebras de Lie de dimensión 3.

Sea $L$ un álgebra de Lie de dimensión 3 sobre $\Bbbk$ no abeliana. Sabemos que el álgebra derivada $L'$ es no nula, esta puede tener dimensión 1, 2 o 3. También sabemos que el centro $Z(L)$ es un ideal propio de $L$. Estudiemos a continuación que podemos decir de $L$ cuando la dimensión de su álgebra derivada es 1, 2 o 3.

- $\dim L' = 1$ y $L' \subset Z(L)$.

En el Ejemplo 1,8 se presentó el álgebra de Heisenberg de dimensión 3, $\mathcal{H}_1$, con base $\{P, Q, C\}$ tal que $[P, Q] = C$ y $[C, P] = [C, Q] = 0$, es decir $C \in Z(\mathcal{H}_1)$.

Veamos que, a menos de isomorfismos, esta es la única álgebra de Lie de dimensión 3 tal que $L' \subset Z(L)$ y $\dim L' = 1$.

Consideremos un álgebra de Lie $L$ con estas condiciones y tomemos $f, g \in L$ tales que $[f, g] \neq 0$. Como estamos asumiendo que $\dim L' = 1$ el corchete $[f, g]$ genera a $L'$. Además, como $L' \subset Z(L)$ tenemos que $[f, g]$ conmuta con todo elemento de $L$. Si definimos $z = [f, g]$, entonces $\{f, g, z\}$ es un conjunto LI y por lo tanto una base de $L$.

- $\dim L' = 1$ y $L' \not\subset Z(L)$.

Tomemos $L = L_1 \oplus L_2$, donde $L_1$ tiene dimensión 2 y es no abeliana y $L_2$ tiene dimensión 1.

Aplicando el Ejercicio 1 tenemos que $L' = L_1' \oplus L_2' = L_1'$, entonces $L'$ tiene dimensión 1. Además, $Z(L) = Z(L_1) \oplus Z(L_2) = L_2$ por lo tanto $L'$ no está contenido en $L_2$.

No hay otras álgebras de Lie con esta propiedad.

**Teorema 19.** *Sea $\Bbbk$ un cuerpo cualquiera. Existe una única álgebra de Lie sobre $\Bbbk$ de dimensión 3 tal que $\dim L' = 1$ y $L' \not\subset Z(L)$. Esta álgebra de Lie es la suma directa del álgebra de Lie no abeliana de dimensión 2 con el álgebra de Lie de dimensión 1.*

*Demostración.* Tomemos un elemento no nulo $x \in L'$, como $x$ no es central existe $y \in L$ tal que $[x, y] \neq 0$.

**Afirmación 1:** Si $x, y \in L$: $[x, y] \neq 0$ entonces $\{x, y\}$ es LI sobre $\Bbbk$. La prueba de esta afirmación queda como ejercicio para el lector.

Como $x$ es un generador de $L'$ tenemos que $[x, y]$ es múltiplo de $x$. Como antes podemos asumir que $[x, y] = x$. Extendemos $\{x, y\}$ a $\{x, y, w\}$ una base de $L$. Nuevamente, como $x$ genera a $L'$ existen escalares $a, b \in \Bbbk$ tales que

$$[x, w] = ax, \quad [y, w] = bx.$$

**Afirmación 2:** $L$ contiene un elemento central no nulo $z$ que no está en el subespacio generado por $x$ e $y$: sea $z = \lambda x + \mu y + \nu w \in L$ entonces

$$[x, z] = [x, \lambda x + \mu y + \nu w] = \mu[x, y] + \nu[x, w] = \mu x + \nu a x,$$

$$[y, z] = [y, \lambda x + \mu y + \nu w] = \lambda[y, x] + \nu[y, w] = -\lambda x + \nu b x,$$

$$[w, z] = [w, \lambda x + \mu y + \nu w] = \lambda[w, x] + \mu[w, y] = -\lambda a x - \mu b x.$$

Si tomamos $\lambda = b$, $\mu = -a$ y $\nu = 1$ entonces $[x, z] = [y, z] = [w, z] = 0$ y $z \notin \operatorname{span}\{x, y\}$.

Por lo cual $L = \operatorname{span}\{x, y\} \oplus \operatorname{span}\{z\}$ suma directa de álgebras de Lie de la forma requerida. □

- $\dim L' = 2$.

Estudiemos ahora el caso $\dim L = 3$ y $\dim L' = 2$. Veamos que sobre $\mathbb{C}$ hay infinitas de estas álgebras de Lie no isomorfas.

Tomemos una base $\{y, z\}$ de $L'$ y extendamos a una base $\{x, y, z\}$ de $L$.

**Lema 20.** *Sea $L'$ de dimensión 2, $\{y, z\}$ base de $L'$ y $\{x, y, z\}$ base de $L$ entonces*

(a) *$L'$ es abeliana.*
(b) *El mapa $\operatorname{ad} x : L' \to L'$ es un isomorfismo.*

*Demostración.* (a) Basta probar que $[y, z] = 0$: sabemos que $[y, z] \in L'$ entonces existen $\alpha, \beta \in \Bbbk$: $[y, z] = \alpha y + \beta z$.
La matriz asociada a $\operatorname{ad} y$ con respecto a la base $\{x, y, z\}$ tiene la forma

$$\begin{pmatrix} 0 & 0 & 0 \\ * & 0 & \alpha \\ * & 0 & \beta \end{pmatrix} : \ * \in \Bbbk$$

Veamos que como $y \in L'$ tenemos que $\operatorname{tr}(\operatorname{ad} y) = 0$: dado que $\operatorname{ad} : L \to \operatorname{gl}(L)$ es un homomorfismo de álgebras de Lie tenemos que

$$\operatorname{ad}[v, w] = \operatorname{ad} v \circ \operatorname{ad} w - \operatorname{ad} w \circ \operatorname{ad} v, \ \forall v, w \in L.$$

Entonces $\operatorname{tr}\big(\operatorname{ad}[v, w]\big) = \operatorname{tr}\big(\operatorname{ad} v \circ \operatorname{ad} w - \operatorname{ad} w \circ \operatorname{ad} v\big) = 0$.
Como $\operatorname{tr}(\operatorname{ad} y) = \beta$ concluimos que $\beta = 0$. De igual forma, si consideramos la matriz para $\operatorname{ad} z$ concluimos que $\alpha = 0$, entonces $[y, z] = 0$.

(b) $L'$ está generada por $\big\{[x, y], [x, z], [y, z]\big\}$. Como $[y, z] = 0$ deducimos que $\big\{[x, y], [x, z]\big\}$ es una base de $L'$. Entonces, la imagen de $\operatorname{ad} x$ tiene dimensión 2, por lo tanto $\operatorname{ad} x : L' \to L'$ es un isomorfismo.

□

**Caso 1:** Existe $x \notin L'$ tal que $\operatorname{ad} x : L' \to L'$ es diagonalizable.
Podemos asumir que $y, z$ son vectores propios de $\operatorname{ad} x$, por (b) del Lema 20 los valores propios asociados deben ser no nulos. Supongamos $[x, y] = \lambda y$. Podemos asumir que $\lambda = 1$, si reescalamos $x$ por $\lambda^{-1}$, tenemos que $[\lambda^{-1}x, y] = y$. Con respecto a la base $\{y, z\}$ de $L'$ el mapa lineal $\operatorname{ad} x : L' \to L'$ tiene matriz

$$\begin{pmatrix} 1 & 0 \\ 0 & \mu \end{pmatrix} : \ \mu \in \mathbb{C} \text{ no nulo.}$$

**Ejercicio 3:** Sean $V$ espacio vectorial y $\varphi$ un endomorfismo de $V$. Sea $L = V \oplus \operatorname{span}\{x\}$. Probar que si definimos en $L$ el corchete $[y, z] = 0$ para todos $y, z \in V$, $[x, y] = \varphi(y)$ para todo $y \in V$, entonces $L$ es un álgebra de Lie y $\dim L' = \operatorname{rango}(\varphi)$. Llamemos $L_\mu$ a esta álgebra de Lie. Probar que $L_\mu$ es isomorfa a $L_\nu$ si y sólo si $\mu = \nu$ o $\mu = \nu^{-1}$.

A partir del ejercicio 3 concluimos que hay infinitas de estas álgebras de Lie no isomorfas.

**Caso 2:** Para todo $x \notin L'$ el mapa $\operatorname{ad} x : L' \to L'$ no es diagonalizable.
Tomemos $x \notin L'$. Como estamos trabajando sobre los complejos, $\operatorname{ad} x : L' \to L'$ debe tener un vector propio, llamémoslo $y \in L'$. Como antes, podemos renormalizar $x$ y asumir que $[x, y] = y$, extendemos $y$ a una base $\{y, z\}$ de $L'$. Tenemos que $[x, z] = \lambda y + \mu z$, con $\lambda \neq 0$ (de lo contrario $\operatorname{ad} x$ sería diagonalizable). Reescalando

$z$ podemos suponer $\lambda = 1$. Entonces la matriz asociada a $\operatorname{ad} x : L' \to L'$ tiene la forma

$$A = \begin{pmatrix} 1 & 1 \\ 0 & \mu \end{pmatrix}.$$

Como $A$ no es diagonalizable no puede tener dos valores propios diferentes, entonces $\mu = 1$.
Nuevamente, esto determina completamente el álgebra de Lie con las propiedades requeridas. A menos de isomorfismos tenemos una sola de estas álgebras.

- $L' = L$.

Supongamos que $L$ es un álgebra de Lie compleja de dimensión 3 tal que $L = L'$. Veamos que a menos de isomorfismo $\mathrm{sl}(2, \mathbb{C})$ es la única álgebra de Lie con esta propiedad.

**Ejercicio 4:** $\mathrm{sl}(2, \mathbb{C}) = \mathrm{sl}(2, \mathbb{C})'$.

**Paso 1:** Sea $x \in L$ no nulo entonces $\operatorname{ad} x$ tiene rango 2.
Extendamos $x$ a una base $\{x, y, z\}$ de $L$. Entonces $L'$ está generado por el conjunto $\big\{[x, y], [x, z], [y, z]\big\}$, como $L = L'$ este conjunto es LI. Por lo cual la imagen de $\operatorname{ad} x$ tiene por base $\big\{[x, y], [x, z]\big\}$.

**Paso 2:** Existe $h \in L$ tal que $\operatorname{ad} h : L \to L$ tiene un vector propio con valor propio no nulo.
Consideremos $x \in L$ no nulo, si $\operatorname{ad} x$ tiene un valor propio no nulo podemos tomar $h = x$. Si $\operatorname{ad} x : L \to L$ no tiene valores propios no nulos, como tiene rango 2, su forma de Jordan es

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix},$$

entonces existe una base $\{x, y, z\}$ de $L$ extendiendo $x$, tal que $[x, y] = x$, $[x, z] = y$. Por lo tanto $\operatorname{ad} y$ tiene a $x$ como vector propio asociado a $-1$. Entonces podemos tomar $h = y$.

**Paso 3:** Por los pasos anteriores podemos encontrar $h, x \in L$ tales que $[h, x] = \alpha x \neq 0$. Como $h \in L = L'$ sabemos que $\operatorname{ad} h$ tiene traza nula. Entonces $\operatorname{ad} h$ puede tener tres valores propios distintos, $-\alpha$, 0 y $\alpha$. Si $y$ es vector propio asociado a $-\alpha$, entonces $\{h, x, y\}$ es base de $L$. En esta base $\operatorname{ad} h$ se representa por una matriz diagonal.

**Paso 4:** Para completar la descripción del álgebra de Lie $L$ necesitamos determinar $[x, y]$.

$$[h, [x, y]] = [[h, x], y] + [x, [h, y]] = \alpha[x, y] - \alpha[x, y] = 0.$$

Tenemos dos aplicaciones del paso 1. Primero $N(\operatorname{ad} h) = \operatorname{span}\{h\}$, entonces $[x, y] = \lambda h$ para algún $\lambda \in \Bbbk$. Segundo, $\lambda \neq 0$ porque de lo contrario $N(\operatorname{ad} h)$ tendría dimensión 2. Normalizando $x$ por $\lambda^{-1}x$ podemos suponer $\lambda = 1$.

Si sustituimos $h$ por un múltiplo no nulo podemos tomar cualquier valor de $\alpha$ como queremos. En particular si tomamos $\alpha = 2$ la estructura constante de $L$ que

respeta la base $\{x, y, h\}$ coincide con la estructura constante de $\mathrm{sl}(2, \mathbb{C})$ respecto la base

$$\left\{ \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ -1 & 0 \end{pmatrix} \right\}.$$

Entonces $L \cong \mathrm{sl}(2, \mathbb{C})$. Por lo tanto, hay una única álgebra de Lie de dimensión tres que coincide con su álgebra derivada.

## 3. Aplicaciones a economía y finanzas

En esta última sección presentaremos algunos problemas y tópicos en Economía y finanzas que se resuelven empleando la teoría de Lie, además desarrollaremos uno de ellos para familiarizar al lector con las técnicas.

En los trabajos de Lo y Hui [6], [7], [8] se estudió la valoración de derivados multiactivos introduciendo diversas técnicas basadas en álgebras de Lie. Previamente, Lo y Hui ya habían empleado la teoría de Lie para estudiar ecuaciones en derivadas parciales con coeficientes que dependen del tiempo, modelos CEV (elasticidad constante de varianza) y opciones con barrera.

Independientemente, empleando igualmente las álgebras de Lie, Björk y Landén en el trabajo [2] estudiaron problemas financieros correspondientes a la toma de decisiones bajo riesgo por parte de los agentes en el marco de la teoría de las funciones de utilidad. Para este estudio se emplearon los grupos de Lie nilpotentes.

En [1], Basov describió algunos métodos basados en las propiedades de los grupos de Lie para resolver el problema de cribado multidimensional.

En [5] Gaspar obtuvo un modelo general para la estructura de los precios a plazos basándose en la metodología dada por Björk y aplicando las álgebras de Lie.

### 3.1. Nociones básicas en economía y finanzas.

Se denomina *derivado financiero* a cualquier producto financiero cuyo valor está basado en el precio que posee un determinado activo. Consisten en operaciones hipotéticas cuya liquidación se realiza mediante la diferencia existente entre el precio de mercado del activo y el precio pactado en la operación hipotética.

Los derivados financieros tienen como función eliminar o reducir las consecuencias adversas producidas por cambios desfavorables en el activo sobre el que se define el derivado (es decir, eliminar el riesgo en las operaciones financieras). Además se emplean como un producto financiero basado en la especulación con los precios del activo.

Los *derivados financieros multiactivos*, consisten en productos financieros cuyo valor se basa en el precio que poseen varios activos (y no solamente uno).

Uno de los derivados financieros se denomina *opción*, que es el derecho a comprar o vender un activo en el futuro a un precio pactado. Debe tenerse en cuenta que, al comprar una opción, el comprador paga una prima por disfrutar del derecho adquirido mientras que el vendedor cobra dicha prima. Por tanto, se realiza una transacción en el instante de la contratación de la opción.

Existen dos tipos de opciones estándar: las opciones de *estilo americano* y las de *estilo europeo*. Las primeras son aquellas en las que es posible ejercer derecho de

compra-venta en cualquier momento anterior a la fecha de vencimiento del contrato mientras que en las segundas sólo se puede ejercer dicho derecho en la fecha de vencimiento. Cualquier otro tipo de opción se denomina *exótica*. Un caso particular de opciones exóticas son las opciones con barreras. Se denomina *opción con barrera* a toda opción cuya cancelación o activación depende del valor alcanzado durante un período de tiempo determinado por el precio del activo subyacente. Este valor será independiente del valor del activo en la fecha de vencimiento de la opción. Es decir, la activación o cancelación de la opción depende de que el precio del activo alcance unos determinados valores umbrales (de ahí la denominación de opciones con barrera).

Principales tipos y subtipos de opciones de barrera:

1. Opciones con barrera *de entrada* (knock-in): la opción pasa a activarse y a ser estándar si el precio del activo subyacente alcanza el valor fijado en la barrera durante el período acordado.
   *a*) Opciones *abajo y de entrada* (down-in): la barrera se fija por debajo del precio inicial del activo, activándose la opción cuando el precio llega a ser inferior a la barrera.
   *b*) Opciones *arriba y de entrada* (up-in): la barrera se fija por encima del precio inicial del activo, activándose la opción cuando el precio es superior a la barrera.
2. Opciones con barrera *de salida* (knock-out): la opción deja de existir o expira sin valor cuando se alcanza el valor
   fijado en la barrera para el precio del activo.
   *a*) Opciones *abajo y de salida* (down-out): la barrera se fija por debajo del precio inicial del activo, expirando la opción cuando el precio llega a ser inferior a la barrera.
   *b*) Opciones *arriba y de salida* (up-out): la barrera se fija por encima del precio inicial del activo, expirando la opción cuando el precio llega a ser superior a la barrera.

Las opciones con barrera pueden contratarse de tal modo que la barrera sea doble (es decir, que sea arriba y abajo a la vez) e incluso puede establecerse una barrera móvil, que vaya ajustándose durante toda la vida de la opción hasta alcanzar la fecha de vencimiento.

Cualquier producto financiero presenta la problemática de la fijación de precios. En la fijación de precios, la empresa debe considerar tanto las necesidades del mercado hacia el producto ofertado como el proceso productivo. Para obtener el máximo beneficio posible, se debe buscar el equilibrio entre elegir un precio competitivo y un precio que permita unos márgenes más amplios de ganancia.

A la hora de determinar los precios de un producto financiero suele considerarse el modelo CEV. Este modelo, introducido por Cox (1975), extiende el de Black-Scholes para la fijación de precios e introduce la posibilidad de considerar una volatilidad estocástica. En el modelo CEV, se supone que el precio $S(t)$ del activo sigue el siguiente proceso de difusión en función del tiempo $t$:

$$dS(t) = \mu(t)S(t)dt + \sigma(t)S(t)^{\beta/2}dZ(t)$$

donde $\mu(t)$ indica la tasa de crecimiento, $\sigma(t)$ es el parámetro de volatilidad, $\beta$ es el parámetro de elasticidad de la función de volatilidad local y $Z(t)$ es un proceso de Wiener (un proceso de Wiener es un proceso estocástico dependiendo

continuamente del tiempo. El ejemplo más conocido de proceso de Wiener es el movimiento Browniano).

### 3.2. Teoría de Lie y fijación de precios para opciones con barrera móvil y con parámetros temporales.

En esta sección desarrollaremos parcialmente el trabajo de Lo y Hui [8], este trabajo se basa en el Teorema de Wei-Norman ([10]), que nunca se había aplicado al campo de las Finanzas. Lo y Hui ajustan y aplican este modelo basado en las álgebras de Lie al problema de la fijación de precios de las opciones con barrera móvil y con parámetros dependientes del tiempo. Para realizar dicha evaluación supusieron que el valor del activo subyacente sigue el siguiente proceso de difusión CEV:

$$dS(t) = \mu(t)S(t)dt + \sigma(t)S(t)^{\beta/2}dZ(t), \quad 0 \leq \beta < 0. \tag{1}$$

Donde $\mu(t)$ es la media del precio de las acciones en el instante $t$, $\sigma(t)S(t)^{\beta/2}$ es la varianza instantánea de dicho precio, $dZ(t)$ es un proceso de Wiener y $\beta$ es el factor de elasticidad.

Partiendo de la ecuación (1), la varianza instantánea del cambio porcentual en el precio se define como $\sigma(t)^2/S(t)^{2-\beta}$.

**Teorema de Wei-Norman**. Considerar la ecuación diferencial de primer orden de operadores lineales

$$\frac{dU(t)}{dt} = H(t)U(t), \quad U(0) = 1. \tag{2}$$

Donde $H$ y $U$ son operadores lineales dependientes del tiempo en un espacio de Banach o uno de dimensión finita. De acuerdo al Teorema de Wei-Norman si el operador $H$ puede expresarse como

$$H(t) = \sum_{n=1}^{N} a_n(t)L_n, \tag{3}$$

donde $a_n$ son funciones escalares dependientes del tiempo y $L_n$ son generadores de un álgebra de Lie resoluble de dimensión $N$ o generadores del álgebra de Lie simple real split. Entonces, el operador $U$ se expresa como:

$$U(t) = \prod_{n=1}^{N} \exp\big[g_n(t)L_n\big]. \tag{4}$$

Siendo $g_n$ funciones escalares dependientes de la variable t. Para hallar estas funciones simplemente sustituimos las ecuaciones (3) y (4) en (2) y comparamos los dos lados de la igualdad, obteniendo las siguientes ecuaciones diferenciales no lineales:

$$\frac{dg_n(t)}{dt} = \sum_{m=1}^{N} \eta_{nm}a_m(t), \quad g_n(0) = 0, \tag{5}$$

donde $\eta_{nm}$ son funciones no lineales de las $g_n$. Entonces, transformamos la ecuación diferencial lineal (2) en un sistema de ecuaciones no lineales acopladas de funciones escalares (5).

Consideremos el siguiente ejemplo para ilustrar esto: los generadores $L_n$ forman el álgebra de Lie de Heisenberg-Weyl definida por las relaciones del corchete

$$[L_1, L_2] = L_3, [L_1, L_3] = [L_2, L_3] = 0, \tag{6}$$

entonces $H$ está dada por

$$H(t) = a_1(t)J_1 + a_2(t)L_2 + a_3(t)L_3. \tag{7}$$

Por el teorema de Wei-Norman $U(t)$ se expresa como

$$U(t) = \operatorname{ext}\big(g_1(t)L_1\big)\operatorname{ext}\big(g_2(t)L_2\big)\operatorname{ext}\big(g_3(t)L_3\big), \tag{8}$$

derivando

$$\frac{dU(t)}{dt}U(t)^{-1} = \frac{dg_1(t)}{dt}L_1 + \frac{dg_2(t)}{dt}L_2 + \left[\frac{dg_3(t)}{dt} + g_1(t)\frac{dg_2(t)}{dt}\right]L_3 \tag{9}$$

Comparando (7) y (9) obtenemos que:

$$\frac{dg_1(t)}{dt} = a_1(t), \quad \frac{dg_2(t)}{dt} = a_2(t), \quad \frac{dg_3(t)}{dt} + g_1(t)\frac{dg_2(t)}{dt} = a_3.$$

Entonces

$$g_1(t) = \int_0^t a_1(\tau)d\tau, \quad g_2(t) = \int_0^t a_2(\tau)d\tau,$$

$$g_3(t) = \int_0^t \big[a_3(\tau) - a_2(\tau)g_1(\tau)\big]d\tau.$$

Lo cual termina de determinar $U$.

**CEV Opción europea**. El modelo CEV con parámetros dependientes del tiempo para opciones europeas estándar se describe por la ecuación diferencial parcial

$$\frac{\partial P(S,\tau)}{\partial \tau} = \frac{1}{2}\sigma(\tau)^2 S^\beta \frac{\partial^2 P(S,\tau)}{\partial S^2} + \big[r(\tau) - d(\tau)\big]S\frac{\partial P(S,\tau)}{\partial S} - r(\tau)P(S,\tau), \tag{10}$$

para $0 \leq \beta < 0$. En esta ecuación $P$ es el valor de la opción, $S$ es el precio del activo subyacente, $\tau$ es el tiempo al vencimiento, $\sigma$ es la volatilidad, $r$ es la tasa de interés libre de riesgo y $d$ son los dividendos generados. La ecuación (10) puede reescribirse como sigue, mediante el cambio de variable $x = \sqrt{S^{2-\beta}}$:

$$\begin{aligned}\frac{\partial u(x,\tau)}{\partial \tau} = {} & \frac{1}{8}\tilde{\sigma}(\tau)^2\frac{\partial^2 u(x,\tau)}{\partial x^2} + \frac{1}{2}\left[\tilde{\mu}(\tau)x - \frac{(4-\beta)\tilde{\sigma}(\tau)^2}{4(2-\beta)x}\right]\frac{\partial u(x,\tau)}{\partial x} + \\ & \left[\frac{(4-\beta)\tilde{\sigma}(\tau)^2}{8(2-\beta)x} - r(\tau)\right]u(x,\tau) \equiv H(\tau)u(x,\tau),\end{aligned} \tag{11}$$

siendo $\tilde{\sigma}(\tau) = (2-\beta)\sigma(\tau)$, $\tilde{\mu}(\tau) = (2-\beta)\big[r(\tau) - d(\tau)\big]$ y $u(x,\tau) = xP(S,\tau)$.

Esta nueva forma de expresar (10) les permitió dar a Lo y Hui una expresión del operador $H$

$$H(t) = a_1(t)K_+ + a_2(t)K_0 + a_3(t)K_- + b(t), \tag{12}$$

donde

$$K_- = \frac{1}{2}\left[\frac{\partial^2}{\partial x^2} - \frac{4-\beta}{(2-\beta)x}\frac{\partial}{\partial x} + \frac{4-\beta}{(2-\beta)x^2}\right],$$

$$K_0 = \frac{1}{2}\left(x\frac{\partial}{\partial x} - \frac{1}{2-\beta}\right), \quad K_+ = \frac{1}{2}x^2,$$

$$a_3(t) = \frac{1}{4}\tilde{\sigma}(t)^2, \quad a_2(t) = \tilde{\mu}(t), \quad a_1(t) = 0,$$

$$b(t) = \frac{1-\beta}{2(\beta-2)}\tilde{\mu}(t) - r(t).$$

Los operadores $K_+$, $K_0$ y $K_-$ son los generadores del álgebra de Lie $su(1,1)$:

$$\big[K_+, K_-\big] = -2K_0,$$
$$\big[K_0, K_\pm\big] = \pm K_\pm.$$

Podemos definir el operador evolución $U(t,0)$ tal que

$$u(x,t) = \exp\left[\int_0^t b(\tau)d\tau\right] U(t,0)u(x,0). \tag{13}$$

Sustituyendo esta ecuación en (11) obtenemos la ecuación de evolución

$$\frac{\partial U(t,0)}{\partial t} = H_I(t)U(t,0), \quad U(0,0) = 1, \tag{14}$$

donde $H_I(t) = a_1(t)K_+ + a_2(t)K_0 + a_3(t)K_-$.

Como $su(1,1)$ es un álgebra de Lie real de dimensión 3 simple split, el teorema de Wei-Norman dice que

$$U(t,0) = \exp\big[c_1(t)K_+\big] \exp\big[c_2(t)K_+0\big] \exp\big[c_3(t)K_-\big], \tag{15}$$

donde los coeficientes $c_i(t)$ están dados por

$$c_1(t) = 0, \quad c_2(t) = \int_0^t \tilde{\mu}(\tau)d\tau$$

y

$$c_3(t) = \frac{1}{4}\int_0^t \tilde{\sigma}(\tau)^2 \exp\big[c_2(\tau)\big]d\tau.$$

Entonces, tenemos como hallar el operador de evolución $U(t,0)$ que permite determinar $u(x,t)$ solución de (11).

## Referencias


[1] Suren Basov, *Lie Groups of Partial Differential Equations and Their Application to the Multidimensional Screening Problems*, University of Melbourne Economics Working Paper No. 895, 2004.

[2] Tomas Björk and Camilla Landén, *On the construction of finite dimensional realizations for nonlinear forward rate models*. Finance and Stochastics, Vol. 6 Issue 3, pp. 303-331, 2002.

[3] Erdmann and Wildon, *Introduction to Lie algebras*. Springer Undergraduate Mathematics Series, 2011.

[4] Fuentes and López Sandoval, *Introducción a las álgebras de Lie*. 2013.

[5] *Finite dimensional Markovian realizations for forward price term structure models*, Stochastic Finance, M. Grossinho, A. Shyriaev, M. Esquvel, P. Oliveira, eds. , Chapter 10, pp. 265-320, Springer Verlag Publisher, 2006.

[6] Lo and Hui, *Valuation of financial derivatives with time-dependent parameters: Lie-algebraic approach*. Quantitative Finance, Vol. 1, No. 1, pp. 73-78, 2001.

[7] Lo and Hui, *Pricing multi-asset financial derivatives with time-dependent parameters—Lie algebraic approach*. International Journal of Mathematics and Mathematical Sciences Volume 32, Issue 7, pp. 401-410, 2002.

[8] Lo and Hui, *Lie-algebraic approach for pricing moving barrier options with time-dependent parameters*. Journal of Mathematical Analysis and Applications, vol. 323, no. 2, pp. 1455–1464, 2006.

[9] Mozgovoy, *Lecture Notes MA3415: Introduction to Lie algebras*. 2017.

[10] James Wei and Eduard Norman, *Lie algebraic solution of linear Differential Equations*. Journal of Mathematical Physics Vol. 4, pp. 575-581, 1963.



IMERL, Facultad de Ingeniería, Universidad de la República. Montevideo, Uruguay.
*Email address*: akgdeloss@gmail.com

MINICURSO

# LA REGULARIDAD DE LOS POLIEDROS EN LA MIRADA MODERNA

ÁNGEL PEREYRA

Publicaciones Matemáticas del Uruguay
Volumen 17, Julio 2019, Páginas 43–55
S 0797-1443

# LA REGULARIDAD DE LOS POLIEDROS EN LA MIRADA MODERNA

ÁNGEL PEREYRA (FIGURAS: ANALÍA CHANQUET)

Resumen. Estas notas reflejan el cursillo aludido en el título, aprovechamos la oportunidad para ampliar ligeramente el contenido de la última clase. En síntesis el objetivo del cursillo fue: analizar algunas variaciones de la noción clásica de poliedro regular, teniendo como concepto moderno unificador el de grupo de simetrías. Con este cometido comezamos revisando la noción de poliedro convexo para precisar el concepto de sólido platónico, a estos les buscamos sus grupos de simetrías. La segunda clase la dedicamos a presentar una visión panorámica de la obtención de todos los poliedros semirregulares, también observamos a sus duales (los sólidos de Catalán). Finalmente ampliando adecuadamente la noción de poliedro mostramos cómo caracterizar la familia de todos los poliedros totalmente transitivos, se consigue una familia que consta de catorce poliedros entre los cuales están los cinco sólidos platónicos.

## 1. Introducción

Estas notas mantienen el espíritu de su versión preliminar: servir de apoyo y promover en los cursillistas -o lectores- una actitud imaginativa y constructiva, conociendo de antemano el propósito y sentido de cada clase, antes que el desarrollo detallado de las mismas. Así que una debilidad -¿o no?- de este trabajo es que las demostraciones sólo están esbozadas o directamente dejadas a cargo del lector. Espero que este material sea de lectura fluida y genere ganas de estudiar más sobre un tema bonito, antiguo y a la vez susceptible de revisiones y generalizaciones.

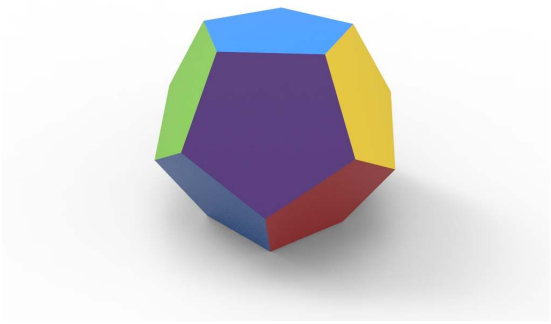


Figura 1. El Dodecaedro, uno de los cinco sólidos platónicos.

©2019 PMU

Como ya dijimos vamos a analizar algunas variaciones de la noción clásica de poliedro regular, teniendo como concepto moderno unificador el de grupo de simetrías. Cabe agregar, simplificando mucho, que paralelamente a la exploración de la regularidad, la noción de poliedro ha ido evolucionando hasta el punto en que el objeto geométrico visualizable es la realización ("proyección" a $\mathbb{R}^3$) de un objeto combinatorio abstracto: una estructura de caras que verifica ciertas propiedades de incidencia en relación al conjunto de las aristas y vértices de cierto grafo.

FIGURA 2. Tres volúmenes poliédricos "muy regulares", el primero convexo y los otros no.

La segunda clase la dedicamos a mostrar cómo construir exhaustivamente la familia de los poliedros semirregulares (o arquimedianos). Contrariamente a lo que se podría suponer no todo poliedro semirregular se obtiene truncando algún sólido platónico, la forma natural para obtener a varios de ellos es por "deformación". Además vimos cómo la noción de grupo de simetría permite arribar a una formulación precisa del concepto de poliedro semirregular y superar ciertas ambigüedades históricas. Al finalizar la clase dimos un vistazo rápido a la familia dual de la de los poliedros semirregulares, dejando de lado a los prismas y a los antiprismas, se obtienen los sólidos de Catalán, fue una oportunidad más de ver la utilidad de la dualidad.

En la tercera clase usamos en profundidad el concepto de transitividad total en los poliedros, este concepto es un debilitamiento de la noción estándar de regularidad. Un poliedro (no necesariamente convexo, eventualmente compuesto) se dice totalmente transitivo si su grupo de simetrías directas lleva cualquier vértice (arista o cara) en cualquier otro vértice (arista o cara, respectivamente). Apoyándonos en cuestiones básicas de las acciones de grupos, esbozamos la demostración de que son 14 estos poliedros (los platónicos, cuatro estrellados y cinco compuestos). Finalizamos el cursillo haciendo algunos comentarios sobre la evolución de la noción de poliedro, mencionando alguna de sus formulaciones más recientes.

## 2. Primera clase: Poliedros convexos. Sólidos platónicos y sus grupos de simetría.

Antes que nada, convendría tener al alcance de las manos realizaciones 3D de los cinco sólidos platónicos, la manipulación de los mismos mejora nuestra capacidad de visualización y argumentación.

Lo primero será hacer una revisión rápida del concepto de poliedro convexo, asumiré que los cursillistas (o lectores) tienen cierto conocimento básico del asunto,

FIGURA 3. Dos poliedros convexos "parcialmente regulares".

no abundaré en detalles. Podemos pensar a un poliedro convexo como la intersección de un número finito de semiespacios cerrados de $\mathbb{R}^3$ tal que esta intersección resulte de interior no vacío y acotada. En definitiva, lo esencial ahora es comprender que dado un poliedro convexo están perfectamente determinadas sus caras, sus aristas y sus vértices. Los vértices, aristas y caras son las intersecciones 0, 1 y 2 dimensionales (respectivamente) de los planos de apoyo del poliedro con el propio poliedro. No necesitamos un conocimiento detallado de esto para el cursillo, con una idea general intuitiva alcanza. Vale el resultado de Euler $V - A + C = 2$ que relaciona el número de vértices, aristas y caras de cualquier poliedro convexo. Si no lo han hecho antes, pueden verificar este teorema para el icosaedro y el dodecaedro. Más adelante apelaremos a una consecuencia de este resultado, por eso lo menciono.

Es muy conocido el hecho de que existen exactamente cinco tipos de poliedros regulares convexos: el tetraedro, hexaedro (cubo), octaedro, dodecaedro e icosaedro. Para ser más precisos deberíamos decir tetraedro regular, etc..

Acá hay varias cuestiones que conviene considerar, no vamos a profundizar en todas ellas:

1. ¿Qué significa que un poliedro convexo sea regular?

2. Si respondimos i. surge otra pregunta,¿cómo estamos seguros de que los cinco poliedros mencionados arriba efectivamente existen y que no fuimos "engañados" por sugerentes dibujos o volúmenes?

3. ¿Cómo estamos seguros de que no hay más que cinco poliedros regulares convexos?

Sobre 1: seguramente acordaremos rápidamente que las caras de un tal poliedro deben ser todas polígonos regulares convexos congruentes, es un buen ejercicio obtener ejemplos en los que se vea que esta condición no sería suficiente. Una familia interesante de observar es la de los deltaedros, está formada por cinco tipos de poliedros convexos no regulares cuyas caras son triángulos equiláteros congruentes, algunos de estos poliedros están representados en las figuras 4 y 8. Una condición adicional que se puede imponer para obtener la regularidad es que todos los ángulos sólidos del poliedro sean congruentes. Observen esto en los poliedros platónicos. Ocurre que existen diferentes condiciones equivalentes que aseguran la regularidad tal como se la conceptualizó "inicialmente", algunas de ellas se presentan en el siguiente enunciado.

**Teorema 1.** *Si $\mathcal{P}$ es un poliedro convexo, cuyas caras son polígonos regulares congruentes, entonces son equivalentes:*

*(a) todos los vértices de $\mathcal{P}$ están en una esfera;*
*(b) todos los ángulos formados por caras adyacentes son congruentes;*
*(c) todos los ángulos sólidos son congruentes;*
*(d) todos los vértices están rodeados por el mismo número de caras.*

Sobre la demostración: es esclarecedor observar que la condición (a) es de naturaleza global y en cambio las demás son de naturaleza local, las implicaciones (a) $\Rightarrow$ (b) $\Rightarrow$ (c) $\Rightarrow$ (d) pueden ser demostradas de manera relativamente sencilla usando conocimientos básicos de geometría euclidiana del espacio tridimensional, en cambio la implicación (d) $\Rightarrow$ (a) -el pasaje de lo local a lo global-, es sustancialmente más difícil.

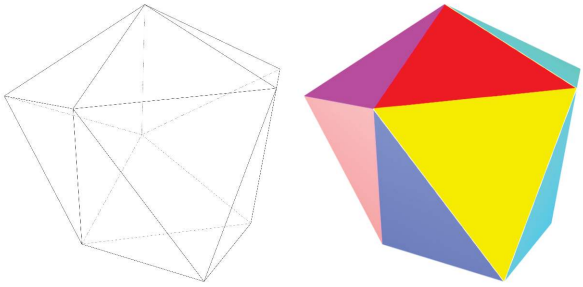


FIGURA 4. El deltaedro de 14 caras.

Sobre 2 y 3: es sabido que la suma de los ángulos planos que forman un ángulo sólido estrictamente convexo -que no contiene rectas- es menor que cuatro rectos, de aquí se deduce que un poliedro convexo regular sólo puede tener como caras triángulos (equiláteros), cuadrados o pentágonos (regulares). Los cinco sólidos platónicos son realizaciones de estas opciones, hay distintos procedimientos geométricos que permitem probar la existencia de todos ellos. Por ejemplo un análisis atento del lado izquierdo de la figura 7 permite probar la existencia del dodecaedro partiendo de un cubo. La figura 5 ilustra la construcción del icosaedro a partir del odecaedro -y a la inversa- usando un método que se explica a continuación.

Es posible dado un poliedro convexo $\mathcal{P}$ construir su poliedro dual $\mathcal{P}^*$, para el caso en el que el poliedro tenga todos sus vértices en una esfera y todas sus aristas congruentes, su dual puede obtenerse así : dado un vértice $v$ de $\mathcal{P}$ se toman los puntos medios de las aristas incidentes en $v$ estos puntos son los vértices de un polígono $D$ inscriptible (en una circunferencia $C$), las tangentes a esta circunferencia $C$ en los vértices de $D$ definen (envuelven) un polígono convexo que es la cara de $\mathcal{P}^*$ que está asociada a $v$; haciendo esto con cada vértice de $\mathcal{P}$ conseguimos todas las caras de $\mathcal{P}^*$. Es sencillo verificar que: el tetraedro es autodual, el octaedro y el cubo están en dualidad, el icosaedro y el dodecaedro también están en dualidad. Es instructivo comprender la conveniencia de las condiciones que le impusimos al poliedro $\mathcal{P}$.

El último tópico de esta primera clase: la noción de grupo de simetrías. Como dijimos en la introducción el grupo de simetrías de un conjunto de puntos $S$ de $\mathbb{R}^3$, está formado por los movimientos que dejan invariante a $S$. Se entiende que

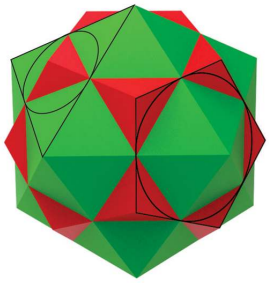


FIGURA 5. El Dodecaedro (rojo) y el Icosaedro (verde) son duales.

la operación del grupo es la composición de movimientos. Es un ejercicio sencillo verificar que efectivamente el enunciado anterior define un grupo. Este grupo, que puede ser finito o no, es un subgrupo del grupo de todos los movimientos de $\mathbb{R}^3$. A veces nos restringiremos a considerar movimientos directos solamente.

Nuestro objetivo ahora es dar alguna descripción de los grupos de simetrías directas de los poliedros platónicos. Un primera observación que nos simplifica el trabajo es que aquellos poliedros que estén en dualidad tendrán el mismo grupo de simetrías. La estrategia es sencilla inicialmente, nos concentramos en los movimientos directos que dejan invariante al tetraedro, los identificamos explícitamente y los contamos: son 12 (identidad incluída), estos movimientos corresponden a las permutaciones pares de cuatro elementos (los vértices); si consideramos movimientos directos y no directos son 24. La última afirmación se deduce aplicando el resultado siguiente.

**Ejercicio 1.** *Si $G$ es un grupo finito de movimientos que contiene algún movimiento no directo y $H$ es el subgrupo de los movimientos directos de $G$, entonces $G$ tiene el doble de elementos que $H$.*

FIGURA 6. El Octaedro.

El siguiente paso es considerar el octaedro, se constata que su grupo de simetrías directas tiene 24 elementos (lo mismo le ocurre al cubo por dualidad).

Para el dodecaedro el conteo de sus simetrías directas da 60 (lo mismo le ocurre al icosaedro por dualidad).

Con un poco más de trabajo puede decirse bastante más sobre la estructura de los grupos de simetrías directas del octaedro y del dodecaedro.

En el caso del octaedro podemos observar que si coloreamos con 4 colores sus caras, usando un color por cara y poniendo el mismo color en caras opuestas (las que son paralelas), resulta que cada simetría directa del octaedro se corresponde biunívocamente con una permutación de los colores, así que abstractamente el grupo es el de las permutaciones de 4 elementos. Es un ejercicio instructivo visualizar la correspondencia mencionada.

El caso del dodecaedro, es un poquito más sutil, se observa que en el dodecaedro se pueden inscribir cinco cubos (cuyos vértices son vértices del dodecaedro), esto se representa en la figura 7 . Con un análisis cuidadoso se verifica que las simetrías directas del dodecaedro se corresponden biunívocamente con ciertas permutaciones (las pares) de los cubos. Para comprender esto sugiero dibujar sobre las caras de un dodecaedro (en 3D) las aristas de los cubos y verificar las siguientes afirmaciones: una diagonal de una cara del dodecaedro determina unívocamente un cubo inscripto; las rotaciones de orden 5 (de ejes que pasan los centros de caras opuestas del dodecaedro) se corresponden con las permutaciones circulares de los cinco cubos; las rotaciones de orden 3 (de ejes que contienen vértices opuestos del dodecaedro) se corresponden con las permutaciones circulares de tres cubos -los cubos que tienen vértices en el eje de rotación quedan invariantes-; las rotaciones de orden 2 (de ejes que pasan por los puntos medios de aristas opuestas del dodecaedro) se corresponden con las permutaciones tales que dadas dos parejas disjuntas de cubos intercambian los cubos en cada pareja y dejan invariante al quinto cubo.

Las consideraciones anteriores dan lugar a la pregunta: ¿hay otros grupos finitos de simetrías directas tridimensionales? En la introducción adelantamos que un resultado debido a F. Klein da una respuesta concluyente a esa pregunta.

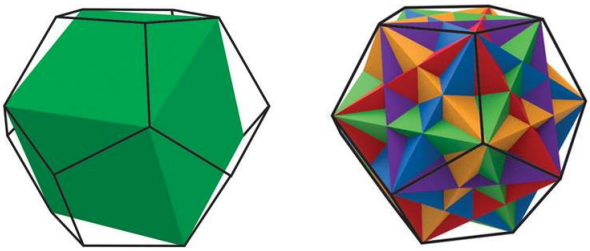


FIGURA 7. Un cubo y cinco cubos inscriptos en el Dodecaedro.

**Teorema 2.** *Un grupo finito de simetrías directas de* $\mathbb{R}^3$ *o es el grupo de las simetrías directas de un sólido plátonico, o es generado por una rotación de orden n (grupo cíclico), o es generado por dos rotaciones de ejes perpendiculares, una de orden n y otra de orden 2 (grupo diedral).*

Es un ejercicio instructivo dar ejemplos de poliedros cuyos grupos de simetrías sean cíclicos o diedrales. También sugerimos determinar (en el sentido del teorema de Klein) los grupos de simetrías (directas) de los poliedros de la figura 9.

## 3. Segunda clase: Poliedros semirregulares. Sólidos de Catalán.

Pappus le atribuye a Arquímedes el descubrimento de los poliedros semirregulares. El análisis que vamos a esbozar se basa en [1] que a su vez es una reformulación de lo escrito por Kepler (1571-1630) sobre esta cuestión.

Vamos a comenzar con una definición provisoria de poliedro semirregular, la cual la mejoraremos al final de la clase. Empezaremos por buscar los poliedros convexos cuyas caras sean todas polígonos regulares y tal que la configuración de caras que rodean un vértice sea la misma para todos los vértices. Lo primero será analizar las posibles configuraciones de polígonos alrededor de un vértice. Usaremos la siguiente notación: la terna $(a, b, c)$ es la configuración en un vértice si en el mismo inciden un polígono regular de $a$ lados, uno de $b$ lados y un tercero de $c$ lados, en ese orden cíclico -de izquierda a derecha o a la inversa-; de manera análoga una $t$-upla da la configuración correspondiente a $t$ polígonos regulares. Es importante tener presente que un ángulo sólido convexo no queda determinado por sus caras, salvo que estas sean tres.

Para familiarizarse con el concepto anterior pueden elegir algunos de los sólidos arquimedianos y escribir la configuración de sus vértices. De gran utilidad, en el sitio `http://bit.ly/1Uh8ePp` se representan todos los poliedros semirregulares y se muestra de forma dinámica cómo se pueden obtener a partir de los sólidos platónicos.

Pasamos a considerar las limitaciones que surgen para las configuraciones, estas restricciones acotarán suficientemente los casos factibles.

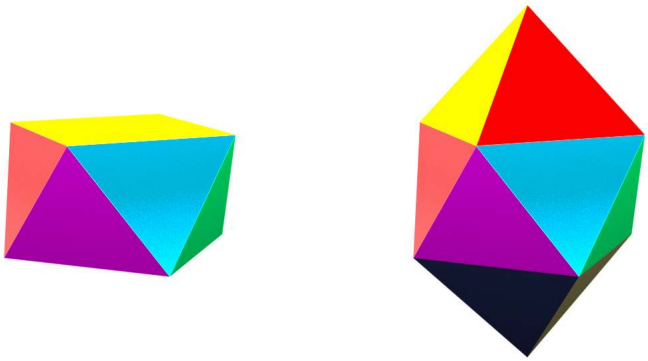


FIGURA 8. Un antiprisma de base cuadrada y un deltaedro de 16 caras.

**Lema 1.** *Si todas las caras de un poliedro convexo son polígonos regulares, entonces a lo sumo tres tipos de polígonos pueden aparecer rodeando un vértice.*

Demostración. Se deduce fácilmente teniendo en cuenta que los ángulos que forman un ángulo sólido estrictamente convexo deben sumar menos que dos rectos.

El siguiente lema tiene en su hipótesis la igualdad de configuración en los vértices e infiere fuertes restricciones sobre ellas.

**Lema 2.** *En un poliedro convexo en el cual los ángulos sólidos tienen la misma configuración, no son posibles las siguientes configuraciones:* $(a, b, c)$ *con* $a$ *impar y* $b \neq c$ *y* $(a, b, c, 3)$ *con* $a \neq c$.

Demostración. Es un ejercicio sencillo.

Basados en los lemas anteriores y haciendo un análisis cuidadoso (no es difícil, es largo) se puede probar el siguiente resultado.

**Teorema 3.** *Si $\mathcal{P}$ es un poliedro convexo, tal que todas sus caras son polígonos regulares y tal que la configuración de todos sus vértices es la misma, entonces las posibles configuraciones de los vértices de $\mathcal{P}$ son:* $(4,4,n)$, $(3,3,3,n)$ *y otras 13 que se explicitan al hacer la demostración. (La configuración* $(4,4,n)$ *corresponde a un prisma con base en un polígono de n lados y la configuración* $(3,3,3,n)$ *corresponde a un antiprisma.)*

Demostración. Las configuraciones de las que habla el teorema son exactamente aquellas que respetan los lemas anteriores. Es decir se encuentran analizando de forma exhaustiva todos los casos posibles a la luz de los lemas. Veamos a modo de ejemplo el caso en que las caras de nuestro hipotético poliedro son pentágonos y triángulos. De acuerdo al primer lema podemos tener un pentágono y hasta cuatro triángulos o dos pentágonos y hasta dos triángulos. La configuración (3,3,3,3,5) es factible -de hecho la realiza el Dodecaedro Romo-, la configuración (3,3,3,5) es realizada por un antiprisma de base pentagonal, la configuración (3,3,5) queda descartada por el segundo lema, la configuración (3,5,5) y (3,3,5,5) también son invalidadas por el segundo lema, la configuración (3,5,3,5) no es excluída por los lemas y de hecho es realizada por el Icosidodecaedro que está representado en la figura 10.

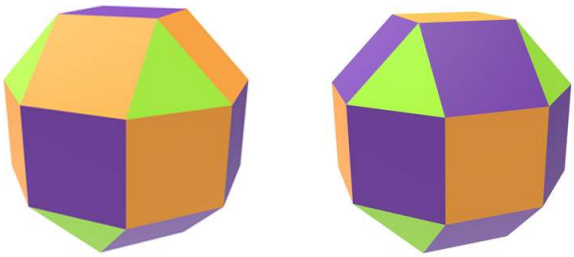


Figura 9. Poliedro de Miller y Rombicubo.

El teorema anterior excluyendo a los prismas y a los antiprismas nos da trece posibles configuraciones, podría ocurrir que algunas configuraciones no se realicen por poliedro alguno y también podría ocurrir que alguna configuración se realizara con poliedros diferentes. Kepler en su trabajo exhibió representaciones gráficas de trece poliedros alcanzando las trece configuraciones. Sin embargo la certeza de la unicidad -a menos de isometrías y homotecias- de la realización de cada configuración surge de un trabajoso teorema de rigidez probado por Cauchy, ese resultado asegura que dado un conjunto de caras con sus relaciones de incidencia, a lo sumo hay un poliedro convexo con esas caras que realiza las relaciones dadas.

Alrededor de 1960 Miller presentó un poliedro (al que a veces se le da su nombre) que parecía que podía calificar como décimocuarto sólido arquimediano -no prismático ni antiprismático-, ese poliedro está representado a la izquierda en la figura 9, es muy similar -pero no congruente- al poliedro que está a su derecha. Ambos poliedros tienen 24 vértices, pero el de Miller tiene menos simetrías -tiene 16 entre directas y no directas-, de manera que aunque todos los vértices de este poliedro tienen la misma configuración no hay forma de llevar un vértice dado en cualquier otro mediante una simetría del poliedro, esa es la explicación moderna de porqué no corresponde darle a este poliedro la categoria de semirregular. Así tenemos una

definición satisfactoria de poliedro semirregular: es un poliedro convexo cuyas caras son polígonos regulares y tal que que sus simetrías (directas y no directas) actúan de forma transitiva en sus vértices. Se constata que estos poliedros son exactamente los que enumeró Kepler.

Varios de los poliedros semirregulares se obtienen por truncamiento o rectificación de los poliedros platónicos -es un buen ejercicio obtener algunos de ellos-, sin embargo no todos ellos se obtienen de esa forma, en el sitio de internet ya referido (`http://bit.ly/1Uh8ePp`) se ilustra este hecho de manera muy persuasiva. Analizando la construcción de todos los poliedros semirregulares se concluye que todos ellos son inscriptibles en una esfera.

**Ejercicio 2.** *¿Es posible distinguir conceptualmente a los prismas y antiprismas de los demás poliedros semirregulares?*

FIGURA 10. El Icosidodecaedro (semirregular) y su dual (de Catalán).

Ya habíamos visto en la primera clase un procedimento para dualizar a los sólidos platónicos. Ese mismo método se puede aplicar para dualizar a los poliedros semirregulares, ¿porqué? Dualizando los poliedros semirregulares que no son prismas ni antiprismas, se obtiene la familia que se conoce con el nombre de Sólidos de Catalán, uno de ellos aparece en la figura 10. Recomiendo buscar y observar representaciones de todos ellos pensando en qué propiedades remarcables poseen.

## 4. Tercera clase: Poliedros totalmente transitivos.

El propósito de esta clase es encontrar todos los poliedros (eventualmente compuestos) que son totalmente transitivos respecto a su grupo de simetrías directas.

Para que esto nos lleve más allá de los poliedros platónicos debemos ampliar nuestra noción de poliedro, hasta ahora muy vaga si se trata de poliedros no convexos. Vamos a permitir que los poliedros no sean convexos y que sus caras sean polígonos no necesariamente convexos ni simples, podrían ser estrellas por ejemplo. También permitiremos que las caras se atraviesen, por ejemplo dos estrellas que se unen compartiendo un lado, incluso las caras podrían atravesarse y que su intersección no fuese un lado. La propiedad crucial que sí mantenemos es que cada arista pertenezca exactamente a dos caras. No queremos al inicio de esta clase formular una definición precisa de poliedro, podemos avanzar con eso pendiente, al final presentaremos alguna definición moderna.

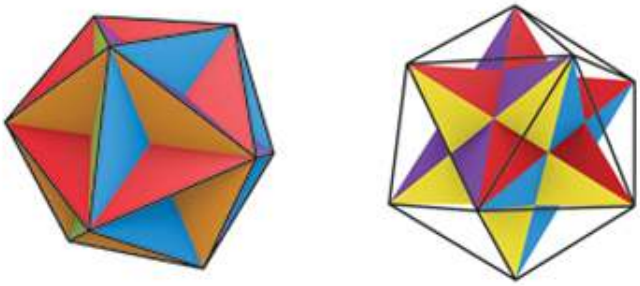


FIGURA 11. Dos poliedros estrellados con sus envolventes convexas. El gran dodecaedro y el pequeño dodecaedro estrellado.

La envolvente convexa de un subconjunto del espacio es el menor convexo que lo contiene. El núcleo de un poliedro $\mathcal{P}$ está formado por las "partes interiores" de las caras de $\mathcal{P}$, el núcleo de un poliedro convexo es vacío -consideren esta noción sólo a nivel intuitivo-. Antes de continuar con la lectura sugiero observar con atención al pequeño dodecaedro estrellado y al gran dodecaedro, en alguna representación que permita definir y observar totalmente sus caras y sus aristas, la figura 11 puede ser de ayuda, ¿cuáles serían sus caras si prentendemos que todas ellas sean congruentes?, ¿qué podemos decir de sus envolventes convexas?, ¿qué de sus núcleos?

Recordamos que un poliedro totalmente transitivo es uno que es transitivo en los vértices, en las aristas y en las caras. Ser transitivo en los vértices es que dados dos vértices existe una simetría directa del poliedro que lleva un vértice en el otro, análogamente para las aristas y las caras. Para chequear que se ha comprendido esta definición propongo que se responda la siguiente pregunta: ¿qué tipo de transitividad verifican los poliedros arquimedianos?, cabe distinguir dos casos, el de todas las simetrías o el de sólo las simetrías directas.

También nos conviene considerar por un momento la estrella octángula (figura 12), está formada por dos tetraedros regulares congruentes cuyas aristas se cortan perpendicularmente en sus puntos medios, consideramos que sus caras son las caras de los tetraedros, ¿qué clase de transitividad tiene este poliedro compuesto?

Ahora que está más o menos claro el tipo de objetos que buscamos, vamos a iniciar la búsqueda exhaustiva de ellos.

Así que partimos de un poliedro (eventualmente compuesto) $\mathcal{P}$ que es totalmente transitivo. Lo primero que observamos es que los vértices de $\mathcal{P}$ están en la superficie de una esfera, esto es consecuencia del teorema de Klein explicado en la primera clase. Las caras de $\mathcal{P}$ (polígonos asumidos planos) deben ser todas congruentes, esto es por la transitividad en las caras, además las caras son inscriptibles (en cfas.) y dada la transitividad en las aristas, se deduce que las caras son polígonos regulares. Le llamamos $\hat{\mathcal{P}}$ a la envolvente convexa de $\mathcal{P}$, ambos poliedros tienen los mismos vértices, esto es debido a que $\hat{\mathcal{P}}$ es la envolvente convexa de puntos en una esfera, por lo tanto el poliedro $\hat{\mathcal{P}}$ es transitivo en sus vértices, puesto que $\mathcal{P}$ lo es, luego todos los vértices de $\hat{\mathcal{P}}$ tienen la misma valencia (número de vértices adyacentes a un vértice dado), es un ejercicio sencillo (usando el teorema de Euler) verificar que la valencia de los vértices de $\hat{\mathcal{P}}$ es menor o igual a 5.

La estrategia de la búsqueda en la que estamos consiste en determinar los posibles poliedros convexos $\hat{\mathcal{P}}$ donde aparecen las caras de $\mathcal{P}$ como polígonos planos (convexos o no) cuyos vértices son vértices de $\hat{\mathcal{P}}$.

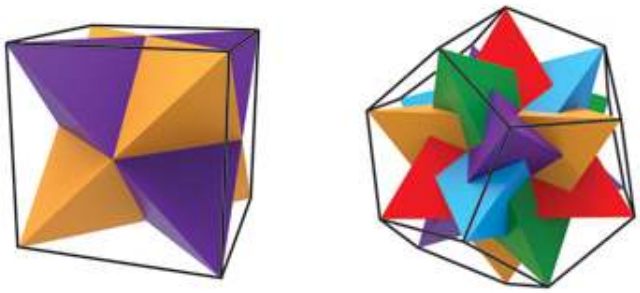


FIGURA 12. Dos poliedros compuestos (por dos y cinco tetraedros) con sus envolventes.

Bien, consideramos ahora la noción de estabilizador de un vértice, arista o cara de $\mathcal{P}$, es el sugbrupo de las simetrías del poliedro que deja invariante al vértice, arista o cara de $\mathcal{P}$ en cuestión. Para entender este concepto busquen determinar el estabilizador de un vértice, arista o cara del dodecaedro regular para su grupo de las simetrías directas.

Ahora vamos a hacer un pequeño cálculo que nos permitirá limitar los posibles $\hat{\mathcal{P}}$. Llamemos $N$ al número de simetrías de $\mathcal{P}$, $E$ a su número de aristas, $V$ a su número de vértices, $\eta$ al cardinal del estabilizador de una arista de $\mathcal{P}$ (es el mismo para todas las aristas) y $\psi$ al cardinal del estabilizador de un vértice de $\mathcal{P}$ (es el mismo para todos los vértices de $\mathcal{P}$). Las simetrías de $\mathcal{P}$ actúan sobre los vértices, aristas y caras de $\mathcal{P}$ de forma transitiva, en cualquiera de los tres casos tenemos una única órbita y sabemos que en cualquier acción de un grupo finito el cardinal de una órbita es el cociente del orden del grupo por el orden del estabilizador de un punto cualquiera de la órbita. Esto aplicado a los vértices y a las aristas nos da $N = V\psi = E\eta$. Por otro lado si $m$ es la valencia de los vértices de $\mathcal{P}$ resulta que $mV = 2E$. En definitiva $m\eta = 2\psi$ y dado que $m > 2$ resulta que a $\psi \geq 2$, lo cual implica que el estabilizador de los vértices de $\mathcal{P}$ no es trivial, estos estabilizadores deben contener una rotación de orden mayor que 1. Ahora las simetrías de $\mathcal{P}$ son simetrías de $\hat{\mathcal{P}}$. En conclusión los vértices de $\hat{\mathcal{P}}$ están en ejes de rotaciones -simetrías de $\hat{\mathcal{P}}$- de orden 2, 3, 4 o 5.

Claramente tenemos a los sólidos platónicos como posibles $\hat{\mathcal{P}}$, sin embargo rápidamente se descartan el tetraedro y el octaedro por no tener polígonos inscriptos adecuados. ¿Habrá otros posibles $\hat{\mathcal{P}}$?

Una estrategia para encontrar los $\hat{\mathcal{P}}$: tomamos un grupo de los del teorema de Klein y un punto $v$ (a posteriori vértice de $\hat{\mathcal{P}}$) en un eje de rotación (de orden 2, 3, 4 o 5), le aplicamos a $v$ todas las simetrías del grupo, salvo que el grupo sea cíclico la órbita de $v$ está formada exactamente por los vértices de $\hat{\mathcal{P}}$. Veamos un ejemplo, tomamos el grupo del cubo y elegimos $v$ en la mitad de una de las aristas del cubo (por $v$ pasa un eje de rotación de orden 2 del cubo), el $\hat{\mathcal{P}}$ resultante es el cubo-octaedro, el cual no tiene caras inscriptas adecuadas para generar ningún $\mathcal{P}$. Otro ejemplo, ahora tomamos el grupo del dodecaedro y $v$ en la mitad de una

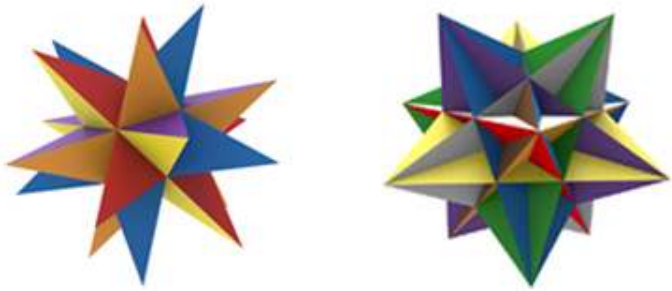


FIGURA 13. El gran dodecaedro estrellado y el gran icosaedro

de sus aristas, observamos que el $\hat{\mathcal{P}}$ resultante es el icosidodecaedro (figura 10). El icosidodecaedro sí tiene adecuados triángulos equiláteros inscriptos como para conformar un poliedro compuesto por cinco octaedros. Es sencillo verificar que los candidatos $\hat{\mathcal{P}}$ útiles a nuestros efectos son: el cubo, el dodecaedro, el icosaedro y el icosidodecaedro. El cubo da lugar a la estrella octángula (compuesta por dos tetraedros). El dodecaedro da lugar: al gran dodecaedro estrellado, a un poliedro compuesto de cinco cubos, a un poliedro compuesto de cinco tetraedros y a un poliedro compuesto por diez tetraedros. El icosaedro da lugar al pequeño dodecaedro estrellado, al gran dodecaedro y al gran icosaedro. Es decir, son catorce en total los poliedros (eventualmente compuestos) que son totalmente transitivos -incluyendo a los sólidos platónicos-. Todos ellos, salvo los platónicos, están representados en las figuras: 11, 12, 13 y 14.

FIGURA 14. Poliedros compuestos: cinco octaedros, cinco cubos y diez tetraedros.

Para terminar dos comentarios:

1. A lo largo del siglo pasado se fue dando un proceso de revisión de la noción de poliedro que condujo a que actualmente algunos autores destacados formulen que un poliedro es un grafo que satisface ciertas condiciones específicas. En ese sentido podemos comentar que Grünbaum en 1977 en [3] definió que un poliedro en $\mathbb{R}^3$ es una colección de polígonos (no necesariamente planos ni finitos) que verifican: i. cada arista es arista de exactamente dos caras; ii. dadas dos aristas $e$ y $e'$ existe una cadena de caras $f_1, ..., f_n$ tales que $e$ es arista de $f_1$, $e'$ es arista de $f_n$ y dos caras sucesivas

cualesquiera comparten una arista; iii. cada conjunto compacto corta sólo un número finito de caras. Un tiempo después en 1988 Farris en [2] propusó agregar una cuarta condición: que las caras que rodean cada vértice estén en ciclo. En 2007 el mismo Grünbaum en [4] formuló una nueva definición abstracta en la que los polígonos son grafos cíclicos finitos, a su vez agregó algunas condiciones más (además de la de Farris) que permiten la dualización de estos poliedros abstractos. Los poliedros compuestos que han aparecido en este trabajo no son poliedros propiamente en el sentido ni siquiera de la primera definición referida, pero ello no los vuelve menos interesantes.

2. Hay una noción de regularidad más estándar que la de transitividad total: un poliedro es regular si sus simetrías (directas e indirectas) actúan de forma transitiva en sus banderas. Una bandera es una terna $(v, a, c)$ donde $v$ es vértice de la arista $a$ y $a$ es arista de la cara $c$. Observen que ninguno de los poliedros que obtuvimos antes es regular (transitivo en las banderas) si sólo usamos simetrías directas. En el trabajo [3] se da una descripción completa de todos los poliedros regulares, como ya comentamos, permitiendo eventualmente que las caras sean polígonos infinitos no nesariamente planos.

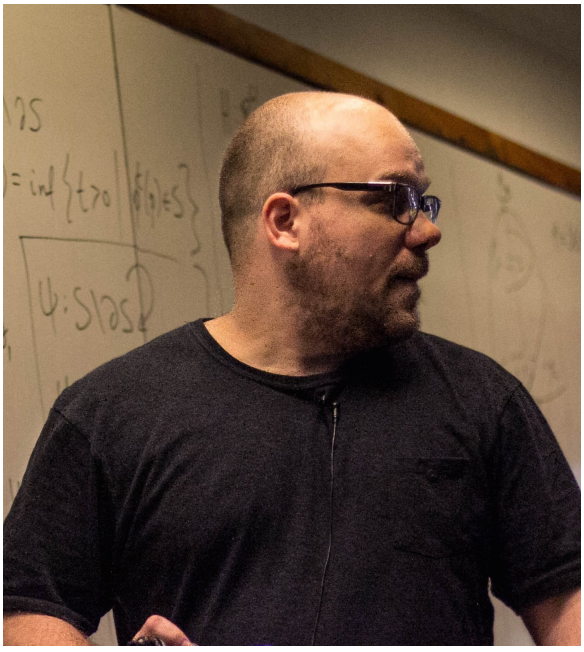
## Referencias

[1] P.R. Cromwell, Polyhedra, Cambridge University Press, 1999.
[2] S. L. Farris, Completely classifying all vertex-transitive and edge-transitive polyhedra, Geom. Dedicata 26 (1988) 111-124.
[3] B. Grünbaum, Regular polyhedra–old and new, Aequationes Math. 16 (1977) 1-20.
[4] B. Grünbaum, Graphs of polyhedra; polyhedra as graphs, Discrete Mathematics 307 (2007) 445-463.

CMat, Facultad de Ciencias, Universidad de la República. Montevideo, Uruguay.
*Email address*: angelperwyz@gmail.com

# CHARLAS PLENARIAS

CHARLA PLENARIA



# HOLOMORPHIC CURVES AND CELESTIAL MECHANICS

UMBERTO L. HRYNIEWICZ

Publicaciones Matemáticas del Uruguay
Volumen 17, Julio 2019, Páginas 61–77
S 0797-1443

# HOLOMORPHIC CURVES AND CELESTIAL MECHANICS

UMBERTO L. HRYNIEWICZ

Abstract. This expository article has two purposes. The first is to explain the connection between Poincaré's work on Celestial Mechanics and the Arnold Conjecture on the minimal number of fixed points of Hamiltonian diffeomorphisms. The second is to outline how holomorphic curve methods can be used to study Hamiltonian dynamics.

## 1. From Celestial Mechanics to Arnold Conjecture

In order to reach the origins of the problems that motivated Floer Theory we need to revisit the end of the XIX century. During this time Celestial Mechanics served as stage for Poincaré to implement some of the revolutionary ideas that laid the foundations of the field of Dynamical Systems. Poincaré was led to make a fundamental statement, known today as the Poincaré-Birkhoff Theorem [32], proved by Birkhoff [5] in 1913. It is a fixed point theorem for certain area-preserving annulus homeomorphisms, the precise statement will be recalled below. For decades mathematicians looked for a better understanding of the theorem, for generalizations to higher dimensions and to other phase spaces. We had to wait until the 1960's when ideas of Arnold finally allowed for proper understanding and generalization. The main purpose of this introduction is to provide more details about this remarkable chain of events.

Poincaré investigated a simplification of the 3-body problem: the *planar circular restricted* 3*-body problem* (PCR3BP). Two massive particles evolve in a relatively circular trajectory of the 2-body problem. A third particle (satellite) moves in the same plane of the first two, under their influence according to Newton's Law of Gravitation. However, the satellite is assumed not to disturb the movement of the massive particles. It turns out that in a rotating coordinate system, where the massive particles remain fixed, the satellite's movement is described by an *autonomous* Hamiltonian; see Section 4 for precise formulas.

Low energy levels (below lowest critical level) have three connected components, they project onto three so-called *Hill regions* of the configuration plane: two of which are bounded neighborhoods of the massive particles while the third is a neighborhood of infinity. We choose one bounded Hill region. The *mass ratio* $\mu \in (0, 1)$ is defined as the ratio between the mass of the particle outside the chosen Hill region and the total mass. The Hamiltonian turns out to have a smooth limit as $\mu \to 0$. The limiting Hamiltonian models the *rotating Kepler problem*, which is just the Kepler problem viewed in a rotating coordinate system.

The rotating Kepler problem is integrable, the integrals being the energy and the angular momentum. The Hill region around the light particle degenerates as $\mu \to 0$, while the boundaries of the other Hill regions converge to circles around the

©2019 PMU

limiting position of the heavy particle, which we call *center*. Using integrability it is not difficult to find circular orbits of the rotating Kepler problem. There are three of them: two inside the bounded Hill region, and a third one inside the unbounded Hill region. One of the two circular orbits in the bounded Hill region is *retrograde* while the other is *direct*, according to which they rotate around the center in the opposite or in the same sense of the relative position of the massive particles.

Poincaré was able to find continuations of the circular orbits when $\mu > 0$ is small enough. For this he devised a method, called the *continuation method*, which we now briefly describe. Choose a transverse section at a point of one of the orbits (recall that energy levels are three-dimensional). There is a well-defined local return map, and the orbit corresponds to a fixed point. The transverse section remains, of course, transverse as we slightly perturb $\mu$. If the derivative of the return map at this fixed point is non-singular for $\mu = 0$ then the fixed point is isolated, and the implicit function theorem gives isolated fixed points of the local return map for all $\mu > 0$ small enough. These correspond to continued periodic orbits. They retain the geometric properties of the unperturbed orbits, namely one is retrograde and the other is direct.

Now, using a method introduced by Levi-Civita [31] which we shall illustrate in Section 4, collisions with the massive particle can be regularized to obtain an energy level diffeomorphic to $S^3$. The movement evolves smoothly there. In doing so, an ambiguity in the representation of states is introduced: the energy level is antipodal symmetric, each state being represented twice as a pair of antipodal points. We ignore this problem for now, keep the ambiguity and say that the regularized dynamics is a two-to-one lift of the unregularized dynamics. The same regularization process applies of course to the rotating Kepler problem. The direct and retrograde orbits lift under this process to a pair of periodic orbits forming a Hopf link. When $\mu = 0$ Poincaré uses the angular momentum to show that these two lifted orbits bound a very special embedded annulus, a so-called *global surface of section* (GSS).

**Definition 1.1.** Let $\phi^t$ be a smooth flow on a smooth closed 3-manifold $M$. A global surface of section for $\phi^t$ is an embedded compact surface $S \hookrightarrow M$ such that

(i) The boundary $\partial S$ consists of periodic orbits, and the interior $S \setminus \partial S$ is transverse to $\phi^t$.
(ii) For every $p \in M \setminus \partial S$ there exist $t_- < 0 < t_+$ such that $\phi^{t_\pm}(p) \in S$.

In the presence of a global surface of section the dynamical properties of the flow $\phi^t$ get encoded in the first return map

$$\psi : S \setminus \partial S \to S \setminus \partial S \qquad \psi(p) = \phi^{\tau(p)}(p) \tag{1}$$

where the first return time $\tau : S \setminus \partial S \to (0, +\infty)$ is defined as

$$\tau(p) = \inf\{t > 0 \mid \phi^t(p) \in S\}. \tag{2}$$

Poincaré showed that the annular global surface of section for $\mu = 0$ could be continued for $\mu > 0$ small enough. In fact, the continuation method applies to continue its boundary, as explained above. Using a $C^\infty$-small isotopy, one continues the annulus "by hand". The task is now to show that the continued annulus is again a global surface of section. By transversality, the return map is well-defined on compact subsets of the interior, and globally encodes the dynamics on arbitrarily large compact subsets of the complement of the boundary orbits. The situation

near the boundary is more delicate, but it can be nicely controlled by the robust positivity of the transverse rotation numbers of the boundary orbits.

A more detailed discussion about global surfaces of section is found in Section 3, where the connection to holomorphic curves will be made.

Why is it nice to have a global surface of section? Because it allows for two-dimensional methods to come into play and shed light onto the three-dimensional flow. This is precisely what Poincaré did, and in doing so he was able to leave the seed for Symplectic Topology and Floer Theory. Let us explain this point, which is main goal of this introduction.

Using the linearized dynamics of the direct and retrograde orbits, one can smoothly extend the first return map up to boundary. Each boundary component is mapped onto itself. The symplectic nature of the flow allows us to find a smooth 2-form on the annulus which is preserved by the extended return map. This 2-form defines an area form in the interior, but vanishes on the boundary. It turns out that there is a homeomorphism between Poincaré's annulus and $\mathbb{R}/\mathbb{Z} \times [0,1]$ that is smooth in the interior and pulls the 2-form back to the standard area element. Such a conjugating homeomorphism is not differentiable on the boundary.

Thus, Poincaré was led to the problem of studying qualitative properties of area- and orientation-preserving homeomorphisms of the closed annulus that preserve boundary components. Consider such a homeomorphism $f$ on $\mathbb{R}/\mathbb{Z} \times [0,1]$. It can be lifted to a homeomorphism $F$ of $\mathbb{R} \times [0,1]$. If we denote its components by $F(x,y) = (X(x,y), Y(x,y))$ then

$$f(x+\mathbb{Z}, y) = (X(x,y)+\mathbb{Z}, Y(x,y)) \qquad \forall (x,y) \in \mathbb{R} \times [0,1]. \tag{3}$$

Of course, there are infinitely many choices of $F$, and they all differ by translation by an integer. It follows that

$$X(x+1,y) = X(x,y)+1 \qquad Y(x+1,y) = Y(x,y) \tag{4}$$

holds for every $(x,y) \in \mathbb{R} \times [0,1]$.

Poincaré's seminal contribution to Symplectic Dynamics and Topology was the formulation of the following statement.

**Theorem 1.2** (Poincaré's last geometric theorem [32])**.** *Suppose that*

$$x \mapsto X(x,1) - x \qquad \textit{and} \qquad x \mapsto X(x,0) - x$$

*are non-vanishing functions with opposite signs. Then $F$ has at least two fixed points which are not integer translations of each other. In particular, they project onto distinct fixed points of $f$.*

This is a truly remarkable statement. Note that $X(x,1) - x$ and $X(x,0) - x$ are 1-periodic functions of $x$ in view of (4). The assumption that they have definite and opposite signs means that $f$ moves boundary components in opposite directions, i.e. $f$ "twists the annulus".

Theorem 1.2 can be strengthened as follows. Let

$$\pi_{\mathbb{R}} : \mathbb{R} \times [0,1] \to \mathbb{R} \qquad (x,y) \mapsto x$$

denote the projection onto the first component, and consider rotation numbers

$$\rho_0 = \lim_{n \to +\infty} \frac{\pi_{\mathbb{R}} \circ F^n(x,0)}{n} \qquad \rho_1 = \lim_{n \to +\infty} \frac{\pi_{\mathbb{R}} \circ F^n(x,1)}{n} \tag{5}$$

of $F$ restricted to boundary components. The *twist condition* asks that

$$\rho_0 \neq \rho_1. \tag{6}$$

A different choice of $F$ will change $\rho_0, \rho_1$ by addition of a common integer, hence (6) is really a condition on $f$. The assumptions of Theorem 1.2 force $\rho_0$ and $\rho_1$ to have opposite signs, in particular (6) holds.

**Theorem 1.3** (Poincaré-Birkhoff)**.** *If $\rho_0 \neq \rho_1$ then for every $p/q$ between $\rho_0$ and $\rho_1$ there exist $P, P' \in \mathbb{R} \times [0,1]$ satisfying*

$$F^q(P) = P + (p, 0) \qquad F^q(P') = P' + (p, 0)$$

*Moreover, the orbits of these points project to distinct periodic orbits of the map $f$. In particular, $f$ admits infinitely many periodic points.*

The above form of the theorem forces existence of infinitely many periodic orbits in the PCR3BP when $\mu > 0$ is small enough and the energy is low. This amounts for one of the fist major victories of Poincaré's ideas.

Birkhoff's proof [5] did not shed light onto generalizations to higher dimensions or to different phase spaces. However, Poincaré was already able to prove Theorem 1.2 in special cases, and his arguments did lend to such generalizations. This was remarked by V. I. Arnold in the 1960's; the following discussion is based on [3, appendix 9]. Arnold realized that, in the smooth case, Theorem 1.2 would be consequence of a certain fixed point theorem for a special class of diffeomorphisms of the torus. Arnold's statement can be generalized to all symplectic manifolds. This was, in fact, only one of many statements that came to be known as *Arnold Conjectures*. These conjectures led Floer [10, 11, 12] to develop what is nowadays known as *Floer theory*.

Arnold considered certain diffeomorphisms of the $2n$-dimensional torus $\mathbb{R}^{2n}/\mathbb{Z}^{2n}$. The space $\mathbb{R}^{2n}$ carries a standard symplectic form

$$\omega_0 = \sum_{j=1}^{n} dx_j \wedge dy_j$$

where coordinates are denoted by $(x_1, \dots, x_n, y_1, \dots, y_n)$. Since $\omega_0$ is invariant by translations, it descends to a symplectic form on $\mathbb{R}^{2n}/\mathbb{Z}^{2n}$ again denoted by $\omega_0$. Arnold considers diffeomorphisms $\Psi : \mathbb{R}^{2n}/\mathbb{Z}^{2n} \to \mathbb{R}^{2n}/\mathbb{Z}^{2n}$ with the following properties:

(H1) $\Psi$ is isotopic to the identity.
(H2) $\Psi$ is symplectic: $\Psi^* \omega_0 = \omega_0$.
(H3) $\Psi$ is exact: it admits a lift to $\mathbb{R}^{2n}$ whose center of gravity vanishes.

Let us explain these in more detail. By (H1) the map $\Psi$ lifts via the projection $\mathbb{R}^{2n} \to \mathbb{R}^{2n}/\mathbb{Z}^{2n}$ to a symplectic diffeomorphism $\widetilde{\Psi}$ of $(\mathbb{R}^{2n}, \omega_0)$ satisfying

$$\widetilde{\Psi}(z + k) = \widetilde{\Psi}(z) + k \qquad \forall z \in \mathbb{R}^{2n},\ \forall k \in \mathbb{Z}^{2n}. \tag{7}$$

Hence we can write $\widetilde{\Psi}(z) = z + \Delta(z)$ for some $\mathbb{Z}^{2n}$-periodic function $\Delta$. The center of mass of $\widetilde{\Psi}$ is defined by the integral

$$\int_{[0,1]^{2n}} \Delta(z) \tag{8}$$

There are of course infinitely many lifts and they all differ by translation by a vector in $\mathbb{Z}^{2n}$. In particular, the same holds for their centers of mass. Hence condition

(H3) is equivalent to saying that the center of mass of any lift of $\Psi$ belongs to $\mathbb{Z}^{2n}$. We may simply refer to $\Psi$ as an *exact symplectic diffeomorphism* of $(\mathbb{R}^{2n}/\mathbb{Z}^{2n}, \omega_0)$.

*Arnold Conjecture (Particular case): Exact symplectic diffeomorphisms of the standard symplectic torus $(\mathbb{R}^{2n}/\mathbb{Z}^{2n}, \omega_0)$ must have at least $2n+1$ fixed points. Morevoer, if all fixed points are non-degenerate then there must be at least $2^{2n}$ fixed points.*

To illustrate the ideas we prove the conjecture in case the exact symplectic diffeomorphism is $C^1$-close to the identity. Let $\widetilde{\Psi}$ be the unique lift to $\mathbb{R}^{2n}$ that is $C^1$-close to the identity. Identifying $\mathbb{R}^{2n} = \mathbb{R}^n_x \times \mathbb{R}^n_y$, we can write in components

$$\widetilde{\Psi}(x, y) = (X(x, y), Y(x, y)).$$

The identity $\widetilde{\Psi}^* \omega_0 = \omega_0$ is equivalent to saying that

$$\eta = \sum_{j=1}^{n} (y_j - Y_j) dx_j + (X_j - x_j) dY_j$$

is a closed 1-form. $C^1$-closeness to the identity ensures that the map $(x, y) \mapsto (x, Y)$ is a diffeomorphism of $\mathbb{R}^{2n}$. Hence we may view $\eta$ in $(x, Y)$-space and find a real-valued function $F(x, Y)$ such that $dF = \eta$. In other words

$$\begin{aligned} X - x &= D_2 F(x, Y) \\ y - Y &= D_1 F(x, Y) \end{aligned} \tag{9}$$

holds, and defines the map implicitly. Finally we make use of the vanishing of the center of mass to conclude that $F$ is $\mathbb{Z}^{2n}$-periodic. Hence, by Morse theory, the function on $\mathbb{R}^{2n}/\mathbb{Z}^{2n}$ induced by $F$ has at least $2n + 1$ critical points. By (9) this means that $\Psi$ has at least $2n + 1$ fixed points. The non-degenerate case follows similarly by the well-known estimates on the number of critical points of a Morse function in terms of sum of Betti numbers. The proof is complete.

The Arnold conjecture for standard symplectic tori was confirmed by a celebrated result due to Conley and Zehnder [8]. Their arguments make use of degree theory in infinite dimensions in order to take advantage of the fact that the classical action functional is a compact perturbation of a quadratic form of infinite index and co-index.

The generalization to arbitrary symplectic manifolds requires the notion of Hamiltonian diffeomorphism. Let $(W, \omega)$ be a symplectic manifold and let $\{\psi^t\}_{t\in[0,1]}$ be a smooth isotopy of $W$ starting at the identity $\psi^0 = id$, consisting of symplectic diffeomorphisms. Let $X_t$ be the vector field generating $\psi^t$, i.e. $\frac{d}{dt}\psi^t = X_t \circ \psi^t$. Differentiating we obtain

$$0 = \frac{d}{dt}(\psi^t)^* \omega = (\psi^t)^* \mathscr{L}_{X_t} \omega = (\psi^t)^* d i_{X_t} \omega \tag{10}$$

from where we see that $i_{X_t}\omega$ is closed, for every $t$. A diffeomorphism $\phi$ of $W$ isotopic to the identity is a *Hamiltonian diffeomorphism* of $(W, \omega)$ if it can be written as $\phi = \psi^1$ for some isotopy $\psi^t$ as above such that $i_{X_t}\omega$ is exact for every $t$. The set of Hamiltonian diffeomorphisms of $(W, \omega)$ will be denoted by $\mathrm{Ham}(W, \omega)$.

*Arnold Conjecture: Let $(W, \omega)$ be a closed symplectic manifold and let $\phi \in \mathrm{Ham}(W, \omega)$. Then*

$$\#\mathrm{Fix}(\phi) \geq \inf\{\#\mathrm{Crit}(f) \mid f \in C^\infty(W, \mathbb{R})\}$$

*and, moreover, if all fixed points of $\phi$ are non-degenerate then*

$$\#\mathrm{Fix}(\phi) \geq \sum_i b_i(W)$$

*where $b_i(W)$ denotes the rank of $H_i(W, \mathbb{Z})$.*

The connection with the previously stated conjecture is that the set of exact symplectic diffeomorphisms of $(\mathbb{R}^{2n}/\mathbb{Z}^{2n}, \omega_0)$ coincides precisely with $\mathrm{Ham}(\mathbb{R}^{2n}/\mathbb{Z}^{2n}, \omega_0)$.

To close the circle of ideas and conclude this introduction we need to explain why Arnold's conjecture on $(\mathbb{R}^2/\mathbb{Z}^2, \omega_0)$ implies Theorem 1.2 (in the smooth case). Consider an area and orientation preserving diffeomorphism $f$ of $\mathbb{R}/\mathbb{Z} \times [0, 1]$ that preserves boundary components. We make a simplifying assumption for the sake of exposition:

$(*)$ $f$ is a rotation near the boundary.

If this assumption is dropped then the discussion below can be modified but the details would draw our attention to unimportant technical issues.

By the assumptions of Theorem 1.2 and by hypothesis $(*)$, there is a lift $F(x, y) = (X, Y)$ of $f$ to $\mathbb{R} \times [0, 1]$ satisfying

$$\begin{aligned} F(x, y) &= (x + \alpha_0, y) \ \text{ if } y \sim 0 \\ F(x, y) &= (x + \alpha_1, y) \ \text{ if } y \sim 1 \end{aligned} \tag{11}$$

where $\alpha_0 \alpha_1 < 0$.

Consider $d_1, d_2 > 0$ to be fixed *a posteriori*, and set $d := d_1 + d_2$. On the thicker strip $\mathbb{R} \times [0, 2 + d]$ consider the map $\Psi$ defined by:

- If $y \in [0, 1]$ then $\Psi(x, y) = F(x, y)$,
- If $y \in [1, 1 + d_1]$ then $\Psi(x, y) = (x + \alpha_1, y)$,
- If $y \in [1 + d_1, 2 + d_1]$ then

$$\Psi(x, y) = (X(x, 2 + d_1 - y), 2 + d_1 - Y(x, 2 + d_1 - y)),$$

- If $y \in [2 + d_1, 2 + d]$ then $\Psi(x, y) = (x + \alpha_0, y)$.

It follows that $\Psi$ is a diffeomorphism of $\mathbb{R} \times [0, 2 + d]$ which agrees with the map $(x, y) \mapsto (x + \alpha_0, y)$ near the boundary.

We extend $\Psi$ to a map $\Psi : \mathbb{R}^2 \to \mathbb{R}^2$ in a $(2 + d)$-periodic fashion in the $y$-variable. Then the extended $\Psi$ is an area and orientation preserving diffeomorphism of $\mathbb{R}^2$ that commutes with the additive action of the lattice $\Gamma = \mathbb{Z} \times (2 + d)\mathbb{Z}$. As such, it descends to a diffeomorphism on the torus $\mathbb{R}^2/\Gamma$.

If we write $\Psi(x, y) = (x + g(x, y), y + h(x, y))$ then the center of mass of $\Psi$ is

$$\left( \int_{[0,1]\times[0,2+d]} g(x, y), \int_{[0,1]\times[0,2+d]} h(x, y) \right)$$

Now we wish to show that if $d_1, d_2$ are suitably chosen then the center of mass of $\Psi$ vanishes. To see this we first note that the second coordinate vanishes: on $[0, 1] \times ([1, 1 + d_1] \cup [2 + d_1, 2 + d])$ the function $h(x, y)$ vanishes, and by the change

of variables formula

$$\begin{aligned}
&\int_{[0,1]\times[1+d_1,2+d_1]} h(x,y)\\
&= \int_{[0,1]\times[1+d_1,2+d_1]} 2+d_1-Y(x,2+d_1-y)-y\\
&= \int_{[0,1]\times[1+d_1,2+d_1]} -(Y(x,2+d_1-y)-(2+d_1-y))\\
&= \int_{[0,1]\times[1+d_1,2+d_1]} -h(x,2+d_1-y) = -\int_{[0,1]^2} h(x,y).
\end{aligned}$$

As for the first coordinate, note that

$$\int_{[0,1]\times[1+d_1,2+d_1]} g(x,y) = \int_{[0,1]^2} g(x,y)$$

from where we get

$$\int_{[0,1]\times[0,2+d]} g(x,y) = \alpha_1 d_1 + \alpha_0 d_2 + 2\int_{[0,1]^2} g(x,y)$$

Since $\alpha_0\alpha_1 < 0$ we can choose $d_1, d_2$ positive in such a way that

$$\alpha_1 d_1 + \alpha_0 d_2 = -2\int_{[0,1]^2} g(x,y)$$

forcing the first coordinate of the center of mass to vanish, as desired.

Hence, the validity of Arnold's conjecture on the 2-torus will force $\Psi$ to have at least three fixed points. By symmetry, $F$ has at least two fixed points, as claimed by Theorem 1.2.

## 2. Hamiltonian Dynamics and PDEs

An important tool for understanding Hamiltonian dynamics is the underlying variational structure, which incarnates in many forms. This section is devoted to explaining the origins of the PDE methods used by Floer to explore the variational structure and attack the Arnold Conjecture. These methods are, of course, based on the seminal work of Gromov [19] where holomorphic curves were first introduced as a tool to study symplectic geometry.

We start with a quick revision of the basics on symplectic geometry. Phase spaces of Hamiltonian systems are *symplectic manifolds*. These are pairs $(W,\omega)$ consisting of a smooth manifold $W$ and a closed 2-form $\omega$ which is *non-degenerate*: at every point, $\ker\omega$ is the trivial vector space. Such forms are called *symplectic forms*.

Symplectic manifolds are necessarily even dimensional, but in fact more is true: locally they are just copies of open subsets of the *standard symplectic vector space* $(\mathbb{R}^{2n},\omega_0)$ where

$$\omega_0 = \sum_{j=1}^{n} dx_j \wedge dy_j. \tag{12}$$

Here the coordinates on $\mathbb{R}^{2n}$ are denoted by $(x_1,\dots,x_n,y_1,\dots,y_n)$. This is the content of *Darboux's theorem*.

Corrections to the discussion on Floer's chain complex and a misuse of the expression "center of mass" have been added as erratum at the end of the volume.

Hamiltonian systems on $(W,\omega)$ are determined by (possibly time dependent) functions $H_t : W \to \mathbb{R}$ as follows. The *symplectic gradient*, also called the *Hamiltonian vector field*, of $H_t$ is the non-autonomous vector field $X_{H_t}$ defined by

$$\omega(X_{H_t}, \cdot) = dH_t$$

for each $t$. The function $H_t$ is called the *Hamiltonian*. In symplectic coordinates provided by Darboux theorem, the ODE

$$\dot{x}(t) = X_{H_t}(x(t)) \tag{13}$$

is nothing but Hamilton's equations of motion. The dynamics on $(W,\omega)$ induced by this ODE is *Hamiltonian dynamics*.

When the Hamiltonian $H$ is time independent then its values are preserved by the flow of $X_H$ since $dH(X_H) = \omega(X_H, X_H) = 0$. Then $H$ can be thought of as *energy*, and the system is conservative.

**The non-autonomous case.** Let us assume that $H_t$ is periodic in time, say of period 1. This imposes no loss of generality, as it turns out. Consider the loop space $\mathscr{L}$ of $W$, defined as a space of maps $\mathbb{R}/\mathbb{Z} \to W$ (regularity is not specified as this is an informal discussion). If $\omega$ is exact then one considers on $\mathscr{L}$ the *action functional* from classical mechanics

$$c \mapsto \int_{\mathbb{R}/\mathbb{Z}} c^*\lambda + \int_0^1 H_t(c(t))\, dt$$

where $\lambda$ is a primitive of $\omega$. Since $\omega$ might not be exact – and it will never be when $W$ is closed – we need to make additional assumptions.

Assume that $(W,\omega)$ is aespherical in the sense that $\omega$ and $c_1(TW,\omega)$ vanish on $\pi_2(M)$. Here $c_1(TW,\omega)$ denotes the first Chern class of the symplectic vector bundle $(TW,\omega)$. For the reader not familiar with Chern classes, note that its vanishing on $\pi_2(M)$ is equivalent to the following condition: for every map $f : S^2 \to W$ the symplectic vector bundle $f^*(TW,\omega)$ over $S^2$ is trivial. Let $\mathscr{L}_0 \subset \mathscr{L}$ be the connected component consisting of contractible loops. We define

$$\mathcal{A}_H : \mathscr{L}_0 \to \mathbb{R} \qquad c \mapsto \int_{\mathbb{D}} v^*\omega + \int_{\mathbb{R}/\mathbb{Z}} H_t(c(t))dt$$

where $v : \mathbb{D} \to W$ is a smooth map satisfying $v(e^{i2\pi t}) = c(t)$. The aesphericity assumption guarantees that $\mathcal{A}_H(c)$ does not depend on the choice of $v$. The Euler-Lagrange equations are precisely Hamilton's equations of motion. Hence, finding (contractible) periodic motions amounts to finding critical points of $\mathcal{A}_H$.

Deep results in Symplectic Topology have been established by exploiting this variational structure. Sometimes one can define so-called *symplectic capacities* as special critical values of $\mathcal{A}_H$. These invariants shed new light onto the theory of Hamiltonian systems. In Gromov's seminal work [19] one finds the first symplectic capacity, so-called *Gromov width*, although the appropriate formalism was introduced later by Ekeland and Hofer who also defined many new symplectic capacities.

There are major difficulties in the analysis of $\mathcal{A}_H$. Critical points have infinite index and co-index, in any reasonable sense. Moreover, $\mathcal{A}_H$ is unbounded from above and below. If one follows any kind of anti-gradient flow of $\mathcal{A}_H$ then critical points and values are expected to be missed, and sublevel sets do not change topology as we pass critical levels.

Floer's idea is to look at the $L^2$ anti-gradient flow equation, even knowing that it does not define a flow. To be more concrete, consider $\omega$-compatible and 1-periodic almost complex structures $J_t$ on $W$. These are 1-parameter families of fields of endomorphisms $J_t$ of $TW$ such that for every $t$ there holds $J_{t+1} = J_t$, $J_t^2 = -I$ and $\omega(\cdot, J_t\cdot)$ is a Riemannian metric. Then

$$\langle X, Y\rangle_{L^2} = \int_{\mathbb{R}/\mathbb{Z}} \omega(c(t))(X(t), J_t(c(t))Y(t))\ dt$$

defines an inner-product on vector fields along $c$ (thought of as tangent vectors of the loop space, based at $c$). The $L^2$ anti-gradient of $\mathcal{A}_H$ at $c$ is the vector field

$$-J_t(c(t))[\dot{c}(t) - X_{H_t}(c(t))]$$

along $c$. If $u = u(s,t)$ is a map $\mathbb{R}\times\mathbb{R}/\mathbb{Z} \to W$, thought of as a path of loops, then the anti-gradient flow equation is *Floer's equation*

$$u_s + J_t(u)[u_t - X_{H_t}(u)] = 0. \tag{14}$$

It is a non-linear elliptic equation, a compact perturbation of the Cauchy-Riemann equation associated to $J$. Its analytical properties suffice to define some kind of Morse homology of $\mathcal{A}_H$, called *Floer homology*. Note that if $u(s,t)$ solves (14) then $u(s+s_0,t)$ is also a soluation, for every $s_0$, but they will be declared equivalent for the purpose of counting solutions.

Let us pause and describe a simple instance of Floer's construction. By the aesphericity condition every contractible 1-periodic solution $x : \mathbb{R}/\mathbb{Z} \to W$ of (13) has a well-defined Conley-Zehnder index $\mathrm{CZ}(x) \in \mathbb{Z}$. We shall not describe the Conley-Zehnder index, referring to [4] for details; we shall only say that it is related to winding properties of the linearization of the ODE (13).

Floer assumes that contractible 1-periodic solutions of (13) are non-degenerate in the sense that 1 is not an eigenvalue of the corresponding linearized flow. This is like saying that the action functional is Morse. He considers the vector space $\mathrm{CF}_*(H)$ over $\mathbb{Z}/2\mathbb{Z}$ freely generated by the contractible 1-periodic solutions of (13), graded by the Conley-Zehnder index. It turns out that when $J_t$ is generic – in some precise sense that we will not explain here – the following holds: if $x, y$ are generators of $\mathrm{CF}(H)$ satisfying $\mathrm{CZ}(x) - \mathrm{CZ}(y) = 1$ then the number of equivalence classes of solutions $u(s,t)$ of Floer's equation (14) satisfying

$$\lim_{s\to-\infty} u(s,t) = x(t) \qquad \lim_{s\to+\infty} u(s,t) = y(t)$$

is finite. Denote the number of such equivalence classes of solutions by $n(x,y)$. Floer defines a degree $-1$ differential

$$\delta_{H,J}(x) = \sum_{\mathrm{CZ}(y)=\mathrm{CZ}(x)-1} (n(x,y) \mod 2)\ y \tag{15}$$

on $\mathrm{CF}(H)$. It turns out that the homology of the chain complex $(\mathrm{CF}_*(H), \delta_{H,J})$ is independent of the pair $(H,J)$ and is canonically isomorphic to the singular homology $H_*(W, \mathbb{Z}/2\mathbb{Z})$. This explains why the sum of Betti numbers is a lower bound on the number of contractible 1-periodic solutions of (13), at least in the non-degenerate case.

This homology theory carries a richer structure: a filtration by action values. This fact is used to define special critical values, which in turn are used to construct invariants. Hence the variational structure of $\mathcal{A}_H$ sheds light onto the geometry of

the phase space. But we are also very much interested in the reverse direction: the properties of $(W, \omega)$ might force $\mathcal{A}_H$ to have special critical points. These periodic orbits can be used to study global properties of the underlying Hamiltonian system. For the basics in Floer homology we refer to the book [4] by Audin and Damian, and the notes [33] by Salamon.

**The autonomous case.** This story has a beautiful analogue for autonomous systems, which is related to contact geometry and is due to Hofer [20]. Assume $H$ is time-independent and look at a regular level $M \subset W$. There is no loss of regularity to suppose that $M = H^{-1}(0)$ since we may add a constant to $H$ without altering the dynamics.

Assume that $\omega$ has a primitive $\lambda$. From now on we also write $\lambda$ to denote the pull-back of this primitive to $M$ by the inclusion map $M \hookrightarrow W$, with no fear of ambiguity. The crucial assumption to be made is that $\lambda$ defines a *contact form* on $M$, i.e. $\omega$ is pointwise non-degenerate on the associated *contact structure* $\xi := \ker \lambda \subset TM$.

The action functional on loops in $M$ is now just $\int \lambda$. Hofer considered the so-called *symplectization*. It is defined as $\mathbb{R} \times M$ equipped with the symplectic form $d(e^a \lambda)$, where $a$ denotes the $\mathbb{R}$-component and we see $\lambda$ as an $\mathbb{R}$-invariant 1-form on $\mathbb{R} \times M$ (with respect to the action of $(\mathbb{R}, +)$ on the first component).

We don't insist in parametrizing the dynamics on $M$ by $X_H$. Instead, we parametrize it by the unique non-vanishing multiple $X_\lambda$ of $X_H$ satisfying $\lambda(X_\lambda) = 1$. It is called the *Reeb vector field* associated to $\lambda$.

In [20] Hofer considered $\mathbb{R}$-invariant almost complex structures $\tilde{J}$ on $\mathbb{R} \times M$ that send $\frac{\partial}{\partial a} \mapsto X_\lambda$, and have the property that $\omega(\cdot, \tilde{J}\cdot)$ is an inner-product on $\xi$. As before, we see here $X_\lambda$ and $\xi$ as $\mathbb{R}$-invariant objects on $\mathbb{R} \times M$. Such a $\tilde{J}$ is compatible with $d(e^a \lambda)$ in the sense explained before.

The $L^2$ gradient flow equation is now written as the Cauchy-Riemann equation $\bar{\partial}_{\tilde{J}} = 0$ associated to $\tilde{J}$. Its solutions are *holomorphic curves*, but here the situation differs drastically from the one dealt by Gromov in [19]: domains necessarily need to be punctured Riemann surfaces. The behavior of these curves near their ends, under a certain finite-energy condition introduced by Hofer, allowed for Morse-homological constructions, giving rise to *Contact Homology* and *Symplectic Field Theory* [9].

## 3. Global surfaces of section

As explained in the introduction, Poincaré constructed annulus-like global surfaces of section in the context of the PCR3BP. Another early and powerful existence result is the following.

**Theorem 3.1** (Birkhoff [6])**.** *The geodesic flow on the unit tangent bundle of any positively curved Riemannian metric on $S^2$ admits a global surface of section.*

In fact, we can be more precise. Consider an embedded closed geodesic $\gamma(t)$ of length $L$ in a positively curved Riemannian 2-sphere, parametrized with unit speed. Choose an ambient orientation and for each $t \in \mathbb{R}/L\mathbb{Z}$ let $n(t)$ be the unit vector at $\gamma(t)$ such that $\{\dot{\gamma}(t), n(t)\}$ is a positive orthonormal frame. The Birkhoff annulus associated to $\gamma$ is

$$A_\gamma = \{\cos\theta\ \dot{\gamma}(t) + \sin\theta\ n(t) \mid t \in \mathbb{R}/L\mathbb{Z},\ \theta \in [0, \pi]\}. \tag{16}$$

This is an embedded annulus in the unit sphere bundle, and Birkhoff showed that it is a global surface of section whenever the Gaussian curvature is positive everywhere.

Birkhoff's proof is very specific to geodesic flows and does not shed much light into the general existence problem.

A very general theory of global surfaces of section exists, for arbitrary flows on 3-manifolds. It is the outcome of the work of many mathematicians during the XX century, including Schwartzman [34], Fried [16], Ghys [18] and Sullivan [35]. However, most of the statements make use of dynamical hypotheses which are very hard to be checked. This in contrast to Theorem 3.1.

Here we focus on existence statements based on the theory developed by Hofer, Wysocki and Zehnder (HWZ) since they resonate with Birkhoff's statement in the sense that their hypotheses are quite concrete and geometric.

Consider $\mathbb{R}^4$ with coordinates $(x_1, y_1, x_2, y_2)$ and its standard symplectic form $\omega_0 = dx_1 \wedge dy_1 + dx_2 \wedge dy_2$. Let $C \subset \mathbb{R}^4$ be a compact convex set with strictly convex smooth boundary $M = \partial C$: this means that the boundary is smooth and has positive sectional curvatures everywhere. Note that any Hamiltonian having $M$ as a regular level will define a smooth flow on $M$ without stationary points, and another choice of such Hamiltonian will change this flow only by reparametrizing it. Hence, one can talk about Hamiltonian dynamics on $M$ up to time-parametrization.

The first and main result is the following remarkable statement.

**Theorem 3.2** (HWZ [24])**.** *Hamiltonian dynamics on strictly convex hypersurfaces inside* $(\mathbb{R}^4, \omega_0)$ *always admit disk-like global surfaces of section.*

Theorem 3.2 is proved using holomorphic curves. Consider $M = \partial C$ the strictly convex smooth boundary of the convex body $C$, and the 1-form

$$\lambda_0 = \frac{1}{2} \sum_{j=1}^{2} x_j dy_j - y_j dx_j$$

which is a primitive of $\omega_0$. We denote $\alpha = \iota^* \lambda_0$ where $\iota : M \hookrightarrow \mathbb{R}^4$ is the inclusion. We parametrize Hamiltonian dynamics on $M$ by the unique vector field $X$ satisfying

$$d\alpha(X, \cdot) = 0 \qquad \alpha(X) = 1$$

which is called the Reeb vector field.

The kernel $\xi = \ker \alpha$ is a contact structure. We represent the symplectization of $(M, \xi)$ as the symplectic manifold $(\mathbb{R} \times M, d(e^a \alpha))$ where the $\mathbb{R}$-coordinate is denoted by $a$. We view $\alpha$, $X$ and $\xi$ as $\mathbb{R}$-invariant objects in the symplectization. One checks that $d\alpha$ turns $\xi$ into a symplectic vector bundle.

Choose a complex structure $J : \xi \to \xi$ such that $d\alpha(\cdot, J\cdot) > 0$. Hofer considers the almost complex structure $\widetilde{J}$ on $\mathbb{R} \times M$ defined by

$$\widetilde{J} : \partial_a \mapsto X, \qquad \widetilde{J}|_\xi = J.$$

The proof in [24] is based on the construction of finite-energy holomorphic planes in $(\mathbb{R} \times M, \widetilde{J})$. These are smooth maps

$$\widetilde{u} = (a, u) : (\mathbb{C}, i) \to (\mathbb{R} \times M, \widetilde{J})$$

satisfying

$$\bar{\partial}(\widetilde{u}) = \frac{1}{2}(d\widetilde{u} + \widetilde{J}(\widetilde{u}) \circ d\widetilde{u} \circ i) = 0. \tag{17}$$

and a finite-energy condition introduced by Hofer [20]. HWZ construct such a plane asymptotic to a periodic trajectory $x : \mathbb{R}/T\mathbb{Z} \to M$ of $X$ in the sense that

$$\lim_{s\to+\infty} u(e^{2\pi(s+it)}) = x(Tt) \qquad \text{in } C^\infty(\mathbb{R}/\mathbb{Z}, M).$$

The trajectory $x$ has very special properties from dynamical and contact topological points of view: its Conley-Zehnder index is equal to 3 and its self-linking number is equal to $-1$. Using this information, results from [22] tell us that the $M$-component $u : \mathbb{C} \to M$ is an embedding transverse to $X$. The elliptic nature of (17) and the asymptotic behavior established in [21] allow for a Fredholm theory [23], which is used by HWZ to view such a plane as one leaf of a local foliation in $M \setminus x(\mathbb{R})$. Finally, one needs compactness results for punctured holomorphic curves to obtain a foliation on the whole of $M \setminus x(\mathbb{R})$ by planes transverse to $X$. This transversality and the positivity of the linearized dynamics at $x$ are the reason why each leaf in this foliation will be the interior of a disk-like global surface of section.

The reader might wonder where is the strict convexity of $M$ used in HWZ's argument. The answer is: both in the construction of the first holomorphic plane and in the passage from local to global foliations. The idea is that when one of these steps fail, the compactness results will produce other holomorphic curves which are asymptotic to periodic trajectories with Conley-Zehnder index less than 3. However, HWZ show that strict convexity forces the periodic Reeb trajectories to have Conley-Zehnder index at least equal to 3, a property called *dynamical convexity* by HWZ.

At this point HWZ were able to obtain the following remarkable corollary.

**Theorem 3.3** (HWZ [24])**.** *Strictly convex energy levels in* $(\mathbb{R}^4, \omega_0)$ *admit either two or infinitely many periodic trajectories.*

To prove this, note that the return map to the disk provided by Theorem 3.2 preserves an area form with finite total area. Brouwer's translation theorem gives a fixed point, corresponding to a second periodic trajectory. Once this fixed point is removed from the disk, we are left with a return map on an open annulus. Frank's results [13, 14] imply that there are either none or infinitely many periodic points.

In the search of finer global structures of the flow, one might ask if there are more global surfaces of section on strictly convex energy levels. One might also ask if non-convex levels admit global sections. The statements below were proved using the methods of HWZ.

**Theorem 3.4** ([27, 28])**.** *A periodic orbit on a strictly convex energy level bounds a disk-like global surface of section if, and only if, it is unknotted and has self-linking number* $-1$*.*

**Theorem 3.5** ([29, 30])**.** *A periodic orbit on a star-shaped energy level bounds a disk-like global surface of section if it is unknotted, its self-linking number is* $-1$*, its Conley-Zehnder index is at least* 3*, and it is linked to all periodic orbits with transverse rotation number equal to* 1*. Conversely, these assumptions are* $C^\infty$*-generically necessary.*

Excellent introductions to pseudo-holomorphic curve theory in symplectizations and symplectic cobordisms can be found in the book [1] by Abbas and the book [36] by Wendl.

## 4. Quick introduction to the PCR3BP

We end this note by giving some details on this which is a central problem in Celestial Mechanics. This also serves as an excuse to point towards a different proof of a result from [7]; see Theorem 4.2. An excellent introduction to the PCR3BP are the books [15] by Frauenfelder and van Koert, and [17] by Geiges. The book of Frauenfelder and van Koert also serves as an introduction to holomorphic curve methods and their use in Celestial Mechanics.

At first one considers three massive bodies, their positions and masses being denoted by $z_1, z_2, z_3$ and $m_1, m_2, m_3$ respectively. According to Newton's Law of Gravitation, motion is described by the following system of second order ODEs

$$\ddot{z}_1 = m_2 \frac{z_2 - z_1}{|z_2 - z_1|^3} + m_3 \frac{z_3 - z_1}{|z_3 - z_1|^3} \tag{18}$$

$$\ddot{z}_2 = m_1 \frac{z_1 - z_2}{|z_1 - z_2|^3} + m_3 \frac{z_3 - z_2}{|z_3 - z_2|^3} \tag{19}$$

$$\ddot{z}_3 = m_1 \frac{z_1 - z_3}{|z_1 - z_3|^3} + m_2 \frac{z_2 - z_3}{|z_2 - z_3|^3} \tag{20}$$

in suitably normalized units. Setting $m_3 = 0$ in (18)-(19) one obtains the *restricted* three-body problem: the first two particles (primaries) move independently of the third particle according to the two-body problem, and the third particle (massless satellite) moves according to (20). Requiring that all particles move on a plane one obtains the *planar* problem; we make this assumption here, and from now on the $z_k$ belong to the complex plane $\mathbb{C}$. The relative position of the primaries $\zeta = z_2 - z_1$ solves Kepler's equation $\ddot{\zeta} = -(m_1 + m_2)\zeta/|\zeta|^3$ and the adjective *circular* refers to the case when $\zeta$ describes a circle. The PCR3BP is then the problem of describing the movement of the satellite.

One can see (20) as a non-autonomous non-linear second order differential equation for $z_3(t)$, which may lead the reader to think that it is intractable. However, a well-known miracle happens: in a certain non-inertial coordinate system the equations of motion not only retain their Hamiltonian form, but the Hamiltonian function turns out to be time-independent!

More precisely, consider a inertial coordinate system where the center of mass rests at the origin. Since $\zeta$ describes circular motion, we find $\omega \neq 0$ such that

$$z_1(t) = r_1 e^{i\omega t} \qquad z_2(t) = -r_2 e^{i\omega t} \tag{21}$$

for some constants $r_1, r_2 > 0$. The condition on the center of mass reads $m_1 r_1 - m_2 r_2 = 0$, and from Kepler's equation we extract the identity

$$(r_1 + r_2)^3 \omega^2 = m_1 + m_2.$$

There is no loss of generality to assume that the angular velocity $\omega$ is positive. In the rotating coordinate system with angular velocity $\omega$, the position $q(t)$ of the satellite relative to the second primary is defined by

$$z_3(t) = (q(t) - r_2)e^{i\omega t}$$

from where it follows, together with (20) and (21), that $q(t)$ solves

$$\ddot{q} + 2i\omega\dot{q} - \omega^2(q - r_2) = -m_1 \frac{q - r_1 - r_2}{|q - r_1 - r_2|^3} - m_2 \frac{q}{|q|^3}. \tag{22}$$

The first primary rests at the point $r_1 + r_2$ and the second primary rests at the origin.

It is convenient to introduce the mass ratio

$$\mu = \frac{m_1}{m_1 + m_2} \in [0, 1]$$

because if $\omega$ and $m_1 + m_2 > 0$ are fixed once and for all, then all constants $m_1, m_2, r_1, r_2$ become functions of $\mu$. Finally we set

$$p = \dot{q} + i\omega(q - r_2)$$

and consider the Hamiltonian

$$H_\mu(q,p) = \frac{1}{2}|p|^2 + \omega \langle q - r_2, ip \rangle - \frac{m_1}{|q - r_1 - r_2|} - \frac{m_2}{|q|} \tag{23}$$

where $\langle \cdot, \cdot \rangle$ denotes the euclidean inner-product. The miracle that materializes in front of our eyes is that (22) is equivalent to Hamilton equations

$$\dot{q} = \nabla_p H_\mu \qquad\qquad \dot{p} = -\nabla_q H_\mu$$

where the surprising fact is that this is an autonomous Hamiltonian system. As is well known, the value of $H_\mu$ is preserved along trajectories:

$$\frac{d}{dt} H_\mu = \langle \nabla_q H_\mu, \dot{q} \rangle + \langle \nabla_p H_\mu, \dot{p} \rangle = \langle \nabla_q H_\mu, \nabla_p H_\mu \rangle - \langle \nabla_p H_\mu, \nabla_q H_\mu \rangle = 0.$$

This is one reason why $H_\mu$ may be called *energy*, and we can say that energy is preserved. The value of $c$ defined by the identity

$$-\frac{1}{2}c = H_\mu \tag{24}$$

is historically called the *Jacobi constant*.

Let us understand a bit the geometry behind some of the energy levels of $H_\mu$. The function $H_\mu$ is unbounded from above and below. Completing the squares in (23) we get

$$H_\mu(q,p) = \frac{1}{2}|\dot{q}|^2 - U_\mu(q)$$

where

$$U_\mu(q) = \frac{1}{2}|\omega(q - r_2)|^2 + \frac{m_1}{|q - r_1 - r_2|} + \frac{m_2}{|q|} \tag{25}$$

is the *effective potential*. The projection to configuration space, i.e. to the $q$-plane, of the sublevel set $\{H_\mu \leq -c/2\}$ coincides with the superlevel set $\{U_\mu \geq c/2\}$.

If $c \sim +\infty$ then $\{U_\mu \geq c/2\}$ has three connected components: two disk-like regions around the positions 0 and $r_1 + r_2$ of the primaries, and an unbounded component given as the complement of a large disk-like region. These are the *Hill's regions*. Their boundaries are called *ovals of zero velocity* for the following reason: if the energy of a satellite inside a bounded Hill region in $\{U_\mu \geq c/2\}$ is constrained by $H_\mu \leq -c/2$ then the satellite can not leave this region, and can only touch its boundary with zero velocity $\dot{q} = 0$ and only if its energy is precisely $-c/2$. In other words, the satellite does not have enough energy to "escape" the influence of the corresponding primary.

This is the situation when the Jacobi constant $-c/2$ increases from $-\infty$ until $c/2$ reaches the highest critical critical value $c_1/2$ of $U_\mu$. The only critical point of $U_\mu$ at this level lies strictly between the primaries, and we see the two bounded Hill regions getting "glued" together. For energy $-c/2 = -c_1/2 + \epsilon$ slightly higher

than $-c_1/2$ the satellite may get arbitrarily close to both primaries, i.e. now there is enough energy to escape. If the Jacobi constant decreases to $-\infty$ then $c/2$ will cross three more critical values of $U_\mu$ corresponding to further modification of the topology of the Hill regions.

Throughout the remaining of this section we stick to values $c > c_1$ and focus on the Hill region around the origin. We make normalization assumptions $\omega = m_1 + m_2 = 1$, without loss of generality. It follows that $m_1 = r_2 = \mu$, $m_2 = r_1 = 1 - \mu$ and $H_\mu$ is written as

$$H_\mu = \frac{|p|^2}{2} + \langle ip, q - \mu \rangle - \frac{\mu}{|q-1|} - \frac{1-\mu}{|q|}$$

the unique parameter being $\mu$. Simple calculations show that $c_1 \to 3$ both when $\mu \to 0^+$ and when $\mu \to 1^-$.

The method of Levi-Civita [31], which we shall now describe, regularizes collisions with the primary at the origin. We introduce new complex coordinates $u, v$ by

$$q = 2v^2 \qquad p = -\frac{u}{\bar{v}}.$$

This amounts to a two-to-one transformation of 4-space which is symplectic up to a constant factor. In fact,

$$\begin{aligned} dq_1 \wedge dp_1 + dq_2 \wedge dp_2 &= \text{Re } [dq \wedge d\bar{p}] \\ &= -2 \text{ Re } \left[ d(v^2) \wedge d\left(\frac{\bar{u}}{v}\right) \right] \\ &= -2 \text{ Re } \left[ 2v dv \wedge \left( \frac{v d\bar{u} - \bar{u} dv}{v^2} \right) \right] \\ &= -4 \text{ Re } [dv \wedge d\bar{u}] \\ &= 4 \left( du_1 \wedge dv_1 + du_2 \wedge dv_2 \right). \end{aligned}$$

It follows that Hamiltonian flows on $(q,p)$-space correspond to Hamiltonian flows on $(u,v)$-space up to a constant time-reparametrization.

We then reparametrize the Hamiltonian flow of $H_\mu$ on the level $\{H_\mu = -\frac{c}{2}\}$ as the Hamiltonian flow of $K_{\mu,c} = \frac{1}{2}|q|(H_\mu + \frac{c}{2}) = |v|^2(H_\mu + \frac{c}{2})$ on the level $\{K_{\mu,c} = 0\}$. Writing $K_{\mu,c}$ in terms of $u, v$ we get

$$K_{\mu,c}(u,v) = \frac{1}{2}|u|^2 + 2|v|^2 \langle u, iv \rangle - \mu \text{ Im}[uv] - \frac{1-\mu}{2} - \mu \frac{|v|^2}{|2v^2 - 1|} + \frac{c}{2}|v|^2 \tag{26}$$

The upshot here is that $|q-1| = |2v^2 - 1| > 0$ as long as $c > c_1$, hence the component $\Sigma_{\mu,c}$ of $\{K_{\mu,c} = 0\}$ corresponding to our chosen Hill region is a smooth hypersurface of $\mathbb{R}^4$ where the Hamiltonian flow is smooth. At this point it is worth pointing out the following nice statement.

**Theorem 4.1** (Albers, Fish, Frauenfelder, Hofer and van Koert [2])**.** *For every $c > 3$ there exists $\mu_0(c) \in (0,1)$ such that if $\mu > \mu_0(c)$ then $\Sigma_{\mu,c}$ is strictly convex, in the sense that its sectional derivatives are strictly positive. In particular Theorem 3.2 applies to give disk-like global surfaces of section.*

Consider a new set of coordinates

$$\hat{u} = u \qquad \hat{v} = \sqrt{c}\, v$$

which are again symplectic up to a constant factor, and write

$$K_{\mu,c}(\hat{u},\hat{v}) = \frac{1}{2}|\hat{u}|^2 + 2\frac{|\hat{v}|^2\,\langle \hat{u}, i\hat{v}\rangle}{c\sqrt{c}} - \mu\,\frac{\mathrm{Im}[\hat{u}\hat{v}]}{\sqrt{c}} - \frac{1-\mu}{2} - \mu\frac{|\hat{v}|^2}{c|\frac{2}{c}\hat{v}^2 - 1|} + \frac{1}{2}|\hat{v}|^2 \tag{27}$$

in terms of these variables. If $c \to +\infty$ then $K_{\mu,c}(\hat{u},\hat{v})$ converges in $C^\infty$ to

$$\frac{1}{2}|\hat{u}|^2 + \frac{1}{2}|\hat{v}|^2 - \frac{1-\mu}{2}$$

which is the Hamiltonian of two uncoupled harmonic oscillators with equal frequencies. In particular, the flow converges to a periodic flow whose trajectories are the Hopf fibers in the sphere of radius $\sqrt{1-\mu}$. In fact, the $C^\infty$-converge is uniform when $\mu$ ranges on a compact subset of $[0,1)$. This simple observation and standard arguments can be used to prove

**Theorem 4.2** (Conley [7])**.** *If $\mu \in (0,1)$ is fixed and the Jacobi constant is large enough then there are annulus-like global surfaces of section for the Hamiltonian flow on the component of the energy level $\{K_{\mu,c} = 0\}$ corresponding to the Hill region around the origin. The return map to these annuli satisfy the twist condition of the Poincaré-Birkhoff Theorem, in particular there are infinitely many periodic orbits for the PCR3BP in these situations.*

## References


[1] C. Abbas, *An introduction to compactness results in symplectic field theory.* Springer, Heidelberg, 2014. viii+252 pp. ISBN: 978-3-642-31542-8; 978-3-642-31543-5.

[2] P. Albers, J. W. Fish, U. Frauenfelder, H. Hofer, and O. van Koert. *Global surfaces of section in the planar restricted 3-body problem.* Archive for Rational Mechanics and Analysis, **204** (2012), no.1, 273–284.

[3] V.I. Arnold, *Mathematical methods of classical mechanics.* Springer-Verlag, Berlin and New York, 1978.

[4] M. Audin and M. Damian, *Morse theory and Floer homology.* Translated from the 2010 French original by Reinie Erné. Universitext. Springer, London; EDP Sciences, Les Ulis, 2014. xiv+596 pp. ISBN: 978-1-4471-5495-2; 978-1-4471-5496-9; 978-2-7598-0704-8.

[5] G. D. Birkhoff, *Proof of Poincaré's last geometric theorem*, Trans. Amer. Math. Soc., **14** (1913), 14–22.

[6] G. D. Birkhoff. *Dynamical Systems*, (Colloq. Publ., Am. Math. Soc., vol. 9) Providence, RI.: Amer. Math. Soc. 1927.

[7] C. Conley. *On some new long periodic solutions of the plane restricted three body problem.* Commun. Pure Appl. Math. 16, (1963) pp. 449–467.

[8] C. Conley and E. Zehnder, *The Birkhoff-Lewis fixed point theorem and a conjecture of V. I.Arnold.* Invent. Math., 73:33–49, 1983.

[9] Y. Eliashberg, A. Givental, H. Hofer. *Introduction to Symplectic Field Theory.* Geom. Funct. Anal., Special volume, Part II (2000), 560-673.

[10] A. Floer, *Symplectic fixed points and holomorphic spheres.* Comm. Math. Phys. **120** (1989), no. 4, 575–611.

[11] A. Floer, *Morse theory for Lagrangian intersections.* J. Differential Geom. **28** (1988), no. 3, 513–547.

[12] A. Floer, *The unregularized gradient flow of the symplectic action.* Comm. Pure Appl. Math. **41** (1988), no. 6, 775–813.

[13] J. Franks. *Recurrence and fixed points of surface homeomorphisms.* Erg. Th. Dyn. Sys., **8**, (1988), 99–107.

[14] J. Franks. *Geodesics on $S^2$ and periodic points of annulus homeomorphisms*, Invent. Math. 108 (1992), no. 2, 403–418.

[15] U. Frauenfelder and O. van Koert, *The Restricted Three-Body Problem and Holomorphic Curves* (Pathways in Mathematics). Birkhäuser Basel (2018).

[16] D. Fried, *The geometry of cross sections to flows*, Topology, 21 (1982), 353–371.

[17] H. Geiges, *The Geometry of Celestial Mechanics* (London Mathematical Society Student Texts). Cambridge: Cambridge University Press (2016).
[18] E. Ghys, *Right-handed vector fields & the Lorenz attractor*, Japan. J. Math. 4, 47–61 (2009).
[19] M. Gromov. *Pseudo-holomorphic curves in symplectic manifolds.* Invent. Math. 82 (1985), 307–347.
[20] H. Hofer. *Pseudoholomorphic curves in symplectisations with application to the Weinstein conjecture in dimension three.* Invent. Math. **114** (1993), 515-563.
[21] H. Hofer, K. Wysocki and E. Zehnder. *Properties of pseudoholomorphic curves in symplectisations I: Asymptotics.* Ann. Inst. H. Poincaré Anal. Non Linéaire **13** (1996), 337–379.
[22] H. Hofer, K. Wysocki and E. Zehnder. *Properties of pseudoholomorphic curves in symplectisations II: Embedding controls and algebraic invariants.* Geom. Funct. Anal. **5** (1995), no. 2, 270–328.
[23] H. Hofer, K. Wysocki and E. Zehnder. *Properties of pseudoholomorphic curves in symplectizations III: Fredholm theory.* Topics in nonlinear analysis, Birkhäuser, Basel, (1999), 381–475.
[24] H. Hofer, K. Wysocki and E. Zehnder. *The dynamics on three-dimensional strictly convex energy surfaces.* Ann. of Math. **148** (1998), 197–289.
[25] H. Hofer, K. Wysocki and E. Zehnder. *Finite energy foliations of tight three-spheres and Hamiltonian dynamics.* Ann. Math **157** (2003), 125–255.
[26] H. Hofer and E. Zehnder, *Symplectic Invariants and Hamiltonian Dynamics.* Birkhäuser, Basel, 1994.
[27] U. Hryniewicz. *Fast finite-energy planes in symplectizations and applications.* Trans. Amer. Math. Soc. **364** (2012), 1859–1931.
[28] U. Hryniewicz. *Systems of global surfaces of section for dynamically convex Reeb flows on the* 3*-sphere,* Journal of Symplectic Geometry **12**, 4, (2014).
[29] U. Hryniewicz, J. E. Licata and Pedro A. S. Salomão. *A dynamical characterization of universally tight lens spaces.* arXiv:1306.6617v2.
[30] U. Hryniewicz and Pedro A. S. Salomão. *On the existence of disk-like global sections for Reeb flows on the tight 3-sphere.* Duke Math. J. **160** (2011), no. 3, 415–465.
[31] T. Levi-Civita, *Sur la régularisation du problème des trois corps.* Acta Math. 42(1), 99–144 (1920).
[32] H. Poincaré. *Sur un théorème de géométrie.* Rend. Circ. Mat. Palermo, **33** (1912), 375–407.
[33] D. Salamon, *Lectures on Floer homology.* Symplectic geometry and topology (Park City, UT, 1997), 143?229, IAS/Park City Math. Ser., 7, Amer. Math. Soc., Providence, RI, 1999.
[34] S. Schwartzman, *Asymptotic cycles*, Ann. of Math. (2) 66 (1957), 270–284.
[35] D. Sullivan, *Cycles for the dynamical study of foliated manifolds and complex manifolds*, Invent. Math., 36 (1976), 225–255.
[36] C. Wendl, *Holomorphic Curves in Low Dimensions.* Springer International Publishing, 2018. xiii+294 pp. ISBN: 978-3-319-91369-8

Universidade Federal do Rio de Janeiro, Departamento de Matematica Aplicada, Av. Athos da Silveira Ramos 149, Rio de Janeiro RJ, Brazil 21941-909.
*Email address*: umbertolh@gmail.com

CHARLA PLENARIA

# SOLUCIÓN DE LAS ECUACIONES DE HAMILTON-JACOBI COMO LÍMITE CLÁSICO DE INTEGRALES CON RESPECTO AL MOVIMIENTO BROWNIANO

JOSÉ R. LEÓN

Publicaciones Matemáticas del Uruguay
Volumen 17, Julio 2019, Páginas 81–92
S 0797-1443

# THE HAMILTON JACOBI EQUATION, THE FEYNMAN-KAC FORMULA AND THE CLASSICAL LIMIT

HENRYK GZYL AND JOSÉ R. LEÓN

Abstract. We present a simplified approach to the problem of obtaining classical mechanics from a classical diffusion equation. It so happens that when the diffusion constant tends to 0 the diffusing particles move along the trajectory of a Newtonian particle. At the end we present a list of examples including particles in a static electro-magnetic field

## 1. Introduction and preliminaries

Our goal in this review is to give some insight to what is now a well known theme: To obtain classical mechanics as the asymptotic limit $\hbar \to 0$ of non-relativistic quantum mechanics. We collected a series of results that were obtained by a large number of researchers over time. Perhaps we can say that the originality of this presentation lies on the simplification of proofs.

Let us begin with some introductory preliminaries. When the quantum mechanical description of microscopic phenomena based on Schrödinger's equation replaced the classical description based on the Newtonian-Hamiltonian formalism, the natural question to ask was: how is the later to be understood with regards to the former.

This was the birth of the $\hbar \to 0$ asymptotic procedures by means of which classical mechanics was to be recovered from quantum mechanics. A few exposés of techniques and interconnection with a variety of problems can be seen in [1], [6], [16], [17] and [24]. By the way, [17] contains an analytic approach to what we do here.

The simplest interesting Schrödinger's equation, that which describes the quantum behavior of a point particle of mass one in a potential $V(x)$, is

$$i\hbar\frac{\partial\psi}{\partial t} = -\frac{\hbar^2}{2}\Delta\psi + V\psi. \tag{1.1}$$

where $\Delta$ is the Laplace-Beltrami operator in the Euclidean metric in $\mathbb{R}^n$ and $\hbar = \frac{h}{2\pi}$, with $h$ being the Planck's constant. This equation can be transformed into a diffusion equation by means of the replacement of $t$ by $\frac{t}{i}$, thus obtaining

$$\hbar\frac{\partial\rho}{\partial t} = \frac{\hbar^2}{2}\Delta\rho - V\rho, \tag{1.2}$$

where $\rho(\cdot, t) = \psi(\cdot, \frac{t}{i})$. This (formal) replacement is interesting for the simple reason that a well developed theory of functional integration exists, and which allows to represent solutions to (1.2) as path integrals, whereas path integration starting from

2000 *Mathematics Subject Classification.* 60F10 70H20.

©2019 PMU

(1.1) is not a clear affair. How to relate both techniques? Different approaches have been tried and most listed in the references [1], [6], [16], [17] and [24], but we may single out a few, namely [5], [8], [11], [14] and [18].

A few of the references deal with the $\hbar \to 0$ limit from various related points of view are [3], [8], [9], [10], [12], [20] and we are not going to improve on the results presented there in various degrees of generality. What we are going to do instead is to play around the initial step, namely the representation of the solution of (1.2) as a path integral and take things to the point where the $\hbar \to 0$ limit becomes an obvious and clear fact, and the conditions under which it is valid become quite transparent.

The end product is similar but not identical to the approaches developed in [9] and [10], and we feel this makes a good preamble to that line of work. Let us now to prepare the stage.

We want to relate the solutions to the Hamilton-Jacobi equation, [2],

$$\frac{\partial S_0}{\partial t} + \frac{1}{2}(\nabla S_0)^2 + V(x) = 0, \qquad S_0(x,P,0) =< x,P >, \tag{1.3}$$

to the asymptotic behavior of the solutions of (1.4) as $h \to 0$. Where $h$ is a positive real number and $\nabla$ is the gradient operator. It is important to remark above the sign of the function $V$, that $< \cdot,\cdot >$ denotes the Euclidean scalar product in $\mathbb{R}^n$ and as in physical literature we use $(x)^2 = ||x||^2$ for $x \in \mathbb{R}^n$.

$$h\frac{\partial \psi_h}{\partial t} = \frac{h^2}{2}\Delta\psi_h + V(x)\psi_h \qquad \psi_h(x,0) = e^{-\frac{<x,P>}{h}}, \tag{1.4}$$

The connection between (1.3) and (1.4) being via the substitution

$$\psi_h(x,t) = e^{-\frac{S_h(x,P,t)}{h}}, \tag{1.5}$$

which requires $S_h(x,P,t)$ to satisfy

$$\frac{\partial S_h}{\partial t} = \frac{1}{2}(\nabla S_h)^2 + V(x) = \frac{h}{2}\Delta S_h, \qquad S_h(x,P,t) =< x,P > . \tag{1.6}$$

Why (1.5)? In the trivially integrable case, this is that of a particle in a Hamiltonian $H(p)$. Below in section 2 we will study in more generality the movement of such a particle and in the particular case that we are considering we get the equation

$$\frac{\partial S_0}{\partial t} + H(\nabla S_0) = 0, \qquad S_0(x,P,0) =< x,P > .$$

Which has $S_0(x,P,t) =< x,P > -H(P)t$, as solution. Consider now the solution of

$$h\frac{\partial \psi_h}{\partial t} = H(-h\nabla)\psi_h, \qquad \psi_h(x,0) = e^{\frac{<x,P>}{h}},$$

which is by the theory of semigroups

$$\psi_h(x,t) = (e^{\frac{tH(-h\nabla)}{h}}\psi_h(\cdot,0))(x) = e^{-\frac{S_0(x,P,t)}{h}}.$$

Then the function $S_0(x,P,t)$ can be recovered from the above solution by means of

$$S_0(x,P,t) = \lim_{h\to 0} -h\log\psi_h(x,t).$$

Another reason is that the solution of equation (1.4) admits a representation in terms of path integrals via the Feynman-Kac formula, see for example [21],

$$\psi_h(x,t) = \mathbb{E}^{\,x}_h[e^{(-\frac{1}{h}(<P,B_h(t)>-\int_0^t V(B_h(s))ds))}]. \tag{1.7}$$

This expression can be computed in a variety of simple cases and is a good starting point for theoretical analysis. Here, as usual, the symbol $\mathbb{E}_h^x$ denotes the expectation with respect to the Brownian motion starting at the point $x \in \mathbb{R}^n$ and $B_h(t)$ denotes this process at time $t$ with variance $h$.

## 2. Some classical mechanics

For completeness we are going to take a small digression into classical mechanics. The reason for the initial condition $S_0(x, P, 0) =< x, P >$ is the following. For a mechanical system on $\mathbb{R}^n$, the function $S_0(x, P, t)$ generates a canonical transformation that brings the system to rest, that is, it maps the current state of the system (denoted by $(x, p)$ to the initial state (denoted by $(X, P)$).

When $S_0(x, P, t)$ is known the state of the system at time $t$ can be obtained from the state $t = 0$ by solving the transformation equations

$$X_i = \frac{\partial S_0}{\partial P_i}, \qquad p_i = \frac{\partial S_0}{\partial x_i}. \tag{2.1}$$

We shall assume that there exists an appropriate time interval $[0, T]$ in which $S_0(x, P, t)$ is defined and the transformation (2.1) is well defined. It is plain to check that for a dynamical system with Hamiltonian $H(p)$ we have $S_0(x, P, t) =< x, P > -H(P)t$ with $S_0(x, p, 0) =< x, P >$, and (2.1) hold.

In general, see [2], a mechanical system on $\mathbb{R}^{2n}$ is specified by giving the Hamiltonian $H(x, p)$, and the trajectories of the system are obtained by solving the Hamilton equations.

$$\dot{x}_i = \frac{\partial H}{\partial p_i}, \quad \dot{p}_i = -\frac{\partial H}{\partial x_i}, \quad i = 1, \dots, n. \tag{2.2}$$

which is always possible, when the derivatives appearing in (2.2) are locally Lipschitz, bounded, in a local sense.

In general the problem of solving (2.2) can be translated into the problem of solving

$$\frac{\partial S_0}{\partial t} + H(x, \nabla S_0) = 0, \quad S_0(x, P, 0) =< x, P > . \tag{2.3}$$

The connection and much more can be seen in [2].

The problem with (2.3) being the apparition of singularities which make the global existence of solutions a problem. Even though the curves $(x(t), y(t))$ may exist and be unique as curves in $\mathbb{R}^{2n}$, their projection $x(t)$ on $\mathbb{R}^n$ may cross which makes it impossible in general to propagate an initial condition for (2.3) without the appearance of singularities.

Quite a lot on these matters can be found in [17].

What we do next is quite straight forward and is similar to, but shrouded in a less sophisticated jargon, the material in [10]. To begin with, suppose that $S_0(x, P, t)$ satisfies (1.3), and consider the curves defined for $0 \le s \le t$ as the solutions to:

$$\dot{\gamma}(s) = (\nabla S_0)(\gamma(s), P, s) \quad \gamma(t) = x, \tag{2.4}$$

which will always possible when there is a $T \le \infty$, such a $\gamma$ satisfying (2.4) exists if $\nabla S_0$ is Lipschitz and bounded as function in $x$ for $0 \le t \le T$. When we differentiate (2.4) once more and make use of (1.3) we obtain that

$$\ddot{\gamma}(s) = -\nabla V(\gamma(s)). \tag{2.5}$$

Consider for each $t$, the curve $x(s)$ defined on $[0,T]$ by $x(s)=\gamma(t-s)$. Then $x(0)=x$, $\dot{x}(t)=-\gamma(0)=-P$. Form $F(s)=S_0(x(s),P,t-s)$, differentiate with respect to $s$ to obtain

$$\frac{dF}{ds}=-\{\frac{1}{2}(\nabla S_0)^2(x(s),P,t-s)-V(x(s))\},$$

and now integrate from 0 to $t$ to obtain

$$S_0(x,P,t)=<x(t),P>+\int_0^t(\frac{1}{2}\dot{x}^2(s)-V(x(s)))ds. \tag{2.6}$$

For some of the results in section 3 we need the following

**Lemma 1.** *Let $x(s)$ be as above and define*

$$U(x,P,t)=<x(t),P>+\int_0^t(\frac{1}{2}\dot{x}^2(s)-V(x(s)))ds,$$

*then we have $U(x,P,t)=S_0(x,P,t)$.*

**Remark 1.** *This is implicit in (2.6), but the conceptual meaning is different, for in order to compute the right hand side of (2.6) one only has to solve $\ddot{x}(s)=-\nabla V(x(s))$, $x(0)=x$, $\dot{x}(t)=-P$, and then integrate to obtain the solution to (1.3).*

**Proof.** Notice to begin with that

$$\frac{\partial U}{\partial t}=<\dot{x}(t),P>+\frac{1}{2}\dot{x}^2(t)-V(x(t))=-\{\frac{1}{2}\dot{x}^2(t)+V(x(t))\},$$

since $\dot{x}(t)=-P$. Compute now

$$\begin{aligned}\frac{\partial U}{\partial x_i} &= \sum_{j=1}^n P_j\frac{\partial x_j(t)}{\partial x_i}+\int_0^t\sum_{j=1}^n\{\dot{x}_j(t)\frac{\partial\dot{x}_j(s)}{\partial x_i}-\frac{\partial V}{\partial x_j}(x(s))\frac{\partial x_j(s)}{\partial x_i}\}ds\\ &= \sum_{j=1}^n\{P_j\frac{\partial x_j(t)}{\partial x_i}+\int_0^t\{\frac{d}{ds}(\dot{x}_j(s)\frac{\partial x_j(s)}{\partial x_i})ds\}\}\\ &= \sum_{j=1}^n\{P_j\frac{\partial x_j(t)}{\partial x_i}+\dot{x}_j(t)\frac{\partial x_j(s)}{\partial x_i}-\dot{x}_j(0)\delta_{ij}\}=-\dot{x}_i(0),\end{aligned}$$

where in the second step we used $\ddot{x}_j(s)=-(\frac{\partial V}{\partial x_j}(x(s))$ and in the third we again used $\dot{x}(t)=-P$.

Given that $x(0)=\gamma(t)=x$. Now a simple application of the law of the conservation of energy yields,

$$\begin{aligned}&\frac{\partial U}{\partial t}(x,t)+\frac{1}{2}(\nabla U)^2(x,t)+V(x)\\ &\qquad=-(\frac{1}{2}\dot{x}^2(t)+V(x(t)))+(\frac{1}{2}\dot{x}^2(0)+V(x(0)))=0,\end{aligned}$$

which together with the fact that at $t=0$ , $U(x,P,0)=<x,P>$ yields the desired result.

## 3. The Classical limit

As we said above, what do here is similar in spirit to the methods in [8] and [10]. We shall proceed with formal manipulations to the very end, point at which we state the necessary qualifiers. We are interested in the behavior of (1.7) this is, of

$$\psi_h(x,t) = \mathbb{E}^{\,x}_h[e^{-\frac{1}{h}(<P,B_h(t)>-\int_0^t V(B_h(s))ds)}],$$

as $h$ becomes small.

Consider $S_0(B_h(s), P, t-s)$ for $0 \leq s \leq t$. From Itô formula

$$\begin{aligned} &S_0(B_h(s),P,t-s) \\ &\quad = S_0(B_h(0),P,t) + \int_0^s (\frac{\partial S_0}{\partial t}(-) + \frac{h}{2}\Delta S_0(-)du + \int_0^s < \nabla S_0(-), dB_h(u) >, \end{aligned}$$

where $(-)$ stands for $(B_h(u), P, t-u)$. Now set $s=t$, use $S_0(x,P,0) =< x, P >$ and (1.3) to rewrite (1.7) as

$$\psi_h(x,t) = e^{-\frac{1}{h}S_0(x,P,t)}\mathbb{E}^{\,x}_h[Z_h(t)e^{-\frac{1}{2}\int_0^t(\Delta S_0)(B_h(s),P,t-s)ds}] \tag{3.1}$$

where $Z_h(s)$ is the exponential martingale given by

$$Z_h(t) = e^{-\frac{1}{h}\left(\int_0^t <(\nabla S_0)(B_h(s),P,t.s),dB_h(s)>+\frac{1}{2}\int_0^t(\nabla S_0)^2(B_h,P,t-s)ds\right)}. \tag{3.2}$$

A different way to reach (3.1) is to make $\psi_h(x,t) = \Lambda(x,t)e^{-\frac{1}{h}S_0(x,P,t)}$ and verify that $\Lambda(x,t)$ has to satisfy

$$\frac{\partial \Lambda}{\partial t} = \frac{h}{2}\Delta\Lambda - < \nabla S_0, \nabla \Lambda > - \frac{\Delta S_0}{2}\Lambda, \quad \Lambda(x,0) = 1 \tag{3.3}$$

and we see that (3.1) expresses the solution to (3.3) as a path integral combining the Cameron-Martin-Girsanov and Feynman-Kac formula, see [21] for example.

Had we rigged things up so Schilder's conditions for his theorem A in [20] be satisfied we could at this point conclude that

$$\lim_{h\to 0}\mathbb{E}^{\,x}_h[Z_h(t)e^{-\frac{1}{2}\int_0^t \Delta S_0(B_h,P,t-s)ds}] = e^{-\frac{1}{2}\int_0^t \Delta S_0(x(s),P,t-s)ds}. \tag{3.4}$$

where $x(s)$ is the curve maximizing the following functional

$$\begin{aligned} \mathcal{L}(\xi) &= -\int_0^t \big( < \nabla S_0(\xi(s),P,t-s), \dot{\xi}(s) > \\ &\qquad + \frac{1}{2}(\nabla S_0)^2(\xi(s),P,t-s) + \frac{1}{2}\dot{\xi}^2(s)\big)ds \\ &= -\frac{1}{2}\int_0^t \big(\dot{\xi}(s) + \nabla S_0(\xi(s),P,t-s)\big)^2 ds \end{aligned} \tag{3.5}$$

which is obviously the curve $x(s)$ introduced in section 2, this is the curve solving

$$\dot{x}(s) = -(\nabla S_0)(x(s),P,t-s) \quad x(0) = x, \tag{3.6}$$

which makes (3.5) attain its maximum value.

We would thus obtain

$$\lim_{h\to 0} -h\log\psi_h(x,t) = S_0(x,P,t),$$

as well as the limit

$$\lim_{h\to 0} e^{\frac{1}{h}S_0(x,P,t)}\psi_h(x,t) = e^{-\int_0^t \Delta S_0(x(s),P,t-s)ds}.$$

But if we did not know about Schilder's result we could proceed as follows. Let $X_h(t)$ be the solution of the stochastic equation

$$dX_h(t)(s) = -(\nabla S_0)(X_h(s), P, t-s)ds + \sqrt{h}dB_1(s) \tag{3.7}$$

and now use the Cameron-Martin-Girsanov formula to rewrite $\psi_h(x,t)$ as

$$\psi_h(x,t) = e^{-\frac{1}{h}S_0(x,P,t)}\mathbb{E}^{\,x}_1[e^{-\frac{1}{2}\int_0^t(\Delta S_0)(X_h(s),P,t-s)ds}] \tag{3.8}$$

and we can now state the

**Theorem 1.** *Assume that $V(x)$ is such that for $t$ in $[0,T]$, (1.7) is finite and that the solution to (1.3) exits, is smooth and $\nabla S_0$ is Lipschitz or bounded and that $\Delta S_0$ is smooth and bounded. Then $X_h$ defined by (3.7) exists and*

$$\lim_{h\to 0} -h\log\psi_h(x,t) = S_0(x,P,t),$$

$$\lim_{h\to 0} e^{\frac{1}{h}S_0(x,P,t)}\psi_h(x,t) = e^{-\frac{1}{2}\int_0^t \Delta S_0(x(s),P,t-s)ds}.$$

**Proof.** The existence and uniqueness of $X_h(s)$ is standard and $X_h(s)$ tends uniformly in probability to $x(s)$ as $h$ tends to zero. All this can be seen in [12]. The first limit is not a problem and for the second use the smoothness of $\Delta S_0$.

**Remark 2.** *In* [12] *we can also find the expansion of $X_h(s)$ in powers of $\sqrt{h}$ and this combined with the smoothness of $\Delta S_0$ allows us to obtain asymptotic expansions of $\psi_h(x,t)$. The whole point in doing things this way is to go around the heavy limiting procedures as appearing in* [9], [10] *and* [20], *since all that is involved are estimates in* [12] *that are typical of the theory of ordinary differential equations.*

What come next is a rehash of a standard theme, see [9] but we say it differently. As above let start from (1.7) and rewrite it by means of the Cameron-Martin-Girsanov translation formula as

$$e^{-\frac{1}{h}b(t)}\mathbb{E}^{\,x}_h[e^{-\frac{1}{h}\int_0^t\sum_{ij}B_h^i(s)B_h^j(s)H_{ij}(B_h(s))ds}], \tag{3.9}$$

where

$$b(t) = < P, x(t) > + \int_0^t (\frac{1}{2}\dot{x}^2(s) - V(x(s)))ds.$$

We have seen in Lemma 2.6, equals $S_0(x,P,t)$, and where

$$H_{ij}(B_h(s)) = \int_0^1 (1-u)\frac{\partial^2 V}{\partial x_i \partial x_j}(u + B_h(s) + x(s))du,$$

here $x(t)$ has been introduced in section 2. To arrive at (3.9) we use the fact that $x(s)$ maximizes the following functional

$$-(< P, \xi(t) > + \int_0^t (\frac{1}{2}\dot{\xi}^2(s) - V(\xi(s)))ds),$$

over the set of trajectories that have square integrable velocity, $\xi(0) = 0$ and $\xi(t) = -P$ which can be seen in [10]. Then the first variation of this functional vanishes at $x(s)$ and we also make use of the fact that for a $f \in C^2(\mathbb{R}^n)$, and any $b \in \mathbb{R}^n$ we have

$$f(b+y) = f(y) + \sum_{i=1}^n b^i \frac{\partial f}{\partial x_i}(y) + \sum_{i,j} b^i b^j \int_0^1 (1-u)\frac{\partial^2 f}{\partial x_i \partial x_j}(u+b+y)du.$$

Now rescale the Brownian motion in (3.9) to obtain

$$\psi_h(x,t) = e^{-\frac{b(t)}{h}}\mathbb{E}_1^x[e^{-\int_0^t \sum_{i,j} B_1^i(s)B_1^j(s)H_{ij}(\sqrt{h}B_1(s))ds}]. \tag{3.10}$$

Then we see that any condition insuring integrability of (3.10) and smoothness of $H_{ij}$ would yield what we want.

For example, when $H_{ij}(x)$ is a positive form for every $x$ and is continuous we are in business. Also when $H_{ij}(x)$ has negative eigenvalues, we may pull things through by restricting $t$ to an interval such that (3.10) is finite. It suffices to take

$$4\inf|\sum_{ij}\xi^i\xi^j H_{ij}(x)| < T.$$

The connection with the previous result being that

$$e^{-\frac{1}{2}\int_0^t \Delta S_0(x(s),P,t-s)ds} = \mathbb{E}_1^x[e^{-\frac{1}{2}\int_0^t B^i(s)B^j(s)\frac{\partial^2 V}{\partial x_i \partial x_j}(x(s))ds}].$$

As can be seem in [15] pages 13-14.

## 4. More general conditions

We may want to consider (1.4) with the initial condition

$$\psi_h(x,0) = f(x)e^{-\frac{1}{h}S(x)}. \tag{4.1}$$

as in [8] say. Proceeding as above, we would have to start studying

$$\frac{\partial S_0}{\partial t} + \frac{1}{2}(\nabla S_0)^2 + V(x) = 0, \quad S_0(x,0) = S(x),$$

and we may assume the same hypothesis as in Theorem 1 to hold. As before, we can write the solution to (1.4) with initial condition (4.1) as

$$\psi_h(x,t) = \mathbb{E}_h^x[f(B_h(t))e^{-\frac{1}{h}\left(S(B_h(t))+\int_0^t \Delta S_0(B_h(s),P,t-s)ds\right)}],$$

which can, as in section 3, be transformed into

$$\psi_h(x,t) = e^{-\frac{1}{h}S_0(x,P,t)}\mathbb{E}_h^x[f(X_h(t))e^{-\frac{1}{h}\int_0^t \Delta S_0(X_h(s),P,t-s)ds}],$$

where $X_h(s)$ was introduced in (3.7). At this point, with boundedness and continuity assumptions on $f$, the standard limit and asymptotic expansions can be obtained.

The alternative procedure described in section 3, goes along the same lines. The only difference is that now we have to maximize the functional

$$\int_0^t \big(V(\xi(s)) - \frac{1}{2}\dot{\xi}^2(s)\big)ds - S(\xi(t)), \tag{4.2}$$

over the curves with square integrable derivative in $[0,T]$, for a convenient $T$. Such curves would satisfy

$$\ddot{x}(s) = -\nabla V(x(s)), \quad x(0) = x,\ \dot{x}(s) = -\nabla S(x(t)),$$

for $0 \le s \le t$. That they maximize (4.2) can be obtained by standard methods of calculus of variations.

Considering now

$$\psi_h(x,t) = \mathbb{E}_h^x[f(B_h(t))e^{-\frac{1}{h}\left(S(B_h(t))+\int_0^t V(B_h(s))ds\right)}]$$

and applying the argument in section 3, it can be transformed into

$$\psi_h(x,t) = e^{\frac{1}{h}b(t)}\mathbb{E}_h^x[f(x(t)+B_h(t))e^{-\frac{1}{h}\int_0^t \sum_{i,j} B_h^i(u)B_h^j(u)H_{i,j}(B_h(u))du}],$$

by means of the Cameron-Martin-Girsanov formula and a Taylor expansion about $x(u)$ up to second order. The term of first order in $B_h(\cdot)$ drop away due to the maximizing property of $x(s)$.

From this point on, the same comments as in the section 3 apply.

## 5. Particle in a static electro magnetic field

The description of a particle moving through an electro-magnetic field is described by the Hamiltonian

$$H(x,p) = \frac{1}{2}(p - A(x))^2 + V(x), \tag{5.1}$$

with $A(x)$ being a smooth vector field on $\mathbb{R}^3$ and $V(x)$ being a smooth scalar function. We assume that $\nabla \times A(x)$ and $\nabla V(x)$ are either bounded or satisfy a growth condition such that either Newton's equations

$$\dot{x}(s) = \dot{x}(s) \times (\nabla \times A(x(s))) - \nabla V(x(s)), \tag{5.2}$$

or Hamilton's equations

$$\dot{x}_i = p_i - A_i(x), \quad \dot{p}_i = -\frac{\partial V}{\partial x_i} + (p_j - A_j)\frac{\partial A_j}{\partial x_j} \tag{5.3}$$

have a unique solution for every initial condition $(x(0), p(0))$. The Hamilton-Jacobi equation corresponding to (5.1) is

$$\frac{\partial S_0}{\partial t} = +\frac{1}{2}(\nabla S_0 - A(x))^2 + V(x) = 0, \quad S_0(x,0) = S(x). \tag{5.4}$$

We assume that for all $t \in [0,T]$ it exists, has enough smoothness and $\nabla S_0$, $\Delta S_0$ are smooth and bounded.

We shall furthermore make the simplifying assumption (usually called the transversality condition or gauge in the physical literature) that

$$< \nabla, A(x) >= 0. \tag{5.5}$$

Consider now the diffusion satisfying

$$h\frac{\partial \psi_h}{\partial t} = \frac{1}{2}(h\nabla + A)^2\psi_h + V\psi_h, \quad \psi(x,0) = f(x)e^{-\frac{1}{h}S(x)}, \tag{5.6}$$

for $x \in \mathbb{R}^3$ and $t > 0$. Under a variety of assumptions on $V(x)$ and $A(x)$, see [21] for example, the solution of (5.6) can be written as

$$\psi_h(x,t) = \mathbb{E}^{\,x}_h[f(B_h(t))e^{\frac{1}{h}\int_0^t < A(B_h), dB_h(s) > + \frac{1}{h}\int_0^t V(B_h(s))ds - S_h(B_h(t))}] \tag{5.7}$$

by means of a combination of the Feynman-Kac and Cameron-Martin-Girsanov translation formula.

**Remark 3.** *Let us explain how one can obtain (5.7). For this, it suffice to compute the infinitesimal generator of the semigroup defined by :*

$$P_t(F)(x) = \mathbb{E}^{\,x}_h[F(B_h(t))e^{\frac{1}{h}\int_0^t < A(B_h), dB_h(s) > + \frac{1}{h}\int_0^t V(B_h(s))ds}],$$

*and* $P_0(F)(x) = f(x)e^{-\frac{1}{h}S(x)}$. *See (5.7) for more on this. From the definitions and rearranging the terms, we obtain*

$$\frac{1}{t}[P_t(F)(x) - F(x)]$$
$$= \frac{1}{t}\mathbb{E}^{\,x}_h[(e^{\frac{1}{2h}\int_0^t A^2(B_h(s))ds + \frac{1}{h}\int_0^t V(B_h(s))ds}$$
$$\times (F(B_h(t))e^{\frac{1}{h}\int_0^t <A(B_h),dB_h(s)> - \frac{1}{2h}\int_0^t A^2(B_h(s))ds} - F(x))$$
$$+ e^{\frac{1}{h}\int_0^t <A(B_h),dB_h(s)> - \frac{1}{2h}\int_0^t A^2(B_h(s))ds}[F(x)e^{\frac{1}{2h}\int_0^t A^2(B_h(s))ds + \frac{1}{h}\int_0^t V(B_h(s))ds} - F(x)]$$
$$\to_{t\to 0} \frac{h}{2}\Delta F(x) + A(x)\nabla F(x) + \frac{1}{2h}A^2(x)F(x) + \frac{1}{h}V(x)F(x)$$
$$= \frac{1}{2h}(h\nabla + A(x))^2 F(x) + \frac{1}{h}V(x)F(x).$$

*Recall that* $< \nabla, A >= 0$. *Moreover, we have used above the following facts. Let* $A$ *be a smooth vector field and consider the stochastic differential equation*

$$dX_h(s) = A(X_h(s))ds + dB_h(t) \stackrel{\mathcal{L}}{=} A(X_h(s))ds + \sqrt{h}dB_1(t),$$

*then*

$$dY_h(s) := \frac{1}{\sqrt{h}}X_h(s) = \frac{1}{\sqrt{h}}A(\sqrt{h}Y_h(s)) + dB_1(t).$$

*Now by using Girsanov's formula we get*

$$\mathbb{E}^{\,x}_h[G(X_h(\cdot))] = \mathbb{E}^{\,x}_h[G(\sqrt{h}Y_h(\cdot))]$$
$$= \mathbb{E}^{\,x}_1[G(\sqrt{h}B_1(\cdot))e^{\frac{1}{\sqrt{h}}\int_0^t <A(\sqrt{h}B_1(s)),dB_1(s)> - \frac{1}{2h}\int_0^t A^2(\sqrt{h}B_1(s))ds}]$$
$$= \mathbb{E}^{\,x}_h[G(B_h(\cdot))e^{\frac{1}{h}\left(\int_0^t <A(B_h(s)),dB_h(s)> - \frac{1}{2}\int_0^t A^2(B_h(s))ds\right)}].$$

We can now proceed as in section 3. That is, we compute $dS_0(B_h(s), t-s)$ as function of $s$ for each $t \in [0,T]$ and then transform (5.7) into

$$\psi_h(x,t) = \mathbb{E}^{\,x}_h[f(B_h(t))Z_h(t)e^{-\frac{1}{2}\int_0^t \Delta S_0(B_h(s),t-s)ds}]e^{-\frac{1}{h}S_0(x,t)}, \tag{5.8}$$

with $Z_h(t)$ being

$$Z_h(t) = e^{-\frac{1}{h}\left(\int_0^t <(\nabla S_0 + A)(B_h(s),t-s),dB_h(s)> + \frac{1}{2}\int_0^t (\nabla S_0 + A)^2(B_h(s),t-s)ds\right)ds},$$

which is a martingale by the Novikov's criterion whenever

$$\mathbb{E}^{\,x}_h[e^{\frac{1}{2}\int_0^t (\nabla S_0 + A)^2(B_h(s),t-s)ds\big)ds}] < \infty.$$

Rescale $B_h(t)$ as to have a fixed set of probability laws and consider the stochastic differential equation

$$dX_h(s) = -[\nabla S_0(X_h(s), t-s) + A(X_h(s))]ds + \sqrt{h}dB_1(s), \tag{5.9}$$

with initial condition $X_h(0) = x$. From our assumptions on $A$ and $\nabla S_0$ we obtain the uniqueness and existence of $X_h(s)$. Now rewrite (5.8) as

$$\psi_h(x,t) = e^{-\frac{1}{h}S_0(x,t)}\mathbb{E}^{\,x}_h[f(X_h(t))e^{-\frac{1}{2}\int_0^t \Delta S_0(X_h(s),t-s)ds}] \tag{5.10}$$

by making use of the Cameron-Martin-Girsanov theorem. Again, from the convergence of $X_h(s)$ as $h$ tends to zero to the solution of

$$\dot{x}(s) = -(\nabla S_0(x(s), t-s) + A(x(s))), \quad x(0) = 0,$$

we obtain the desired limiting behavior of (5.8) and using the expansion of the solutions of $X_h(s)$ of (5.9) in terms of $\sqrt{h}$ we can get asymptotic expansions of $\psi_h(x,t)$.

## 6. Examples

In this section we present a few examples, some of which are rather simple, so much seemingly nobody has cared to write them down. We do so never the less.

(1) The case of the free particle with Hamiltonian $H(p) = \frac{p^2}{2}$ was mentioned in section (3). In this case

$$\psi_h(x,t) = E_h^x[e^{-<P,B_h(t)>}] = e^{-\frac{1}{h}(<x,P>-\frac{1}{2}tP^2)}.$$

Then we obtain that

$$-\lim_{h\to 0} h\log\psi_h(x,t) =< x,P > -\frac{1}{2}tP^2 = S_0(x,P,t).$$

(2) Consider now the case $V(x) = ax$ in $\mathbb{R}$. The general case $V(x) =< a,x >$ in $\mathbb{R}^n$ can be deduced from this by rotation. In this case the solution of (1.4) is given by

$$\psi_h(x,t) = e^{-\frac{1}{h}\left(xP+axt+\frac{a^2t^3}{6}+\frac{P^2t}{2}+\frac{at^2P}{2}\right)},$$

from which it follows that

$$S_0(x,P,t) = xP - axt - \frac{a^2t^3}{6} - \frac{P^2t}{2} - \frac{at^2P}{2}.$$

To finish with this short list of well known cases consider now the model in $\mathbb{R}$ when $V(x) = -\frac{x^2}{2}$. Thus

$$\psi_h(x,t) = e^{-\frac{1}{h}\left(\frac{x^2}{2}\coth t-(P-\frac{x}{\sinh t})^2\frac{\tanh t}{2}-h\log(\cosh t)^{\frac{1}{2}}\right)},$$

and we see as $h$ tends to zero we get

$$-\lim_{h\to 0} h\log\psi_h(x,t) = S_0(x,P,t) = \frac{xP}{\cosh t} - \frac{P^2-x^2}{2}\tanh t.$$

(3) In this part we will consider the second order partial differential operator

$$\Delta = e^{2x}\frac{\partial^2}{\partial x^2} + e^{2x}\frac{\partial}{\partial x}.$$

Acting on the class of smooth real valued functions. That is the Laplace-Beltrami operator for the metric $g_{11} = e^{-2x}$. And it is associated to a geodesic flow on the line, with Hamiltonian $H(p,x) = e^{2x}\frac{p^2}{2}$ see [10] for diffusion on manifolds. The corresponding Hamilton-Jacobi equation is

$$\frac{\partial S_0}{\partial t} + \frac{e^{2x}}{2}(\frac{\partial S_0}{\partial x})^2 = 0. \tag{6.1}$$

Being the associated diffusion equation

$$\frac{\partial \psi_h}{\partial t} = \frac{h}{2}\Delta\psi_h, \tag{6.2}$$

We shall take as initial condition for $S_0(x,t)$ the datum $s(x) = e^{-2x}p$ where $p$ is a real number and put $\psi_h(x,0) = \frac{s(x)}{h}$. A simple computation shows that the solution to (6.1) is given by

$$S_0(x,t) = e^{-2x}\frac{p}{1+2pt}.$$

Therefore the solution is defined when $p < 0$ only if $t < \frac{1}{2|p|}$. To find the diffusion with generator $\frac{h}{2}\Delta$ we have to solve

$$dX_h(t) = e^{X_h(t)} dB_h(t) + \frac{h}{2} e^{2X_h(t)} dt, \tag{6.3}$$

which for the law $P^x$ is given by

$$X_h(t) = -\log(e^{-x} + x - B_h(t)).$$

The last assertion is readily verified by using the Itô's formula. Hence the solution is defined until an explosion time

$$T_h(e^{-x}) = \inf\{t \geq 0 : B_h(t) \geq e^{-x}\} = \frac{1}{h} T_1(e^{-x}),$$

the last equality holds by scaling and $T_1$ refers to the same object for standard Brownian motion.

From all this we see that

$$\begin{aligned}
&\psi_h(x,t) \\
&= \mathbb{E}^{\,x}_h[e^{-\frac{p}{h}e^{-2X_h(t)}}] = \mathbb{E}^{\,0}[e^{-\frac{p}{h}(e^{-x}-B_h(t))^2}; t < \frac{1}{h}T_1(e^{-x})\}] \\
&= \mathbb{E}^{\,0}[e^{-\frac{p}{h}(e^{-x}-B_h(t))^2}] - \mathbb{E}^{\,0}[e^{-\frac{p}{h}(e^{-x}-B_h(t))^2}; t \geq \frac{1}{h}T_1(e^{-x})] \\
&= \mathbb{E}^{\,0}[e^{-\frac{p}{h}(e^{-x}-B_h(t))^2}] + o(1) = e^{-\frac{1}{h}S_0(x,t)} + o(1).
\end{aligned}$$

For general conditions two problems arise: first there appear singularities in $S_0(x,t)$ besides the explosions in $X_h(t)$, and a general statement would have take both of these facts into account.

## 7. Final Comments

- Things seem so much easier than in [20] only because here the functionals involved are much simpler than those considered by Schilder.
- Apart from the important problem related to the limit $\log -h\psi_h(x,t)$, there is a class of problems related to the study of the eigenfunctions of $\frac{h}{2}\Delta + V$ as $h$ tends to zero. See for instance [17] and [19].
- When a path integral representation for the solution of (1.1) exists see [3], a similar analysis can be performed.
- It is necessary to add that as far as asymptotic expansions goes, it is easier to start with an asymptotic expansion for $A$ in (3.3) and solve the simple, linear, recursive set of equations that appears when an expansion in powers of $h$ is considered.
- We want to observe that this exposition is a lengthening and an update of a previous article [13] written by the authors, that appeared in the already distant 1989.
- Last but non least. In statistics it is important to consider the transition density for a general diffusion when the step of discretization $h$ tends to zero. This type of problem was considered for the first time by Dacunha-Castelle & Florens-Zmirou in [7] in the case of a general diffusion and $n = 1$. Recently this has become a subject of active research, because of its implications in approaching the likelihood in several applied problems. The difference with the approach developed here consists of, that instead of functionals

of Brownian motion, there it is necessary to study functionals of Brownian bridges.

## References

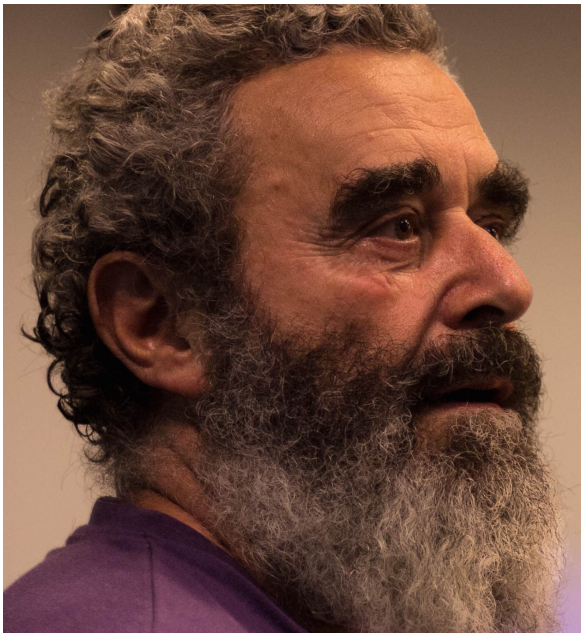

[1] S. Albeverio and T. Arede. The relation between quantum mechanics and classical mechanics. A survey of some mathematical aspects. Bibos prepint 28 (1983), Bielefeld.
[2] V.I. Arnold, Méthodes Mathématiques de la Mechanique Classique. MIR (1976), Moscow.
[3] R. Azencott, A. Doss. L'équation de Schrodinger quand $h$ tend vers zéro, une approche probabilistique. LNM Vol 1109 (1985).
[4] J.M. Bismut. Large deviations and Malliavin Calculus. Birhäuser (1989), Berlin.
[5] R.H. Cameron. A family of integrals serving to connect the Wiener and Feynman integrals. Jour. Math Phys. 34, 126-141 (1961).
[6] R.D. Carlitz, D.A. Nicole. Classical paths and quantum mechanics. Annals of Physics, V. 164, Issue 2, Pages 411-462 (1985).
[7] D. Dacunha-Castelle and D. Florens-Zmirou. Estimation of the coefficients of a diffusion from discrete observations, Stochastics. Volume 19, 1986-Issue 4 263-284 (1986).
[8] H. Doss. Sur une resolution stochastique de l'équation de Schroedinger à coefficients analytiques. Comm. Math Phys. V. 73 247-269 (1980).
[9] H. Doss. Démonstration probabiliste de certains développements asymptotiques quasi classiques. Bull. Sci. Math., V. 109 (No. 2), pp. 179-208 (1985).
[10] D. Elworthy and A Truman. Classical mechanics, the diffusion (heat) equation and the Schrödinger equation on a Riemannian manifold. Jour. of Math Physics 22, 2144 (1981).
[11] J. Feldman. On the Schrödinger and heat equations for nonnegative potentials. Tran Am Math Soc, V. 108,251-269 (1963).
[12] M. Freindlim and A.D. Wentzel. Radom perturbations of dynamical systems. Springer-Verlag. Berlin. (1984).
[13] H. Gzyl and J.R. León. Classical limits via Brownian Motion. Notas de Matemáticas del Departamento de Matemáticas de la Universidad de los Andes. Venezuela No. 100 (1989).
[14] K. Itô. Wiener integral and Feynman integral. Proc. Fourth Berkeley Symposium on Prob. and Stat. V. 2, 277-238. Univ. California Press (1961).
[15] J.R. León, S. Yosleff. High frequency approximation for the Helmholtz equation a probabilistic approach. Reports on Mathematical Physics. V. 40 1-16 (1997).
[16] R.G Littlejohn. The semiclassical evolution of wave packets. Phys. Reports 138. 4,5, 193-291 (1986).
[17] V.P. Maslov and M,V. Fedoniuk. Semiclassical approximation in Quantum Mechanics. D. Reidel Pubs. Co, (1981), Dordrecht.
[18] E. Nelson. Feynman integrals and the Schroedinger equation. Jour. Math. Phys. V 5, 332-343 (1969).
[19] D. Robert. Autour de l'approximation semi-classique. Progress in Mathematics No. 68, Birkhäuser, (1987).
[20] M. Schilder. Some asymptotic formulas for Wiener integrals. Trans. Am. Math. Soc. V.125 63-85 (1966).
[21] B. Simon. Functional integration and quantum physics. Academic Press, (1979), New York
[22] D.W. Strook. An introduction to the Theory of Large Deviations. S'ringer-Verlag,(1985), Berlin.
[23] S.R.S. Varadhan. Diffusion problems and partial differential equations. Tata Lecture Series. Springer Verlag,(1980), Berlin.
[24] A. Voros. Semiclassical approximations. Ann. IHP V. XXIV. No 1. 31-90 (1986).



Centro de Finanzas, IESA, Caracas, Venezuela.
*Email address*: henryk.gzyl@gmail.com

IMERL, Facultad de Ingeniería, Universidad de la República. Montevideo, Uruguay.
Escuela de Matemática UCV, Venezuela.
*Email address*: rlramos@fing.edu.uy

CHARLA PLENARIA



# DINÁMICA ELEMENTAL DE CUBRIMIENTOS DEL ANILLO

ALVARO ROVELLA

Publicaciones Matemáticas del Uruguay
Volumen 17, Julio 2019, Páginas 95–104
S 0797-1443

# DINÁMICA ELEMENTAL DE CUBRIMIENTOS DEL ANILLO

ALVARO ROVELLA

Resumen. Se trata de introducir algunas ideas para estudiar la dinámica de mapas de cubrimiento del anillo abierto, con la intención de dar una mínima muestra de las preguntas que se plantean y de las técnicas que se usan.

## 1. Introducción.

El objeto de estudio de esta nota es la dinámica de los cubrimientos no invertibles del anillo abierto $A$ en sí mismo. Si $f$ es un cubrimiento de grado $d$, donde $d$ se supone de valor absoluto mayor que uno, entonces $f$ es no invertible ya que cada punto tendrá $|d|$ primágenes. El caso $|d| = 1$ refiere obviamente a homeomorfismos del anillo y la teoría es completamente diferente: desde los resultados a las técnicas, no hay similitudes. Sin embargo, sí puede pensarse el presente trabajo como una generalización de la dinámica en el círculo $S^1$. La generalización es (en muchos aspectos) bastante simple cuando se trabaja en el anillo cerrado, ya que en este caso, una transformación de cubrimiento debe llevar el borde en sí mismo, dando origen a una transformación de cubrimiento de $S^1$. Uno de los objetivos es generalizar al caso del anillo abierto $A$ la existencia de semiconjugaciones. En efecto, todo cubrimiento de grado $d$ de $S^1$ es semiconjugado a uno particular, el mapa $m_d(z) = z^d$, donde $d$ es el grado de $f$. Una semiconjugación de un mapa $f : S^1 \to S^1$ de grado $d$ es una función continua y de grado uno tal que $hf = m_d h$; la existencia de tal $h$ es posible siempre que $|d| > 1$, lo que marca una gran diferencia con el caso de los homeomorfismos. Cada Plateau (o intervalo maximal en el que $h$ es constante) es un intervalo errante o periódico de $f$. Por lo tanto se puede describir la dinámica de $f$ en términos de la función $h$ y el mapa $m_d$. Este resultado cimienta la afirmación de que $m_d$ es la dinámica más simple que un cubrimiento de grado $d$ del círculo puede presentar (si bien eso no dice que la dinámica de $m_d$ sea elemental). Muchos resultados en dinámica unidimensional se obtienen a partir de la existencia de semiconjugaciones, los más remarcables: la clasificación de cubrimientos expansivos y la clasificación por conjugación de todos los cubrimientos.

Para dimensión dos el problema se traduce así : sea $f$ un cubrimiento de grado $d$ del anillo $A$, donde $|d| > 1$; se plantea si existe una semiconjugación $h : A \to S^1$ de $f$ a $m_d$, es decir, una $h$ continua que cumpla $hf = m_d h$. En el caso en que $A$ es el anillo cerrado la respuesta es afirmativa, e impica una simplificación importante en el estudio de la dinámica de $f$. Pero cuando $A$ es el anillo abierto, entonces el problema es más intrincado: es el objetivo de esta nota explicar este punto, las equivalencias de la existencia de semiconjugaciones y las posibles obstrucciones.

**1.1. Definiciones básicas.** A partir de ahora, $A$ es el anillo abierto y $f$ es un cubrimiento de $A$ en sí mismo. Se trata de describir el comportamiento a largo

©2019 PMU

plazo de las órbitas de la transformación. Se puede interpretar el anillo de varias maneras: una es pensar $A$ como el producto $S^1 \times (0,1)$, otra como el conjunto $\mathbb{C}^* = \{z \in \mathbb{C} : z \neq 0\}$. Con esta última realización de $A$ se puede definir el índice de una curva cerrada $\gamma \subset A$ como el número de vueltas que la curva da al 0. Una curva cerrada se dice esencial si su índice es no nulo y trivial en caso contrario. Se define el grado del cubrimiento $f$ como el índice de la curva $f \circ \gamma$ donde $\gamma$ es una curva de índice uno. Es claro que el grado así definido no depende de la elección de la curva $\gamma$. También es fácil probar a partir de estas definiciones que cada punto de $A$ tiene exactamente $|d|$ preimágenes por $f$.

Si $k > 0$, se define $f^k$ como la composición de $f$ consigo misma $k$ veces. Es obviamente una aplicación de $A$ en $A$. Cuando $k = 0$ entonces $f^k$ es la función identidad y cuando $k > 0$ la notación $f^{-k}$ denota la preimagen de $f^k$, o sea $(f^k)^{-1}$.

Si $\tilde{A} = \mathbb{R} \times (0,1)$ y $P : \tilde{A} \to A$ es $P(x,y) = (\exp(2\pi i x), y)$ entonces $(\tilde{A}, P)$ es el cubrimiento universal de $A$. Cualquier cubrimiento $f$ de grado $d$ de $A$ tiene un *levantamiento* $F : \tilde{A} \to \tilde{A}$, es un homeomorfismo que cumple $PF = fP$, y $F(x+1,y) = F(x,y)+(d,0)$ para todo $(x,y)$. Una vez conocido un levantamiento $F$ de $f$ los demás se obtienen tralsadando $F$ en la dirección horizontal: $F_j = F + (j,0)$, $j$ entero. En cierto sentido se ve que $F$ tiene una expansión en la dirección horizontal, ya que dos puntos $(x,y)$ y $(x',y)$ a distancia $k$ se transforman por $F$ en dos puntos a distancia $|d|k$. En este hecho reside buena parte de las razones que hacen a esta teoría tan diferente a la de homemorfismos de $A$ y se aproxima a una generalización de los cubrimientos de $S^1 = \{z \in \mathbb{C} : |z| = 1\}$. Se hará mención repetidas veces a la dinámica de cubrimientos de $S^1$. El cubrimiento universal de $S^1$ se denota por $(\mathbb{R}, P_0)$ donde $P_0(x) = \exp(2\pi i x)$.

**1.2. Antecedentes.** Sin la mínima intención ni posibilidad de dar una descripción completa de los resultados conocidos, podemos acaso rastrear algunos de las primeras apariciones de mapas de este tipo en la literatura. Uno de los primeros resultados en esta teoría aparecen en las tesis de F.Przytycki, (ver [5]) que, en sus investigaciones al respecto de la estabilidad de endomorfismos de clase $C^1$ utiliza ciertas transformaciones del toro $T^2$ que provienen de extender un cubrimiento del anillo y que proporcionan ejemplos de mapas Axioma A que no son $\Omega$-estables, distinguiendose así de los resultados válidos para homeomorfismos. M ás adelante, Tsujii encontró transformaciones de clase $C^\infty$ del anillo abierto que poseen un atractor global que soporta una medida invariante absolutamente continua respecto a Lebesgue (ver [6]). Esta propiedad es imposible para mapas invertibles. En [1], las mismas transformaciones se estudian desde un punto de vista topológico, demostrándose que el atractor tiene *frecuentemente* interior no vacío. En la mayoría de los casos, estas transformaciones aparecen como ejemplos anómalos. Estas notas están están basadas en [2], ver también [3] y [4] donde se estudian condiciones para la existencia de puntos periódicos. Estas son las únicas referencias donde el objeto de estudio son los mapas de grado $d$ del anillo (y no aparecen como ejemplos o contraejemplos como en los trabajos antes mencionados).

## 2. Ingredientes.

En esta sección definimos los tres objetos que desde diferentes ángulos muestran la complejidad que presenta el asunto. Se introduce primero el *número de rotación* para construir *semiconjugaciones*, que para mapas de cubrimiento de $S^1$ constituye una herramienta muy útil. Sin embargo, como se verá al final de la sección, no es

posible generalizar estas construcciones a aplicaciones del anillo. Posteriormente se definen *conectores* y se muestra su relevancia en la búsqueda de semiconjugaciones.

**2.1. Número de rotación.** Se estudia primero la definición para cubrimientos del círculo $S^1$, donde la técnica permite llegar a una clasificación bastante satisfactoria. Sea $f$ una transformación de cubrimiento de grado $d$ de $S^1$. Notar que si $F$ es un levantado de $f$ y $d > 0$ entonces $F(x+1) = F(x) + d$ para todo $x$, de donde se deduce que $F(x+n) = F(x) + dn$ y esto implica que $\lim_{x\to\pm\infty} F(x) = \pm\infty$, y que $F^n(x)$ debe crecer exponencialmente si $x$ es grande.

**Teorema 2.1.** *Sea $f$ un cubrimiento de grado $d$ de $S^1$ donde $d$ es un entero de valor absoluto mayor que 1, y $F : \mathbb{R} \to \mathbb{R}$ un levantamiento de $f$. Entonces para cada $x \in \mathbb{R}$ existe el límite:*

$$\lim_{n\to+\infty} \frac{F^n(x)}{d^n}$$

*Además, denotando por $\rho(x)$ a ese límite, la función $\rho$ cumple las siguientes propiedades*

1. *$\rho$ es continua*
2. $\rho(x+1) = \rho(x) + 1$
3. $\rho(F(x)) = d\rho(x)$

Se da a seguir una idea de la prueba. Hay un hecho básico que la sustenta:
**Afirmación:** Si $F : \mathbb{R} \to \mathbb{R}$ es una función continua tal que $F(x+1) = F(x) + d$ para todo $x$, entonces la función $F(x) - dx$ está acotada.
Prueba: la segunda hipótesis sobre $F$ implica que $F(x) - x$ es periódica de período igual a uno; la continuidad de $F$ implica que $F(x) - x$ es acotada en el intervalo $[0, 1]$.
Para la prueba del teorema, se considera el espacio $\mathcal{H}$ de todas las funciones $H : \mathbb{R} \to \mathbb{R}$ tales que: (1) $H$ es continua (2) La función que a cada $x \in \mathbb{R}$ asocia $|H(x) - x|$ es una función acotada y (3) $H(x+1) = H(x) + 1$ para todo $x$.
En el espacio $\mathcal{H}$ se considera la distancia $D(H_1, H_2) = \sup |H_1(x) - H_2(x)|$. Notar que $D$ está bien definida porque $H_1 - H_2$ está acotado en virtud de la propiedad (2), si bien cada $H_i$ no es acotada. Es sencillo probar que $D$ es una distancia en $\mathcal{H}$ y que con esta distancia $\mathcal{H}$ es un espacio métrico completo. Se considera el operador $\Phi$ definido en $\mathcal{H}$ por:

$$\Phi(H) = \frac{1}{d} H \circ F$$

Se prueba fácilmente (usando la afirmación) que $\Phi$ lleva $\mathcal{H}$ en sí mismo y que es una contracción. Luego su punto fijo $H$ es una función que cumple las tres propiedades enumeradas en la tesis del teorema. Resta ver que el límite existe y es igual a $H$. Notar que

$$\frac{F^n(x)}{d^n} = \frac{F^n(x) - H(F^n(x))}{d^n} + \frac{H(F^n(x))}{d^n}$$

El primer sumando tiende a 0 ya que $|H(y) - y|$ es una función accotada. El segundo término es igual $\Phi^n(H)$ calculado en $x$; como $H$ es punto fijo de $\Phi$ también lo es de $\Phi^n$: se deduce que el límite existe y es igual a $H(x)$. En definitiva, es $\rho = H$.

**2.2. Semiconjugaciones.** Antes de comenzar con esta sección es conveniente pensar un poco acerca de la equivalencia entre dinámicas. Diremos que dos sistemas dinámicos $f$ y $g$ son conjugados o equivalentes cuando existe un homeomorfismo $h$ del dominio de $f$ al dominio de $g$ tal que $hf = gh$. Esto implica que desde el punto de vista dinámico los sistemas $f$ y $g$ son indistinguibles, ya que el comportamiento dinámico de $f$ se traduce por el homeomorfismo $h$ en un comportamiento idéntico para el sistema $g$. Sería lógico, pero como pronto se verá, muy ambicioso, proponerse el siguiente plan: establecer una familia de sistemas conocidos, más fáciles de entender y luego probar que cualquier otro es conjugado a uno de estos sistemas más familiares.
Dentro del conjunto de todos los sistemas que se están considerando en estas notas, hay una subfamilia particularmente interesante y rica: es la clase de todos los cubrimientos $m_d(z) = z^d$ de $S^1$ con $d \in \mathbb{Z}$. Es un representante en cada clase de homotopía de mapas de $S^1$. Es claro que pocos cubrimientos de $S^1$ y ningún cubrimiento de $A$ es conjugado a algún $m_d$, pero sí se puede definir un concepto un poco más débil que permite decir mucho acerca de la dinámica de aquellos cubrimientos del anillo que cumplan con la propiedad.

**Definición 2.1.** Sea $f$ un cubrimiento de grado $d$ de $X$, donde $X$ puede ser $S^1$ o $A$. Una función $h : X \to S^1$ es una semiconjugación de $f$ si se cumple que $hf = m_d h$ y que la imagen por $h$ de una curva de índice uno es una curva de índice $\pm 1$. En caso de existir tal $h$, se dice que $f$ es semiconjugada a $m_d$.

Notar antes que nada que la segunda condición impuesta a la semiconjugación es simplemente para evitar trivialidades, ya que si $f$ es cualquier cubrimiento de $X$ y $h(x) = 1$ para todo $x$ entonces la ecuación $hf = m_d h$ se cumple, pero significa nada. Como consecuencia del teorema 2.1 se tiene el siguiente hecho:

**Corolario 2.1.** *Si $f$ es un cubrimiento de grado $d$ de $S^1$ y $|d| > 1$, entonces $f$ es semiconjugado a $m_d$.*

La función $\rho$ obtenida en el teorema 2.1 cumple $\rho(x+1) = \rho(x) + 1$. Esto implica que se puede bajar por la proyección universal $P_0$ de $S^1$ a una función $h$ de $S^1$ en $S^1$. Las demás propiedades de $h$ se deducen de las de $\rho$. □

Continuando con los cubrimientos de $S^1$ se verá a seguir una aplicación de la existencia de semiconjugaciones. Antes de esto, una observación acerca de las semiconjugaciones de $m_d$:
Sea $h$ una semiconjugación de $m_d$. Se puede probar que $h$ es una conjugación, es decir, es un homeomorfismo. Además vale que el conjunto de todos las semiconjugaciones de $m_d$ es un grupo con la composición, isomorfo al grupo dihedral con $|d-1|$ elementos, $D_{d-1}$. Si $h$ es una semiconjugación de un cubrimiento $f$ entonces toda otra semiconjugación $h'$ de $f$ se obtiene componiendo $h$ con un elemento de $D_{d-1}$.

**Definición 2.2.** Sea $f$ un cubrimiento de $S^1$ y defina $\Lambda_f$ como el conjunto de los puntos $x$ tales que $h^{-1}(h(x))$ contiene algún punto diferente de $x$.

**Teorema 2.2.** *Si $f$ es un cubrimiento de $S^1$, entonces el conjunto $\Lambda_f$, que fue definido a partir de $h$ (una semiconjugación de $f$) no depende en realidad de $h$. Además se tienen las siguientes propiedades:*

1. *$\Lambda_f$ es completamente invariante (es decir, $f^{-1}(\Lambda_f) = \Lambda_f$).*
2. *Cada $h^{-1}(h(x))$ es un intervalo (puede ser un punto).*

3. *Si $L$ es una componente no trivial de $\Lambda_f$, entonces el interior de $L$ es periódico o errante y la restricción de $f$ a esa componente es inyectiva.*
4. *$\Lambda_f$ es vacío sii $f$ es conjugada a $m_d$. Si no es vacío es denso.*

Sea $h'$ otra semiconjugación de $f$; por el ejercicio anterior, existe un homeomorfismo $c$ tal que $h' = ch$. De esto se deduce la primera afirmación. Las afirmaciones de (1) a (4) no son difíciles de probar. Además se puede progresar en la clasificación usando la siguiente idea:

Cuando $\Lambda_f$ es vacío entonces $f$ es conjugado a $m_d$ y cuando no, cada componente de $\Lambda_f$ es un intervalo, donde $f$ es inyectiva y o bien es errante, o bien es periódica. Si dos funciones tienen el mismo conjunto $\Lambda$ entonces serán conjugadas sii sus restricciones a los intervalos periódicos son homeomorfismos del intervalo conjugados. Esta idea conduce a una clasificación de los cubrimientos de $S^1$. Ver el resultado preciso en [2].

**2.3. Conectores.** Ya se definió número de rotación y se vio como sirve para obtener semiconjugaciones para cubrimientos de $S^1$. La cuestión ahora es ver qué de todo esto puede generalizarse a cubrimientos de $A$. En esta parte se introduce un tercer ingrediente que es una propiedad geométrica determinante. Se puede pensar $A$ como la esfera menos los polos; la compactificación con dos puntos de $A$ es la esfera $S^2$. Se denota por $A^*$ a la compactificación, por $N$ y $S$ los puntos agregados.

**Definición 2.3.** Se llama conector a cualquier subconjunto cerrado y conexo $K$ contenido en $A$, que cumple dos propiedades:
(1) $K \cup \{N, S\}$ es conexo en $A^*$,
(2) $A \setminus K$ es conexo.

Equivalentemente, $K$ es un conector si $K^* = K \cup \{N, S\}$ es compacto y conexo en $A^*$ y su complemento en $A^*$ es conexo. O también un conector es un conjunto que conecta $N$ con $S$ y que no separa $A$. Por ejemplo, $K_1 := \{(x, 1) : x \in (0, 1)\}$ es un conector en $A$. También es un conector la curva $K_2 = \{P(h(x), x) \in S^1 \times (0, 1) \ : \ x \in (0, 1)\}$ donde $h$ es un homeomorfismo de $(0, 1)$ en $\mathbb{R}$: esta curva se enrosca en ambas componentes de la frontera del anillo. En general una curva simple $\gamma : (0, 1) \to A$ tal que $\gamma(t) \to S$ cuando $t \to 0$ y $\gamma(t) \to N$ cuando $t \to 1$ es un conector. Pero no todo conector contiene necesariamente una curva tal.
No es un conector si separa el anillo, por ejemplo si a una curva como la $\gamma$ de recién se le une una curva esencial cualquiera del anillo.
La prueba de las siguientes propiedades y ejemplos de conectores es sencilla

En esta afirmaciones $K$ denota un conector de $A$ y $f$ un cubrimiento de grado $d$ de $A$.

1. La preimagen por $f$ de un conector es igual a la unión de $|d|$ conectores.
2. $f$ es una biyección de cada componente de $A \setminus f^{-1}(K)$ sobre $A \setminus K$.
3. Un conexo $K$ conecta $N$ con $S$ sii corta toda curva esencial en $A$. Por lo tanto $K$ es un conector sii es conexo, cerrado, su complemento es conexo y corta toda curva esencial.

## 3. Equivalencias.

El resultado fundamental de esta sección relaciona una condición puramente analítica como es la existencia del número de rotación con una condición más bien geométrica. Un conector $K$ es libre si $f(K) \cap K = \emptyset$, y es invariante si $f(K) = K$.

Sea $(x,y) \in \tilde{A} = \mathbb{R} \times (0,1)$, $F$ un levantamiento de $f$ y $(x_n, y_n) = F^n(x,y)$. Se define el número de rotación como

$$\rho_F(x,y) = \lim_{n \to +\infty} \frac{x_n}{d^n}$$

en caso que el límite exista. Si existe para todo punto de $\tilde{A}$ y es una función continua, se dice que $f$ tiene número de rotación. Usando que $(F+j)^n = F^n + j(d^{n-1} + \cdots + d+1)$ se prueba que $\rho_{F+j} = \rho_F + \frac{j}{d-1}$. Luego la definición no depende del levantado que se haya elegido.

**Teorema 3.1.** *Sea $f$ un cubrimiento de grado $d$ del anillo abierto $A$, donde $|d| > 1$. Se consideran las siguiente propiedades:*
*(a) Para cualquier levantamiento $F$ de $f$ existe la función número de rotación y es una función continua.*
*(b) Existe una semiconjugación de $f$.*
*(c) Existe un conector libre para $f$.*
*(d) Existe un conector invariante para $f$.*

*Entonces (b) (c) y (d) son equivalentes y (a) implica cualquiera de ellas.*
*Si la función que a cada $(x,y) \in \tilde{A}$ asocia $x_1 - dx \in \mathbb{R}$ es acotada (donde $(x_1, y_1) = F(x,y)$), entonces $f$ cumple las 4 condiciones.*

Ver el último ejemplo de estas notas donde se muestra que (a) no es equivalente al resto. La implicación de (a) a (b) es más bien obvia, y ya se explicó en la sección anterior (ojo que la hipótesis de que el número de rotación es continuo se necesita para que sea fácil, comparar con el problema planteado al final). Para probar la última de las afirmaciones del teorema se copia la prueba del teorema unidimensional, ya que la hipótesis impone la validez de la afirmación previa. La siguiente secuencia de lemas implicará el resto del teorema.

**Lema 3.1.** *Si $K$ es un conector libre para $f$ entonces existen $|d-1|$ conectores invariantes.*

Se indica un camino para la prueba en el caso $d = 2$ para ahorrar discusiones. Notar que $f^{-1}(K) \cap K = \emptyset$ (caso contrario $K$ no es libre). Además $f^{-1}(K)$ es la unión de 2 conectores. Como $K$ es conexo y no corta $f^{-1}(K)$, resulta que una sola de las dos componentes de $A \setminus f^{-1}(K)$ contiene $K$; se denota por $U$ a esta componente y $V$ a la otra. Entonces resulta que $f(V)$ contiene a la clausura de $V$ y $f$ es inyectiva en $V$. Para cada $N \in \mathbb{N}$ se define:

$$V_N = \bigcap_{n=0}^{N} f^{-n}(V).$$

Es conexo porque es intersección de conexos: no porque cada $f^{-n}(V)$ sea conexo, sino porque sólo una componente de $f^{-n}(V)$ está contenida en $V_{n-1}$. Además cada $V_N$ es abierto y contiene a la clausura de $V_{N+1}$. Se define $K'$ como la intersección de todos los $V_N$. Se deduce que $K'$ es conexo por ser intersección de conexos encajados, es cerrado porque es también la intersección de las clausuras de los $V_N$, conecta $S$ y $N$ porque cada $V_N$ los conecta y su complemento tiene una sola componente porque lo mismo ocurre para cada $V_N$. Se concluye que $K'$ es un conector, y es invariante por construcción. □

**Lema 3.2.** *Si un cubrimiento de grado d de A (con $|d| > 1$) tiene un conector invariante, entonces es semiconjugado a $m_d$.*

La idea (para $d = 2$ es ir definiendo la $h$ así : $h$ vale 1 en el conector invariante y ha de valer $-1$ en su preimagen, si se espera que la ecuación $hf = m_2 h$ se cumpla. Tomando $f$-preimágenes sucesivas de $K$ y $m_2$-preimágenes del 1 y ordenándolas se va definiendo la $h$. Esto define $h$ apenas en un subconjunto de $A$, exactamente el subconjunto $R$ formado por aquellos puntos que alguna iterada positiva de $f$ lleva a $K$. Un instante de observación permite concluir que hay una única manera de definir $h$ en cada componente del complemento de $R$ de manera que sea continua. □

**Lema 3.3.** *Sea $h$ una semiconjugación de $f \in Cov_d(A)$. Entonces $h^{-1}(1)$ contiene un conector invariante.*

Es claro que la preimagen de 1 es un conjunto invariante, no tiene que ser conexo necesariamente pero es seguro que contiene un conector. Si no fuese así, usando la tercera de las propiedades de los conectores, se tiene una curva esencial $\gamma$ que no corta la preimagen de 1. Quiere decir que $h\gamma$ tiene índice 0 lo que contradice la definición de semiconjugación. □

## 4. Problemas y ejemplos.

**Ejemplo:** Si $\bar{A}$ es el anillo cerrado y $f$ es un cubrimiento de $\bar{A}$, entonces $f$ cumple las condiciones del teorema 3.1. Para demostrarlo, se puede repetir la demostración hecha en el caso unidimensional, ya que el dominio fundamental del cubrimiento universal tiene clausura compacta y por lo tanto vale la propiedad de acotación que sustentaba la prueba de la existencia del número de rotación en aquel caso.

El siguiente problema está en la base de las discusiones anteriores.
Se considera $\mathbb{G}_d$ el espacio de gérmenes de mapas $d$ a 1 con la topología de Whitney. Esto es, cada elemento de $\mathbb{G}_d$ está representado por una transformación $f$ definida en un entorno $U$ de 0 de forma que lleva 0 en 0 y es un cubrimiento de grado $d$ del anillo $U \setminus \{0\}$ en el anillo $f(U) \setminus \{0\}$. Dos transformaciones son equivalentes si coinciden en un entorno de 0; el cociente de esta relación es el espacio de gérmenes, $\mathbb{G}_d$. La topología está dada por las bases de entornos

$$\mathcal{U}(f;\epsilon) = \{g : |g(x) - f(x)| < \epsilon(x)\ \forall x\},$$

donde $\epsilon$ es una función positiva y continua definida en $U \setminus \{0\}$, y $U$ es un entorno de 0.
*Gérmenes conjugados.* Dos transformaciones $f$ y $g$ son conjugadas si existe un entorno de 0 y un homeomorfismo $h$ definido en ese entorno tal que $fh = hg$. En este caso, es obvio que si $f'$ y $g'$ son transformaciones equivalentes a $f$ y $g$, estas también serán conjugadas; así se define la conjugación de gérmenes. Es claro que la relación de conjugación es de equivalencia en $\mathbb{G}_d$, por lo tanto vale plantearse:

**Problema:** Hallar un representante de cada clase de conjugación de gérmenes.

Esta es una pregunta demasiado ambiciosa. Para internarse un poco en su significado puede venir bien atender el próximo lema y los ejemplos que le siguen.

Se restringirá la pregunta a dos clases bien particulares de gérmenes:

El germen representado por $f$ es *repulsor* si existe un entorno abierto $U$ de 0 tal que la clausura de $f^{-1}(U)$ está contenida en $U$ y para todo $x \in U$ existe un $n \geq 0$ tal que $f^n(x) \in U \setminus f^{-1}(U)$.
El germen representado por $f$ es *atractor* si existe un entorno abierto $U$ de 0 tal que la clausura de $f(U)$ está contenido en $U$ y para todo $x \in U$ se cumple que $\lim_{n\to+\infty} f^n(x) = 0$.

**Lema 4.1.** *Hay una sola clase de conjugación de gérmenes repulsores.*

Conviene introducir otra notación:
Sea $\gamma$ una curva simple cerrada y esencial en $U \setminus \{0\}$; defina $A_\gamma$ como la componente de $U \setminus \gamma$ que contiene al 0. Para la prueba del lema pueden seguirse los siguientes pasos:

1. Notar que si $\gamma$ es como arriba, entonces $f^{-1}(\gamma)$ es también una curva simple cerrada y esencial.
2. Si $f$ es representante de un germen repulsor entonces existe una curva $\gamma$ simple cerrada y esencial en $U \setminus \{0\}$ tal que $f^{-1}(\gamma) \subset A_\gamma$ y
$$\cap_n A_{\gamma_{-n}} = \{0\},$$
donde $\gamma_{-n} = f^{-n}(\gamma)$.
3. Sean $f_1$ y $f_2$ representantes de gérmenes repulsores y para $i = 1, 2$, sea $\gamma^i$ la curva dada en la definición. Sea $h$ un homeomorfismo de $\gamma^1$ en $\gamma^2$, defina $h$ en $x \in f_1^{-1}(\gamma^1)$ de forma que $h(f_1(x)) = f_2(h(x))$. Notar luego que es posible extender $h$ a un homeomorfismo de $A_{\gamma^1} \setminus f_1^{-1}(A_{\gamma^1})$ en $A_{\gamma^2} \setminus f_2^{-1}(A_{\gamma^2})$. Luego usando la propiedad de repulsor y la ecuación de conjugación se ve que $h$ se extiende de una única forma al resto de $A_{\gamma^1}$.
4. Una vez definida $h$ en este dominio fundamental, se extiende de la única manera posible (para que cumpla la ecuación de conjugación) a su preimagen. Y así sucesivamente a todas las preimágenes, cuya unión cubre $U \setminus \{0\}$ como consecuencia de la definición de repulsor.

Un germen es *topológicamente estable* si tiene un entorno tal que todos los gérmenes en ese entorno son conjugados.

**Corolario 4.1.** *Todo germen repulsor es topológicamente estable.*

Esto se deduce del lema anterior una vez que se demuestra que todo perturbado de un germen repulsor es también repulsor.

**Ejemplo.** La transformación $m_2(z) = z^2$ no es estable.
Se puede escribir $m_2(r, x) = (r^2, 2x)$, donde las coordenadas son $r \in (0, 1)$ y $x$ es un número real módulo 1, o sea $x \in \mathbb{R}/\mathbb{Z}$. Sea $\epsilon(r, x)$ una función continua y positiva definida en $(0, 1)$. Se puede suponer que $\epsilon$ no depende de $x$, tomando en vez de $\epsilon$ la función $\epsilon'(r) = \text{mín}_x\{\epsilon(r, x)\}$. Es posible encontrar funciones continuas $\rho(r)$ y $\phi(r, x)$ tales que $\rho(r) < \epsilon(r)$ para todo $r$ y $\phi(r, x) = 2x$ para $|x| > \rho(r)$, de forma que la transformación $f(r, x) = (r^2, \phi(r, x))$, cumple las siguientes propiedades:
1. el conjunto $R = \{(r, x) : r < 1\ x < \rho(r)\}$ es un conjunto invariante para $f$
2. $|\phi(r, x) - 2x| < \epsilon(r)$ para todo $(r, x)$.

Se deduce que $f$ está en el entorno $\mathcal{U}(m_2, \epsilon)$ y no puede ser conjugada a $m_2$ porque $f$ tiene un conjunto invariante de interior no vacío donde $f$ es inyectiva, pero $m_2$ no. Ver los detalles en [2].

Más o menos con la misma idea se puede probar lo siguiente, en contraposición al corolario 4.1

**Corolario 4.2.** *Ningún germen atractor es topológicamente estable.*

El próximo ejemplo muestra que no todo cubrimiento del anillo abierto cumple las condiciones equivalentes del teorema 3.1. Su construcción es bastante más difícil, puede leerse el detalle en [2].

**Contraejemplo.** Existe un cubrimiento $f$ de grado $d$ del anillo abierto $A$ que no es semiconjugado a $m_d$. La transformación $f$ se construye a partir de un cubrimiento de $A$ por anillos $A_n$ ($n \in \mathbb{Z}$) donde los anillos se tocan solo en su frontera y de forma tal que $f : A_n \to A_{n+1}$ es un cubrimiento de grado 2 para todo $n \in \mathbb{Z}$. Fijando el anillo $A_0$ se puede conseguir que se cumplan las siguientes propiedades, para cada $n > 0$:

1. Existen dos curvas $\gamma_n^i$ ($i = 1, 2$) conectores de $A_0$ que tienen la misma imagen por $f^n$.
2. Existen puntos $P_i^n \in \gamma_i^n$ que tienen la misma imagen por $f^n$ y cumplen que $P_1^n$ tiende al borde interior de $A_0$ y $P_2^n$ tiende al borde exterior de $A_0$ cuando $n$ tiende a $+\infty$.
3. Cada una de las curvas $\gamma_i^n$ corta al menos $n$ veces y en el mismo sentido a la curva $\{(x, y) \in A_0 : x = 1\}$ contenida en $A_0$.

Esta construcción no es difícil, se prueba por inducción en $n$ suponiendo que se cumple hasta $n$ y luego construyendo con cuidado la restricción de $f$ al anillo $A_n$.

Una vez construida $f$, observe que no puede haber un conector invariante: En efecto, si $K$ es un conector, tómese una componente $K_0$ de $A_0 \cap K$ que sea un conector de $A_0$. Note que si $n$ es grande, $A_0$ debe intersectar $\gamma_1^n$ después de $P_1^n$ y $\gamma_n^2$ antes de $P_n^2$. Esto implica que $f^n(K_0)$ tiene autointersección esencial. Por lo tanto, $f(K) = K$ implica que $K$ no es un conector ya que $f^n(K_0)$ es esencial.

Resulta entonces que el número de rotación no existe para este cubrimiento. Y que no tiene un conector libre.

**Problema.** Clasificar todos los cubrimientos que no poseen número de rotación.

**Ejemplo.** Puede existir una semiconjugación pero el número de rotación no. Sea $F(x, y) = (2x/y, y^2)$ definida en $\mathbb{R} \times (0, 1)$. Se verifica que $F$ es el levantado de un cubrimiento de grado 2 del anillo abierto $A$. Se puede calcular explícitamente el límite de $x_n/2^n$ y da $+\infty$ para cualquier $(x, y)$ si $(x_n, y_n) = F^n(x, y)$. Sin embargo, es fácil construir un conector invariante, y por lo tanto $f$ es semiconjugada a $m_2$.

**Problema.** Sea $f$ tal que para todo $(x, y) \in \tilde{A}$ existe el número de rotación. ¿Es necesariamente una función continua?

Notar que si $x$ es un levantado cualquiera de un punto periódico de $f$, entonces el número de rotación existe para este $x$. Podría plantearse una pregunta menos ambiciosa:

**Problema.** Sea $f$ un cubrimiento de $A$. ¿Es la función número de rotación continua en los periódicos de $f$?

La propiedad de la existencia de conector libre se puede generalizar a ámbitos mucho más amplios. Sea $X$ un espacio y $f : X \to X$. Dos puntos $P$ y $Q$ se dicen libremente conectados si existe un continuo $K$ que contiene ambos puntos y tal que $f(K) \cap K \subset \{P, Q\}$. Se permite la intersección en $P$ o $Q$ para no descartar de primera a los puntos fijos de $f$. El contraejemplo que se expuso arriba muestra que los puntos $N$ y $S$ de la compactificación de $A$ no son libremente conectados para el cubrimiento $f$ allí construido. Sin embargo, no es difícil ver que todo par de puntos de $A$ están libremente conectados si $f$ es un cubrimiento de $A$. El último problema aquí planteado es:

**Problema.** Describir todos los cubrimientos ramificados de la esfera $S^2$ tales que todo par de puntos están libremente conectados.

Por ejemplo, si $f$ es un cubrimiento de $A$ y es semiconjugado, entonces la extensión de $f$ a la compactificación con dos puntos de $A$ cumple la propiedad anterior.
Una respuesta satisfactoria a esta pregunta puede implicar interesantes consecuencias en dinámica bidimensional.

## Referencias

[1] R.Bamón, J.Kiwi, J.Rivera-Letelier, R.Urzúa. *On the topology of solenoidal attractors on the cylinder.* Ann. del'Inst. H.Poincaré (C) 23-2 (2006) 209-236.
[2] J. Iglesias, A. Portela, A. Rovella, J. Xavier. *Dynamics of annulus maps I: semiconjugacies.* Math Z. DOI 10.1007/s00209-016-1653-6 (2014)
[3] J. Iglesias, A. Portela, A. Rovella, J. Xavier. *Dynamics of annulus maps II: periodic points for coverings.* Fund. Math. DOI 10.464/fm 89-6-2016
[4] J. Iglesias, A. Portela, A. Rovella, J. Xavier. *Dynamics of annulus maps III: completeness.* Nonlinearity 29(9) 2641-2655 (2016).
[5] F.Przytycki. *On $\Omega$-stability and structural stability of endomorphisms satisfying Axiom A.* Studia mathematica 60 (1977) 61-77.
[6] M.Tsujii. *Fat solenoidal attractors.* Nonlinearity 14 (2001) 1011-1027

CMat, Facultad de Ciencias, Universidad de la República. Montevideo, Uruguay.
*Email address*: leva@cmat.edu.uy

# PONENCIAS

PONENCIA

# COMPLEJIDAD EN ANÁLISIS NUMÉRICO Y EL PROBLEMA DE VALORES PROPIOS

DIEGO ARMENTANO

Publicaciones Matemáticas del Uruguay
Volumen 17, Julio 2019, Páginas 109–118
S 0797-1443

# COMPLEJIDAD EN ANÁLISIS NUMÉRICO Y EL PROBLEMA DE VALORES PROPIOS

DIEGO ARMENTANO

Abstract. La complejidad de un algoritmo es el número de pasos requeridos para pasar de una entrada a una salida. El conocimiento de este número nos permite comparar distintos métodos y poder determinar cuáles son más eficientes. Sin embargo el estudio de la complejidad es un tema muy complicado del cuál se sabe muy poco, aún en problemas básicos como encontrar raíces de polinomios o valores propios de matrices. En esta artículo daremos un paseo por distintos problemas y métodos, discutiendo en cada caso qué se sabe (o no se sabe) sobre la complejidad. (La exposición del artículo sigue la charla dictada por del autor para el 6to Coloquio Uruguayo de Matemática.)

## 1. Introducción

El *análisis numérico* se ocupa del diseño y estudio de *algoritmos* diseñados para resolver problemas clásicos de análisis. Algunos ejemplos clásicos de estos problemas son: interpolación de funciones, integración y diferenciación numérica, aproximación de soluciones de sistemas de ecuaciones y de ecuaciones diferenciales, optimización, etc.

Por algoritmo entendemos una secuencia de instrucciones que son ejecutadas a partir de una entrada (*input*), la cual eventualmente producirá una salida (*output*).

Existen dos clases muy diferenciadas de algoritmos en análisis numérico. Los más sencillos son los algoritmos denominados *métodos directos* que funcionan en una cantidad acotada de pasos. Ejemplos de este tipo son: la eliminación gaussiana, o cualquier problema en el que exista una solución con fórmula explícita (e implementable).

Por el otro lado están los algoritmos *indirectos* que consisten en un sistema dinámico que aproxima a la solución buscada, pero que requiere (aún en un modelo sin errores) infinitos pasos para dar con la solución exacta. Para estos casos es necesario dar una condición de parada a nuestro algoritmo, como por ejemplo que el output esté suficientemente cerca de la solución exacta (ver Sección 3.1). Un ejemplo fundamental, como consecuencia de los trabajos de Abel y Galois, es el de encontrar raíces de polinomios con grado mayor a 4. Los métodos indirectos son los más complicados e interesantes de estudiar como veremos a continuación.

## 2. Condicionamiento

La primer dificultad para el análisis de un algoritmo son los *errores de redondeo*. Las computadoras actuales trabajan con el sistema de aritmética de *punto flotante*, y entonces todas las operaciones aritméticas son realizadas en un subconjunto finito $\mathscr{E}$ de los números reales $\mathbb{R}$. Unas de las características de este sistema aritmético es

©2019 PMU

la existencia de una función $r : \mathbb{R} \to \mathscr{E}$ llamada *función de redondeo* que aproxima el dato ingresado (en algún sentido le da el valor del elemento más próximo de $\mathscr{E}$). De esta manera todas las operaciones aritméticas en $\mathbb{R}$ pasan a operaciones aritméticas en $\mathscr{E}$ a través de la función $r$.

En general, como se comentó en el punto anterior, nuestros datos y operaciones aritmética están sujetos a pequeños errores de redondeos. Por lo tanto, dado un algoritmo que intente resolver nuestro problema, el resultado será una solución de un problema perturbado. Existe una cantidad denominada *número de condición* que cuantifica qué tan sensible es nuestro problema a pequeñas perturbaciones en la entrada. (Es importante destacar que el condicionamiento es una característica intrínseca del problema que estamos mirando, y no depende del algoritmo en cuestión.)

A continuación veremos un ejemplo que ilustra las dificultades mencionadas.

*Ejemplo* 2.1. Supongamos que queremos resolver el sistema de ecuaciones lineales

$$Ax = b, \tag{2.1}$$

donde $A \in \mathbb{R}^{m\times m}$ y $b \in \mathbb{R}^m$ son las entradas. Al ingresar estas entradas a la computadora obtenemos que $A$ y $b$ son sustituidos por

$$\widehat{A} := A + \Delta A, \quad \widehat{b} := b + \Delta b,$$

donde $\Delta A$ y $\Delta b$ son errores producidos por la aritmética de la computadora. (Para simplificar la exposición supongamos que no hay error de aproximación en $b$, i.e. $\Delta b = 0$.)

Luego, aún en el caso de estar en un modelo ideal de algoritmo que funcione sin errores, tenemos que nuestro output tiene la forma

$$\widehat{x} := x + \Delta x := (\widehat{A})^{-1} b,$$

siendo $x := A^{-1}b$ la verdadera solución a nuestro problema.

El número de condición, asociado al input $A$, controla el tamaño del vector $\Delta x$. Se prueba que

$$\frac{\|\Delta x\|}{\|x\|} \leq \kappa(A) \frac{\|\Delta A\|}{\|A\|},$$

siendo $\kappa(A) = \|A\| \cdot \|A^{-1}\|$ el número de condición asociado a la entrada $A$, y $\|\cdot\|$ la norma euclideana. (Ver por ejemplo Blum et al. [9, Capítulo 11].)

En particular esto sugiere que un número de condición grande conspira contra la precisión de nuestro output. (Observar que en este ejemplo no hemos especificado ningún algoritmo.)

El número de condición tiene asociado una interpretación geométrica importante que viene dada por el *Teorema de número de condición*. Si sobre el espacio de matrices $\mathbb{R}^{n\times n}$ consideramos la norma de *Frobenius*, $\|A\|_F := \sqrt{\sum_{i,j} a_{ij}^2}$, entonces se tiene el siguiente resultado.

*Teorema* 1 (Eckart-Young). *Sea $A \in \mathbb{R}^{n\times n}$ invertible. Entonces*

$$\kappa(A) = \frac{\|A\|}{\operatorname{dist}_F(A, \Sigma)}$$

*siendo $\Sigma \subset \mathbb{R}^{n\times n}$ el conjunto de matrices de determinante cero, y $d_F$ la distancia inducida por la norma de Frobenius. (Ver* [9, página 202]*.)*

El conjunto $\Sigma$ es exactamente el conjunto de entradas problemáticas dado que el sistema no tiene solución. (En el contexto del problema de resolver sistemas lineales, $\Sigma$ es el conjunto de *entradas mal condicionadas*.)

El teorema anterior dice que el número de condición es inversamente proporcional a la distancia de la matriz al conjunto de matrices no invertibles.

*Observación* 2.2. El teorema de número de condición vale en muchos contextos, por ejemplo ver Shub-Smale [24] para el caso de ceros de polinomios, Armentano [1] para el problema de valores propios, Cucker et al. [15] para el problema de conteo de raíces reales. Hasta ahora no hay una teoría general sobre el número de condición en problemas computacionales. En cada caso particular se estudia si hay un resultado similar.

## 3. Complejidad algorítmica

3.1. **Ceros aproximados.** Una pregunta importante es saber cuándo nuestro algoritmo debe parar, y para esto es necesario establecer un criterio.

En el siguiente párrafo describiremos dos criterios diferentes.

- *$\varepsilon$-aproximación:* Fijado un $\varepsilon > 0$ de antemano, una *$\varepsilon$-aproximación* es un output que está a una distancia menor o igual a $\varepsilon$ de una solución verdadera.
- *Aproximación à la Smale:* Un output es un *cero aproximado* en el sentido de Smale si una versión apropiada del método de Newton comenzando en él converge cuadraticamente, de manera inmediata, a una solución verdadera.

Observar que ser un cero aproximado es más fuerte que ser una $\varepsilon$-aproximación. Esto resulta fácilmente del hecho de que podemos convertir un cero aproximado en una $\varepsilon$-aproximación simplemente iterando una cantidad prefijada (que depende de $\varepsilon$) de veces el método de Newton.

En el otro sentido, no es verdad que toda $\varepsilon$-aproximación puede ser convertida en un cero aproximado. Por ejemplo, con el $\varepsilon$ prefijado, es posible que existan más de una solución en el entorno de tamaño $\varepsilon$. O peor aún, la solución puede tener "multiplicidad" y en tal caso el método de Newton converge de manera lineal (ver [9]).

3.2. **Complejidad media.** Una buena medida de la eficiencia de un algoritmo es su *complejidad*. *Grosso modo* la complejidad de un algoritmo, diseñado para resolver un problema de análisis numérico, es el *tiempo* requerido por el mismo para pasar de la entrada a la salida. Por tiempo se entiende la cantidad de operaciones aritméticas básicas (sumas, multiplicaciones, inversiones, etc.) realizadas por el algoritmo.

La complejidad de un algoritmo depende de la entrada que estemos trabajando. Por ejemplo, es claro de la discusión anterior sobre el número de condición que una entrada mal condicionada lleve más tiempo que una entrada bien condicionada. En este sentido la complejidad de un algoritmo depende fuertemente de la entrada que estemos mirando.[1]

Para evitar la dependencia en la entrada, Smale [29] propone una manera sistemática de analizar la complejidad de un algoritmo considerando un análisis probabilístico del mismo. Si consideramos una medida de probabilidad $\nu$ sobre el

[1]Además también depende de la solución asociada a nuestra salida. Esto se puede ver claramente en el caso de encontrar raíces de un polinomio donde hay diferentes soluciones y algunas pueden presentar más dificultades que otras para localizarlas.

espacio de entradas $\mathfrak{I}$, ¿cuál es el valor esperado de la complejidad? Este número es lo que se define por *complejidad media*. Conocer la complejidad media nos permite, no sólo comparar algoritmos, sino poder decidir si el problema computacional tiene *complejidad media polinomial*. Con esto nos referimos a problemas tales que existe un algoritmo diseñado para resolverlo con complejidad $K$, donde su promedio sobre todas las entradas satisface

$$\int_{a\in\mathfrak{I}} K(a)\, d\nu(a) \leq C(\dim \mathfrak{I})^q,$$

donde $\dim \mathfrak{I}$ es la dimensión espacio del espacio de entradas, y $q$ un natural.

3.3. **Complejidad del teorema de Bézout.** En las últimas décadas, motivado por lo trabajos pioneros Shub-Smale [24, 25, 26, 27], se han realizados muchos trabajos con el objetivo de diseñar algoritmos –para el problema de resolver sistemas de ecuaciones– con complejidad media polinomial. El problema 17 de la lista de problemas para el siglo XXI de Steve Smale dice lo siguiente:

> *"Can a zero of n complex polynomial equations in n unknowns be found approximately, on the average, in polynomial time with a uniform algorithm?"* S. Smale [31].

Una respuesta afirmativa a esta pregunta se obtuvo en 2017 por P. Lairez [19], mediante una derandomización del algoritmo randomizado con complejidad media polinomial propuesto por Beltrán-Pardo [7]. (Ver también Beltrán-Pardo[8] y Bürgisser-Cucker [10].)

## 4. Complejidad del problemas de valores propios

En esta sección daremos un breve paseo sobre la complejidad del problema de valores propios.

El problema de valores propios es el siguiente: dada una matriz $A$, de tamaño $m \times m$, encontrar un escalar $\lambda$ y un vector no nulo $v$ tales que

$$(\lambda \mathsf{Id}_m - A)v = 0. \tag{4.1}$$

Nos restringiremos a trabajar sobre el cuerpo de los reales o complejos. Además, observar que el sistema (4.1) es homogéneo en $v$ y por lo tanto podemos trabajar sobre el espacio proyectivo asociado.

Este es un problema común para los científicos que día a día "resuelven" este problema mediante una computadora. Sin embargo, el problema de estudiar su complejidad estuvo abierto por más de 30 años, convirtiéndolo en un de los grandes desafíos para la comunidad de álgebra lineal numérica.

> *"So the problem of devising an algorithm [for the eigenvalue problem] that is numerically stable and globally (and quickly!) convergent remains open."* J. Demmel [14, page 139]

En lo que sigue comentaremos de manera sucinta qué se sabe sobre la complejidad de los algoritmos más conocidos para el problema de valores propios.

4.1. **Polinomio característico.** Es probable que el primer algoritmo que se nos ocurra para atacar el problema de encontrar valores propios de una matriz $A$, es calcular las raíces del polinomio característico $\chi_A$, para luego encontrar (aproximar) sus raíces. Sin embargo este método parece no ser el apropiado debido a que es

*numéricamente inestable* como veremos a continuación. El polinomio de Wilkinson (de aspecto inofensivo)

$$W(x) := \prod_{i=1}^{20}(x-i) = x^{20} + w_{19}x^{19} + \cdots + w_1 x + w_0$$

se utiliza como ejemplo de este fenómeno. Supongamos que $A$ es la matriz diagonal con entradas $1, 2, \ldots, 20$ (y por lo tanto $\chi_A(x) = W(x)$). Luego un error del orden $2^{-23}$ en el coeficiente $w_{19} = -210$ produce, aún si los demás coeficientes son calculados de manera exacta, una variación enorme en los ceros de $\chi_D$. Por ejemplo, el cero 18 y el cero 19 se unen para crear una raíz doble en 18.62, la cual se convertirá en dos raíces complejas conjugadas si el error es aún mayor. (En otras palabras, esto muestra que el polinomio $W(x)$ está mal condicionado en la raíces 18 y 19.)

Sin embargo, se puede probar que la matriz $A$ está bien condicionada para el problema de valores propios (cf. Sección 2), i.e., una pequeña perturbación en las entradas de la matriz $A$ provocan una pequeña alteración en los valores propios. (Ver por ejemplo Armentano [1, Sección 3]).

*Observación* 4.1. El fenómeno como el que ocurre con el polinomio de Wilkinson motivó a los analistas numéricos a descartar este procedimiento para el problema de valores propios. Más aún, la comunidad no sólo descarta este método para atacar el problema sino que por el contrario, para encontrar raíces de polinomios se transforma el problema a encontrar valores propios de una matriz con las raíces del polinomio, conocida en la literatura como la *matriz compañera*. (Ver por ejemplo Trefethen-Bau [32].) (En Bürgisser-Cucker-Rocha [12] se da una posible explicación del porqué de este fenómeno.)

4.2. **Método de la potencia.** Dada una matriz $A$, el *método de la potencia* se define por la iteración

$$x_k := \frac{Ax_{k-1}}{\|Ax_{k-1}\|}, \quad (k = 1, 2, \ldots). \tag{4.2}$$

donde $x_0$ es cierto vector unitario inicial. (Observar que el método está definido siempre que $x_k$ no esté en el núcleo de $A$).

El método de la potencia es un algoritmo sencillo para encontrar el vector propio dominante. Su simplicidad, desde el punto de vista de sistemas dinámicos, lo hace muy atractivo. Sin embargo este método tiene un lado negativo: no es eficiente desde el punto de vista del análisis numérico, y además sólo encuentra un vector propio.

Si existe un valor propio dominante, entonces es un ejercicio verificar que para casi todo vector inicial la sucesión 4.2 converge a una dirección dominante.

Kostlan [17] obtuvo los primeros resultados finos sobre la complejidad de este método para encontrar $\varepsilon$-aproximaciones del vector propio dominante. En particular se tiene el siguiente resultado.

**Teorema 2** ([17])**.** *Sea $A$ una matriz aleatoria $m \times m$ simétrica con entradas independientes gaussianas estándar fuera de la diagonal, y gaussianas con varianza 2 en la diagonal. Entonces la complejidad media del método de la potencia es infinito.*

La randomización elegida resulta de la distribución *gaussian orthogonal ensamble* (GOE) en matrices simétricas $m \times m$, siendo esta la única medida de probabilidad invariante bajo la acción por conjugación del grupo ortogonal.

Si uno excluye un entorno tubular de las matrices mal condicionadas para este problema, o refina el método de la potencia, puede obtener resultados de complejidad media finita para aproximar el vector propio dominante. Ver Kostlan [17] y [18].

4.3. **Iteración de cociente Rayleigh.** Un método más eficiente es la *iteración cociente de Rayleigh* que consiste en: dado un vector inicial unitario $x_0$, se genera una sucesión $\{x_k\}$ dada por

$$\lambda := x_{k-1}^T A x_{k-1}, \quad (A - \lambda \mathsf{Id})y = x_{k-1}, \quad x_k := \frac{y}{\|y\|}, \qquad (k = 1, 2, \ldots). \tag{4.3}$$

La idea detrás de este método está en usar el método de la potencia para la matriz $(A - \lambda \mathsf{Id})^{-1}$ cuando $\lambda$ está cerca de ser un valor propio, para luego actualizarlo.

Es conocido que este método es convergente (y rápido) para el caso de matrices simétricas. (Ver por ejemplo Ostrowski [20], Parlett-Kahan [21], y Batterson-Smillie [5]

Sin embargo, no existen resultados relacionados a la complejidad media del mismo.

En el caso de matrices no simétricas, Batterson-Smillie [6] probaron que la situación es muy distinta dado que no existe convergencia global.

**Teorema 3** ([6])**.** *Si $m \geq 3$, entonces hay un abierto de matrices para las cuales cada una tiene una abierto de condiciones iniciales donde el método iterativo del cociente de Rayleigh no converge.*

Es claro que resulta entonces que la complejidad media es $\infty$.

4.4. **Algoritmo QR.** El algoritmo QR, desarrollado por J. Francis en la década del 50, es el método más conocido para encontrar valores propios de una matriz. Este método se basa en la descomposición QR en la cual toda matriz $A$ se puede factorizar de la forma $A = QR$, donde $Q$ es ortogonal y $R$ triangular superior (análogo en el caso complejo).

El algoritmo QR genera una sucesión $\{A_k\}$ que se construye de la siguiente manera: con matrices iniciales de $A_0 := A = Q_0 R_0$, definimos

$$A_k := R_{k-1} Q_{k-1} = Q_k R_k, \qquad (k = 1, 2, \ldots).$$

Es un ejercicio verificar que las matrices $A_k$ son ortogonalmente conjugadas a $A$, y por lo tanto sus espectros coinciden.

Resulta que en condiciones genéricas $A_k$ tiende a una matriz triangular superior, y por lo tanto en la diagonal están sus valores propios. (Ver Dedieu [13].)

Se puede probar que el primer vector de la matriz $A_k$ coincide (a menos de un escalar) con la sucesión (4.2) generada por el método de la potencia para la matriz $A$. Por lo tanto, en el caso simétrico, resulta del Teorema 2 que la complejidad media del algoritmo QR para el caso GOE es infinito.

Por esta razón se utilizan variantes del método QR, denominados algoritmos QR con shift, los cuales potencian la velocidad trasladando la matriz $A$ por múltiplos de la identidad (cf. (4.3).)

Actualmente el método más utilizado es al algoritmo QR de Francis con shift (dobles) (ver Batterson [4]). Sin embargo no hay teoremas que aseguren su convergencia en casos no estructurados, y menos aún un estudio de la complejidad media del mismo.

4.5. **Flujos isoespectrales.** Otro tipo de algoritmos son los asociados a flujos isoespectrales. Sobre el espacio de matrices simétricas se considera un flujo $\phi_t$ que satisface las siguientes propiedades:

- (isoespectral) espectro de $\phi_t(A)$ es constante, $t \geq 0$;
- (convergencia) el flujo $\phi_t(A)$ converge a una matriz diagonal cuando $t \to +\infty$.

Un ejemplo de esto es el flujo generado por la ecuación diferencial de corchete doble de Brockett

$$\dot{A}(t) = [A(t), [A(t), D]], \quad A(0) = A, \quad D = \operatorname{diag}(1, 2, \ldots, m).$$

(Ver [16, Capitulo 2].)

Discretizaciones de estos algoritmos inducen algoritmos para resolver el problema de valores propios. Sin embargo, no existen resultados de complejidad media para estos métodos.

Sería interesante tener resultados que vincularan la longitud de las lineas de flujos con la complejidad de estos algoritmos.

4.6. **Métodos de homotopía.** En esta sección desarrollaremos algoritmos de continuación, o llamados también *métodos de homotopía*, orientado para el problema de valores propios de matrices complejas.

(Nos concentraremos en el problema de valores propios, pero se destaca que *mutatis mutandis* se puede extender a otros problemas numéricos.)

Brevemente, los métodos de homotopía pueden describirse de la siguiente manera. El sistema de ecuaciones $(\lambda I_m - A)v = 0$, con $v \in \mathbb{P}(\mathbb{C}^m)$, siendo $\mathbb{P}$ el espacio proyectivo asociado, es el punto final de un camino de problemas

$$(\lambda(t)I_n - A(t))v(t) = 0, \ v(t) \neq 0, \ 0 \leq t \leq 1, \tag{4.4}$$

con $(A(1), \lambda(1), v(1)) = (A, \lambda, v)$.

Comenzando desde una terna $(A(0), \lambda(0), v(0))$ "continuamos" este camino hasta el sistema objetivo $(\lambda I_n - A)v = 0$. La manera algorítmica de hacerlo es construir un número finito de ternas

$$(A_k, \lambda_k, v_k), \quad 0 \leq k \leq K, \tag{4.5}$$

con $A_k := A(t_k)$, y $0 = t_0 < t_1 < \cdots < t_K = 1$, y donde $(\lambda_k, v_k)$ es la aproximación del valor y vector propio $(\lambda(t_k), v(t_k))$ de $A_k$.

Las ternas dadas en (4.5) se construyen utilizando un método de *predicción-corrección* de la siguiente manera: dado la partición $0 = t_0 < t_1 < \cdots < t_K = 1$, y el par $(\lambda_0, v_0) \in \mathbb{C} \times \mathbb{P}(\mathbb{C}^m)$, se define

$$(\lambda_k, v_k) := N_{A_k}(\lambda_{k-1}, v_{k-1}), \quad 1 \leq k \leq K,$$

donde $N_A$ es el operador de Newton $N_A$ definido en $\mathbb{C} \times \mathbb{P}(\mathbb{C}^m)$ (ver Armentano [1]).

La partición $0 = t_0 < t_1 < \cdots < t_K = 1$ se construye de tal manera que el par $(\lambda_k, v_k)$ es siempre un cero aproximado de $A_k$.

La complejidad de este algoritmo es el número $K$ de pasos suficientes para validar esta aproximación.

Un resultado importante para los métodos de homotopía es vincular la complejidad $K$ con alguna propiedad analítico geométrica del problema.

**Teorema 4** ([1]). *Para $t \in [0,1]$, sea $\Gamma(t) = (A(t), \lambda(t), v(t))$ satisfaciendo (4.4), y $\|A(t)\| = 1$. Entonces, la complejidad $K$ satisface*

$$K \leq \int_0^1 \left\|\dot{\Gamma}(t)\right\| \mu\left(\Gamma(t)\right) \, dt,$$

*siendo $\left\|\dot{\Gamma}(t)\right\|$ la velocidad de la curva en $\mathbb{C}^{m\times m} \times \mathbb{C} \times \mathbb{P}(\mathbb{C}^m)$, y $\mu$ el número de condición que está dado por*

$$\mu(A, \lambda, v) = \|A\|_F \cdot \left\|\left(\Pi_{v^\perp}(\lambda \mathsf{Id} - A)|_{v^\perp}\right)^{-1}\right\|,$$

*donde denotamos por $\Pi_{v^\perp}$ es la proyección ortogonal sobre el ortogonal a $v$.*

Este resultado fue motivado por Shub [23], que es el resultado análogo para el caso de la complejidad del problema de Bézout.

El Teorema 4 motiva dos preguntas importantes.

- ¿Cómo elegir la terna inicial $(A_0, \lambda_0, v_0)$ satisfaciendo (4.1)?;
- ¿Cómo conocer *a priori* caminos $\{A(t)\}$ tales que el levantado $\Gamma(t)$ satisface $\mu(\Gamma(t))$ es pequeño?

La primer pregunta tiene fácil respuesta, dado que hay muchas formas distintas de tomar ternas iniciales. Aunque es importante destacar que cada elección influenciará la complejidad, y por lo tanto es razonable preguntarse si hay una terna óptima. (Ver Problema 7 de la lista de Smale [31].)

Para responder la segunda pregunta vale la pena mencionar que en este caso también hay un teorema de número de condición (cf. Observación 2.2). Por lo tanto basta considerar curvas que se mantengan alejadas de los problemas mal condicionados, i.e., matrices con valores propios múltiples (ver [1]). Dado que no se conoce bien la geometría del problema, tener una respuesta a la segunda pregunta es de suma dificultad. En general se opta por tomar segmentos de recta, o realizar caminos que preserven el condicionamiento. Esto último es viable dado que en general el número de condición tiene ciertos invariantes conocidos *a priori*, como por ejemplo en este caso, el número de condición es invariante por conjugaciones por matrices unitarias.

El primer resultado sobre complejidad para este problema, para el caso de matrices hermitianas, se encuentra en Armentano-Cucker [3] donde los autores dan un algoritmo randomizado con complejidad media polinomial.

El primer resultado satisfactorio sobre la complejidad del problemas de valores propios es reciente, y es el siguiente.

**Teorema 5** ([2]). *Hay algoritmos de homotopía, tales que si el input es una matriz $A$ compleja $m \times m$, con entradas gaussianas estándar independientes, entonces el output es un cero aproximado de la matriz $A$ con complejidad media polinomial en $m$.*

Las cotas de la complejidad media para encontrar un valor y vector propio de matrices complejas gaussianas son de orden $O(m^7)$. Este resultado es una cota superior, pero es posible que pueda ser mejorada.

En el artículo [2] se detallan cuáles son los puntos de partida y las homotopías utilizadas.

*Observación* 4.2. Es posible que los métodos de homotopía no sean tan rápidos como sus competidores para este problema. Lo más destacable de este resultado es que da una respuesta positiva a un problema abierto desde hace décadas en álgebra lineal numérica.

## References


[1] D. Armentano. Complexity of path-following methods for the eigenvalue problem. *Found. Comput. Math.*, 14(2):185–236, 2014.

[2] D. Armentano, C. Beltrán, P. Bürgisser, F. Cucker, M. Shub A stable, polynomial-time algorithm for the eigenpair problem. aceptado para publicación en *JEMS*.

[3] D. Armentano, F. Cucker. A randomized homotopy for the Hermitian eigenpair problem. *Found. Comput. Math.* 15 (2015), no. 1, 281–312.

[4] S. Batterson. Convergence of the Francis shifted QR algorithm on normal matrices. *Linear Algebra Appl.* 207 (1994), 181–195.

[5] S. Batterson and J. Smillie. The dynamics of Rayleigh quotient iteration. SIAM J. Numer. Anal. 26, 624-636, 1989.

[6] S. Batterson and J. Smillie. Rayleigh quotient iteration for nonsymmetric matrices. *Math.Comp.*, 55(191):169–178, 1990.

[7] C. Beltrán and L.M. Pardo. Smale's 17th problem: average polynomial time to compute affine and projective solutions. *J. Amer. Math. Soc.*, 22(2):363–385, 2009.

[8] C. Beltrán and L.M. Pardo. Fast linear homotopy to find approximate zeros of polynomial systems. *Found. Comput. Math.*, 11(1):95–129, 2011.

[9] L. Blum, F. Cucker, M. Shub, and S. Smale, *Complexity and Real Computation*, Springer-Verlag, New York, 1998.

[10] P. Bürgisser and F. Cucker. On a problem posed by Steve Smale. *Annals of Mathematics*, 174:1785–1836, 2011.

[11] P. Bürgisser and F. Cucker. *Condition*, volume 349 of *Grundlehren der mathematischen Wissenschaften*. Springer-Verlag, Berlin, 2013.

[12] P. Bürgisser, F. Cucker, E. Rocha. On the condition of the zeros of characteristic polynomials. *J. Complexity* 42 (2017), 72–84.

[13] JP. Dedieu. Points fixes, zéros et la méthode de Newton. *Mathématiques & Applications [Mathematics & Applications]*, 54. Springer, Berlin, 2006.

[14] J.W. Demmel. *Applied Numerical Linear Algebra*. SIAM, 1997.

[15] F. Cucker, T. Krick, G. Malajovich, M. Wschebor. A numerical algorithm for zero counting II: Distance to Ill-posedness and Smoothed Analysis, *J. fixed point theory appl.*, 6 (2009)

[16] U. Helmke, J. Moore. Optimization and dynamical systems. (With a foreword by R. Brockett.) *Communications and Control Engineering Series*, Springer-Verlag London, London, 1994. xiv+391 pp.

[17] E. Kostlan. *Complexity theory of numerical linear algebra* J. of Computational and Applied Mathematics 22, 219–230, (1988)

[18] E. Kostlan. *Statistical complexity of dominant eigenvector calculation.* Journal of Complexity 7, 371-379, (1991).

[19] P. Lairez. A deterministic algorithm to compute approximate roots of polynomial systems in polynomial average time. *Found. Comput. Math.* 17 (2017), no. 5, 1265–1292.

[20] A. Ostrowski. On the convergence of Rayleigh quotient iteration for the computa- tion of the characteristic roots and vectors, I. Arch. Rational Mech. Anal. 1, 233-241, (1958)

[21] B. Parlett, W. Kahan. On the convergence of a practical QR algorithm. *Inform. Process. Lett.* 68, 114-118, (1969).

[22] C.W. Pfrang, P. Deift, and G. Menon. *How long does it take to compute the eigenvalues of a random symmetric matrix.* arXiv:1203.4635.

[23] M. Shub. Complexity of Bézout's Theorem VI: geodesics in the condition (number) metric. *Found. Comput. Math.*, 9(2):171–178, 2009.

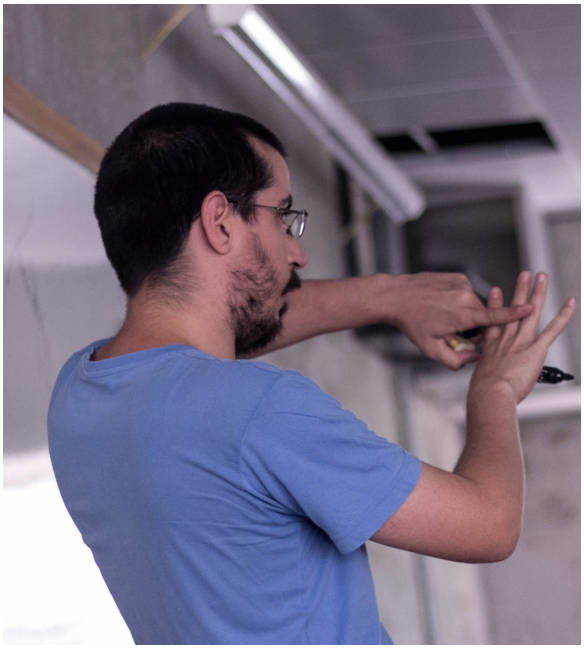
[24] M. Shub and S. Smale. Complexity of Bézout's Theorem I: geometric aspects. *Journal of the Amer. Math. Soc.*, 6:459–501, 1993.

[25] M. Shub and S. Smale. Complexity of Bézout's Theorem II: volumes and probabilities. In F. Eyssette and A. Galligo, editors, *Computational Algebraic Geometry*, volume 109 of *Progress in Mathematics*, pages 267–285. Birkhäuser, 1993.

[26] M. Shub and S. Smale. Complexity of Bézout's Theorem III: condition number and packing. *Journal of Complexity*, 9:4–14, 1993.
[27] M. Shub and S. Smale. Complexity of Bézout's Theorem IV: probability of success; extensions. *SIAM J. of Numer. Anal.*, 33:128–148, 1996.
[28] M. Shub and S. Smale. Complexity of Bézout's Theorem V: polynomial time. *Theoretical Computer Science*, 133:141–164, 1994.
[29] S. Smale. The fundamental theorem of algebra and complexity theory, *Bull., New Ser., Am. Math. Soc.*, 4(1), 1–36 (1981).
[30] S. Smale. Complexity theory and numerical analysis. In A. Iserles, editor, *Acta Numerica*, pages 523–551. Cambridge University Press, 1997.
[31] S. Smale. Mathematical problems for the next century. *Mathematical Intelligencer*, 20:7–15, 1998.
[32] LN. Trefethen, D. Bau. Numerical linear algebra. *Society for Industrial and Applied Mathematics (SIAM)*, Philadelphia, PA, 1997.

CMat, Facultad de Ciencias, Universidad de la República. Montevideo, Uruguay.
*Email address*: diego@cmat.edu.uy

PONENCIA



# DENDRITACIONES DE SUPERFICIES

ALFONSO ARTIGUE

Publicaciones Matemáticas del Uruguay
Volumen 17, Julio 2019, Páginas 121–123
S 0797-1443

# DENDRITACIONES DE SUPERFICIES

ALFONSO ARTIGUE

Resumen. Presentaremos una generalización del concepto de foliación para espacios métricos. En particular estudiamos dendritaciones de superficies, que se definen como atlas maximales de descomposiciones locales semicontinuas en dendritas. Esta estructura tiene aplicaciones en el modelado de conjuntos estable e inestables de sistemas dinámicos topológicos con alguna forma de expansividad.

El propósito de este breve artículo es presentar una *teoría de foliaciones* desde un punto de vista de la *teoría de continuos*. Los detalles se pueden encontrar en [2]. Esta estructura, que en su forma más general es llamada *cw-foliación*, contendrá como casos particulares a las foliaciones de variedades, laminaciones, foliaciones singulares tipo pseudo Anosov y las foliaciones definidas por Hiraide [7] para estudiar a los expansivos con sombreado. La clave de nuestro enfoque es considerar descomposiciones semicontinuas como cartas locales. No se asume que las placas se distribuyan como una estuctura producto, como en las foliaciones de variedades. Incluso, dos placas en una misma carta local pueden no ser homeomorfas.

Las cw-foliaciones forman un marco conceptual para comprender la distribución de los continuos estables e inestables en los sistemas dinámicos que consideraremos. Una motivación para este trabajo es la clasificación de los homeomorfismos cw-expansivos en superficies. Diremos que un homeomorfismo $f$ es *cw-expansivo* si existe $\delta > 0$ tal que si $\operatorname{diam}(f^n(A)) < \delta$ para todo $n \in \mathbb{Z}$ y $A$ es conexo entonces $A$ contiene sólo un punto. Ejemplos importantes de dinámicas cw-expansivas son los difeomorfismos de Anosov de variedades compactas de cualquier dimensión y los pseudo Anosov de superficies compactas.

Para los difeomorfismos de Anosov los conjutos estables locales forman cartas foliadas, al menos de clase $C^0$, ver [9]. Si consideramos homeomorfismos de superficies, via [8, 11], sabemos que los conjuntos estables locales forman foliaciones con puntos singulares. Decimos que un homeomorfismo $f$ es *expansivo* si existe $\delta > 0$ tal que si $\operatorname{dist}(f^n(x), f^n(y)) < \delta$ para todo $n \in \mathbb{Z}$ entonces $y = x$. Hiraide y Lewowicz de forma independiente demostraron que los expansivos de superficies compactas son conjugados a difeomorfismos pseudo Anosov. Si consideramos la cw-expansividad en superficies tenemos que los conjuntos estables locales están lejos de formar foliaciones, incluso singulares. Por ejemplo, pueden no ser una unión finita de arcos, en [1] se construye un homeomorfismo cw-expansivo de una superficie compacta con un punto fijo cuyo conjuto estable local es conexo pero no localmente conexo. Estos ejemplos sugieren que el objetivo de clasificar todos los homeomorfismos cw-expansivos de superficies requiere nuevos métodos.

©2019 PMU

En [2] introducimos algunas definiciones que se ubican entre la expansividad y la cw-expansividad, que serán llamadas *cwN-expansividad*. En cierto modo, es una versión de la $N$-expansividad desde un punto de vista de la teoría de continuos. Dado un número natural $N \geq 1$, decimos que $f$ es *$N$-expansivo* [14] si existe $\delta > 0$ tal que si $A \subset X$ (un subconjunto arbitrario) y $\operatorname{diam}(f^n(A)) \leq \delta$ para todo $n \in \mathbb{Z}$ entonces $A$ tiene a lo sumo $N$ puntos. Decimos que $f$ es *cwN-expansivo* si existe $\delta > 0$ tal que si $A, B \subset X$ son continuos, $\operatorname{diam}(f^n(A)) < \delta$ para todo $n \geq 0$ y $\operatorname{diam}(f^n(B)) < \delta$ para todo $n \leq 0$ entonces $A \cap B$ tiene a lo sumo $N$ puntos. Además, un homeomorfismo es *$cw_F$-expansivo* (donde $F$ significa *finito*) si existe $\delta > 0$ tal que si $A, B \subset X$ son continuos, $\operatorname{diam}(f^n(A)) < \delta$ para todo $n \geq 0$ y $\operatorname{diam}(f^n(B)) < \delta$ para todo $n \leq 0$ entonces $A \cap B$ es finito. Estas formas de expansividad nos permiten demostrar la conexión local de los conjuntos estables locales y concluir que son dendritas, ver [2, Theorem 6.7.1]. Recordemos que una dendrita es un continuo de Peano que no contiene curvas cerradas simples, un continuo es un espacio métrico compacto y conexo, y un continuo es de Peano si es localmente conexo. Es sabido que tales dendritas poseen un tamaño uniforme. De esta manera llegamos naturalmente al concepto de descomposición dendrítica.

El tema de las descomposiciones topológicas tiene una literatura diversa, ver por ejemplo [5, 6, 15, 17, 19]. También existen aplicaciones en sistemas dinámicos [1, 3, 4, 10, 16]. Un resultado notable de Moore [12] afirma que el cociente de una descomposición semicontinua de la esfera de dimensión 2 en continuos que no separan es nuevamente la esfera (asumiendo que la descomposición tiene al menos dos elementos, en caso contrario el cociente daría un punto). En cierto sentido, las decomposiciones de Moore generalizan a las foliaciones de dimensión 0, es decir, la descomposición en puntos. Una carta foliada en el sentido usual del cuadrado $[0, 1] \times [0, 1]$ por lineas horizontales tiene dos propiedades: cada placa es un arco y el cociente es un arco. En una carta de una dendritación sucede que: cada placa es una dendrita y a su vez el cociente es una dendrita.

El concepto de cwN-expansividad aparece naturalmente del estudio de los artículos [8, 11]. Muchos de sus argumentos únicamente utilizan la cw1 expansividad. Este concepto fue esencialmente considerado previemente en [18]. En [2, Theorem 6.8.5] demostramos que todo homeomorfismo cw1-expansivo de una superficie compacta es expansivo. En espacios no localmente conexos dicho resultado no es cierto, ver [2, Example 5.3.2]. En [2, §2.2.1] se muestra que existen homeomorfismos cw2-expansivos de la esfera de dimensión 2 que no son 2-expansivos.

Para culminar esta breve presentación mencionaremos algunos resultados generales sobre dendritaciones de superficies. En [2, Theorem 6.2.7] se demuestra que para toda dendritación de una superficie cerrada existe un subconjunto residual de puntos sin ramificaciones. En particular, una hoja genérica es una variedad inmersa de dimensión 1. Este resultado es consecuencia de otro resultado de Moore [13], que garantiza que a lo sumo se puede encajar una cantidad numerable de trípodes disjuntos en el plano. En [2, Theorem 6.7.1] consideramos homeomorfismos $\mathrm{cw_F}$-expansivos y demostramos que los conjuntos estable e inestables forman dendritaciones. Además, ninguna hoja es un continuo de Peano, las hojas genéricas son variedades no compactas de dimensión 1 y existe un conjunto denso de puntos de la superficie en los cuales los arcos estables e inestables se cortan transversalmente. En cierto modo, estos resultados ofrecen una imagen global cualitativa de los homeomorfismos $\mathrm{cw_F}$-expansivos de superficies.

## Referencias


[1] A. Artigue, *Anomalous cw-expansive surface homeomorphisms*, Discrete and continuous dynamical systems **36** (2016), 3511–3518.

[2] ———, *Dendritations of surfaces*, Ergodic Theory and Dynamical Systems, posted on 2017, DOI 10.1017/etds.2017.14.

[3] M. Barge and B.F. Martensen, *Classification of expansive attractors on surfaces*, Ergodic Theory and Dynamical Systems **31** (2011), 1619–1639.

[4] A. de Carvalho and M. Paternain, *Monotone quotients of surface diffeomorphisms*, Math. Research Letters **10** (2003), 603–619.

[5] J.J. Charatonik and W.J. Charatonik, *Connected subsets of dendrites and separators of the plane*, Topology and its Applications **36** (1990), 233–245.

[6] R.J. Daverman, *Decompositions of manifolds*, Pure and applied mathematics, Academic Press. Inc, 1986.

[7] K. Hiraide, *Expansive homeomorphisms with the pseudo-orbit tracing property of n-tori*, J. Math. Soc. Japan **41** (1989), 357–389.

[8] ———, *Expansive homeomorphisms of compact surfaces are pseudo-Anosov*, Osaka J. Math. **27** (1990), 117–162.

[9] M. Hirsch, C. Pugh, and M. Shub, *Invariant manifolds*, Lecture notes in mathematics, vol. 583, Springer-Verlag, 1977.

[10] K. Kawamura, H.M. Tuncali, and E.D. Tymchatyn, *Expansive homeomorphisms on Peano curves*, Houston Journal of Mathemtaics **21** (1995), 573–583.

[11] J. Lewowicz, *Expansive homeomorphisms of surfaces*, Bol. Soc. Bras. Mat. **20** (1989), 113-133.

[12] R.L. Moore, *Concerning upper-semicontinuous collections of continua*, Transactions of the AMS **4** (1925), 416–428.

[13] ———, *Concerning Triods in the Plane and the Junction Points of Plane Continua*, Proc. Natl. Acad. Sci. USA **14** (1928), 85–88.

[14] C.A. Morales, *A generalization of expansivity*, Disc. and Cont. Dyn. Sys. **32** (2012), 293–301.

[15] S. Nadler Jr., *Continuum Theory*, Pure and Applied Mathematics, vol. 158, Marcel Dekker, New York, 1992.

[16] A. Passeggi and J. Xavier, *A classification of minimal sets for surface homeomorphisms*, Mathematische Zeitschrift **278** (2014), 1153–1177.

[17] J.H. Roberts, *There does not exist an upper semi-continuous decomposition of E into arcs*, Duke Math. J. **2** (1936), 10–19.

[18] M. Sambarino, *Estructura local de conjuntos estables e inestables de homeomorfismos en superficies*, Universidad de la República, Uruguay, Monograph, 1993.

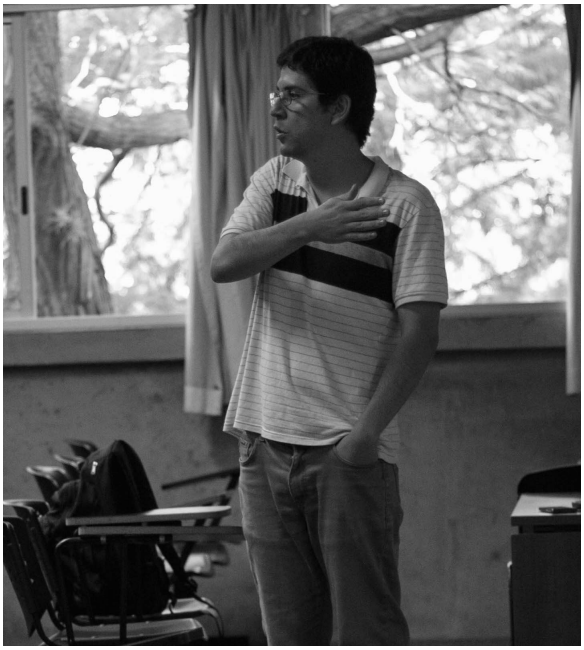
[19] M. Smith, *A theorem on continuous decompositions of the plane into nonseparating continua*, Proceedings of the AMS **55** (1976), 221–222.



Departamento de Matemática y Estadística del Litoral, Universidad de la República. Gral. Rivera 1350, Salto, Uruguay.

*Email address*: artigue@unorte.edu.uy

PONENCIA



# CEROS DE POLINOMIOS Y FUNCIONES ALEATORIAS

FEDERICO DALMAO

Publicaciones Matemáticas del Uruguay
Volumen 17, Julio 2019, Páginas 127–138
S 0797-1443

# CEROS DE POLINOMIOS Y FUNCIONES ALEATORIAS

FEDERICO DALMAO

Resumen. Hacemos un breve repaso de los resultados recientes sobre el número de ceros de polinomios y ondas aleatorias y algunos problemas afines.

## 1. Introducción

El objetivo de este artículo es presentar algunos resultados, sobre todo los que han salido en el último lustro, vinculados a los ceros de polinomios aleatorios y, en menor medida, a las ondas aleatorias. No se incluyen resultados originales.

El estudio de los ceros de diferentes clases de funciones es sempiterno y ubicuo por lo que es difícil decir algo que realmente no sobre para enfatizar su importancia. A pesar de esto, digamos que algunos de los principales teoremas del Cálculo, como los de Bolzano y Rolle, son teoremas de existencia de raíces. El mismo Teorema Fundamental del Álgebra y su versión multivariada, el Teorema de Bézout, y el problema de hallar los valores propios de un operador lineal tratan sobre la existencia de raíces.

En estos días hay una gran actividad entorno a los ceros de funciones aleatorias que involucra a geómetras, algebristas, numeristas, probabilistas, etc. Se estudia el número de ceros (si estos son aislados) o la medida del conjunto de nivel cero (si el dominio tiene mayor dimensión que el codominio) y también se estudian otras características topológicas como el número de componentes conexas, etc.

¿Por qué pensar en funciones aleatorias? Porque es una forma de buscar resultados genéricos, que no valgan para una función concreta sino para una función típica. Una forma habitual de precisar lo típico es justamente la de elegir la función al azar y observar el comportamiento esperado. En otras palabras, el comportamiento para una función concreta puede ser muy llamativo pero quizás sea completamente atípico, después de aleatorizar no deberían observarse tales funciones.

Si la función se elige al azar, sus ceros se vuelven un conjunto aleatorio. En general es muy difícil calcular la distribución de probabilidad de las diferentes características del conjunto de ceros. Por esto usualmente se estudian características parciales como el promedio, la varianza o se estudian los límites cuando algún parámetro de interés tiende a infinito.

Uno puede poner al número de ceros en un marco clásico de la siguiente manera, si se toma una partición del dominio en subconjuntos disjuntos, el número de ceros global es la suma del número de ceros sobre cada subconjunto de la partición, así el número de ceros se puede escribir como una suma. Ya en poder de una suma de variables aleatorias uno se puede preguntar por ejemplo por un teorema central del límite. Sin embargo, las variables en una suma de este tipo son en general dependientes.

©2019 PMU

En este trabajo nos preocupamos acerca de dos tipos de funciones, los polinomios y las ondas aleatorias y estudiamos principalmente sus ceros reales. Ambas clases de funciones se pueden escribir como una combinación (posiblemente infinita) de ciertas funciones básicas, es decir, son de la forma

$$f(t) = \sum_k a_k \phi_k(t).$$

En nuestros casos las funciones básicas $\phi_k$ son

$$t^k, \quad \cos(kt), \quad \operatorname{sen}(kt), \quad J_k(t); \qquad k = 0, 1, \ldots.$$

Aquí $J_k$ es la función de Bessel del primer tipo y orden $k$.

La aleatorización se hace a través de los coeficientes $a_k$ de la combinación, es decir, se toma a estos coeficientes como variables aleatorias. A veces, conviene pensar una función de la forma anterior como un proceso estocástico. Es decir, como una función de dos variables $(\omega, t)$ donde la primera representa el azar y la segunda el tiempo. En particular, al fijar un instante $t$ obtenemos una variable aleatoria y al fijar una realización $\omega$, o sea al fijar los coeficientes, obtenemos una trayectoria.

Como siempre, resulta que tomar coeficientes con distribución normal facilita enormemente el trabajo y habilita al uso de una variedad de herramientas. Recién en segunda instancia se intenta extender cada resultado al mundo no gaussiano.

## 2. Polinomios Aleatorios

Los polinomios son las funciones más sencillas que uno se puede imaginar, en particular, para calcular su valor en un punto dado sólo son necesarias operaciones enteras (sumas y productos). Por otro lado, según el Teorema de Weierstraß, con ellos se puede aproximar tan finamente como se quiera cualquier función continua en un intervalo. Por estas y por otras razones, es natural que sean las primeras funciones a estudiar a cualquier respecto.

**2.1. Polinomios algebraicos.** Los primeros trabajos rigurosos sobre las raíces reales de polinomios aleatorios aparecieron en los años 30 y 40 a manos de Bloch y Pólya, Kac, Littlewood y Offord entre otros, ver el libro de Bharucha-Reid y Sambandham [14] el panorama general hasta los 80. Ellos consideraron los polinomios de Kac

$$P_d(t) = a_0 + a_1 t + a_2 t^2 + \cdots + a_d t^d,$$

siendo $a_0, \ldots, a_d$ variables aleatorias independientes con la misma distribución. Littlewood y Offord [28, 29, 30] en los 40 para coeficientes normales estándar, uniformes en $[0, 1]$ o uniformes en $\{-1, 1\}$, probaron que el número de ceros reales $N_d$ de $P_d$ es *polilogarítmico*; es decir que la probabilidad de

$$\alpha \frac{\log(d)}{\log(\log(d))} \leq N_d \leq \beta \log^2(d),$$

tiende a 1 cuando el grado $d$ tiende a infinito siendo $\alpha$ y $\beta$ constantes. En [14] se pueden encontrar generalizaciones de este resultado.

En el caso de coeficientes normales estándar, Kac [23] dio una fórmula integral para la esperanza de $N_d$ de la forma

$$\mathbb{E}\,(N_d) = \int_{-\infty}^{\infty} \rho_d(t) dt,$$

siendo

$$\rho_d(t) = \frac{\sqrt{A(t)C(t) - B^2(t)}}{A(t)}; A(t) = \sum_{k=0}^{d} t^{2k}, B(t) = \sum_{k=1}^{d} kt^{2k-1}, C(t) = \sum_{k=1}^{d} k^2 t^{2k-2}.$$

Además del artículo original [23] es muy interesante leer el de Edelman y Kostlan [19] en el que se da una versión puramente geométrica cambiando el cálculo de probabilidades por el de áreas en la esfera.

Esta fórmula va más allá del número de raíces, muestra que en el eje real hay una *densidad de ceros*, $\rho_d$, en el sentido de cuántas raíces por unidad de longitud hay a lo largo de la recta real. Esta densidad muestra un comportamiento muy llamativo de los ceros, tanto en el caso de los reales como en el de los complejos, cuando el grado $d$ tiende a infinito los ceros tienden a concentrarse sobre los puntos de módulo 1. Esta acumulación de los ceros no se da en otras familias de polinomios[1].

De la fórmula dada por Kac se deduce, en particular, que

$$\mathbb{E}(N_d) = \frac{2}{\pi}\log(d) + o(\log(d)).$$

Esto es bastante sorprendente, de las $d$ raíces complejas del polinomio $P_d$ menos de $\log(d)$ son reales en promedio. O sea, cualquier polinomio aleatorio de grado $d = 1000$ tiene mil raíces complejas y de ellas sólo 4 son reales en promedio. Si el grado escala a un millón, en promedio habrá sólo 9 raíces reales.

Unos treinta años después, Maslova [34, 35] dio la asintótica para la varianza

$$\mathbf{var}(N_d) \underset{d\to\infty}{\sim} \frac{4}{\pi}\left(1 - \frac{2}{\pi}\right) \cdot \log(d)$$

y probó un Teorema Central del Límite para $N_d$. Esto es,

$$\frac{N_d - \mathbb{E}(N_d)}{\sqrt{\mathbf{var}(N_d)}}$$

converge en distribución a una normal estándar cuando el grado $d \to \infty$. Las hipótesis del resultado de Maslova son la independencia e igualdad en distribución de los coeficientes; que los coeficientes sean centrados[2] y que tengan algún momento de orden superior a dos (que exista $\delta > 0$ tal que $\mathbb{E}(|a_1|^{2+\delta}) < \infty$).

Estos resultados nos dan el primer ejemplo de comportamiento universal, es decir, con tal de que valga este resultado no importa cual es la distribución de los coeficientes sino únicamente que sean centrados y que cumplan la condición de momento. En realidad, la universalidad llega al primer orden de esta aproximación, Nguyen, Nguyen y Vu [36] probaron que el segundo término depende de la distribución. Esto es, existe una constante $C_{dist}$ que depende de la distribución de los coeficientes tal que

$$\mathbb{E}(N_d) = \frac{2}{\pi}\log(d) + C_{dist} + o(1).$$

El método utilizado por Maslova se basa en la presencia de cierta repulsión entre los ceros de los polinomios que permite aproximar el número de ceros del polinomio en un intervalo por el número de cambios de signos de las trayectorias en una grilla suficientemente fina.

[1] En realidad, según [15], se verifica para polinomios palíndromos en el sentido de que verifiquen $\mathbf{var}(a_{d-k}) = \mathbf{var}(a_k)$ para todo $k$.

[2] En el caso de coeficientes $a_k$ no centrados, probó que hay en promedio la mitad de ceros.

Se han estudiado otras formas de aleatorizar los coeficientes, obteniendo resultados principalmente en lo que concierne al número medio de raíces. Los resultados de varianza y de distribución límite son más duros. Entre los diferentes tipos de polinomios aleatorios, vale la pena destacar los llamados polinomios de Kostlan[3], en los cuales se supone que los coeficientes $a_0, \dots, a_d$ son variables aleatorias independientes, normales, centradas y tales que la varianza de $a_k$ es igual al coeficiente binomial $\binom{d}{k}$. En este caso, se sabe desde los 90 [41] que el número medio de raíces reales es $\sqrt{d}$.

En 2015 se probó que la varianza asintótica del número de raíces es del mismo orden que la esperanza y que el número de ceros reales verifica un Teorema Central del Límite [16]. Un resultado análogo acaba de aparecer (agosto de 2017) para los polinomios de Weyl en los cuales la varianza de $a_k$ es $1/\sqrt{k!}$, ver Do y Vu [18].

Los polinomios de Kostlan tienen una versión multivariada, esto es lo que los hace realmente especiales, ver Armentano [3]. Consideramos un sistema cuadrado $\mathbf{P} = (P_1, \dots, P_m) = 0$ de $m$ ecuaciones polinomiales en $m$ variables

$$P_\ell(t) = \sum_{|\mathbf{j}| \le d_\ell} a_{\mathbf{j}}^{(\ell)} t^{\mathbf{j}},$$

donde hemos usado multí índices

1. $\mathbf{j} = (j_1, \dots, j_m) \in \mathbb{N}^m$, con $|\mathbf{j}| = \sum_{k=1}^m j_k$;
2. $a_{\mathbf{b}j}^{(\ell)} = a_{j_1 \dots j_m}^{(\ell)} \in \mathbb{R}$, $\ell = 1, \dots, m$, $|\mathbf{j}| \le d$;
3. $t = (t_1, \dots, t_m)$, con $t^{\mathbf{j}} = \prod_{k=1}^m t_k^{j_k}$.

Los coeficientes $a_{\mathbf{j}}^{(\ell)}$ son normales independientes con media cero y varianzas

$$\mathbf{var}\left(a_{\mathbf{j}}^{(\ell)}\right) = \binom{d_\ell}{\mathbf{j}} = \frac{d_\ell!}{j_1! \dots j_m! (d_\ell - |\mathbf{j}|)!}.$$

La fórmula multinomial implica que la covarianza entre $P_\ell(s)$ y $P_\ell(t)$ es

$$\mathbf{cov}\left(P_\ell(s), P_\ell(t)\right) = (1 + \langle s, t \rangle)^{d_\ell},$$

siendo $\langle \cdot, \cdot \rangle$ el producto interno usual en $\mathbb{R}^m$. Esto implica que la covarianza no cambia si aplicamos una isometría, hecho que es clave para que los cálculos sean manejables. La esperanza del número de raíces de $\mathbf{P}$ es la raiz del número de Bézout $\sqrt{d_1 \dots d_m}$, ver [41].

En el caso en que los polinomios tienen el mismo grado $d$, cuando $d \to \infty$, la varianza es del mismo orden que la esperanza, de forma más precisa en [5] y [27] se probó que

$$\lim_{d \to \infty} \frac{\mathbf{var}\left(N_d\right)}{d^{m/2}} = c \approx 0,55...$$

También se ha estudiado el caso rectangular, ver [26, 27].

En realidad, en el caso de sistemas de ecuaciones polinomiales aleatorias surge una nueva variante de estudio pues se puede considerar que el tamaño del sistema es el que crece a infinito en lugar del grado. Azaïs y Wschebor [9] y Wschebor [43] obtuvieron la varianza límite bajo este régimen.

[3]También se los conoce como polinomios elípticos, etc. Esta es la elección más natural para la distribución de los coeficientes según el propio Kostlan

**2.2. Polinomios trigonométricos.** Los polinomios trigonométricos son funciones de la forma

$$P_d(t) = \sum_{k=1}^{d} a_k \cos(kt), \quad t \in [0, 2\pi];$$

o, de forma simplificada

$$Q_d(t) = \sum_{k=1}^{d} a_k \cos(kt) + b_k \operatorname{sen}(kt), t \in [0, 2\pi].$$

Decimos que la versión $Q_d$ es más simple pues tiene mejores propiedades probabilísticas. De hecho, $Q_d$ es un proceso estocástico estacionario en sentido amplio en general y en el caso de coeficientes normales es estacionario de forma estricta. Esto simplifica en gran medida los cálculos, en particular la varianza de $Q_d(t)$ y la covarianza entre $Q_d(t)$ y $Q_d(t+s)$ no dependen de $t$.

Los polinomios trigonométricos son especialmente importantes en física nuclear donde los valores propios de una matriz aleatoria de gran tamaño están estrechamente vinculados con la energía del sistema, ver los detalles en [15]. Esto conduce rápidamente a preocuparse por las raíces de polinomios aleatorios (los polinomios característicos de tales matrices). El tamaño de la matriz se traduce en el grado del polinomio. Por las simetrías subyacentes al problema físico es de esperar que una gran parte de las raíces de estos polinomios se concentren alrededor del círculo unitario en el plano complejo, de ahí la idea de restringir los polinomios a puntos de la forma

$$e^{ik\theta} \equiv (\cos(kt), \operatorname{sen}(kt)).$$

Por ende, es coherente mantener el mote de polinomios para estas funciones y el de grado para $d$. Los primeros estudios sobre el número de ceros de los polinomios trigonométricos clásicos $P_d$ son de los 50 y en el caso de $Q_d$ de los 60, ver [14] y sus referencias. El número medio de raíces reales de $P_d$ y $Q_d$ en el caso de coeficientes normales estándar independientes, ver [14, 44], es

$$\mathbb{E}\,(N_d) \underset{d\to\infty}{\sim} \frac{2}{\sqrt{3}} d.$$

En este caso el promedio es del mismo orden que la cota determinista $2d$ y muy superior al promedio en el caso de los polinomios algebraicos. Recientemente, entre 2015 y 2016 se ha mostrado la universalidad de este resultado (al primer orden), Flasche [20], Angst y Poly [1], ver también Angst, Dalmao y Poly [2] y la sección siguiente.

También recientemente, en el caso de coeficientes normales estándar, se ha obtenido la varianza límite, ver Granville y Wigman [21], Su y Shao [42]. Cabe destacar que la varianza es del mismo orden que la esperanza. Además, Granville y Wigman [21], ver también Azaïs y León [8], y Azaïs, Dalmao y León [6] obtuvieron los Teoremas Centrales para $Q_d$ y $P_d$ respectivamente.

La universalidad de la distribución límite se conoce sobre intervalos de longitud $1/d$, ver [22] y [7]. Sin embargo, en el caso de la varianza asintótica, en noviembre de 2017 Bally, Caramelino y Poly [11] llegaron al muy llamativo resultado de la no universalidad. De hecho, si los coeficientes de $Q_d$ son independientes centrados y con varianza uno, la varianza límite del número de ceros se escribe como el límite del caso normal más un término de corrección que depende de la curtosis de los

coeficientes, es decir, de la diferencia entre el cuarto momento de los coeficientes y el de una normal estándar.

**2.3. Coeficientes dependientes.** Incluir dependencia entre los coeficientes es algo harto más díficil y en el caso trigonométrico sólo hay resultados sobre la esperanza del número de ceros.

Los primeros trabajos datan de los 80 y abarcan el caso de coeficientes igualmente distribuídos normales con covarianza constante $\rho(k) = \mathbb{E}(a_0 a_k) = \rho$ y geométrica $\rho(k) = \mathbb{E}(a_0 a_k) = \rho^k$, ver Sambandham [40], Sambandham y Renganhatan [39]. Los cálculos se basan en cotas finas para las series involucradas y son bastante específicos.

En 2017, Angst, Dalmao y Poly [2] lograron calcular el límite de la esperanza del número de ceros en el caso gaussiano pero con condiciones bastante generales sobre la dependencia de los coeficientes expresadas en términos de la densidad espectral de los coefcientes. El caso más notable incluído es cuando los coeficientes $a_k$ son representados por incrementos sucesivos del *movimiento Browniano fraccionario*. Esto implica que se admite una *dependencia fuerte*, es decir, tal que $\sum |\rho(k)| = \infty$.

Lo sorprendente es que en todos estos casos se llega al mismo límite que en el caso de coeficientes independientes.

## 3. Ondas aleatorias

El estudio de las ondas, u olas, aleatorias proviene de la física y en especial de los trabajos de Longuet-Higgins [31] y Berry [13]. Aquí llamamos onda a una solución, a valores reales o complejos, de la ecuación de Helmholtz

$$\Delta f + kf = 0,$$

siendo $\Delta$ el operador de Laplace-Beltrami, simplemente Laplaciano de aquí en más. En otras palabras, las ondas son las funciones propias del Laplaciano, los valores propios $k$ son interpretados como los niveles de energía. El interés recae sobre el comportamiento del conjunto de los ceros de estas funciones para valores propios $k$ altos. Los ceros son interpretados como zonas de silencio o de obscuridad, etc. según qué represente la onda.

Dependiendo del dominio (en general una variedad) sobre el que se definan el Laplaciano $\Delta$ y las ondas $f$ se presentan diferentes situaciones, aquí nos limitamos a dimensión dos. En el toro plano $\mathbb{T}_2$ o en la esfera $\mathbb{S}^2$ el espectro de $\Delta$ es puramente discreto y los valores propios tienen multiplicidad mayor a uno y se conocen de forma explícita bases de los correspondientes subespacios propios. Lo usual, es tomar una combinación con coeficientes aleatorios de estas bases y obtener así una onda aleatoria. Si el dominio es una variedad en la cual los valores propios tienen multiplicidad uno, lo usual es combinar funciones propias de diferentes valores con coeficientes aleatorios.

Para ser más precisos, en el caso del toro $\mathbb{T}_2$ los valores propios son de la forma $k = 4\pi^2 n$ siendo $n = a^2 + b^2$ para ciertos $a, b \in \mathbb{Z}$. Para $n$ de esta forma, las funciones

$$e_{a,b}(x) = e^{2\pi i \langle (a,b), x \rangle}; \quad a^2 + b^2 = n$$

forman una base ortonormal del subespacio propio correspondiente al valor $4\pi^2 n$. En este caso, tomando coeficientes complejos $c_{a,b}$ independientes normales estándar[4]

[4]Esto es, $c_{a,b} = c_{a,b}^{(1)} + i c_{a,b}^{(2)}$ siendo $c_{a,b}^{(1)}, c_{a,b}^{(2)}$ variables normales estándar reales independientes.

las ondas aleatorias resultantes

$$T_n(x) = \sum_{a,b\in\mathbb{Z}:\ a^2+b^2=n} c_{a,b}e_{a,b}(x); \quad x \in \mathbb{T}_2,$$

se conocen como *ondas aleatorias aritméticas*. Si se quieren ondas reales se puede tomar parte la real de $T_n$ o imponer que $c_{-a,-b} = \overline{c_{a,b}}$. En el caso de $\mathbb{S}^2$ los valores propios son de la forma $k = n(n+1)$ para $n \in \mathbb{N}$ y las bases de los subespacios propios están formadas por los llamados armónicos esféricos. Luego se procede de la misma manera. Se obtienen así los *armónicos esféricos aleatorios*.

Debido a la forma de los valores propios surgen sutilezas e intervienen propiedades aritméticas, ver [24] por ejemplo. En particular, las dimensiones $N_n$ de los correspondientes subespacios propios fluctuan y como consecuencia no hay una única forma de interpretar el límite cuando el nivel de energía tiende a infinito.

Para las ondas aritméticas en el toro $\mathbb{T}_2$ se conocen los límites de la esperanza y la varianza de la longitud de las curvas de nivel cero [32], o del número de ceros comunes a dos copias independientes de ellas [17] cuando $N_n \to \infty$. Lo interesante es que surgen límites no universales, dependiendo de la sucesión $N_{n_j} \to \infty$ elegida se obtienen diferentes varianzas límites e incluso se obtienen distribuciones límites distintas[5], ver [17, 32] y las referencias allí presentes.

En el caso de los armónicos esféricos aleatorios en $\mathbb{S}^2$, a diferencia del anterior, en [33] han obtenido teoremas centrales del límite para la longitud de la curva de nivel cero.

Berry [13] conjeturó que cuando el valor propio $k$ (equivalentemente $n$ en los modelos anteriores) es grande, las funciones propias del Laplaciano se pueden modelar por una onda aleatoria, que llamó *plana*, que se puede describir en el caso de dimensión dos como un proceso Gaussiano $b_k$ centrado con covarianza dada por

$$\mathbf{cov}\,(b_k(x), b_k(y)) = J_0(\|x-y\|),$$

siendo $J_0$ la función de Bessel del primer tipo y de orden 0. En el plano, expresando $x$ en coordenadas polares $r, \theta$ se puede escribir

$$b_k(x) = \sum_{\ell\in\mathbb{Z}} c_\ell J_{|\ell|}(kr)e^{i\ell\theta},$$

siendo $J_\ell$ la función de Bessel del primer tipo y de orden $\ell$. En [37] se obtiene un Teorema Central del Límite para la longitud de la curva de nivel cero en el caso real y para el número de ceros en el caso complejo sobre regiones del plano cuando éstas crecen hasta cubrir todo el plano.

Las ondas aritméticas aleatorias y los armónicos esféricos aleatorios son coherentes con la conjetura de Berry pues sus covarianzas convergen a la planteada por él.

Aquí nos hemos centrado en el número de ceros (o en la longitud de la curva de nivel cero); pero es justo mencionar que hay una literatura enorme que trata sobre otros aspectos geométricos de los ceros de las ondas aleatorias. Por ejemplo, un aspecto importante es que los ceros de ondas (y polinomios) aleatorias como procesos puntuales son sorprendentemente regulares y sus puntos parecen estar sobre un reticulado, ver [12] donde se compara el conjunto de puntos críticos de una onda aleatoria plana con los puntos generados por un proceso de Poisson entre

[5]Se habla de teoremas límites no centrales cuando la distribución límite no es normal.

otros. Ver también [4] en el caso de polinomios. Por otro lado, se estudia el número de componentes conexas del conjunto de ceros, o el número de componentes conexas en que éste divide al dominio, etc, ver [25] y sus referencias por ejemplo.

## 4. Contando los ceros de una función

En esta última sección queremos echar un vistazo a la sala de máquinas. Es decir, ya no hablaremos sobre los resultados obtenidos acerca de los ceros de polinomios y ondas aleatorias sino que nos preocuparemos sobre cómo se obtienen tales resultados. Para mantenernos en un contexto simple nos restringimos al número de ceros de una función real de variable real. En [10] se encuentra la teoría general.

El punto de partida es la *fórmula de conteo* de Mark Kac que data de la década de 1940. Sea $f : [a, b] \to \mathbb{R}$ una función derivable y con derivada continua en todo punto. Para el *número de ceros* de $f$ en el intervalo $[a, b]$ escribimos $N_f[a, b]$, esto es

$$N_f[a, b] = \#\{t \in [a, b] : f(t) = 0\}.$$

De manera rápida e informal la fórmula de Kac se puede escribir como:

$$N_f[a, b] = \int_a^b \delta_0(f(t))dt,$$

siendo $\delta_0$ la *Delta de Dirac* en 0. La Delta de Dirac en 0 es una "función" que se anula en todos los puntos salvo en cero donde vale infinito de modo que

$$\int \delta_0(x)dt = 1.$$

Para formalizar esta idea hay que recurrir a una *aproximación a la unidad*. Esto es, a una sucesión de funciones $(\psi_n)_{n\in\mathbb{N}}$ no negativas, tales que $\int \psi_n = 1$ para todo $n$ y $\int_{|t|>\varepsilon} \psi_n(t)dt \underset{n\to\infty}{\to} 0$ para todo $\varepsilon > 0$. Se puede probar que si $g$ es continua se cumple

$$\lim_{n\to\infty} \int \psi_n(t)g(t)dt = g(0).$$

Dado un conjunto $A$ denotamos por $\mathbf{I}_A$ su función característica, esto es

$$\mathbf{I}_A(x) = \begin{cases} 1, & \text{si } x \in A; \\ 0, & \text{si } x \notin A. \end{cases}$$

Entonces, eligiendo $\psi_\delta(x) = \frac{1}{2\delta}\mathbf{I}_{\{|x|<\delta\}}$, la fórmula de Kac resulta

$$N_f[a, b] = \lim_{\delta\to 0} \frac{1}{2\delta} \int_a^b |f'(t)|\mathbf{I}_{\{|f(t)|<\delta\}}dt.$$

En realidad hay que tener un poco de cuidado si $f$ tiene puntos de tangencia a nivel cero o si toma este valor en los extremos del intervalo $[a, b]$. Afortunadamente, en nuestro contexto, después de aleatorizar podemos ignorar estos inconvenientes pues ocurren con probabilidad cero.

Es interesante remarcar que si bien aparece un límite en la fórmula, dicho límite se alcanza. Es decir, si $\eta$ es suficientemente chico, entonces

$$N_f[a, b] = \frac{1}{2\eta} \int_a^b |f'(t)|\mathbf{I}_{\{|f(t)|<\eta\}}dt.$$

Pero $\eta$ depende de $f$ y si $f$ es aleatoria, como en los casos que nos interesan, también lo es $\eta$.

Hay dos caminos complementarios para estudiar la distribución del número de ceros de $f$ a partir de la fórmula de conteo de Kac.

**4.1. La fórmula de Rice.** El primero de los caminos consiste en tomar esperanza de ambos lados de la fórmula de Kac. Con ello se obtiene una fórmula integral para el número medio de ceros de $f$ en el intervalo $[a, b]$. Sin preocuparnos de los tecnicismos el camino es

$$\begin{aligned}
\mathbb{E}\left(N_f[a,b]\right) &= \mathbb{E}\left[\lim_{\delta\to 0}\frac{1}{2\delta}\int_a^b |f'(t)|\mathbf{I}_{\{|f(t)|<\delta\}}dt\right] \\
&= \lim_{\delta\to 0}\frac{1}{2\delta}\int_a^b \mathbb{E}\left[|f'(t)|\mathbf{I}_{\{|f(t)|<\delta\}}\right]dt \\
&= \lim_{\delta\to 0}\frac{1}{2\delta}\int_a^b\int_{-\delta}^{\delta} \mathbb{E}\left[|f'(t)| \mid f(t)=x\right]p_{f(t)}(x)dxdt \\
&= \int_a^b \lim_{\delta\to 0}\frac{1}{2\delta}\int_{-\delta}^{\delta} \mathbb{E}\left[|f'(t)| \mid f(t)=x\right]p_{f(t)}(x)dxdt \\
&= \int_a^b \mathbb{E}\left[|f'(t)| \mid f(t)=0\right]p_{f(t)}(0)dt.
\end{aligned}$$

Aquí, el primer factor en el integrando es la esperanza condicional de la variable aleatoria $|f'(t)|$ dado el evento $\{f(t)=0\}$. Lidiar con tal esperanza condicional es en general complicado. El caso gaussiano es el ideal pues la independencia es equivalente a la no correlación y gracias a eso se puede levantar la condición eligiendo $\alpha\in\mathbb{R}$ tal que $f'(t)-\alpha f(t)$ sea independiente de $f(t)$. Luego, si $\mathbb{E}\left(f(t)\right)=0$ se tiene

$$\mathbb{E}\left[|f'(t)| \mid f(t)=0\right] = \mathbb{E}\left(|f'(t)-\alpha f(t)|\right) = \sqrt{\frac{2}{\pi}}\sigma_t;$$

siendo $\sigma_t^2$ la varianza de $f'(t)-\alpha f(t)$. El cálculo directo da $\alpha = \frac{\mathbf{cov}\left(f'(t),f(t)\right)}{\mathbf{var}\left(f(t)\right)}$.

Vale la pena mencionar que la fórmula de Rice permite estudiar una cantidad global, que depende de todo el intervalo, como el número de ceros por medio de cantidades locales dependientes de un único valor de $t$.

De forma similar es posible hallar expresiones para los momentos de cualquier orden. En efecto, sean $k\in\mathbb{N}$ y $N^{[k]} = N\cdot(N-1)\cdots(N-k+1)$, entonces

$$\begin{aligned}
\mathbb{E}\,N_f^{[k]}[a,b] = \int_{[a,b]^k}\mathbb{E}\left[\prod_{j=1}^{k}|f'(t_j)| \mid f(t_1)=\cdots=f(t_k)=0\right] \\
\cdot\, p_{f(t_1),\cdots,f(t_k)}(0,\cdots,0)dt_1\cdots t_k.
\end{aligned}$$

En general, conocer todos los momentos de una variable aleatoria es equivalente a conocer su distribución (ver el problema de los momentos). Esa es la idea detrás del método de los momentos para hallar estimadores o del método homónimo para probar la convergencia en distribución. Sin embargo, en la práctica puede ser muy difícil resolver la fórmula de inversión y pasar de los momentos a la distribución.

**4.2. Caos de Wiener.** El segundo camino que abre la fórmula de conteo de Kac consiste en hacer un desarrollo en serie dentro de la integral en la fórmula de Kac. De nuevo el caso gaussiano es el ideal.

Supongamos entonces que el proceso $f$ es gaussiano, esto es, que los vectores aleatorios $(f(t_1), \cdots, f(t_k))$ tienen distribución normal para cualquier $k$ y para cualesquiera $(t_1, \ldots, t_k)$. Sea $\varphi$ la densidad normal estándar

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad x \in \mathbb{R}.$$

Los polinomios de Hermite $(H_n)_{n\in\mathbb{N}}$, definidos por

$$H_n(x) = (-1)^n e^{\frac{x^2}{2}} \frac{d^n}{dt^n} e^{-\frac{x^2}{2}}.$$

o alternativamente por

$$H_0(x) = 1, \quad H_1(x) = x \text{ y } H_n(x) = xH_{n-1}(x) + (n-1)H_{n-2}(x), \quad n \geq 2,$$

forman un sistema ortogonal completo en el espacio de las funciones de cuadrado integrable con producto interno $\langle f, g\rangle = \int_{\mathbb{R}} f(x)g(x)\varphi(x)dx$. Además $\|H_n\|_2^2 = n!$. Esto implica que las funciones $x \mapsto \psi_n(x)$ e $y \mapsto |y|$ se pueden escribir como

$$\psi_n(x) = \sum_{n=0}^{\infty} a_{2n} H_{2n}(x); \qquad |y| = \sum_{n=0}^{\infty} b_{2n} H_{2n}(x).$$

Los coeficientes $a_{2n}$, $b_{2n}$ son explícitos. La razón por la cual aparecen sólo los términos pares es la paridad de las funciones $\psi_n$ y valor absoluto.

Reemplazando cada factor del integrando de la fórmula de Kac por su desarrollo de Hermite y tomando límite en $n$ se obtiene la llamada expansión en el *Caos de Wiener* del número de ceros de $f$

$$N_f[a,b] = \sum_{n=0}^{\infty} \sum_{k=0}^{\lfloor \frac{n}{2} \rfloor} a_{2k} b_{2n-2k} \int_a^b H_{2k}(f(t)) H_{2n-2k}(f'(t))dt =: \sum_{n=0}^{\infty} I_n.$$

Estas igualdades son en el sentido de $L^2$. Las variables $I_n$, llamadas componentes caóticas de $N_f[a,b]$, son no correlacionadas.

La utilidad de este tipo de desarrollos radica en que, para cada $n$, las variables $I_n$ tienen una estructura muy particular. En realidad, $I_n$ puede escribirse como una integral estocástica múltiple de orden $n$ respecto a un movimiento Browniano estándar, ver [38].

El Teorema del Cuarto Momento de Peccati, Taqqu y Tudor [38], da condiciones para probar la convergencia en distribución a una normal de una sucesión de variables de la forma $I_n$ para un $n$ fijo (o de una suma sobre un número finito de valores de $n$) a partir de la convergencia del cuarto momento al de la normal. Esto es una gran simplificación respecto al tradicional método de los momentos[6] que dice que dada una sucesión de variables aleatorias $X_n$, la convergencia de todos los momentos $\mathbb{E}(X_n^k)$ a los respectivos momentos de una variable normal implica que $X_n$ converge en distribución a una normal. El Teorema del Cuarto Momento dice que en el caso en que las variables $X_n$ están dentro de un mismo caos, basta probar la convergencia del cuarto momento $(\mathbb{E}(X_n^4))$ al cuarto momento de una variable normal. Entonces, en lugar de probar una infinitud de convergencias basta probar una sola.

Otro punto fuerte de este desarrollo se pone de manifiesto al probar la no anulación de la varianza pues en general se puede probar a mano la positividad de la varianza de la primer componente.

[6] hay una formulación equivalente en términos de los llamados cumulantes.

Para finalizar, en general en los teoremas límites no centrales para los ceros de ondas aleatorias hay una componente que domina sobre todas las demás y se puede estudiar directamente su límite.

## Referencias


[1] J. Angst, G. Poly. Universality of the mean number of real zeros of random trigonometric polynomials under a weak Cramer condition. arXiv:1511.08750

[2] J. Angst, F. Dalmao, G. Poly. On the real zeros of random trigonometric polynomials with dependent coefficients. Proc. Amer. Math. Soc. 147 (2019), no. 1, 205-214.

[3] D. Armentano. A review of some recent results on random polynomials over $\mathbb{R}$ and over $\mathbb{C}$. Publicaciones Matemáticas del Uruguay, Proceedings of the IFUM Colloquium, Vol. 12, 1-14, 2011

[4] D. Armentano, C. Beltrán, M. Shub. Minimizing the discrete logarithmic energy on the sphere: the role of random polynomials. Trans. Amer. Math. Soc. 363 (2011), no. 6, 2955-2965.

[5] J-M. Azaïs, D. Armentano, F. Dalmao, J. León. Asymptotic variance of the number of real roots of random polynomial systems. Proc. Amer. Math. Soc. 146 (2018), no. 12, 5437-5449.

[6] J-M. Azaïs, F. Dalmao, J. León, CLT for the zeros of classical random trigonometric polynomials. Ann. Inst. Henri Poincaré Probab. Stat. 52 (2016), no. 2, 804-820.

[7] J-M. Azaïs, F. Dalmao, J.R. León, I. Nourdin, G. Poly. Local universality of the number of zeros of random trigonometric polynomials with continuous coefficients. arXiv:1512.05583

[8] J-M Azaïs, J. R. León. Clt for crossings of random trigonometric polynomials. Electron. J. Probab. 18, no. 68 (2013), 1-17.

[9] J-M Azaïs, M. Wschebor. On the roots of a random system of equations. The theorem on Shub and Smale and some extensions. Found. Comput. Math. 5 (2005), no. 2, 125-144.

[10] J-M Azaïs, M. Wschebor. Level sets and extrema of random processes and fields. John Wiley & Sons Inc., Hoboken, NJ, (2009).

[11] V. Bally, L. Caramellino, G. Poly. Non universality for the variance of the number of real roots of random trigonometric polynomials. arXiv:1711.03316

[12] D. Beliaev, V. Cammarota, I. Wigman Two point function for critical points of a random plane wave. arXiv:1704.04943

[13] M.V. Berry. Regular and irregular semiclassical wave functions. J. Phys. A: Math. Gen. 10 2083-91.

[14] A.T. Bharucha-Reid, M. Sambandham. Random polynomials. Probability and Mathematical Statistics. Academic Press, Inc., Orlando, FL, 1986. xvi+206 pp. ISBN: 0-12-095710-8.

[15] E. Bogomolny, O. Bohigas, P. Lebœuf. Quantum chaotic dynamics and random polynomials. J. Statist. Phys. 85 (1996), no. 5-6, 639-679.

[16] F. Dalmao. Asymptotic variance and CLT for the number of zeros of Kostlan Shub Smale random Polynomials. C.R. Acad. Sci. Paris Ser. I 353 (2015), 1141-1145.

[17] F. Dalmao, I. Nourdin, G. Peccati, M. Rossi. Phase Singularities in Complex Arithmetic Random Waves. arXiv:1608.05631

[18] Y. Do, V. Vu. Central limit theorems for the real zeros of Weyl polynomials. arXiv:1707.09276

[19] A. Edelman, E. Kostlan. How many zeros of a random polynomial are real? Bull. Amer. Math. Soc. (N.S.) 32 (1995), no. 1, 1-37.

[20] H. Flasche. Expected number of real roots of random trigonometric polynomials. Stochastic Process. Appl. 127 (2017), no. 12, 3928-3942.

[21] A. Granville, I. Wigman. The distribution of the zeros of random trigonometric polynomials. Amer. J. Math., 133(2), (2011), 295-357.

[22] A. Iksanov, Z. Kabluchko, A. Marynych. Local universality for real roots of random trigonometric polynomials. Electron. J. Probab. 21 (2016), Paper No. 63, 19 pp.

[23] M. Kac. On the average number of real roots of a random algebraic equation. Bull. Amer. Math. Soc., 49, (1943), 314-320.

[24] P. Kurlberg, I: Wigman. On probability measures arising from lattice points on circles. Math. Ann. 367 (2017), no. 3-4, 1057-1098.

[25] P. Kurlberg, I: Wigman. Variation of the Nazarov-Sodin constant for random plane waves and arithmetic random waves. Adv. Math. 330 (2018), 516-552.

[26] T. Letendre. Expected volume and Euler characteristic of random submanifolds. J. Funct. Anal. 270 (2016), no. 8, 3047-3110.

[27] T. Letendre, M. Puchol. Variance of the volume of random real algebraic submanifolds II. arXiv:1707.09771
[28] J. E. Littlewood, A. C. Offord. On the distribution of the zeros and a-values of a random integral function. I. J. London Math. Soc., 20, 1945, 130-136.
[29] J. E. Littlewood, A. C. Offord. On the distribution of zeros and a- values of a random integral function. II. Ann. of Math. (2), 49:885 - 952; errata 50, (1949), 990-991.
[30] J. E. Littlewood, A. C. Offord. On the number of real roots of a random algebraic equation. III. Rec. Math. [Mat. Sbornik] N.S., 12(54), (1943), 277-286.
[31] M. S. Longuet-Higgins. The statistical analysis of a random, moving surface. Philos. Trans. Roy. Soc. London Ser. A. 249 (1957), 321-387.
[32] D. Marinucci, G. Peccati, M. Rossi, I. Wigman. Non-Universality of Nodal Length Distribution for Arithmetic Random Waves Geom. Funct. Anal. 26 (2016), no. 3, 926–960.
[33] D. Marinucci, M. Rossi, I. Wigman. The Asymptotic Equivalence of the Sample Trispectrum and the Nodal Length for Random Spherical Harmonics. arXiv:1705.05747
[34] N. B. Maslova. The distribution of the number of real roots of random polynomials. Teor. Verojatnost. i Primenen., 19, (1974), 488-500.
[35] N. B. Maslova. The variance of the number of real roots of random polynomials. Teor. Verojatnost. i Primenen., 19, (1974), 36-51.
[36] H. Nguyen, O. Nguyen, V. Vu. On the number of real roots of random polynomials. Commun. Contemp. Math. 18 (2016), no. 4, 1550052, 17 pp.
[37] I. Nourdin, G. Peccati, M. Rossi. Nodal Statistics of Planar Random Waves. arXiv:1708.02281
[38] G. Peccati, M. S. Taqqu. Wiener chaos: moments, cumulants and diagrams, volume 1 of Bocconi & Springer Series. Springer, Milan, 2011. A survey with computer implementation, Supplementary material available online.
[39] N. Renganathan, M. Sambandham. On the average number of real zeros of a random trigonometric polynomial with dependent coefficients. II. Indian J. Pure Appl. Math. 15 (1984), no. 9, 951-956.
[40] M. Sambandham. On the number of real zeros of a random trigonometric polynomial. Trans. Amer. Math. Soc. 238 (1978), 57-70.
[41] M. Shub, S. Smale. Complexity of Bézout's theorem. II. Volumes and probabilities. Computational Algebraic Geometry (Nice, 1992), Progress in Mathematics, vol. 109, Birkhäuser, Boston, MA, 1993, pp. 267-285.
[42] Z. Su, Q. Shao. Asymptotics of the variance of the number of real roots of random trigonometric polynomials. Science China Mathematics. 55(11), (2012), 2347-2366.
[43] M. Wschebor. On the Kostlan-Shub-Smale model for random polynomial systems. Variance of the number of roots. J. Complexity 21 (2005), no. 6, 773-789.
[44] J.E. Wilkins, Jr. Mean number of real zeros of a random trigonometric polynomial. Proc. Amer. Math. Soc., 111(3), (1991), 851-863.

Departamento de Matemática y Estadística del Litoral, Universidad de la República. Gral. Rivera 1350, Salto, Uruguay.
*Email address*: fdalmao@unorte.edu.uy
*URL*: http://dmel.interior.edu.uy/federico-dalmao/

PONENCIA

# EQUIVARIANT ALGEBRAIC KK-THEORY

EUGENIA ELLIS

Publicaciones Matemáticas del Uruguay
Volumen 17, Julio 2019, Páginas 141–150
S 0797-1443

# EQUIVARIANT ALGEBRAIC KK-THEORY

EUGENIA ELLIS

Abstract. In this paper we present results from [6] and [7] about equivariant algebraic kk-theory.

## 1. Kasparov's KK-theory

Kasparov's KK-theory is the major tool in *noncommutative topology*, [13]. The KK-theory of separable $C^*$-algebras is a common generalization both of topological $K$-homology and tolopological $K$-theory as an *additive bivariant functor* Let $A$ and $B$ separable $C^*$-algebras then a *group* $KK(A,B)$ is defined such that

$$KK_*(\mathbb{C},B)\simeq K_*^{top}(B)\qquad KK^*(A,\mathbb{C})\simeq K^*_{hom}(A).$$

An important property of KK-theory is the so-called *Kasparov product*,

$$KK(A,B)\times KK(B,C)\to KK(A,C)$$

which is bilinear with respect to the additive group structures. Denote by $C^*$-Alg to the category of separable $C^*$-algebras. The Kasparov groups $KK(A,B)$ for $A,B\in C^*$-Alg form a morphisms sets $A\to B$ of a *category* $KK$. The composition in $KK$ is given by the Kasparov product and the category $KK$ admits a *triangulated category* structure.

There is a *canonical functor* $k: C^*$-Alg $\to KK$ that acts identically on objects and every *-homomorphism $f: A\to B$ is represented by an element $[f]\in KK(A,B)$. The functor $k: C^*$-Alg $\to KK$

- is homotopy invariant: $f_0\sim f_1$ implies $k(f_0)=k(f_1)$.
- is $C^*$-stable: any corner embedding $A\to A\otimes\mathcal{K}(\ell^2\mathbb{N})$ induces an isomorphism $k(A)=k(A\otimes\mathcal{K}(\ell^2\mathbb{N}))$.
- is split-exact: for every split-extension $I\xrightarrow{f}A\xrightarrow{g}A/I$ (i.e. there exists a *-homomorpfhism $s: A/I\to A$ such that $g\circ s=\mathrm{id}$) then $k(I)\xrightarrow{k(f)}k(A)\xrightarrow{k(g)}k(A/I)$ is part of a distinguished triangle.

The functor $k: C^*$-Alg $\to KK$ is the *universal* homotopy invariant, $C^*$-stable and split exact functor. The main authors who worked in the previous results are: J. Cuntz [4], N. Higson [12], G. Kasparov [13] and R. Meyer [15].

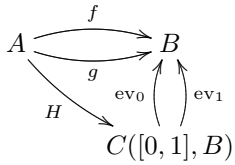
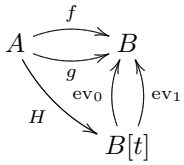
## 2. Algebraic kk-theory

Algebraic kk-theory was introduced in [3] by G. Cortiñas and A. Thom in order to show how methods from K-theory of operator algebras can be applied in completely algebraic setting. Let $\ell$ a commutative ring with unit and Alg the category of $\ell$-algebras (with or without unit). For each pair $(A,B)$ of $\ell$-algebras a group $\mathrm{kk}(A,B)$

©2019 PMU

is defined. A category $\mathfrak{KK}$ is obtained whose objects are $\ell$-algebras and where the morphisms from $A$ to $B$ are the elements of the group $\mathrm{kk}(A,B)$. The category $\mathfrak{KK}$ is triangulated and there is a canonical functor $j : \mathrm{Alg} \to \mathfrak{KK}$ with universal properties. These properties are algebraic homotopy invariance, matrix invariance and excision.

We resume the results with a dictionary between Kasparov's KK-theory and algebraic kk-theory as follows:

| **Kasparov's KK-theory** [13] | $\leftrightarrow$ | **Algebraic kk-theory** [3] |
|---|---|---|
| bivariant K-theory on $C^*$-Alg | | bivariant K-theory on Alg |
| $k : C^*\text{-Alg} \to KK$ | | $j : \mathrm{Alg} \to \mathfrak{KK}$ |
| $k$ is stable w.r.t. compact operators | | $j$ is stable w.r.t. matrices |
| $A \simeq_{KK} A \otimes \mathcal{K}(\ell^2(\mathbb{N}))$ | | $A \simeq_{\mathfrak{KK}} M_\infty(A) = \bigcup_{n\in\mathbb{N}} M_n(A)$ |
| $k$ is continuous homotopy invariant | | $j$ is polynomial homotopy invariant |
|  | |  |
| $B \simeq_{KK} C([0,1],B)$ | | $B \simeq_{\mathfrak{KK}} B[t]$ |
| $k$ is split exact | | $j$ is excisive |
| $k$ is universal for these properties | | $j$ is universal for these properties |
| $\boxed{KK_*(\mathbb{C},A) \simeq K_*^{top}(A)}$ | | $\boxed{kk_*(\ell,A) \simeq \mathrm{KH}_*(A)}$ |



KH is Weibel's homotopy K-theory defined in [22].

**Theorem 2.1.** [3] *The functor $j : \mathrm{Alg} \to \mathfrak{KK}$ is an excisive, homotopy invariant, and $M_\infty$-stable functor and it is the universal functor for these properties.*

Let $\mathcal{X}$ be a infinity set. Consider

$$M_{\mathcal{X}} := \{a : \mathcal{X} \times \mathcal{X} \to \ell : \mathrm{supp}(a) < \infty\}.$$

Let $A$ be an algebra, then $M_{\mathcal{X}}A := M_{\mathcal{X}} \otimes_\ell A$. The category $\mathfrak{R}_{\mathcal{X}}$ is constructed as it is the category $\mathfrak{KK}$ taking $M_{\mathcal{X}}$ instead of $M_\infty$.

**Theorem 2.2.** [16] *The functor $j : \mathrm{Alg} \to \mathfrak{R}_{\mathcal{X}}$ is an excisive, homotopy invariant, and $M_{\mathcal{X}}$-stable functor and it is the universal functor for these properties.*

If $\mathcal{X} = \mathbb{N}$ both theorems are the same.

## 3. Equivariant algebraic kk-theory

We introduce in [7] an algebraic bivariant K-theory for the category of algebras with an action of a group $G$ or $G$-algebras. For each pair $(A,B)$ of $G$-algebras a group $\mathrm{kk}^G(A,B)$ is defined. A category $\mathfrak{KK}^G$ is obtained whose objects are $G$-algebras and where the morphisms from $A$ to $B$ are the elements of the group $\mathrm{kk}^G(A,B)$. The category $\mathfrak{KK}^G$ is triangulated and there is a canonical functor $j : G\text{-Alg} \to \mathfrak{KK}^G$

with universal properties. These properties are algebraic homotopy invariance, equivariant matrix invariance and excision. Let $A$ be a $G$-algebra and

$$M_G := \{a : G \times G \to \ell : \mathrm{supp}(a) < \infty\}.$$

Consider in $M_G \otimes A$ the following action of $G$

$$g \cdot (e_{s,t} \otimes a) = e_{gs,gt} \otimes g \cdot a$$

A *$G$-stable* functor identifies any $G$-algebra $A$ with $M_G \otimes A$. We consider this notion as a equivariant version of the matrix invariance.

An equivariant version of the dictionary stated above is the following:

| **Equivariant Kasparov's KK-theory** | $\leftrightarrow$ | **Equivariant algebraic kk-theory** |
|---|---|---|
| [13] | | [7] |
| bivariant K-theory on $G$-$C^*$-Alg | | bivariant K-theory on $G$-Alg |
| $k : G\text{-}C^*\text{-Alg} \to KK^G$ | | $j : G\text{-Alg} \to \mathfrak{RR}^G$ |
| $k$ is stable w.r.t. compact operators | | $j$ is $G$-stable |
| $A \simeq_{KK^G} A \otimes \mathcal{K}(\ell^2(G \times \mathbb{N}))$ | | $A \simeq_{\mathfrak{RR}^G} M_\infty M_G(A)$ |
| $k$ is continuous homotopy invariant | | $j$ is polynomial homotopy invariant |
| $B \simeq_{KK^G} C([0,1], B)$ | | $B \simeq_{\mathfrak{RR}^G} B[t]$ |
| $k$ is split exact | | $j$ is excisive |
| $k$ is universal for these properties | | $j$ is universal for these properties |
| $G$ compact | | $G$ finite and $\frac{1}{\lvert G\rvert} \in \ell$ |
| $\boxed{KK^G_*(\mathbb{C}, A) \simeq K^{top}_*(A \rtimes G)}$ | | $\boxed{kk^G_*(\ell, A) \simeq KH_*(A \rtimes G)}$ |

**Theorem 3.1** ([7])**.** *The functor $j : G\text{-Alg} \to \mathfrak{RR}^G$ is an excisive, homotopy invariant, and $G$-stable functor and it is the universal functor for these properties.*

We obtain some adjoint functors at the level of bivariant kk-theory which at the level of algebras they are not. We have algebraic versions of

- Green-Julg Theorem, see [15], [11], [1], [20], [21].
- Adjointness property between $\mathrm{Ind}_H^G$ and $\mathrm{Res}_G^H$, see [15].
- Baaj-Skandalis Duality, see [2] and [20].

3.1. **Algebraic Green-Julg Theorem.** Consider the crossed product functor

$$\rtimes : G\text{-Alg} \to \mathrm{Alg} \qquad A \rtimes G = A \otimes \ell G \qquad (a \rtimes g)(b \rtimes h) = a[g \cdot b] \rtimes gh$$

and the trivial action functor $\tau : \mathrm{Alg} \to G\text{-Alg}$. The pair

$$G\text{-Alg} \underset{\tau}{\overset{\rtimes}{\rightleftarrows}} \mathrm{Alg}$$

can be extended to

$$\begin{array}{ccc} G\text{-Alg} & \underset{\tau}{\overset{\rtimes}{\rightleftarrows}} & \mathrm{Alg} \longleftarrow \text{not adjoint functors} \\ {\scriptstyle j^G}\downarrow & & \downarrow{\scriptstyle j} \\ \mathfrak{RR}^G & \underset{\tau}{\overset{\rtimes}{\rightleftarrows}} & \mathfrak{RR} \longleftarrow \text{adjoint functors if } G \text{ is finite and } \frac{1}{\lvert G\rvert} \in \ell \end{array}$$

**Theorem 3.2.** [7] *Let $G$ be a finite group of $n$ elements such that $\frac{1}{n} \in \ell$. Let $B$ be a $G$-algebra and $A$ an algebra. There is an isomorphism*

$$\psi_{GJ} : kk^G(A^\tau, B) \to kk(A, B \rtimes G)$$

We give an example to show that the adjointness between of $\tau$ and $\rtimes$ fails to hold at the algebra level. Let $G = \mathbb{Z}_2 = \{1, \sigma\}$, $A = \ell$ and $B = (\ell G)^*$ the dual algebra of $\ell G$ with the regular action. Note $\hom_{G\text{-Alg}}(A^\tau, B)$ has two elements only:

$$\varphi_i : \ell \to (\ell G)^* \qquad \varphi_0(1) = 0 \qquad \varphi_1(1) = \chi_1 + \chi_\sigma$$

One the other hand $\hom_{\text{Alg}}(A, B \rtimes G) = \hom_{\text{Alg}}(\ell, (\ell G)^* \rtimes G)$ has at least as many elements as $\ell$. For each $\lambda \in \ell$ we can define

$$\varphi_\lambda : \ell \to (\ell G)^* \rtimes G \quad \varphi_\lambda(1) = \chi_1 \rtimes 1 + \lambda(\chi_1 \rtimes \sigma) \quad \lambda \in \ell$$

Note $\varphi_\lambda$ is an algebra morphism because $\chi_1 \rtimes 1 + \lambda(\chi_1 \rtimes \sigma)$ is an idempotent element.

**Corollary 3.3.** [7] *Let $G$ be a finite group of $n$ elements such that $\frac{1}{n} \in \ell$. Then*

$$kk^G(\ell, B) \simeq \text{KH}(B \rtimes G) \qquad kk^G(\ell, \ell) \simeq \text{KH}(\ell G)$$

3.2. **Adjointness between $\text{Ind}_H^G$ and $\text{Res}_G^H$.** Let $H$ be a subgroup of $G$ and $A$ an $H$-algebra. Define

- $A^{(G)} = \{\alpha : G \to A : \alpha \text{ is a function with finite support}\}$
- $\text{Ind}_H^G(A) = \{\alpha \in A^{(G)} : \alpha(s) = h \cdot \alpha(sh) \quad \forall h \in H, s \in G\}$
- $(g \cdot \alpha)(s) = \alpha(g^{-1}s)$

The functors

$$G\text{-Alg} \underset{\text{Ind}_H^G}{\overset{\text{Res}_G^H}{\rightleftarrows}} H\text{-Alg} \longleftarrow \text{Ind}_H^G \text{ is NOT left adjoint to } \text{Res}_G^H$$

can be extended to

$$\mathfrak{KK}^G \underset{\text{Ind}_H^G}{\overset{\text{Res}_G^H}{\rightleftarrows}} \mathfrak{KK}^H \longleftarrow \text{Ind}_H^G \text{ is a left adjoint to } \text{Res}_G^H$$

**Theorem 3.4.** [7] *Let $G$ be a group, $H$ a subgroup of $G$, $B$ an $H$-algebra and $A$ a $G$-algebra. Then there is an isomorphism*

$$\psi_{IR} : kk^G(\text{Ind}_H^G(B), A) \to kk^H(B, \text{Res}_G^H(A))$$

Let us show that this adjunction is not true at the level of algebras. Consider $G = \mathbb{Z}_2 = \{1, \sigma\}$, $H = \{1\}$, $\ell = \mathbb{Q}$, $A$ be any $\mathbb{Q}$-algebra and $A^\tau$ is the algebra $A$ with the trivial action of $G$. Let us compute $\hom_{G\text{-Alg}}(\text{Ind}_H^G(\ell), A^\tau)$ and $\hom_{\text{Alg}}(\ell, A)$. We obtain

$$\text{Ind}_H^G(\ell) = \mathbb{Q}^{\mathbb{Z}_2} = \{a\chi_1 + b\chi_\sigma : a, b \in \mathbb{Q}\}$$

If $\alpha \in \hom_{G\text{-Alg}}(\mathbb{Q}^{\mathbb{Z}_2}, A^\tau)$ then $\alpha(\chi_1) = p$ with $p$ an idempotent element of $A$. We obtain

$$\alpha(\chi_\sigma) = \alpha(\sigma \cdot \chi_1) = \sigma \cdot \alpha(\chi_1) = \sigma \cdot p = p$$

and

$$0 = \alpha(\chi_1\chi_\sigma) = \alpha(\chi_1)\alpha(\chi_\sigma) = p^2 = p$$

Finally $\hom_{G\text{-Alg}}(\mathbb{Q}^{\mathbb{Z}_2}, A^\tau) = 0$. On the other hand $\hom_{\text{Alg}}(\ell, A)$ has at least as many elements as idempotents of $A$.

**Corollary 3.5.** [7]

- $kk^G(\ell^{(G/H)}, A) \simeq kk^H(\ell, \mathrm{Res}^H_G(A))$.
- *If $H$ is finite then* $kk^G(\ell^{(G/H)}, A) \simeq \mathrm{KH}(A \rtimes H)$.
- $kk^G(\ell^{(G)}, A) \simeq \mathrm{KH}(A)$.

3.3. **Algebraic Baaj-Skandalis duality.** A dual notion of $G$-algebra is the concept of $G$-graded algebra. A *$G$-graduation* on an algebra $A$ is a decomposition on submodules

$$A = \bigoplus_{s\in G} A_s \quad A_s A_t \subseteq A_{st} \quad \forall s,t \in G$$

Let $A$ be a $G$-algebra. Then

$$A \rtimes G = \bigoplus_{s\in G} A \rtimes s \qquad \text{and } (A \rtimes s)(A \rtimes t) \subset A \rtimes st$$

thus $A \rtimes G$ is a $G$-graded algebra. Let $B$ be a $G$-graded algebra. Let $G\hat{\rtimes}B$ be the algebra which as a module is $\ell^{(G)} \otimes B$ and the product is the following

$$(\chi_g \rtimes a)(\chi_h \rtimes b) := \chi_g \rtimes a_{g^{-1}h} b. \tag{3.6}$$

Here $b_g$ is the homogeneous element associated to $g$ in the decomposition

$$b = \sum_{g\in G} b_g.$$

One checks that the product (3.6) is associative and the action of $G$

$$s \cdot \chi_g \rtimes a = \chi_{sg} \rtimes a$$

makes it into a $G$-algebra.

The functors

$$G\text{-Alg} \underset{\hat{\rtimes}}{\overset{\rtimes}{\rightleftarrows}} G_{gr}\text{-Alg}$$

can be extended to

$$\begin{array}{ccc} G\text{-Alg} & \underset{\hat{\rtimes}}{\overset{\rtimes}{\rightleftarrows}} & G_{gr}\text{-Alg} \longleftarrow \text{not an equivalence} \\ {\scriptstyle j^G}\downarrow & & \downarrow{\scriptstyle \hat{j}^G} \\ \mathfrak{KK}^G & \underset{\hat{\rtimes}}{\overset{\rtimes}{\rightleftarrows}} & \hat{\mathfrak{KK}}^G \longleftarrow \text{an equivalence} \end{array}$$

## 4. Algebraic quantum kk-theory

In [6] we introduce an equivariant algebraic kk-theory for $\mathcal{G}$-modules algebras where $\mathcal{G}$ is an algebraic quantum group.

4.1. **Van Daele's algebraic quantum groups.** Let $\ell = \mathbb{C}$ and $(\mathcal{G}, \Delta, \varphi)$ be an algebraic quantum group (see [5],[18] and [19]), in other words a regular multiplier Hopf algebra with invariants. That means, $(\mathcal{G}, \Delta)$ is a *multiplier Hopf algebra*:

- $\mathcal{G}$ associative algebra over $\mathbb{C}$ with non-degenerate product.
- $M(\mathcal{G})$ multiplier algebra of $\mathcal{G}$: $(\rho_1, \rho_2) \in M(\mathcal{G})$ if
  - $\rho_i : \mathcal{G} \to \mathcal{G}$ is a linear map $(i = 1, 2)$
  - $\rho_1(hk) = \rho_1(h)k \quad \rho_2(hk) = h\rho_2(k) \quad \rho_2(h)k = h\rho_1(k) \quad \forall h, k \in \mathcal{G}$

The multiplication in $M(\mathcal{G})$ is defined as follows

$$(\rho_1, \rho_2)(\tilde{\rho}_1, \tilde{\rho}_2) = (\rho_1\tilde{\rho}_1, \tilde{\rho}_2\rho_2)$$

- An homomorphism $\Delta : \mathcal{G} \to M(\mathcal{G} \otimes \mathcal{G})$ is a comultiplication if
  - $\Delta(h)(1 \otimes k) \in \mathcal{G} \otimes \mathcal{G} \qquad (h \otimes 1)\Delta(k) \in \mathcal{G} \otimes \mathcal{G} \quad \forall h, k \in \mathcal{G}$
  - The coassociativity property is satisfied:

$$(h \otimes 1 \otimes 1)(\Delta \otimes \mathrm{id}_{\mathcal{G}})(\Delta(k)(1 \otimes r)) = (\mathrm{id}_{\mathcal{G}} \otimes \Delta)((h \otimes 1)\Delta(k))(1 \otimes 1 \otimes r) \quad \forall h, k, r \in \mathcal{G}$$

- The maps $T_i : \mathcal{G} \otimes \mathcal{G} \to \mathcal{G} \otimes \mathcal{G}$ such that

$$T_1(h \otimes k) = \Delta(h)(1 \otimes k) \quad \text{and} \quad T_2(h \otimes k) = (h \otimes 1)\Delta(k)$$

  are bijective.

If $(\mathcal{G}, \Delta)$ is a multiplier Hopf algebra there is a unique homomorphism $\epsilon : \mathcal{G} \to \mathbb{C}$, called counit, such that

$$(\epsilon \otimes \mathrm{id}_{\mathcal{G}})(\Delta(h)(1 \otimes k)) = hk \qquad (\mathrm{id}_{\mathcal{G}} \otimes \epsilon)((h \otimes 1)\Delta(k)) = hk \quad \forall h, k \in \mathcal{G}.$$

There is also a unique anti-homomorphism $S : \mathcal{G} \to M(\mathcal{G})$, called antipode, such that

$$m(S \otimes \mathrm{id}_{\mathcal{G}})(\Delta(h)(1 \otimes k)) = \epsilon(h)k \qquad m(\mathrm{id}_{\mathcal{G}} \otimes S)((h \otimes 1)\Delta(k)) = \epsilon(k)h \quad \forall h, k \in \mathcal{G}$$

here $m$ is the multiplication map. $(\mathcal{G}, \Delta)$ is a *regular multiplier Hopf algebra* if $S(\mathcal{G}) \subseteq \mathcal{G}$ and $S$ is invertible. There is a natural embedding $\iota_{\mathcal{G}} : \mathcal{G} \to M(\mathcal{G})$ which is an homomorphism

$$h \mapsto (L_h, R_h) \qquad L_h(k) = hk \qquad R_h(k) = kh$$

Moreover $\rho h \in \mathcal{G}$ and $h\rho \in \mathcal{G}$ for all $h \in \mathcal{G}$ and $\rho \in M(\mathcal{G})$,

$$\rho h = (L_{\rho_1(h)}, R_{\rho_1(h)}) \qquad h\rho = (L_{\rho_2(h)}, R_{\rho_2(h)})$$

We write $\rho h = \rho_1(h)$ and $h\rho = \rho_2(h)$.

A *right invariant functional* on $\mathcal{G}$ is a non-zero linear map $\psi : \mathcal{G} \to \mathbb{C}$ such that

$$(\psi \otimes \mathrm{id}_{\mathcal{G}})\Delta(h) = \psi(h)1$$

Here $(\psi \otimes \mathrm{id}_{\mathcal{G}})\Delta(h)$ denotes the element $\rho \in M(\mathcal{G})$ such that

$$\rho k = (\psi \otimes \mathrm{id}_{\mathcal{G}})(\Delta(h)(1 \otimes k)) \qquad k\rho = (\psi \otimes \mathrm{id}_{\mathcal{G}})((1 \otimes k)\Delta(h))$$

Similarly, a *left invariant functional* on $\mathcal{G}$ is a non-zero linear map $\varphi : \mathcal{G} \to \mathbb{C}$ such that

$$(\mathrm{id}_{\mathcal{G}} \otimes \varphi)\Delta(h) = \varphi(h)1.$$

Invariant functionals do not always exist. If $\varphi$ is a left invariant functional on $\mathcal{G}$ then it is unique up to scalar multiplication and $\psi = \varphi \circ S$ is a right invariant functional.

The dual of $(\mathcal{G}, \Delta)$ is $(\hat{\mathcal{G}}, \hat{\Delta})$:

- The elements of $\hat{\mathcal{G}}$ are the linear functionals of the form $\varphi(h\cdot)$

$$\hat{\mathcal{G}} = \{\xi_h : \mathcal{G} \to \mathbb{C} : \xi_h(x) = \varphi(hx)\}$$

  The elements of $\hat{\mathcal{G}}$ can also be written as $\varphi(\cdot h)$, $\psi(h\cdot)$, $\psi(\cdot h)$.
- The product on $\hat{\mathcal{G}}$ is defined as follows

$$(\xi_h \cdot \xi_k)(x) = (\varphi \otimes \varphi)(\Delta(x)(h \otimes k))$$

- The coproduct $\hat{\Delta} : \hat{\mathcal{G}} \to M(\hat{\mathcal{G}} \otimes \hat{\mathcal{G}})$ is defined by defining the elements $\hat{\Delta}(\xi_1)(1 \otimes \xi_2)$ and $(\xi_1 \otimes 1)\hat{\Delta}(\xi_2)$ in $\hat{\mathcal{G}} \otimes \hat{\mathcal{G}}$ as follows

$$((\xi_1 \otimes 1)\hat{\Delta}(\xi_2))(h \otimes k) = (\xi_1 \otimes \xi_2)(\Delta(h)(1 \otimes k))$$

$$(\hat{\Delta}(\xi_1)(1 \otimes \xi_2))(h \otimes k) = (\xi_1 \otimes \xi_2)((h \otimes 1)(\Delta(k))$$

- $(\hat{\mathcal{G}}, \hat{\Delta})$ is isomorphic to $(\mathcal{G}, \Delta)$ as algebraic quantum group

4.2. **Examples.**

- $\mathcal{G} = \mathbb{C}G$ with the usual Hopf algebra structure.

$$\varphi = \psi = \chi_e : G \to \mathbb{C} \quad \chi_e(h) = \begin{cases} 1 & e = h \\ 0 & e \neq h \end{cases}$$

$\boxed{\mathcal{G} \text{ is compact type}}$ because $1 \in \mathcal{G}$.

- $\mathcal{G} = \mathbb{C}\hat{G} = \{\sum_{g \in G} a_g \chi_g : a_g \in \mathbb{C}\ a_g \neq 0 \text{ for a finite amount of } g\}$

$$\chi_g \chi_h = \begin{cases} \chi_g & g = h \\ 0 & g \neq h \end{cases}$$

$$\Delta : \mathcal{G} \to M(\mathcal{G} \otimes \mathcal{G})$$

$$\Delta(\chi_g) = \sum_{t \in G} \chi_{gt^{-1}} \otimes \chi_t$$

The integral is $\varphi = \psi : \mathcal{G} \to \mathbb{C} \qquad \varphi(\chi_h) = \psi(\chi_h) = 1$
$\boxed{\mathcal{G} \text{ is discrete type}}$ because exists $k \in \mathcal{G}$ suche that $xk = \epsilon(x)k$.

- $\mathcal{G} = \mathcal{H}$ a finite dimensional Hopf algebra.
$\boxed{\mathcal{G} \text{ is compact and discrete}}$

4.3. **The algebra $\hat{\mathcal{A}}(\mathcal{G})$.** Let $(\mathcal{G}, \Delta)$ be an algebraic quantum group and $A$ be a $\mathcal{G}$-module algebra.

$\boxed{\hat{\mathcal{A}}(\mathcal{G}) := \mathcal{G} \underset{\text{ev}}{\otimes} \hat{\mathcal{G}}}$ $\qquad (g \otimes f)(\tilde{g} \otimes \tilde{f}) = gf(\tilde{g}) \otimes \tilde{f}$

$$t \cdot (g \otimes f) = \sum t_{(1)} \cdot g \otimes t_{(2)} \cdot f$$

$$(t \cdot f)(g) = f(S(t)g)$$

$\boxed{\hat{\mathcal{A}}(\mathcal{G}) \otimes A}$ $\qquad (g \otimes f \otimes a)(\tilde{g} \otimes \tilde{f} \otimes \tilde{a}) = gf(\tilde{g}) \otimes \tilde{f} \otimes a\tilde{a}$

$$t \cdot (g \otimes f \otimes a) = \sum t_{(1)} \cdot g \otimes t_{(3)} \cdot f \otimes t_{(2)} \cdot a$$

A functor $F : \mathcal{G}\text{-Alg} \to \mathcal{D}$ is *$\mathcal{G}$-stable* if $F(\iota_1)$ and $F(\iota_2)$ are isomorphism where

$$\iota_2 : A \to \begin{pmatrix} \hat{\mathcal{A}}(\mathcal{G}) \otimes A & 0 \\ 0 & A \end{pmatrix} \leftarrow \hat{\mathcal{A}}(\mathcal{G}) \otimes A : \iota_1$$

are corner inclusions. In [6] we define a strong and a weak stabilization. The previous definition correspond to the weak stabilization of [6], here we omit the word weak.

4.4. **Algebraic quantum kk-theory.** The following theorem allow us to define a bivariant K-theory on the category of $\mathcal{G}$-module algebras.

**Theorem 4.1.** [6] *Let $\mathcal{X}$ be a set such that* $\mathrm{card}(\mathcal{X}) = \mathbb{N} \times \dim_{\mathbb{C}}(\mathcal{G})$. *Let $F : \mathcal{G}$-*Alg $\to \mathcal{D}$ *be a $M_{\mathcal{X}}$-stable functor. The functor*

$$\hat{F} : \mathcal{G}\text{-Alg} \to \mathcal{D} \qquad A \mapsto F(\hat{\mathcal{A}}(\mathcal{G}) \otimes A)$$

*is $\mathcal{G}$-stable.*

**Theorem 4.2.** [6] *The functor $j^{\mathcal{G}} : \mathcal{G}$-Alg $\to \mathfrak{KK}^{\mathcal{G}}$ is an excisive, homotopy invariant, and $\mathcal{G}$-stable functor. Moreover, it is the universal functor for these properties.*

4.5. **Green-Julg theorem.** Consider the smash product functor

$$\# : \mathcal{G}\text{-Alg} \to \text{Alg} \text{ such that } A\#\mathcal{G} = A \otimes \mathcal{G} \text{ and } (a\#g)(b\#k) = \sum a(g_{(1)} \cdot b)\#g_{(2)}k$$

and the trivial action functor $\tau : \text{Alg} \to \mathcal{G}$-Alg. The pair

$$\mathcal{G}\text{-Alg} \underset{\tau}{\overset{\#}{\rightleftarrows}} \text{Alg}$$

can be extended to

$$\begin{array}{ccc} \mathcal{G}\text{-Alg} & \underset{\tau}{\overset{\#}{\rightleftarrows}} & \text{Alg} \longleftarrow \text{not adjoint functors} \\ \downarrow{\scriptstyle j^G} & & \downarrow{\scriptstyle j} \\ \mathfrak{KK}^{\mathcal{G}} & \underset{\tau}{\overset{\#}{\rightleftarrows}} & \mathfrak{KK} \longleftarrow \text{adjoint functors if } \mathcal{G} \text{ is a semisimple Hopf algebra.} \end{array}$$

**Theorem 4.3.** [6] *Let $\mathcal{H}$ be a semisimple Hopf algebra and $A$ be an $\mathcal{H}$-module algebra then*

$$kk^{\mathcal{H}}(\mathbb{C}, A) \simeq \mathrm{KH}(A\#\mathcal{H})$$

4.6. **Baaj-Skandalis duality.** In the context of algebraic quantum groups we have a good framework for group duality and we obtain a similar theorem of Baaj-Skandalis duality. Let $\mathcal{G}$ be an algebraic quantum group.

- $\hat{\mathcal{G}}$ is a $\mathcal{G}$-module: $(g \rightharpoonup f)(k) = f(kg)$
- $\mathcal{G}$ is a $\hat{\mathcal{G}}$-module: $f \rightharpoonup g = \sum f(g_{(2)})g_{(1)}$

**Theorem 4.4.** [5] *Let $A$ be a $\mathcal{G}$-module algebra then*

$$(A\#\mathcal{G})\#\hat{\mathcal{G}} \simeq \hat{\mathcal{A}}(\mathcal{G}) \otimes A$$

**Theorem 4.5.** [6] *The functors $\#\mathcal{G} : \mathfrak{KK}^{\mathcal{G}} \to \mathfrak{KK}^{\hat{\mathcal{G}}}$ and $\#\hat{\mathcal{G}} : \mathfrak{KK}^{\hat{\mathcal{G}}} \to \mathfrak{KK}^{\mathcal{G}}$ are equivalences and $kk^{\mathcal{G}}(A, B) \simeq kk^{\hat{\mathcal{G}}}(A\#\mathcal{G}, B\#\mathcal{G})$ where $A, B$ are $\mathcal{G}$-module algebras.*

## 5. Conclusion

The equivariant algebraic kk-theory introduced in [7] is a bivariant K-theory in the category of algebras with an action of a group $G$. The equivariant algebraic kk-theory introduced in [6] is a bivariant K-theory in the category of $\mathcal{G}$-module algebras where $\mathcal{G}$ is an algebraic quantum group. We resume properties and adjointness theorems of each case in the following table:

| | $G$ group [7] | $\mathcal{G}$ algebraic quantum group [6] |
|---|---|---|
| equivariant $\mathfrak{KK}$ stability | $\mathfrak{KK}^G$ <br> $\boxed{B \simeq_{\mathfrak{KK}^G} M_G B}$ <br> $\mathfrak{KK}^G \underset{\tau}{\overset{\rtimes}{\rightleftarrows}} \mathfrak{KK}$ | $\mathfrak{KK}^{\mathcal{G}}$ <br> $\boxed{B \simeq_{\mathfrak{KK}^{\mathcal{G}}} \hat{\mathcal{A}}(\mathcal{G}) \otimes B}$ <br> $\mathfrak{KK}^{\mathcal{G}} \underset{\tau}{\overset{\#}{\rightleftarrows}} \mathfrak{KK}$ |
| Green-Julg Theorem | adjoints functors <br> $G$ is finite, $\frac{1}{\lvert G\rvert} \in \ell$ <br> $\boxed{kk^G_*(\ell, A) \simeq KH_*(A \rtimes G)}$ | adjoint functors if $\mathcal{G} = \mathcal{H}$ <br> semisimple Hopf algebra <br> $\boxed{kk^{\mathcal{H}}_*(\ell, A) \simeq KH_*(A\#\mathcal{H})}$ |
| Ind-Res | $\mathfrak{KK}^G \underset{\mathrm{Ind}_H^G}{\overset{\mathrm{Res}_G^H}{\rightleftarrows}} \mathfrak{KK}^H$ <br> adjoint functors | $\boxed{?}$ |
| Imprimitivity | $\boxed{\mathrm{Ind}_H^G(B) \rtimes G \simeq_{\mathfrak{KK}} B \rtimes H}$ | $\boxed{?}$ |
| Baaj-Skandalis duality | $\mathfrak{KK}^G \underset{\hat{\rtimes}}{\overset{\rtimes}{\rightleftarrows}} \hat{\mathfrak{KK}}^G$ <br> equivalences | $\mathfrak{KK}^{\mathcal{G}} \underset{\#}{\overset{\#}{\rightleftarrows}} \mathfrak{KK}^{\hat{\mathcal{G}}}$ <br> equivalences |

## References


[1] R. Akbarpour and M. Khalkhali. Hopf algebra equivariant cyclic homology and cyclic homology of crossed product algebras. J. Reine Angew. Math. 559 (2003), 137–152.

[2] S. Baaj, G. Skandalis. C*-algèbres de Hopf et théorie de Kasparov équivariante (Hopf C*-algebras and equivariant. Kasparov theory). *K-Theory 2 (6) (1989) 683–721.*

[3] G. Cortiñas and A. Thom. Bivariant algebraic $K$-theory. *Journal für die Reine und Angewandte Mathematik (Crelle's Journal)*, 610:267–280, 2007.

[4] J.Cuntz A new look at KK-theory. *K-Theory* 1(1)(1987)31–51.

[5] Drabant, Bernhard; Van Daele, Alfons and Zhang, Yinhuo Actions of multiplier Hopf algebras. *Comm. Algebra* 27 (1999), no. 9, 4117–4172.

[6] E. Ellis. Algebraic quantum kk-theory *Comm. Algebra* 46 (2018), no. 8, 3642–3662.

[7] E. Ellis. Equivariant algebraic *kk*-theory and adjointness theorems. *J. Algebra*, 398 (2014), 200–226.

[8] G. Grarkusha Algebraic Kasparov K-theory I Doc. Math. 19 (2014) 1207-1269.
[9] G. Garkusha Universal bivariant algebraic K-theories. J. Homotopy Relat. Struct. 8 (2013), no. 1, 67-116.
[10] P.G. Goerss and J.F. Jardine. Simplicial homotopy theory Progress in Mathematics (Boston, Mass.), 1999.
[11] E. Guentner, N. Higson, J Trout. Equivariant $E$-theory for $C^*$-algebras. Mem. Amer. Math. Soc. 148 (2000), no. 703, viii+86 pp.
[12] N.Higson. A characterization of KK-theory. *Pacific J. Math.* 126(2)(1987)253–276.
[13] G.G. Kasparov The operator K -functor and extensions of $C*$-algebras. Izv. Akad. Nauk SSSR, Ser. Mat. 44 (3) (1980) 571–636, pp. 719.
[14] S. Mac Lane. Categories for the working mathematician. *2nd ed., Graduate Texts in Mathematics.* 5. New York, NY: Springer, 1998.
[15] R. Meyer. Categorical aspects of bivariant K-theory Cortiñas, Guillermo (ed.) et al., $K$-theory and noncommutative geometry. Proceedings of the ICM 2006 satellite conference, Valladolid, Spain, August 31–September 6, 2006. Zürich: European Mathematical Society (EMS). Series of Congress Reports, 1-39, 2008
[16] E. Rodriguez. Bivariant algebraic K-theory categories and a spectrum for $G$-equivariant bivariant algebraic $K$-theory PhD Thesis (2017). Universidad de Buenos Aires.
[17] S. Montgomery. *Hopf Algebras and Their Actions on Rings*. CBMS Regional Conference Series in Mathematics **82** (1993).
[18] A. Van Daele. Multiplier Hopf Algebras Trans. Amer. Math. Soc. 342 (1994), 917-932.
[19] A. Van Daele. An algebraic framework for group duality Adv. Math. 140 (1998), no. 2, 323-V366
[20] C. Voigt. Cyclic cohomology and Baaj-Skandalis duality. J. K-Theory 13 (2014), no. 1, 115–145
[21] C. Voigt. Equivariant periodic cyclic homology. J. Inst. Math. Jussieu 6 (2007), no. 4, 689–763.
[22] C. Weibel. *Homotopy Algebraic $K$-theory*. Contemporary Math. **83** (1989) 461–488.

IMERL, Facultad de Ingeniería, Universidad de la República. Montevideo, Uruguay.
*Email address*: eellis@fing.edu.uy

PONENCIA

# AMENABILIDAD, ACCIONES DE GRUPOS, Y EQUIVALENCIA ORBITAL

EUSEBIO GARDELLA

Publicaciones Matemáticas del Uruguay
Volumen 17, Julio 2019, Páginas 153–164
S 0797-1443

# AMENABILIDAD, ACCIONES DE GRUPOS, Y EQUIVALENCIA ORBITAL

EUSEBIO GARDELLA

Resumen. Estas notas son una versión expandida de una charla dada en el *Sexto Coloquio Uruguayo de Matemática*, en Diciembre de 2017. El objetivo de las mismas es presentar ciertos resultados, algunos clásicos y otros recientes, referentes a la clasificación de acciones libres y ergódicas de grupos discretos a menos de conjugación y equivalencia orbital.

Dado el rol fundamental que juega la noción de amenabilidad de grupos en la teoría de equivalencia orbital, hemos incluído una breve introducción a este tópico, con particular énfasis en la relación entre amenabilidad y la ausencia de subgrupos libres no conmutativos.

## Índice

1. Introducción 153
2. La paradoja de Banach-Tarski y amenabilidad 154
3. Clasificación de acciones libres y ergódicas 156
4. Esbozo de la prueba del Teorema 3.9 159
Referencias 163

## 1. Introducción

La Teoría Ergódica estudia acciones de grupos, generalmente discretos, en espacios de probabilidad, vía automorfismos que preservan la medida. El *problema de clasificación* en Teoría Ergódica procura buscar un método explícito que permita distinguir ciertas clases de acciones, por ejemplo, aquellas que son libres y ergódicas[1]. Este programa fue iniciado por Halmos en la década de los 50's, y desde entonces ha recibido mucha atención. La relación de equivalencia más natural es la de conjugación, mientras que la equivalencia orbital se convirtió en una de las nociones más relevantes en la teoría a partir del trabajo seminal de Dye. La equivalencia orbital también ha encontrado numerosas aplicaciones en las álgebras de operadores, vía la muy estudiada construcción de productos cruzados de Murray y von Neumann.

La equivalencia orbital es muy flexible, ya que entre otras cosas, se aplica a acciones de grupos diferentes, mientras que conjugación sólo tiene sentido para acciones de un mismo grupo. Sorprendentemente, la amenabilidad de grupos tiene

[1]Las transformaciones ergódicas son aquellas que son *irreducibles*, y el Teorema de Descomposición Ergódica afirma que cualquier acción puede escribirse como una integral directa de acciones ergódicas.

©2019 PMU

un rol central en este contexto: una combinación de resultados de Dye y Ornstein-Weiss muestra que todas las acciones libres y ergódicas de grupos amenables son orbitalmente equivalentes. Los resultados de clasificación a menos de conjugación son menos generales: para acciones de los enteros via shifts de Bernoulli, Ornstein mostró que la entropía es un invariante completo. A pesar de que otros teoremas de clasificación fueron probados en este contexto, el caso general permaneció abierto, hasta que Foreman-Rudolph-Weiss demostraron, de forma rigurosa, que un teorema de clasificación *razonable* para acciones de los enteros no sería posible: la relación de conjugación de transformaciones ergódicas no es Borel. Esto puede interpretarse diciendo que no existe un método o protocolo que involucre una cantidad numerable de información y pasos, que pueda distinguir cuándo dos transformaciones ergódicas son conjugadas.

En décadas recientes, el estudio de la equivalencia orbital se ha concentrado en acciones de grupos no amenables. En ciertos casos, es posible obtener resultados de *rigidez*, que muestran que bajo ciertas hipótesis, equivalencia orbital implica conjugación (y en particular "recuerda" el grupo). En este contexto, la teoría de superrigidez, desarrollada por Zimmer, ha jugado un rol primordial. En los últimos años, la infusión de técnicas provenientes del álgebra de operadores, como la teoría de deformación-rigidez de Popa, le ha dado mayor ímpetu al área.

En estas notas, hacemos un recuento sobre nuestro trabajo reciente en colaboración con Martino Lupini [13], donde hacemos nuevas contribuciones al problema de clasificación, tanto para conjugación como para equivalencia orbital, en el caso no amenable. Nuestros resultados implican una solución completa en el caso de equivalencia orbital, que puede formularse en la siguiente dicotomía: la equivalencia orbital es Borel si y sólo si el grupo es amenable, en cuyo caso existe una única clase de equivalencia. Nuestras técnicas provienen del álgebra de operadores y son esencialmente analíticas, demostrando la rica interacción existente entre las C*-álgebras y álgebras de von Neumann, por un lado, y la Teoría Ergódica, por el otro.

Hemos organizado estas notas de la siguiente manera: en la Sección 2 damos una breve introducción a la amenabilidad para grupos discretos, con especial énfasis en la no amenabilidad del grupo libre $\mathbb{F}_2$ y de los grupos que lo contienen. En la Sección 3 introducimos el problema de clasificación en Teoría Ergódica, y damos un recuento de los resultados más relevantes, incluyendo los aportes de [13]. Finalmente, la Sección 4 contiene un esbozo de la prueba de nuestro teorema principal, así como algunas reflexiones finales.

**Agradecimientos:** Quisiéramos agradecer al comité organizador del Coloquio, y en particular a Martín Rieris, por darnos la oportunidad de presentar nuestro trabajo. También agradecemos a Fernando Abadie y Marina Gardella por haber leído rigurosamente el manuscrito, y por todas sus correcciones y comentarios.

## 2. La paradoja de Banach-Tarski y amenabilidad

En esta sección introducimos el concepto de amenabilidad para grupos. El libro de Paterson [22] es una excelente referencia donde el lector podrá encontrar mucho más de lo que presentamos aquí. Comenzamos con una motivación usando el siguiente teorema, que es popularmente conocido como la *paradoja de Banach-Tarski*:

**Teorema 2.1** (Banach-Tarski, 1924)**.** *Existe una partición finita de la bola* $\mathbb{D}^3 \subseteq \mathbb{R}^3$*, tal que los conjuntos de esta partición pueden rotarse y trasladarse de modo que el resultado de estos movimientos es la unión disjunta de dos copias de* $\mathbb{D}^3$*.*

Este teorema es paradojal porque desafía la intuición: ¿cómo es posible duplicar el volumen? La respuesta es que estos conjuntos no pueden "medirse" con una medida que sea invariante por congruencias y de acuerdo a la cual la bola tenga volumen total 1. Esto motiva la siguiente definición:

**Definición 2.2** (von Neumann, 1930)**.** *Un grupo discreto* $\Gamma$ *es* amenable *si admite una medida de probabilidad finitamente aditiva, definida en todos los subconjuntos de* $\Gamma$*, que es invariante por la acción (izquierda) de* $\Gamma$*.*

La medida en la definición anterior no es una medida en el sentido clásico de Teoría de la Medida, pues no se asume que sea aditiva con respecto a uniones numerables. (En Inglés, se la denomina "mean"). La medida de Haar satisface las condiciones en la definición excepto que no tiene masa total 1, a menos que el grupo sea finito.

La noción de amenabilidad admite una cantidad sorprendente de formulaciones equivalentes, de naturalezas muy variadas: algunas son puramente algebraicas, algunas puramente analíticas, otras se refieren puramente a la teoría de representaciones del grupo, etc. A continuación recordamos tan sólo dos de ellas.

Recordemos que el símbolo $\triangle$ denota la diferencia simétrica de conjuntos, es decir, $E \triangle F = (E \cup F) \cap (E \cap F)^c$.

**Definición 2.3.** *Sea* $\Gamma$ *un grupo discreto numerable. Una* sucesión de Følner *para* $\Gamma$ *es una sucesión* $(F_n)_{n\in\mathbb{N}}$ *de subconjuntos finitos de* $\Gamma$ *que satisfacen*

$$\lim_{n\to\infty} \frac{|\gamma F_n \triangle F_n|}{|F_n|} = 0$$

*para todo* $\gamma \in \Gamma$*.*

**Definición 2.4.** *Sea* $\Gamma$ *un grupo discreto numerable. Decimos que* $\Gamma$ *admite una* descomposición paradójica *si existen subconjuntos disjuntos* $A, B \subseteq \Gamma$*, particiones finitas*

$$A = \bigcup_{j=1}^{n} A_j \quad y \quad B = \bigcup_{j=1}^{n} B_j$$

*y elementos* $\gamma_1^A, \ldots, \gamma_n^A, \gamma_1^B, \ldots, \gamma_n^B \in \Gamma$ *tales que* $\{\gamma_j^A A_j\}_{j=1}^n$ *y* $\{\gamma_j^B B_j\}_{j=1}^n$ *son particiones de* $\Gamma$*.*

**Teorema 2.5.** *Sea* $\Gamma$ *un grupo discreto numerable. Las siguientes afirmaciones son equivalentes:*

*(1)* $\Gamma$ *es amenable.*
*(2)* $\Gamma$ *admite una sucesión de Følner.*
*(3)* $\Gamma$ *no admite una descomposición paradójica.*

La única dirección que es sencilla es (1) $\Rightarrow$ (3), que dejamos como ejercicio para el lector.

Las sucesiones de Følner son una herramienta muy importante, que permiten tomar promedios aproximados en grupos amenables. Por esta razón, los grupos amemables también son llamados *promediables.*

La proposición que sigue nos permite construir muchos ejemplos de grupos amenables. Por ejemplo, deducimos de (2) y (4) que cualquier grupo nilpotente es amenable.

**Proposición 2.6.** *Sea $\Gamma$ un grupo discreto.*

*(1) Si $\Gamma$ es finito, entonces es amenable.*
*(2) Si $\Gamma$ es abeliano, entonces es amenable.*
*(3) Si $\Gamma$ es amenable, entonces cualquier subgrupo y cualquier cociente de $\Gamma$ es también amenable.*
*(4) Si $\Lambda$ es un subgrupo normal y amenable de $\Gamma$ y $\Gamma/\Lambda$ es amenable, entonces $\Gamma$ es amenable.*

El ejemplo prototípico de un grupo no amenable es el grupo libre en dos generadores, $\mathbb{F}_2$[2]. En vista de la parte (3) de la proposición anterior, cualquier grupo que contiene a $\mathbb{F}_2$ tampoco es amenable.

Volviendo a la paradoja de Banach-Tarski, una forma de probarla consiste en seguir los siguientes pasos. Primero, probar que $\mathbb{F}_2$ es un subgrupo del grupo de isometrías de $\mathbb{R}^3$, y luego, usar una descomposición paradójica de $\mathbb{F}_2$ para encontrar una partición de $\mathbb{D}^3$ como en el enunciado del Teorema 2.1.

El hecho de que la paradoja de Banach-Tarski sólo usa la existencia de $\mathbb{F}_2$ como subgrupo, condujo a von Neumann a pensar que la existencia de subgrupos libres podría ser la única obstrucción a la amenabilidad. Más concretamente:

**Pregunta 2.7.** *Sea $\Gamma$ un grupo discreto. ¿Es cierto que $\Gamma$ es amenable si y sólo si no contiene un subgrupo isomorfo a $\mathbb{F}_2$?*

El hecho que esta pregunta tenga una respuesta afirmativa fue conocido como la *conjetura de von Neumann*, y permaneció abierta por más de cuatro décadas, hasta que Olshanski'i encontró en [20] un ejemplo de un grupo no amenable cuyos subgrupos propios son todos finitos. A pesar de que esta conjetura resultó ser falsa en su formulación original, se ha probado recientemente que todo grupo no amenable contiene $\mathbb{F}_2$ "a nivel dinámico". Volveremos sobre esto hacia el final de estas notas.

## 3. Clasificación de acciones libres y ergódicas

Por el resto de este documento, fijamos un grupo discreto, infinito y numerable $\Gamma$, y un espacio de probabilidad estándar[3] sin átomos $(X, \mu)$. (Para fijar ideas, podemos tomar el intervalo $[0, 1]$ con la medida de Lebesgue). Denotamos por $\mathrm{Aut}(X, \mu)$ el grupo de biyecciones medibles de $X$, cuyas inversas también son medibles, y que preservan la medida $\mu$.

Nuestro objeto de estudio son las acciones de $\Gamma$ en $(X, \mu)$ por automorfismos, es decir, los homomorfismos de grupos $\Gamma \to \mathrm{Aut}(X, \mu)$, que abreviamos a $\Gamma \curvearrowright (X, \mu)$.

**Definición 3.1.** *Una acción $\Gamma \curvearrowright (X, \mu)$ se llama:*

*(1)* libre*, si $\mu(\{x \in X : \gamma \cdot x = x\}) = 0$ para todo $\gamma \in \Gamma \setminus \{1\}$.*

[2]Denotemos los generadores de $\mathbb{F}_2$ por $a$ y $b$, y sea $S(a)$ el subconjunto de $\mathbb{F}_2$ de las palabras que comienzan con $a$; lo mismo para $S(a^{-1})$, $S(b)$ y $S(b^{-1})$. Entonces una descomposición paradójica de $\mathbb{F}_2$ se obtiene tomando $A = S(a) \cup S(a^{-1})$ con $\gamma_1^A = 1$ y $\gamma_2^A = a$, y $B = S(b) \cup S(b^{-1})$ con $\gamma_1^B = 1$ y $\gamma_2^B = b$.

[3]Esto es, una medida de probabilidad en la $\sigma$-álgebra de Borel de un espacio métrico completo y separable. Dicha $\sigma$-álgebra es independiente del espacio escogido. A menos de isomorfismo, existe una única medida de probabilidad sin átomos definida en dicha $\sigma$-álgebra.

*(2)* ergódica*, si cuando $E \subseteq X$ es $\Gamma$-invariante y medible, entonces $\mu(E)$ es o bien 0 o 1.*

**Observación 3.2.** *Si $\Gamma$ es infinito y actúa libremente en un espacio de probabilidad, entonces dicho espacio no puede tener átomos, pues la órbita de cualquier átomo sería infinita (por ser la acción libre) y por lo tanto tendría medida infinita, lo cual es absurdo. Por lo tanto, no hay pérdida de generalidad al restringirnos a espacios sin átomos.*

El tema central de estas notas es la clasificación de acciones libres y ergódicas a menos de las relaciones de equivalencia definidas a continuación:

**Definición 3.3.** *Sean $\theta, \kappa \colon \Gamma \curvearrowright (X, \mu)$ acciones. Decimos que*

*(1) $\theta$ es* conjugada *a $\kappa$, escrito $\theta \cong \kappa$, si existe $f \in \mathrm{Aut}(X, \mu)$ tal que $f \circ \theta_\gamma = \kappa_\gamma \circ f$ en $\mu$-casi todo punto $x \in X$, para todo $\gamma \in \Gamma$.*

*(2) $\theta$ es* orbitalmente equivalente *a $\kappa$, escrito $\theta \sim_{\mathrm{OE}} \kappa$, si existe $f \in \mathrm{Aut}(X, \mu)$ tal que $f(\theta(\Gamma) \cdot x) = \kappa(\Gamma) \cdot f(x)$ para $\mu$-casi todo $x \in X$.*

Si bien la conjugación es la relación de equivalencia más natural, es también muy rígida. Por ejemplo, una acción libre y ergódica genérica (es decir, elegida al *azar*) no es conjugada a su inversa[4]. Sin embargo, cualquier acción es siempre orbitalmente equivalente a su inversa. De hecho, la equivalencia orbital admite una descripción muy satisfactoria en términos de álgebras de operadores:

**Teorema 3.4** (Singer [23], 1977)**.** *Sean $\theta, \kappa \colon \Gamma \curvearrowright (X, \mu)$ acciones[5]. Entonces $\theta \sim_{\mathrm{OE}} \kappa$ si y sólo si existe un isomorfismo $\psi \colon L^\infty(X, \mu) \rtimes_\theta \Gamma \to L^\infty(X, \mu) \rtimes_\kappa \Gamma$ de los productos cruzados que satisface $\psi(L^\infty(X, \mu)) = L^\infty(X, \mu)$.*

Las siguientes preguntas de Halmos, planteadas en sus famosas notas en Teoría Ergódica, motivan nuestro trabajo:

**Pregunta 3.5** (Halmos [15], 1956)**.** *¿Existe un* método *que determine si dos acciones (libres y ergódicas) $\Gamma \curvearrowright (X, \mu)$ son conjugadas? ¿Y orbitalmente equivalentes?*

Idealmente, una respuesta afirmativa vendría acompañada de una "receta", como podría ser el cálculo de algún invariante. El mismo Halmos admitió en sus notas que sus preguntas son vagas, y una interpretación formal se puede encontrar en el contexto de la teoría de complejidad de Borel:

**Pregunta 3.6** (Kechris [19], 2006)**.** *Sea $\Gamma$ un grupo discreto numerable. ¿Son Borel las relaciones de conjugación y equivalencia orbital de acciones (libres, ergódicas) de $\Gamma$?*

Tenemos que especificar el significado de la palabra "Borel.[en] la pregunta anterior. Supongamos que $E$ es un espacio métrico completo y separable con una noción de equivalencia $\sim$ en él. Identificamos $\sim$ con un subconjunto de $E \times E$ a través de su grafo:

$$\mathrm{gr}(\sim) = \{(e, f) \in E \times E \colon e \sim f\}.$$

Decimos que $\sim$ es Borel si $\mathrm{gr}(\sim)$ es un subconjunto de Borel de $E \times E$ (con la $\sigma$-álgebra producto). La relación es Borel precisamente si existe un procedimiento uniforme y explícito que, dados dos elementos $e$ y $f$ de $E$, corre por una cantidad

[4]Más precisamente, la probabilidad de que sea conjugada a su inversa es cero.

[5]De hecho, en este teorema uno puede asumir que $\theta$ y $\kappa$ son acciones de grupos distintos.

numerable de pasos, en cada uno de los cuales verifica si el par $(e, f)$ pertenece a ciertos abiertos de $E \times E$, y al final decide si $e$ y $f$ son equivalentes.

Esta discusión se puede adaptar al contexto de acciones de $\Gamma$ de la siguiente forma. Primero, observamos que el espacio $\mathrm{Aut}(X, \mu)$ de automorfismos de $(X, \mu)$ es un espacio métrico completo con la topología débil (convergencia en medida). El espacio producto $\mathrm{Aut}(X, \mu)^\Gamma$ también es métrico completo, y el espacio $\mathrm{Act}_\Gamma(X, \mu)$ de acciones de $\Gamma$ en $(X, \mu)$ es cerrado en él[6], por lo que también es completo. El mismo razonamiento se aplica al subespacio de las acciones que son libres y ergódicas, puesto que éste es también un subespacio cerrado en $\mathrm{Act}_\Gamma(X, \mu)$.

A continuación, hacemos un resumen de lo que se conoce en relación a las Pregunta 3.5 y Pregunta 3.6. Nuestro recuento no es cronológico, sino lógico.

Uno espera que la relación de conjugación sea bastante más complicada que la de equivalencia orbital. En efecto, esta relación ya es complicada para los enteros:

**Teorema 3.7** (Foreman-Rudolph-Weiss [8], 2010)**.** *La relación de conjugación de automorfismos libres y ergódicos no es Borel.*

A pesar de que el teorema anterior sugiere que la relación de conjugación no es Borel para ningún grupo, esto no había sido confirmado, hasta ahora, en ningún otro caso.

Mucho más se sabe con respecto a equivalencia orbital. Una combinación de resultados clásicos de Dye y Ornstein-Weiss da una respuesta completa en el caso amenable:

**Teorema 3.8** (Dye y Ornstein-Weiss [6, 21], 1963 y 1982)**.** *Dos acciones libres y ergódicas cualesquiera de $\mathbb{Z}$ son orbitalmente equivalentes. Más generalmente, dos acciones libres y ergódicas cualesquiera de un grupo amenable son orbitalmente equivalentes*[7]*.*

También existen "modelos" explícitos: los shifts de Bernoulli

$$\beta \colon \Gamma \curvearrowright [0,1]^\Gamma = \{f \colon \Gamma \to [0,1]\},$$

dados por $\beta_\gamma(f)(\delta) = f(\gamma^{-1}\delta)$ para todos $\gamma, \delta \in \Gamma$ y $f \colon \Gamma \to [0,1]$.

Deducimos del Teorema 3.8 que la versión referente a equivalencia orbital en la Pregunta 3.5 tiene una respuesta muy sencilla cuando $\Gamma$ es amenable: dos acciones de $\Gamma$ son *siempre* orbitalmente equivalentes. Por lo tanto, la pregunta es sólo interesante en el caso no amenable. En este contexto, lo único que era conocido eran resultados referentes a la *cardinalidad* de las clases de equivalencia orbital:

- Connes-Weiss [4], 1980: si $\Gamma$ no tiene la propiedad (T) de Kazhdan[8], entonces admite al menos dos acciones libres y ergódicas que no son orbitalmente equivalentes.
- Hjorth [16], 2005: si $\Gamma$ tiene la propiedad (T), entonces admite una cantidad no numerable de acciones libres y ergódicas que no son orbitalmente equivalentes.

(De estos resultados, junto con el Teorema 3.8, deducimos que $\Gamma$ es amenable *si y sólo si* dos acciones libres y ergódicas cualesquiera de $\Gamma$ son orbitalmente equivalentes. Nuestro Corolario 3.10 es la versión definitiva de esta observación).

[6]Esto es fácil de ver: las condiciones que una función $\Gamma \to \mathrm{Aut}(X, \mu)$ debe cumplir para ser un homomorfismo de grupos, son cerradas en la topología de la convergencia puntual.

[7]De hecho, el resultado es incluso más general: dos acciones libres y ergódicas cualesquiera de dos grupos amenables (potencialmente distintos) son siempre orbitalmente equivalentes.

[8]Ver, por ejemplo, la Definición 4.3 y los comentarios que le suceden.

- Gaboriau-Popa [10], 2005: los grupos libres $\mathbb{F}_n$, para $1 < n < \infty$, admiten una cantidad no numerable de acciones libres y ergódicas que no son orbitalmente equivalentes.
- Ioana [18], 2007: si $\Gamma$ contiene $\mathbb{F}_2$, entonces admite una cantidad no numerable de acciones libres y ergódicas que no son orbitalmente equivalentes..
- Epstein [7], 2007: cualquier grupo no amenable admite una cantidad no numerable de acciones libres y ergódicas que no son orbitalmente equivalentes.

Como conjugación es una relación más fuerte que equivalencia orbital, los resultados anteriores también muestran que cualquier grupo no amenable admite una cantidad no numerable de acciones libres y ergódicas que no son conjugadas. Por supuesto, el hecho de que exista una cantidad no numerable de clases de equivalencia no implica que dicha relación no sea Borel. Tampoco alcanza con saber que la relación de equivalencia orbital no es Borel para concluir que la relación de conjugación tampoco lo es, ya que no ser medible no es una propiedad heredada por subconjuntos arbitrarios.

En colaboración con Martino Lupini [13] (véase también [12]), hemos respondido las preguntas de Halmos (en la reformulación de Kechris dada en la Pregunta 3.6) para grupos no amenables, tanto para la relación de conjugación como la de equivalencia orbital:

**Teorema 3.9** (G.-Lupini [13], 2017)**.** *Si $\Gamma$ no es amenable, entonces las relaciones de equivalencia orbital y conjugación de acciones libres y ergódicas de $\Gamma$ no son Borel.*

Junto con el Teorema 3.8, nuestro teorema implica la siguiente dicotomía:

**Corolario 3.10.** *Sea $\Gamma$ un grupo infinito numerable.*

*(1) Si $\Gamma$ es amenable, entonces dos acciones libres y ergódicas cualesquiera de $\Gamma$ son orbitalmente equivalentes.*
*(2) Si $\Gamma$ no es amenable, entonces la equivalencia orbital de acciones libres y ergódicas de $\Gamma$ no es Borel.*

Una aplicación concreta del Teorema 3.9, para quienes no estén necesariamente interesados en complejidad de Borel, es para descartar la validez de ciertos teoremas de clasificación hipotéticos. Por ejemplo, consideremos el siguiente resultado:

**Teorema 3.11** (Bowen [1], 2010)**.** *Los shifts de Bernoulli de grupos libres se clasifican por su entropía.*

Uno podría, un poco inocentemente, pensar que tal vez la entropía clasifica no sólo los shifts de Bernoulli, sino todas las acciones libres y ergódicas. Pero el Teorema 3.9 implica que esto no puede suceder, ya que la entropía es un invariante que se computa de forma Borel, y la igualdad de números reales es una relación de Borel.

## 4. Esbozo de la prueba del Teorema 3.9

A pesar de lo que su enunciado parecería indicar, la prueba del Teorema 3.9 usa únicamente técnicas que provienen del álgebra de operadores, más precisamente de las álgebras de von Neumann. En esta última sección, describimos las principales ideas usadas en dicha prueba. Para ello, necesitamos algunas definiciones y resultados preliminares. Seguimos fijando un espacio de probabilidad estándar sin átomos

$(X,\mu)$ y un grupo infinito numerable $\Gamma$. Denotamos por $\mathcal{U}(L^\infty(X,\mu))$ al grupo de las funciones medibles $X \to S^1$, identificadas cuando coinciden en $\mu$-casi todo punto. Observar que $\mathcal{U}(L^\infty(X,\mu))$ es precisamente el grupo de los unitarios del álgebra de von Neumann $L^\infty(X,\mu)$.

**Definición 4.1.** *Sea $\Gamma \curvearrowright^\theta (X,\mu)$ una acción, y sean $\Delta \leq \Lambda \leq \Gamma$ subgrupos.*

*(1) Un* $\theta$-cociclo *es una función $u\colon \Gamma \to \mathcal{U}(L^\infty(X,\mu))$ que satisface $u_{\gamma\rho} = u_\gamma \theta_\gamma(u_\rho)$ para todo $\gamma, \rho \in \Gamma$.*

*(2) Un $\theta$-cociclo $u$ se llama* $\Delta$-invariante *si $u_\delta = 1$ y $\theta_\delta(u_\gamma) = u_\gamma$ para todo $\delta \in \Delta$ y todo $\gamma \in \Gamma$.*

*(3) Un $\theta$-cociclo $u$ es un* coborde débil $\Lambda$-relativo *si existen un unitario $v \in \mathcal{U}(L^\infty(X,\mu))$ y escalares $z_\lambda \in S^1$, para $\lambda \in \Lambda$, que satisfacen*

$$u_\lambda(x) = z_\lambda v(x)\overline{v(\theta_{\lambda^{-1}}(x))}$$

*para $\mu$-casi todo $x \in X$ y para todo $\lambda \in \Lambda$.*

*Observemos que el grupo de los $\theta$-cociclos $\Delta$-invariantes tiene una estructura natural de grupo abeliano, con la operación dada por multiplicación puntual. Con dicha estructura, el conjunto de los cobordes débiles $\Lambda$-relativos es un subgrupo normal, aunque no necesariamente cerrado.*

*(4) El grupo de* 1-cohomología débil, $\Delta$-invariante, y $\Lambda$-relativa *de $\theta$ es el cociente $H^1_{:\Delta,\Lambda,w}(\theta)$ de los $\theta$-cociclos $\Delta$-invariantes por los cobordes débiles $\Lambda$-relativos.*

El grupo $H^1_{:\Delta,\Lambda,w}(\theta)$ es un invariante de conjugación, pero no de equivalencia orbital.

**Observación 4.2.** *Cuando $\Delta = \{e\}$ y $\Lambda = \Gamma$, recuperamos la definición usual de 1-cohomología. Cuando $\Delta$ es un subgrupo normal de $\Lambda = \Gamma$, entonces $H^1_{:\Delta,\Lambda,w}(\theta)$ es la cohomología de la acción cociente.*

Sea $\pi\colon \Gamma \to \mathcal{U}(\mathcal{H})$ una representación unitaria de $\Gamma$ en un espacio de Hilbert $\mathcal{H}$. Dado un subgrupo $\Omega \subseteq \Gamma$, un vector $\xi \in \mathcal{H}$ se dice $\Omega$-*invariante* si $\pi_\omega(\xi) = \xi$ para todo $\omega \in \Omega$. Decimos que $\pi$ tiene *vectores unitarios casi invariantes* si existe una sucesión $(\xi_n)_{n\in\mathbb{N}}$ de vectores unitarios en $\mathcal{H}$ que satisfacen $\lim_{n\to\infty} \|\pi_\gamma(\xi_n) - \xi_n\| = 0$ para todo $\gamma \in \Gamma$.

**Definición 4.3.** *Decimos que una terna $\Delta \leq \Lambda \leq \Gamma$ tiene la* propiedad (T) *si cualquier representación unitaria de $\Gamma$ que tiene vectores unitarios $\Delta$-invariantes que son casi invariantes por $\Gamma$, tiene vectores unitarios que son $\Lambda$-invariantes.*

Por ejemplo, $\{1\} \leq \Gamma \leq \Gamma$ tiene la propiedad (T) si y sólo si $\Gamma$ tiene la propiedad (T) en el sentido de Kazhdan. Cuando $\Delta = \{1\}$, obtenemos la propiedad (T) relativa estudiada por Margulis. Por último, si $\Delta$ es un subgrupo normal de $\Lambda$ y $\Lambda/\Delta$ tiene la propiedad (T), entonces $\Delta \leq \Lambda \leq \Gamma$ tiene la propiedad (T).

Adaptando métodos muy sofisticados de Popa, probamos un teorema de *superrigidez* para acciones maleables de grupos con la propiedad (T) para ternas. En el contexto de la prueba del Teorema 3.9, usaremos dicho teorema de superrigidez en la siguiente forma:

**Teorema 4.4** (G.-Lupini [13], 2017)**.** *Sea $\Delta \leq \Lambda \leq \Gamma$ una terna con la propiedad (T), y supongamos que $\Delta$ tiene índice infinito en $\Gamma$. Sea $\beta\colon \Gamma \curvearrowright L^\infty(X,\mu)^{\otimes\Gamma/\Delta} \otimes L^\infty(X,\mu)^{\otimes\Gamma}$ el shift canónico. Entonces*

$$H^1_{:\Delta,\Lambda,w}(\beta) = \{1\}.$$

A continuación, describimos la prueba del Teorema 3.9.

**Primer paso:** Supongamos que $\Gamma$ contiene $\mathbb{F}_2$. Comenzamos por fijar una inclusión $\mathbb{F}_2 \leq \mathrm{SL}_2(\mathbb{Z})$, que corresponde a una acción $\mathbb{F}_2 \curvearrowright \mathbb{Z}^2$ por automorfismos de grupos. Usando dualidad, obtenemos una acción de $\mathbb{F}_2$ en $\mathbb{T}^2 = \widehat{\mathbb{Z}^2}$, la cual denotamos por $\rho\colon \mathbb{F}_2 \curvearrowright \mathbb{T}^2$.

**Notación 4.5.** *Podemos (co)inducir $\rho$ a una acción $\Gamma \curvearrowright (Y, \nu)$ de la siguiente manera. Definimos*

$$Y = \{f\colon \Gamma \to \mathbb{T}^2\colon f(\gamma a) = \rho_{a^{-1}}(f(\gamma)) \text{ para todos } \gamma \in \Gamma, a \in \mathbb{F}_2\} \subseteq (\mathbb{T}^2)^\Gamma,$$

*munido con la restricción $\nu$ de la medida producto. Entonces $(Y, \nu)$ es el espacio de probabilidad estándar sin átomos, y definimos $\widehat{\rho}\colon \Gamma \curvearrowright (Y, \nu)$ por $\widehat{\rho}_{\gamma_0}(f)(\gamma_1) = f(\gamma_0^{-1}\gamma_1)$ para todos $\gamma_0, \gamma_1 \in \Gamma$.*

Probaremos lo siguiente:

**Teorema 4.6.** *Existe una asignación de Borel $A \mapsto \theta_A$ de grupos abelianos infinitos numerables sin torsión a acciones libres y ergódicas $\Gamma \curvearrowright (X, \mu)$ que satisface:*

*(1) $A \cong A'$ si y sólo si $\theta_A \cong \theta_{A'}$, y*
*(2) si $\mathcal{A}$ es una colección de grupos abelianos infinitos numerables sin torsión tales que $\{\theta_A\colon A \in \mathcal{A}\}$ son todas orbitalmente equivalentes y dos a dos no conjugadas, entonces $\mathcal{A}$ es numerable.*

El teorema anterior implica que existe una función de Borel, de la relación de isomorfismo de grupos abelianos sin torsión, a la relación de equivalencia orbital de acciones libres y ergódicas de $\Gamma$, que identifica a lo sumo una cantidad numerable de grupos no-isomorfos. Usando que la relación de isomorfismo de dichos grupos no es Borel [17, 5], se puede probar que la relación de equivalencia orbital de las acciones de $\Gamma$ tampoco es de Borel.

*Prueba del Teorema 4.6.* Fijamos un subgrupo normal $\Delta$ de $\mathbb{F}_2$ tal que $\mathbb{F}_2/\Delta$ tiene la propiedad (T), de modo que la terna $\Delta \leq \mathbb{F}_2 \leq \Gamma$ tiene la propiedad (T). Sea $A$ un grupo abeliano infinito numerable sin torsión, y sea $(G, h_G)$ su grupo dual munido con la medida de Haar. Observemos que $(G, h_G)$ es el espacio de probabilidad estándar sin átomos. Sea $M = L^\infty(G, h_G)$, y denotemos por $\mathtt{Lt}\colon G \to \mathrm{Aut}(M)$ la acción inducida por la traslación por izquierda de $G$ en sí mismo. Consideramos las siguientes acciones:

$$\Gamma \curvearrowright^\beta M^{\otimes \Gamma/\Delta} \otimes M^{\otimes \Gamma} \curvearrowleft^\alpha G,$$

donde $\beta$ es el producto tensorial del shift de Bernoulli, y $\alpha$ es el producto tensorial de $\mathtt{Lt}^{\otimes \Gamma/\Lambda}$ y $\mathrm{id}_{M^{\otimes \Gamma}}$.

Sea $M_A$ el álgebra de puntos fijos de $\alpha$. Como $\alpha$ y $\beta$ conmutan, deducimos que $\beta$ induce una acción $\beta_A\colon \Gamma \curvearrowright M_A$.

Definimos $\theta_A = \beta_A \otimes \widehat{\rho}$, que es una acción de $\Gamma$ en el espacio de probabilidad estándar sin átomos. Es libre porque tiene la acción libre $\beta_A$ como factor. Usando que $\Gamma/\Delta$ es infinito, y propiedades de $\widehat{\rho}$, uno puede mostrar que $\theta_A$ es ergódica.

**Afirmación:** *hay un isomorfismo de grupos $H^1_{:\Delta,\mathbb{F}_2,w}(\theta_A|_{\mathbb{F}_2}) \cong A$.* Sea $u$ un cociclo $\Delta$-invariante para $\theta_A|_{\mathbb{F}_2}$. Entonces $u$ es también un cociclo $\Delta$-invariante para $\beta|_{\mathbb{F}_2} \otimes \widehat{\rho}|_{\mathbb{F}_2}$. Además, como $\widehat{\rho}|_{\mathbb{F}_2}$ es esencialmente una amplificación de $\rho$, el Teorema 4.4 se aplica a ella. Deducimos entonces que $u$ es un coborde para $\beta|_{\mathbb{F}_2} \otimes \widehat{\rho}|_{\mathbb{F}_2}$,

por lo que existe un unitario $v \in \mathcal{U}(M^{\otimes \Gamma/\Delta})$ que implementa la trivialidad. Usando ergodicidad, uno muestra que existe un carácter $\chi_u \in \widehat{G} \cong A$ que satisface

$$\alpha_g(z) = \chi_u(g)z$$

para todo $g \in G$. El resto de la prueba de la afirmación consiste en mostrar que la asignación $[u] \mapsto \chi_u$ es un isomorfismo de grupos.

Deducimos entonces que si $A$ y $A'$ son grupos abelianos infinitos sin torsión, entonces $\theta_A \cong \theta_{A'}$ si y sólo si $A \cong A'$. Esto prueba la parte (1) del Teorema 4.6, y también muestra que la relación de conjugación de acciones libres y ergódicas de $\Gamma$ no es Borel.

La segunda parte del teorema es considerablemente más técnica, y hace uso esencial de las propiedades de rigidez de $\rho$. Por razones de espacio, omitimos su prueba. □

La noción de propiedad (T) para una terna de grupos es nueva, y es probablemente la técnica más innovadora de nuestro trabajo. Su uso nos permite tener acceso a resultados de superrigidez (como en el Teorema 4.4) para acciones de grupos libres, mientras que los resultados anteriores acerca de rigidez no se aplican a ellos. Tal vez por esta razón, las pruebas de los resultados en [10, 18, 7] sólo producen una cantidad no numerable de acciones.

Hemos esbozado la prueba del Teorema 4.6 (y por lo tanto del Teorema 3.9) en el caso en que $\Gamma$ contiene a $\mathbb{F}_2$. En vista de los ejemplos encontrados en [20], esto no es suficiente, así que procedemos a considerar el caso general.

**Segundo paso:** $\Gamma$ es un grupo no amenable arbitrario. Como discutimos en la primer sección, la conjetura de von Neumann (Pregunta 2.7) es falsa, y existen grupos no amenables que no contienen a $\mathbb{F}_2$. Sorprendentemente, Gaboriau y Lyons demostraron en [9] que dicha conjetura tiene una respuesta afirmativa en el contexto "dinámico-medible":

**Teorema 4.7** (Gaboriau-Lyons [9], 2009)**.** *Si $\Gamma \curvearrowright (X, \mu)^\Gamma$ es el shift de Bernoulli de un grupo no amenable $\Gamma$, entonces existe una acción libre y ergódica $\mathbb{F}_2 \curvearrowright (X, \mu)^\Gamma$ tal que $\mathbb{F}_2 \cdot x \subseteq \Gamma \cdot x$ para casi todo $x \in X^\Gamma$.*

El resultado anterior nos da una inclusión de las relaciones de equivalencia orbital, concebidas como subconjuntos del espacio producto. El ámbito apropiado para utilizar esto es el de los grupoides étales: en este contexto, desarollamos una noción de propiedad (T) para ternas de grupoides y probamos un resultado de superrigidez análogo al Teorema 4.4. El proceso de coinducción se efectúa al nivel de las relaciones de equivalencia, interpretadas como grupoides. Esta estrategia requiere una cantidad importante de trabajo adicional, el cual omitimos aquí.

El contexto abstracto en el que trabajamos nos permite probar resultados mucho más generales, que en particular se aplican a acciones de grupos que no son necesariamente discretos:

**Teorema 4.8.** *Sea $G$ un grupo localmente compacto, unimodular y no amenable. Entonces las relaciones de conjugación y equivalencia orbital de acciones libres y ergódicas de $G$ no son Borel.*

Para equivalencia orbital, el caso de grupos amenables tiene el mismo resultado que el caso discreto, y esto fue probado por Connes-Feldman-Weiss en [3]. Usando un resultado reciente de Bowen-Hoff-Ioana en [2], nuestra prueba del Teorema 4.8

consiste esencialmente en reducirnos al caso de un grupo discreto que contiene a $\mathbb{F}_2$. En otras palabras, obtenemos el resultado en el caso de grupos localmente compactos "casi" como un corolario del resultado para grupos discretos.

Quedan muchas preguntas interesantes por responder. Por ejemplo:

**Pregunta 4.9.** *¿Es Borel la relación de conjugación de acciones libres y ergódicas de un grupo discreto infinito?*

Para $\mathbb{Z}$ y para grupos no amenables, la respuesta es *no*, pero la pregunta está abierta para todos los demás grupos, y se espera que la respuesta sea también *no* en general.

Los resultados descritos en estas notas se refieren a acciones por automorfismos de espacios de probabilidad. Sería interesante explorar versiones topológicas de algunos de ellos, por ejemplo del Teorema 3.8. No está claro cuál es la relación de equivalencia apropiada (¿equivalencia orbital continua?), ni tampoco el espacio topológico que deba reemplazar al espacio de probabilidad estándar (¿el espacio de Cantor?). El trabajo de Giordano-Putnam-Skau [14] sobre homeomorfismos minimales de espacios de Cantor puede ofrecer pistas al respecto.

Nuestra motivación original para probar el Teorema 3.9 proviene del álgebra de operadores, más precisamente de las $C^*$-álgebras. En este contexto, parece existir un resultado análogo al Corolario 3.10, explícitamente enunciado en la Conjetura A en [11], que reproducimos a continuación (las definiciones relevantes pueden encontrarse en [11]):

**Conjetura 4.10** (G.-Lupini [11], 2016)**.** *Sea $\Gamma$ un grupo discreto numerable sin torsión, y sea $\mathcal{D}$ una $C^*$-álgebra fuertemente autoabsorbente.*

*(1) Si $\Gamma$ es amenable, entonces dos acciones fuertemente externas de $\Gamma$ en $\mathcal{D}$ son cociclo-equivalentes.*
*(2) Si $\Gamma$ no es amenable, entonces la relación de cociclo-equivalencia de acciones fuertemente externas de $\Gamma$ en $\mathcal{D}$ no es Borel.*

El trabajo [13] comenzó como un intento de mejorar nuestro entendimiento de los resultados en el caso conmutativo, para luego poder hacer avances en dicha conjetura. Una solución satisfactoria al siguiente problema será instrumental para probar la parte (2).

**Problema 4.11.** *¿Existe un análogo al Teorema 4.7 en el contexto de las acciones fuertemente externas de grupos no amenables en álgebras fuertemente autoabsorbentes? ¿Y en el contexto de acciones externas en el factor hiperfinito de tipo $II_1$?*

## Referencias

[1] L. P. Bowen, *A measure-conjugacy invariant for free group actions*, Ann. of Math. (2), 171 (2010), pp. 1387–1400.
[2] L. Bowen, D. Hoff, and A. Ioana, *von Neumann's problem and extensions of non-amenable equivalence relations*, Groups, Geometry, and Dynamics, to appear.
[3] A. Connes, J. Feldman, and B. Weiss, *An amenable equivalence relation is generated by a single transformation*, Ergod. Th. Dynam. Sys. 1 (1981), pp. 431–450.
[4] A. Connes and B. Weiss, *Property T and asymptotically invariant sequences*, Israel J. Math. 37 (1980), no. 3, pp. 209–210.
[5] R. Downey and A. Montalbán, *The isomorphism problem for torsion-free abelian groups is analytic complete*, J. Algebra 320 (2008), no. 6, pp. 2291–2300.

[6] H. A. Dye, *On groups of measure preserving transformations. II*, Amer. J. Math., 85 (1963), pp. 551–576.
[7] I. Epstein, *Orbit inequivalent actions of non-amenable groups* (2007) Preprint, arXiv:0707.4215.
[8] M. Foreman, D. Rudolph, and B. Weiss, *The conjugacy problem in ergodic theory*, Ann. Math. 173 (2011), no. 3, pp. 1529–1586.
[9] D. Gaboriau and R. Lyons, *A measurable-group-theoretic solution to von Neumann's problem*, Invent. Math., 177 (2009), pp. 533–540.
[10] D. Gaboriau and S. Popa, *An uncountable family of nonorbit equivalent actions of $\mathbb{F}_n$*, J. Amer. Math. Soc., 18 (2005), pp. 547–559.
[11] E. Gardella and M. Lupini, *Actions of rigid groups on UHF-algebras*, J. Funct. Anal., to appear (2018).
[12] E. Gardella and M. Lupini, *On the classification problem of free ergodic actions of nonamenable groups*, C. R. Math. Acad. Sci. Paris 355 (2017), no. 10, pp. 1037–1040.
[13] E. Gardella and M. Lupini, *The Borel complexity of conjugacy, orbit equivalence, and von Neumann equivalence of actions of nonamenable groups*, (2017). Preprint, arXiv:1708.01327.
[14] T. Giordano, I. Putnam, and C. Skau, *Topological orbit equivalence and $C^*$-crossed products*, J. Reine Angew. Math., 469 (1995), pp. 51–111.
[15] P. R. Halmos, *Lectures on ergodic theory*, Chelsea Publishing Co., New York, 1960.
[16] G. Hjorth, *A converse to Dye's theorem*, Trans. Amer. Math. Soc., 357 (2005), pp. 3083–3103.
[17] G. Hjorth, *The isomorphism relation on countable torsion free abelian groups*, Fund. Math. 175 (2002), no. 3, pp. 241–257.
[18] A. Ioana, *Orbit inequivalent actions for groups containing a copy of $\mathbb{F}_2$*, Invent. Math., 185 (2011), pp. 55–73.
[19] A. S. Kechris, *Global aspects of ergodic group actions*, vol. 160 of Mathematical Surveys and Monographs, American Mathematical Society, Providence, RI, 2010.
[20] A. J.Ol′šanskiĭ, *On the question of the existence of an invariant mean on a group*, Uspekhi Mat. Nauk, 35 (1980), pp. 199–200.
[21] D. S. Ornstein and B. Weiss, *Equivalence of measure preserving transformations*, Mem. Amer. Math. Soc., 37 (1982), pp. xii+116.
[22] A. L. T. Paterson, *Amenability*, Mathematical Surveys and Monographs, 29. American Mathematical Society, Providence, RI, 1988.
[23] I. M. Singer, *Automorphisms of finite factors*, Amer. J. Math., 77 (1977), pp. 117–133.

Westfälische Wilhelms-Universität Münster, Fachbereich Mathematik.
Einsteinstrasse 62, 48149 Münster, Alemania.
*Email address*: gardella@uni-muenster.de
*URL*: http://wwwmath.uni-muenster.de/u/gardella/

PONENCIA

# TRANSFORMACIONES EN LA ENSEÑANZA DE LA MATEMÁTICA PARA ESTUDIANTES INGRESANTES A LA FACULTAD DE ARQUITECTURA, DISEÑO Y URBANISMO

OMAR GIL

Publicaciones Matemáticas del Uruguay
Volumen 17, Julio 2019, Páginas 167–179
S 0797-1443

# TRANSFORMACIONES EN LA ENSEÑANZA DE LA MATEMÁTICA PARA ESTUDIANTES INGRESANTES A LA FACULTAD DE ARQUITECTURA, DISEÑO Y URBANISMO

OMAR GIL

## 1. Resumen

Presentaremos una estrategia de transformación de la enseñanza aplicada a partir de 2013, en el curso de Matemática de la carrera de Arquitectura en la Facultad de Arquitectura Diseño y Urbanismo de la Universidad de la República, articulada en torno a los siguientes ejes

- organización del aula en equipos de trabajo y enseñanza activa;
- aplicación de un sistema de evaluación con instancias individuales, grupales, autoevaluación y evaluación por pares, que ofrece a los estudiantes oportunidades de revisión de su trabajo;
- contextualización de la Matemática en la formación del estudiante;
- atención a los aspectos afectivos y de vínculo en clase;
- sistematización de la gestión y organización de los cursos;
- diversificación de la oferta de actividades;
- retroalimentación de la toma de decisiones a partir del seguimiento de los cursos, el análisis de las características de la población estudiantil y la evaluación de las políticas implementadas.

Entre 2013 y 2016 se constató una importante mejora del rendimiento de los estudiantes, evidenciada en la reducción del tiempo empleado en aprobar el curso de Matemática previsto en el plan de estudios vigente hasta 2016. Por otra parte, en las entrevistas realizadas, los estudiantes expresan satisfacción con los dispositivos grupales y por la mejor vinculación de los temas del curso con otros contenidos de su formación.

En 2017 se implementó un nuevo plan de estudios en la carrera de Arquitectura, que permitió profundizar la estrategia de transformaciones a través de una organización curricular en la que ningún curso de Matemática es obligatorio, pero los estudiantes deben completar seis créditos en esta disciplina, escogiendo al menos uno de tres cursos con objetivos generales comunes pero objetivos específicos y contenidos netamente diferenciados. Los datos disponibles son preliminares y contaminados por contingencias relativas a la aplicación del nuevo plan, pero aún así muestran algunos rasgos alentadores.

Entre los problemas que persisten, destacamos el escaso rendimiento comparativo de la estrategia de volver a cursar cuando la aprobación no se logra en el primer intento. Algunas indagaciones exploratorias sugieren que este fenómeno podría relacionarse con creencias y estrategias de aprendizaje inadecuadas, basadas en

©2019 PMU

concebir la Matemática como una actividad reproductora, en la que el estudiante es instruido en técnicas para la resolución de ejercicios. Por otra parte, experiencias iniciales de orientación de estudiantes realizadas con ingresantes en 2013 y 2015, sugieren que constituyen una población abierta a considerar itinerarios de formación diferentes a los establecidos en las opciones por defecto de los planes de estudio, circunstancia que abre un interesante horizonte de posibilidades a explorar. Sobre el final de esta presentación discutiremos dos propuestas que apuntan a atender estos problemas:

- la creación de un sistema de orientación, ordenamiento y seguimiento de los estudiantes en sus cursos iniciales de Matemática;
- el desarrollo sistemático de acciones de tutoría entre pares para apoyar a estudiantes ingresantes.

En este proceso, la Cátedra buscó un equilibrio entre la aplicación contextualizada de estrategias de enseñanza bien establecidas en la literatura [2, 3] y la identificación de problemas que ameriten el desarrollo de líneas de investigación propias [13]. Este trabajo es una puesta al día de [7].

## 2. Período 2013–2016

Entre 2013 y 2016 se aplicaron paulatina y sostenidamente una serie de reformas informadas por principios bien establecidos en la literatura sobre enseñanza y aprendizaje en ambientes universitarios, por la contextualización de grandes ideas de la disciplina en la formación de futuros arquitectos y por la retroalimentación del proceso a través del seguimiento de los cursos y el análisis de la actividad de los estudiantes. En paralelo, en el contexto de masividad moderada de este curso, se atendió a sus aspectos organizativos apuntando a hacer un uso eficiente de los recursos disponibles y a establecer canales claros y fluidos de comunicación con los estudiantes. El aula se organizó recurriendo a estrategias de enseñanza activa e instrucción entre pares, apuntando especialmente a fortalecer los vínculos que los estudiantes establecen entre sí, con los docentes y con el conocimiento. Se modificó el sistema de evaluación, explicitando tanto su rol formativo como de certificación e incluyendo instancias individuales y grupales que brindaran a los estudiantes oportunidades de revisión y perfeccionamiento de su trabajo. También se avanzó en la documentación del curso, produciendo un libro de texto, registrando en video muchas de las secciones y enriqueciendo la oferta de actividades a través de la plataforma de Moodle de la FADU. Al tiempo que se desarrollaba este programa, la Cátedra ha ido identificando problemas que justifiquen el desarrollo de programas de investigación y oportunidades para propuestas de extensión. Sin embargo, no fue esta la principal prioridad durante el período analizado, ya que la Cátedra se centró en reformar su oferta de actividades de enseñanza. Estas iniciativas han tenido impactos observables sobre el desempeño de los estudiantes en los cursos de Matemática, que se reportan en la sección 3.

Hasta el año lectivo de 2016 la carrera de Arquitectura estuvo organizada por un plan de estudios del año 2002, que preveía un curso de Matemática de 11 créditos (165 horas totales de trabajo, incluyendo el tiempo de clase, estudio y tareas domiciliarias), obligatorio, ubicado en el primer semestre de la carrera. El curso abordaba una introducción al cálculo diferencial e integral de una y dos variables y a la geometría del espacio con técnicas vectoriales. Aproximadamente la mitad del tiempo de trabajo era presencial y la otra mitad domiciliario. El curso

tenía dos modalidades de aprobación, un curso controlado de duración semestral, que se dictaba en los dos semestre lectivos de cada año, y un curso libre, que se reducía a la aprobación de un examen. En este trabajo nos referiremos a esta dos modalidades como "curso" y "examen" respectivamente. El curso y el examen eran instancias completamente independientes entre sí, en el sentido de que no se requería ningún tipo de ganancia de curso para estar habilitado a rendir el examen. En el primer semestre de cada año el curso recibía a la generación de ingreso, de unos 600 estudiantes, y el segundo semestre se dedicaba mayoritariamente a estudiantes de generaciones anteriores, típicamente recursantes. A partir de 2014, a esta oferta se agregó la de un curso intensivo de verano, con prácticamente todas sus horas de trabajo presenciales. La demanda total para el curso de Matemática cada año es del orden de las mil plazas.

Hacia el año 2010 diversos estudios de la Facultad coincidían en diagnosticar que la unidad curricular Matemática tenía problemas en la definición de sus objetivos, en la integración de sus contenidos con el resto de la carrera y en el rendimiento de los estudiantes [8]. La estrategia desplegada a partir de 2012, fundamentalmente entre 2013 y 2016, apuntó a corregir estos problemas. También se debió atender la masividad moderada que implican la relación entre el número de estudiantes que la Cátedra debe atender y los recursos docentes disponibles. A comienzos de 2013 el equipo docente de la Cátedra estaba conformado por un profesor titular con dedicación exclusiva, siete asistentes y seis ayudantes con cargos de doce horas semanales de dedicación.

A fines de 2017 el equipo de la Cátedra está integrado por el mismo profesor titular y seis asistentes y tres ayudantes y ha sido designada por el Consejo de Facultad una profesora adjunta. Los dos profesores de la Cátedra tienen doctorados en Matemática. Los asistentes tienen títulos de grado, tres de ellos en Arquitectura (en uno de los casos todavía no se ha alcanzado la titulación, pero los estudios están casi completados), uno en Ingeniería, uno en Matemática y uno como Profesor de Matemática. Estos dos últimos están cursando posgrados. Los ayudantes son estudiantes de Matemática, de Profesorado en Matemática y de Ingeniería en Computación. Los miembros del equipo tienen formación en educación o especial sensibilidad hacia el tema, una característica compartida también por otros docentes que integraron la Cátedra en el período 2013-2016. Estos rasgos hicieron posible promover y sostener innovaciones importantes en las propuestas de enseñanza, sobre un esquema de organización coordinado centralmente, que buscó ser respetuoso de los niveles de responsabilidad y dedicación que correspondía esperar de cada miembro del equipo, pero al mismo tiempo requirió de todos ellos un alto nivel de compromiso para poder aplicarse.

El objetivo general que orientó este proceso fue contribuir a la mejora continua de los aprendizajes en cursos de Matemática de nivel universitario. Los objetivos específicos que nos planteamos son:

- agregar valor a la formación de estudiantes de Arquitectura promoviendo aprendizajes de Matemática contextualizados y en tiempos acordes con los establecidos por los planes de estudio;
- desarrollar estrategias de enseñanza adecuadas a contextos de masividad;
- desarrollar estrategias docentes que apoyen el tránsito de los estudiantes entre las etapas secundaria y terciaria de sus estudios;

- crear un equipo docente interdisciplinario capaz de promover procesos de transformación en la propuesta de enseñanza;
- sostener un proceso continuo de la mejora de la enseñanza, informado por el estado del arte en Ciencias de la Educación e identificar lineas de investigación;
- generar propuestas transferibles, con la adecuada contextualización, a otros equipos docentes y servicios universitarios.

Con esta inspiración, la Cátedra ha desarrollado una línea de trabajo orientada por los principios de la investigación-acción, en el sentido de una forma de indagación introspectiva colectiva emprendida por participantes en situaciones sociales con el objeto de mejorar la racionalidad y justicia de sus prácticas sociales o educativas, así como la comprensión de esas prácticas y de las situaciones que en estas tienen lugar [11], con un fuerte énfasis hacia la acción orientada por el conocimiento disponible y la experiencia acumulada.

La Cátedra articuló sus intervenciones en una serie de ejes temáticos, que atienden a factores que, a partir del análisis teórico de la situación o de los diagnósticos efectuados, identificamos como importantes para potenciar los aprendizajes en nuestro contexto. En paralelo, se aplicaron mecanismos de seguimiento con un foco relativamente amplio, orientados en primera instancia a captar emergentes que guiaran la toma de decisiones, para ir refinando luego gradualmente la atención y convergiendo hacia núcleos de problemas que ameritaran el desarrollo de programas de investigación. Estos grandes lineamientos se han mantenido tanto en el período de aplicación del plan de estudios 2002 como durante 2017, en que comenzó a implementarse el plan de estudios 2015.

**2.1. Organización del aula en equipos de trabajo y estrategias de enseñanza activa.** La teoría y la evidencia indican que la enseñanza activa y la instrucción entre pares promueven el aprendizaje [5, 15, 16]. A partir de 2013 comenzó un proceso de reorganización del aula que facilita la aplicación de estas estrategias y es compatible con las condiciones de moderada masividad existentes. Se adoptó una modalidad de trabajo en grupos de carácter teórico-práctico, de entre 90 y 126 estudiantes, a cargo de uno o dos docentes acompañados de colaboradores honorarios, en los que los estudiantes son asignados por los docentes a equipos de trabajo integrados por entre cinco y siete estudiantes, que se mantienen a lo largo de todo el curso. Con esta estructuración del aula, la actividad va pasando por diversas modalidades de trabajo: exposición del docente, resolución de ejercicios, proposición de una situación problemática, abordaje de un problema a través de un juego, etcétera, que a su vez pueden alternar momentos de trabajo individual, en equipo y con la totalidad del grupo.

**2.2. Atención a los vínculos de los estudiantes con el conocimiento, con sus pares y con los docentes.** Aunque mientras estuvo vigente el plan 2002 el programa del curso Matemática no se modificó, la presentación fue evolucionando hacia una mayor contextualización en la formación de los estudiantes, organizando los contenidos alrededor de problemas relevantes para los estudios de Arquitectura y la práctica profesional de los arquitectos. En paralelo, buscando también promover un vínculo positivo de los estudiantes con la disciplina Matemática, se buscó explicitar su rol como modo abstracto de representación, capaz de apoyar un proceso de diseño generando variabilidad y ayudando a tomar decisiones sobre objetos aún no existentes. Existe abundante evidencia acerca de que el clima de

clase afecta los aprendizajes. Ver, por ejemplo, el capítulo 6 de [2] y sus referencias. La Cátedra desarrolla sistemáticamente acciones para personalizar la relación de los docentes con sus estudiantes y atenuar, o directamente eliminar, en nuestras aulas la despersonalización asociada a la masividad. En particular, se han desarrollado estretegias para que los docentes sean capaces de identificar a todos sus estudiantes por sus nombres. Cada grupo tiene asociado un grupo en la red social Facebook como ámbito de interacción complementario y se estimula adeAal2010más a los docentes a promover otras instancias de socialización que enriquezcan el clima de clase y los aprendizajes. Una política activa de captación de colaboradores honorarios, permite contar en el aula y el trabajo a través de las redes sociales con pares que actúan, en todos los aspectos, como dinamizadores del trabajo de los estudiantes, los equipos y los grupos. Se apunta así a la conformación de una comunidad de aprendizaje, que contribuye al desarrollo de la responsabilidad individual y las habilidades de cooperación y comunicación, valiosas no solo para el curso de Matemática sino para la formación integral del estudiante.

**2.3. Aplicación de un sistema de evaluación con instancias individuales, grupales, autoevaluación y evaluación por pares.** La práctica orientada por objetivos y evaluada con instrumentos que brinden al estudiante devoluciones oportunas promueve el aprendizaje. Ver, por ejemplo el capítulo 5 de [2] y sus referencias. El sistema de evaluación para la acreditación del curso tiene tres componentes:

- pruebas parciales;
- pruebas cortas en clase;
- evaluación entre pares y autoevaluación.

Las pruebas en clase brindan a los estudiantes puntos de apoyo intermedio en su proceso. Combinan instancias individuales y grupales y comienzan a aplicarse tempranamente en el curso, típicamente en su segunda semana. Habitualmente se toman seis de estas pruebas y tienen el 36 % del peso de la calificación final. En algunas oportunidades se han sustituido pruebas por trabajos en equipo. Los parciales son la componente de mayor peso en la evaluación, un 60 % de la nota final. Cada uno de ellos está acompañado de una prueba de recuperación, a la que los estudiantes pueden acceder solo después de realizar una autocorrección de su trabajo, mediada por su equipo, que incluye la resolución correcta por los equipos de la prueba original. Este mecanismo apunta a capitalizar en aprendizajes el importante esfuerzo que los estudiantes hacen en las pruebas que tienen peso para la certificación del curso, trabajando en un espacio en que, en interacción con sus pares, puedan prestar atención a sus procesos, construir autonomía para su trabajo y valoración de las instancias de evaluación como un recurso para su crecimiento intelectual. La autoevaluación y la evaluación por pares tienen un menor peso relativo para la acreditación, el 4 %, pero permiten explicitar en base a rúbricas las actitudes de responsabilidad individual y frente a los compañeros que entendemos adecuadas para el aprendizaje. Se espera que además potencien el trabajo grupal y se refleje, por lo tanto, en todas las otras instancias de evaluación.

El trabajo en equipo realimenta el trabajo de cada estudiante con información sobre sus progresos, pero además en el aula se promueven otras formas de socializar respuestas, siguiendo los modelos de instrucción por pares [5] o de intercambio a partir de resultados de equipos [12]. Fuera del aula, la Cátedra ha implementado cuestionarios con realimentación a través de la plataforma Moodle de la FADU.

En todas estas acciones se pone especial atención al manejo de los aspectos afectivos y vinculares, explicitando que todas las respuestas y todos los aportes son valiosos y contribuyen a la construcción de aprendizajes.

Todos los instrumentos de evaluación ofrecen devoluciones en tiempos adecuados: la actividad de clases produce una realimentación instantánea y las pruebas se diseñan para tener una corrección rápida, que permite informar a los estudiantes sobre sus resultados en cuanto se agota el tiempo que se les concede para la revisión en equipo de sus procesos.

Se espera también que los estudiantes puedan apreciar el potencial de las instancias de evaluación como apoyo en su proceso de aprendizaje y la intención del equipo docente de que este proceso sea exitoso, al tiempo que se les brinda chances de mostrar y certificar sus progresos, lo que contribuye a mantener la motivación y la dedicación a los cursos.

**2.4. Sistematización de la gestión y organización de los cursos.** Se ha creado un conjunto de notas que evolucionó hasta constituir un primer borrador de un libro para el curso, se han documentado en videos muchos contenidos y enriquecido el espacio de trabajo en la plataforma Moodle de la FADU. En particular, se han creado bancos de preguntas que facilitan el trabajo de los docentes al preparar evaluaciones y permiten construir cuestionarios que los estudiantes pueden usar para medir sus progresos, procesos que están en etapa de automatización [6].

En materia de gestión, la Cátedra ha adoptado una modalidad de organización y comunicación centralizada, que busca tanto crear una infraestructura adecuada para los cursos como comunicar de manera organizada a los estudiantes toda la información sobre las actividades de enseñanza de la Cátedra, con modos y tonos que aporten a la construcción del vínculo entre estudiantes y docentes.

**2.5. Diversificación de la oferta de actividades.** En 2013 se ofreció un recorrido especial, anual, para el curso. Esta experiencia tuvo resultados muy pobres (ver la tabla 1), por lo que no se reiteró. Al cierre del año lectivo de 2013 la Cátedra ofreció durante el receso veraniego una versión intensiva del curso de Matemática, con un formato íntegramente presencial. Esta experiencia tuvo resultados muy positivos, reflejados tanto en los números de aprobación como en las devoluciones de los estudiantes y la percepción de los equipos docentes, y se repitió en los veranos de 2015 y 2016. No se reiteró en el tramo final del año lectivo de 2016 porque la FADU estaba inmersa en la instrumentación del nuevo plan de estudios 2015.

La Cátedra ha tomado también algunas iniciativas de orientación de los estudiantes, buscando mejorar el tránsito por el primer semestre de sus estudios universitarios y el aprovechamiento de sus recursos. En 2013 y 2015 se habilitaron períodos especiales de desistimiento, luego del primer parcial y su recuperación. Las conclusiones preliminares sugieren que estos mecanimos tienen potencial para ordenar mejor los esfuerzos [4]. Se han implementado también acciones para apoyar en distintos momentos a los estudiantes en materia de requisitos previos para los cursos.

**2.6. Seguimiento de los cursos.** La Cátedra reservó parte de sus recursos para el seguimiento de sus cursos, por medio de observaciones de clase y entrevistas a grupos de estudiantes [14]. Entre 2013 y 2014 las entrevistas se acompañaron de observaciones participantes [10]. Ambas acciones se orientaron fundamentalmente a relevar el clima de clase y la calidad de la actividad de los estudiantes en sus equipos de trabajo. En 2015 y 2016 se aplicaron entrevistas para indagar acerca de los recursos

de los estudiantes para abordar el curso de Matemática y sobre sus creencias acerca del aprendizaje de esta disciplina. Estos dispositivos complementaron los sistemas de encuestas que la FADU aplica y el análisis de los datos numéricos, lo que ofreció la oportunidad de enriquecer la mirada del mismo fenómeno educativo a través de diferentes instrumentos [1]. Los resultados se socializaron fundamentalmente a través de las reuniones periódicas del equipo docente.

**2.7. Desarrollo del equipo de trabajo.** Se ha cuidado la conformación de un equipo de trabajo con personas de diferentes formaciones y el mantenimiento de reuniones periódicas que combinan características de coordinación y de seminario. También se ha promovido la asistencia colectiva a cursos, jornadas de prefeccionamiento y congresos, y la posterior socialización de la experiencia. Los procesos seguidos al interior del colectivo docente están orientados por los mismos principios que pretendemos aplicar para estimular el crecimiento intelectual de nuestros estudiantes, en un entorno que tiene en consideración los aspectos afectivos y de vínculo. Ver las conclusiones finales de [2].

## 3. Resultados

Los resultados de estas intervenciones se han reflejado en los índices de aprobación de los cursos y aparecen en las entrevistas en forma de expresiones de satisfacción con la organización grupal, la contextualización de la actividad matemática y el apoyo que ofrecen los equipos docentes, tanto a través de la aplicación de un sistema de evaluación que ofrece contención como del cuidado de los aspectos afectivos y vinculares. Los datos de inscripciones y aprobaciones entre 2008 y 2016 se muestran en la tabla 1.

| AÑO | Semestre 1 | | | Semestre 2 | | | Curso anual | | | Verano | | | Totales | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | I | A | % | I | A | % | I | A | % | I | A | % | I | A |
| 2008 | 368 | 89 | 24 | 208 | 115 | 55 | | | | | | | 576 | 204 |
| 2009 | 387 | 134 | 35 | 185 | 68 | 37 | - | - | - | - | - | - | 572 | 202 |
| 2010 | 370 | 142 | 38 | 190 | 53 | 28 | - | - | - | - | - | - | 560 | 195 |
| 2011 | 398 | 131 | 33 | 188 | 69 | 37 | - | - | - | - | - | - | 586 | 200 |
| 2012 | 389 | 149 | 38 | 208 | 100 | 48 | - | - | - | - | - | - | 597 | 249 |
| 2013 | 403 | 230 | 57 | 416 | 152 | 37 | 81 | 10 | 12 | 112 | 63 | 56 | 1012 | 455 |
| 2014 | 552 | 280 | 51 | 245 | 60 | 24 | - | - | - | 144 | 123 | 85 | 941 | 463 |
| 2015 | 398 | 231 | 58 | 198 | 87 | 44 | - | - | - | 189 | 120 | 63 | 785 | 438 |
| 2016 | 579 | 251 | 43 | 215 | 72 | 33 | - | - | - | - | - | - | 794 | 232 |

Tabla 1. Estudiantes inscriptos (I) y aprobados (A) por semestre, entre 2008 y 2016

Los primeros semestres se distinguen de los segundos, ya que las poblaciones de ingresantes y recursantes no son comparables. En el año 2013 se incluyen también los datos del curso anual y entre 2013 y 2015 los de los cursos de verano. La primera columna de cada categoría recoge el número de estudiantes inscriptos al curso que rindieron el primer parcial. La segunda, el número de estudiantes que aprobaron. La tercera el porcentaje que estos últimos representan sobre los primeros. La inscripción inicial en cada curso es algo superior al dato que aparece en la tabla. La deserción entre el inicio de los cursos y el primer parcial explica esta diferencia, que sitúa el número anual de plazas solicitadas en el orden de las mil. A partir de 2012 se observa una mejora en los resultados de los cursos de primer semestre. Los mecanismos de desistimiento aplicados en 2013 y 2015 llevaron las tasas de

aprobación a niveles próximos al 60 %. También los cursos de verano muestran muy buenos rendimientos. Estas iniciativas diferentes, de algún modo complementarias, impactaron en el rendimiento de los cursos. Los datos muestran un incremento del número de aprobaciones entre 2013 y 2015, para volver luego en 2016 a valores similares a los de 2012. Este comportamiento puede explicarse en términos de la reducción de los tiempos que los estudiantes requieren para aprobar Matemática. Entre 2013 y 2015 se aprecia la transición entre estas dos situaciones, que las figuras 1 y 2 ilustran. La abcisa de cada punto de la figura 1 representa, para las generaciones 2003, 2004, 2005, 2013, 2014 y 2015, el porcentajes de estudiantes de la generación que aprobó Matemática hasta un determinado momento. En la ordenada se representa el tiempo medio de aprobación para ese conjunto de estudiantes, medido en años. El gráfico se ha completado con trazos continuos.

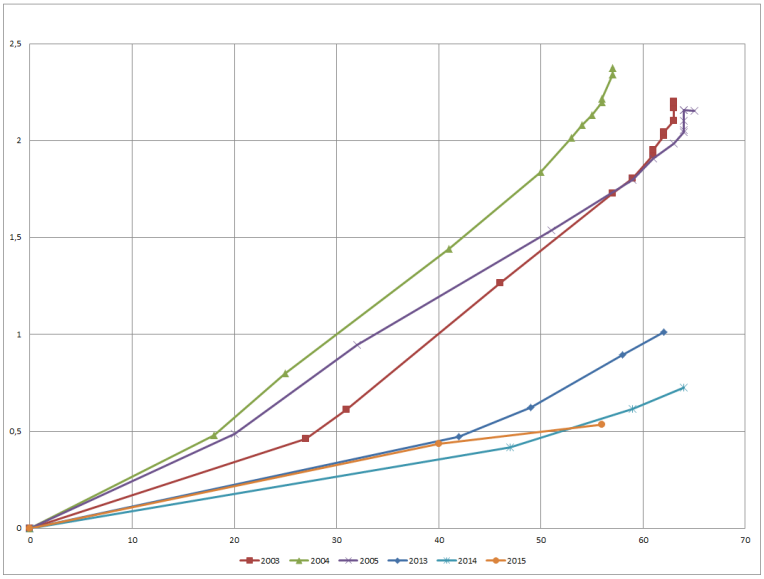



FIGURA 1: tiempos medios de aprobación en función del porcentaje de aprobados.

La figura 2 contiene un extracto de los datos de la figura 1, en un formato que facilita su visualización: las barras indican el tiempo medio de aprobación del 44 % de los estudiantes de cada generación que aparece identificada en el eje horizonal.

Los rendimientos promedio han mejorado, pero las instancias posteriores de cursado continúan mostrándose ineficientes. Una hipótesis explicativa posible es que la población recursante está mayoritariamente integrada por estudiantes regidos por un sistema de creencias que no favorece el aprendizaje y que fue detectado y descrito en [13], por lo que es de presumir que requiera acciones diferentes a las desplegadas hasta el momento. La validez de esta hipótesis y el diseño de estrategias adecuadas para estudiantes que no alcanzan en su primer intento la aprobación del curso, será el objeto de futuros trabajos. En este sentido, las propuestas de la Cátedra han alcanzado ya cierto grado de madurez que justifica comenzar a desarrollar líneas de investigación acerca de sus resultados, los obstáculos para su profundización y la planificación de posibles alternativas.

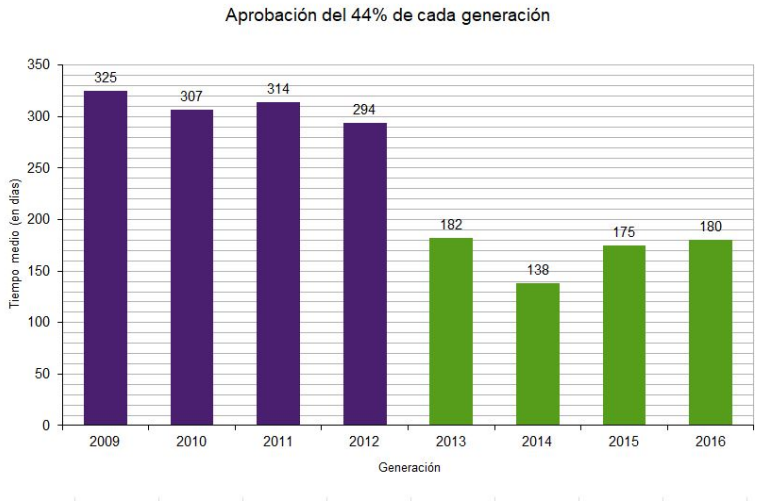



FIGURA 2: tiempos medios de aprobación del 44 % de los estudiantes de cada generación.

## 4. 2017: PRIMER AÑO DE APLICACIÓN DEL PLAN 2015

El plan 2015 tiene una estructura flexible, organizada en torno a tres grandes áreas que hacen a la formación del Arquitecto: Historia, Teoría y Crítica; Proyecto y Representación; Tecnología. El plan prevé mínimos de créditos en cada área que habilitan importantes espacios de opcionalidad y actividades transversales a más de un área. La implementación que se comenzó a instrumentar en 2017 incluyó a la Matemática dentro del área de la Tecnología y redujo los créditos obligatorios de Matemática de los once créditos que tenía el curso de Matemática previsto en el plan 2002 a seis créditos. Al mismo tiempo, se habilitaba la posibilidad de ofrecer créditos optativos. La Cátedra instrumentó el plano ofreciendo tres cursos de seis créditos, con los mismo objetivos generales de aproximación al pensamiento matemático, manejo de estructuras formales, modelización de problemas de otras disciplinas y adquisición de competencias transversales en materia de trabajo colaborativo, y cursos específicos bien diferenciados:

- GEOMETRÍA DE LAS REPRESENTACIONES PLANAS DEL ESPACIO. Este curso trata de la geometría del espacio 3D con técnicas vectoriales y analíticas, que se aplican al estudio de problemas de representación como el análisis de cortes, de distintos tipos de representación en perspectiva y de la proyección estereográfica;
- PRINCIPIOS MATEMÁTICOS DE LA ESTABILIDAD DE LAS COSNTRUCCIONES. Es un curso introductorio al cálculo diferencial de una variable y al cálculo integral en una y dos variables, organizado alrededor del problema de determinar las tensiones ý deformaciones en una viga horizontal con vínculos isostáticos que debe soportar un sistema de cargas verticales, y utilizar este conocimiento para resolver tareas sencillas de dimensionamiento por criterios de tensión y/o de flecha máxima.
- SIMETRÍA. Una aproximación abstracta al estudio de la forma, en la que se caracterizan los distintos tipos de simetría que pueden tener diseños planos

como rosetones, frisos y mosaicos y cuerpos en el espacio como poliedros con algún grado de regularidad. Este conocimiento permite conocer y explotar las posibilidades y restricciones que la estructura matemática implica para este tipo de objetos.

Estos cursos se ditactaron con una mayor carga de horas presenciales: dos de cada tres horas previstas se realizaban en el aula, con la guia de docentes del equipo de la Cátedra y colaboradores honorarios. La aplicación de esta organización de los cursos durante 2017 sufrió las vicisitudes propias de la implementación de un nuevo plan de estudios y algunas particulares que afectaron al colectivo docente de la Cátedra. Sin embargo, los resultados de aprobación fueron mejores a los de años anteriores, alcanzaron en el promedio de todos los cursos al 56 % de los estudiantes inscriptos. Este guarismo sólo se había alcanzado antes en el marco del curso de verano, con todas las particularidades favorables y el esfuerzo adicional que representan, y en los primeros semestre de 2013 y 2015, en que se habilitaron períodos especiales de desistimiento luego del primer parcial [4]. A su vez, los cursos de verano de 2018, en el cierre del año lectivo de 2017, alcanzaron globalmente un rendimiento de 82 % de aprobación. Los datos preliminares son, por lo tanto, alentadores, aunque insuficientes para sacar conclusiones.

La Cátedra amplió además su oferta implementando, en colaboración con PROGRESA, el Programa para el Respaldo de los Aprendizajes de la Comisión Sectorial de Enseñanza dos módulos de Tutoría Entre Pares/Matemática (TEP/Mat 1 y 2) orientados a una tutoría de carácter académico en el contexto de aulas de Matemática organizadas sobre principios de enseñanza activa e instrucción colaborativa entre pares. Estos dos módulos se pueden validar como unidades curriculares optativas, asignan cuatro créditos cada uno y se planea ofrecerlos todos los semstres. TEP/Mat 1 tiene un carácter más teórico y de reflexión sobre la experiencia propia de cada estudiante y se dictó por primera vez el segundo semestre de 2017, con la presencia de doce estudiantes que siguieron los talleres con compromiso y excelente aprovechamiento. TEP/Mat 2 tiene TEP/Mat 1 como requisito previo, además de seis créditos en Matemática y comenzará a dictarse en 2018.

La Cátedra ha integrado el paquete TEP/Mat 1 y 2 a las actividades de formación de su plantel de Estudiantes Colaboradores Honorarios, al tiempo que prevé ampliar su plantel de Colaboradores y dar más difusión a la oferta de los TEP/Mat. En la sección 5 se prefiguran algunas posibles acciones relacionadas con la mayor presencia de estudiantes pares en las actividades de la Cátedra.

## 5. Posibles lineas futuras

Aunque la Cátedra ha ido avanzando sostenidamente en mejorar el rendimiento de sus cursos y la contextualización de sus contenidos en la formación de los estudiantes de Arquitectura, encontramos todavía muchas áreas de oportunidad para nuevas intervenciones. Destacamos dos que entendemos que mereceran atención prioritaria en el corto plazo.

**5.1. Orientación de estudiantes.** Aunque el rendimiento de los estudiantes ha mostrado una mejora significativa que parece consolidada, cuando los resultados de aprobación se miden en términos de la matrícula inicial de los cursos, las cifras de 2017 caen al 48 %. Algunos emergentes sugieren además que cuando los estudiantes no aprueban un curso no alcanzan a valorar los objetivos de aprendizaje

que sí alcanzaron. Indicaciones de este fenómeno aparecen en el pequeño número de estudiantes que optan por rendir examen (sería de esperar que estudiantes con aprovechamiento correcto en el curso culminen pronto su proceso rindiendo examen en la primera ocasión disponible), en un rendimiento en segunda cursadas comparables al de las primeras cursadas, a la persistencia luego de varias cursadas de errores conceptuales en cuestiones básicas y, de manera poco formalizada pero frecuentemente elocuente, en el discurso de los propios estudiantes que relatan cómo se sienten o qué hacen cuando no alcanzan el objetivo de aprobación del curso. En los primeros semestres lectivos de los años 2013 y 2015 la Cátedra realizó la experiencia de ampliar los plazos de desistimiento, para ofrecer a los estudiantes cuyo desempeño indicaba riesgo alto de fracaso la posibilidad de abandonar sin penalización los cursos de Matemática y reorganizar su actividad. El resultado se tradujo en un mejor rendimiento de los cursos de Matemática (ver en la tabla 1 los porcentajes de aprobación de esos cursos) y en un mejor rendimiento de los estudiantes en los cursos que efectivamente siguieron [4].

Esta experiencia, en combinación con el buen desempeño de los cursos de verano, abre la puerta a considerar estrategias de orientación de los estudiantes para que tomen Matemática en el momento más adecuado a sus reales posibilidades de aprovecharla, articulando la siguiente batería de acciones, con el propósito de que tanto la Cátedra como los estudiantes apliquen más eficaz y eficientemente sus recursos y mejorar el desempeño de cada estudiante en particular y del sistema en su conjunto:

- dar orientación temprana a los estudiantes acerca de las posibilidades de desistir de los cursos en su tramo inicial y ajustar su plan de trabajo a sus reales posibilidades;
- dar un tratamiento diferencial a los estudiantes recursantes, que requiera la revisión y puesta al día de las experiencias vividas y los conocimientos adquiridos en cursadas anteriores, que serán evaluados en las etapas iniciales de su incorporación a un nuevo curso;
- mantener activo, con carácter permanente, el programa de cursos de verano, como una opción con la que los estudiantes puedan contar a la hora de reprogramar sus actividades;
- establecer mecanismos de apoyo y eventuales modificaciones en los mecanismos de evaluación, para que los estudiantes que logren objetivos parciales pero no completen la aprobación del curso puedan completar el proceso en el primer período de examen siguiente a su participación en cursos controlados;
- crear y mantener activos permanentemente mecanismos de apoyo para que los estudiantes puedan revisar en el momento que lo necesiten requisitos previos de nuestros cursos que no son materia de trabajo durante ellos, pero cuya pobre conceptualización o manejo constituya un obstáculo para progresar en los cursos de la Cátedra.

Como apoyo a la implementación de esta política, la Cátedra se propone ampliar la incorporación de Estudiantes Colaboradores Honorarios, en el marco de un programa de formación que incluya la participación en los cursos de Tutoría Entre Pares TEP/Mat 1 y 2.

**5.2. Diseño paramétrico y Matemática.** La segunda dirección de trabajo que la Cátedra pretende priorizar es incluir en su oferta de actividades un curso

organizado alrededor de problemas geométricos propios del diseño en Arquitectura, presumiblemente tomando el Diseño Paramétrico [9] como eje de la actividad. Es una línea de trabajo incipiente, en la que esperamos colaborar con el equipo de trabajo que está desarrollando propuestas de similar tenor en las universidades argentinas de Buenos Aires, La Matanza y Mar del Plata.

## 6. Conclusiones y contribuciones

Es posible conseguir mejoras en los resultados de un curso masivo de Matemática ubicado al inicio de los estudios universitarios, aplicando de manera coherente y sistemática una batería de acciones orientada por principios bien establecidos por las Ciencias de la Educación. Para ello se ha requerido reorientar parcialmente los recursos disponibles, tendiendo a integrar un equipo interdisciplinario, pero no fue necesario un aumento de recursos. La experiencia sugiere también que es posible incursionar en el terreno de la orientación de estudiantes ingresantes, en la medida que se establezcan marcos de confianza entre los estudiantes y los equipos docentes y se habiliten alternativas genuinas que brinden a los estudiantes oportunidades de avance en la carrera. Al mismo tiempo, muestra que es posible identificar y avanzar en líneas de investigación sobre los aprendizajes con impacto sobre las prácticas desde estadios tempranos de desarrollo.

## 7. Agradecimientos

Federico Giménez elaboró los datos y los gráficos que se incluyen en tablas y figuras.

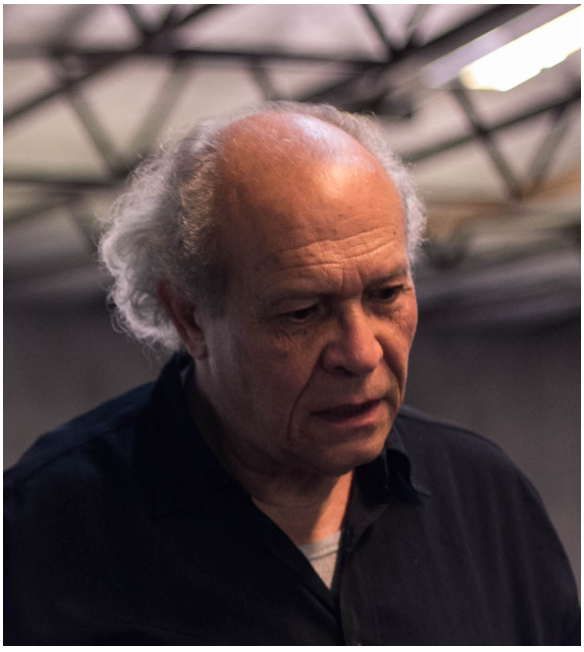
## Referencias

[1] Aguilar, S. y Barroso, J. (2015) La triangulación de datos como estrategia en investigación educativa, Pixel-Bit: Revista de Medios y Educación 47 73-88.
[2] Ambrose, S., Bridges, M.W., DiPietro, M., Lovett, M.C., Norman, M.K. (2010). How learning works. Seven research-based principles for smart teaching. San Francisco, John Wiley & Sons.
[3] Bain, K. (2004). What the best college teachers do. Londres, Harvard University Press.
[4] Cátedra de Matemática (2015). Informe de desistimiento de Matemática. Montevideo.
[5] Crouch, C.H., Mazur, E. (2001) Peer Instruction: Ten years of experience and results, American Journal of Physics 69 (9) 970-977;
[6] Gil, O., Motz, R., Sánchez, E. (2017) Exercise specification for collaboration, reusing and automatically generating evaluations, presentado a LACLO 2017.
[7] Gil, O., Nuin, A.L. (2017) Transformación de la enseñanza en un curso de Matemática de primer año para estudiantes de Arquitectura (2013-2016), comunicación presentada en JIES, Jornadas de Investigación en Educación Superior, Montevideo, Uruguay.
[8] Gil, O., Pantaleón, C., Patiño, L., Perchman, M., Seijo, R., Zino, L. (2006). Una propuesta para la Cátedra de Matemática de la Facultad de Arquitectura. Montevideo.
[9] Jabi, W. (2013) Parametric Design for Architecture, Laurence King Publishing.
[10] Kawulich, B. (2006). La observación participante como método de recolección de datos, Forum Qualitative Sozialforschung / Forum: Qualitative Social Research 6 (2), Art. 43, recuperado de `http://nbn-resolving.de/urn:nbn:de:0114-fqs0502430`.
[11] Kemmis, S. y R. McTaggart (1988). Cómo planificar la investigación-acción. Barcelona, Alertes.
[12] Michaelsen, L.K., Bauman Knight, A., Dee Fink, L. (editores) (2002). Team Based Learning. A transformative use of small groups. Westport, Praeger Publishers.
[13] Nuin, A.L., (2015, julio). Las creencias en relación al aprendizaje de la matemática: un estudio realizado con estudiantes de arquitectura de la Universidad de la República en Uruguay. Comunicación presentada en el XXII Congreso Internacional de Educación y Aprendizaje, Madrid, España.

[14] Ruiz, J., (2016, julio). Focus group y grupo de discusión: similitudes y diferencias. Comunicación presentada en el XXII Congreso Español de Sociología, Gijón, España.
[15] Turn to your neighbour. The official peer instruction blog (n.d) Recuperado de `https://blog.peerinstruction.net/`.
[16] Vygotsky, L. S. (1978) Mind in Society. Cambridge, MA: Harvard University Press

Facultad de Arquitectura Diseño y Urbanismo, UdelaR, Montevideo, Uruguay.
*Email address*: `omargil@fing.edu.uy`

PONENCIA



# EL NÚMERO DE ORO, GEOMETRÍA, ÁLGEBRA, ARITMÉTICA

GERARDO GONZÁLEZ SPRINBERG

Publicaciones Matemáticas del Uruguay
Volumen 17, Julio 2019, Páginas 183–192
S 0797-1443

# EL NÚMERO DE ORO, GEOMETRÍA, ÁLGEBRA, ARITMÉTICA

GERARDO GONZÁLEZ SPRINBERG

Abstract. A brief history of the golden number, its continued fraction and the Fibonacci sequence. Relations with regular polygons and polyhedra. Continued fractions of rational and irrational numbers, Lagrange periodicity theorem for quadratic irrationals, and Klein's geometric interpretation. Jung-Hirzebruch continued fractions and two dimensional singularities. Higher dimensional generalizations and a new result on periodicity for a cubic irrational.

## 1. Introducción histórica

El número de oro (o número áureo) es probablemente uno de los más emblemáticos ejemplos de varias propiedades

$$\varphi = \frac{1+\sqrt{5}}{2} = 1.618\,033\,988\,749\,894\,848\,204\,...$$

Los tres puntos suspensivos representan una infinidad de decimales sin periodicidad o regularidad conocidas, por lo que llevaría un tiempo y espacio indefinido explicitarlos! El número de oro es usualmente denotado por la letra giega $\varphi$ en honor al escultor y arquitecto Fidias quien construyó el Partenón en Atenas. Es ya conocido por los Pitagóricos (en el siglo -6). Estudiado posteriormente por Platón y Eudoxio de Cnide es presentado geométricamente en la Definición 3 del Libro VI de los Elementos de Euclides (en el siglo -3):

"Un segmento de recta se dice estar dividido en extrema y media razón si todo el segmento es al mayor como el mayor es al menor."

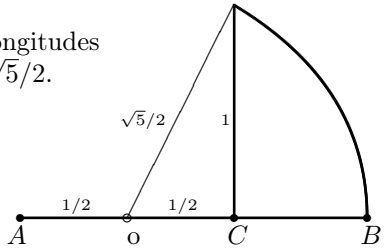
$$\frac{a+b}{a} = \frac{a}{b}$$

Este cociente es independiente de la longitud del segmento y es denotado $\varphi$ (o $\tau$ de la palabra griega τομη, sección). En la Proposición 30 del Libro VI, Euclides da un método de construcción de la sección en extrema y media de un segmento.
Puede ser probado que el número de oro $\varphi$ no es el cociente de dos números enteros, i.e. es un número irracional. Este hecho parece haber creado pánico y crisis política entre los Pitagóricos pues para ellos todo era representable por números pero solo conocían los números racionales (esto puede haber sido el origen del nombre "irracional" para estos números). El número $\varphi$ es un irracional cuadrático, i.e. un número algebraico irracional, raíz de un polinomio de grado 2 con coeficientes enteros. Denotando $x$ al cociente $\frac{a}{b}$ la igualdad $\frac{a+b}{a} = \frac{a}{b}$ es equivalente a $1 + \frac{1}{x} = x$. Multiplicando esta igualdad por $x$ se obtiene la ecuación polinómica $x^2 - x - 1 = 0$ y $\varphi$ es la raíz positiva de esta ecuación. La otra raíz (conjugada) es $-1/\varphi$, pues

©2019 PMU

el producto de las dos raices es $-1$ por la relación entre raíces y coeficientes del polinomio de la ecuación. Notar que $1/\varphi = \varphi - 1$ es la parte fraccionaria de $\varphi$.
Un segmento de longitud $\varphi$ es obtenido usando el teorema de Pitágoras (Euclides, Libro IV.10):

Un triángulo rectángulo con catetos de longitudes 1 y 1/2 tiene la hipotenusa de longitud $\sqrt{5}/2$. El segmento $AB$ tiene longitud $\varphi$ si $oB$ es $\sqrt{5}/2$, donde $o$ es el punto medio de $AC$ de longitud 1.




Relaciones entre $\varphi$, Fibonacci, polígonos y poliedros:

Por iteración de la igualdad $\varphi = 1 + 1/\varphi$ se obtiene $\varphi = 1 + 1/(1 + 1/\varphi)$ y así sucesivamente. Se obtiene una expresión de $\varphi$ llamada *fracción continua*:

$$\varphi = 1 + \cfrac{1}{1 + \cfrac{1}{\ddots + \cfrac{1}{1 + \cfrac{1}{\ddots}}}}$$

O más concisa $\varphi = 1 + \underset{1}{\overset{1}{\lfloor}} + \cdots + \underset{1}{\overset{1}{\lfloor}} + \cdots$. Notar que todos los coeficientes son iguales a 1, con una regularidad que contrasta con la caótica expresión decimal. Conservando solo una parte finita inicial de la fracción continua se obtiene un número racional llamado *convergente* o *reducida* de la fracción continua.
Denotando $F_n/F_{n-1}$ el convergente obtenido con los primeros $n$ términos de la fracción continua de $\varphi$, entonces $F_1/F_0 = 1$ , $F_2/F_1 = 1 + 1/1$ , $F_3/F_2 = 1 + 1/(1 + 1/1)$ , $\cdots$ , $F_{n+1}/F_n = 1 + 1/(1 + F_n/F_{n-1})$, así que $F_0 = 1$ , $F_1 = 1$ , $F_2 = 2$ , $\cdots$ , $F_{n+1} = F_n + F_{n-1}$ si $n \geq 1$.
Esta sucesión de enteros $1$ , $1$ , $2$ , $3$ , $5$ , $8$ , $13$ , $21$ , ... donde cada término (a partir del tercero) es la suma de los dos precedentes fue introducida en el tratado de aritmética "Liber Abaci" (Libro de cálculos) en 1202 por Leonardo de Pisa, llamado *Fibonacci* (hijo de Bonacci), como solución del célebre problema sobre el número de parejas de conejos que dan nacimiento a una nueva pareja cada mes. [1] Fibonacci también popularizó la notación indo-arabiga de los números usada hasta hoy en día. Los cocientes $F_n/F_{n-1}$ : $(1\ ,\ 2\ ,\ 3/2 = 1.5\ , 5/3 = 1,\overline{6}\ , 8/5 = 1.6\ ,\ 13/8 = 1.625\ ,\ 21/13 = 1.615384615384615\ ,\ \cdots)$ convergen hacia $\varphi = 1.618\cdots$ con valores alternativamente inferiores y superiores.

Algunas propiedades de la sucesión de Fibonacci :

$$(F_n, F_{n+1}) = 1 \quad \text{for} \quad n \geq 0 \quad ; \quad F_n^2 - F_{n-1}F_{n+1} = (-1)^n \quad \text{para} \quad n \geq 1$$

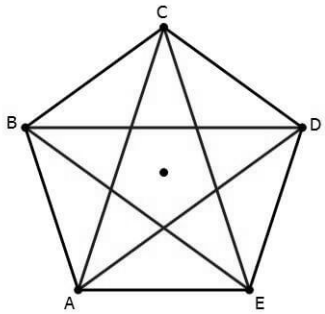
$$|F_{n+1}/F_n - \varphi| \leq 1/F_n^2 \quad ; \quad F_n = \frac{1}{\sqrt{5}}(\varphi^{n+1} - \varphi'^{n+1}) \text{ donde } \varphi' = \frac{1-\sqrt{5}}{2}$$

[1] *Nota.* Cada nueva pareja de conejos es un hermano y una hermana, pero la iglesia no criticó este incestuoso problema matemático, aunque Galileo fue excomulgado durante más de tres siglos por "e pur si muove".

1
1 1
1 2 1
1 3 3 1
1 4 6 4 1
1 5 10 10 5 1
1 6 15 20 15 6 1
1 7 21 25 25 21 7 1

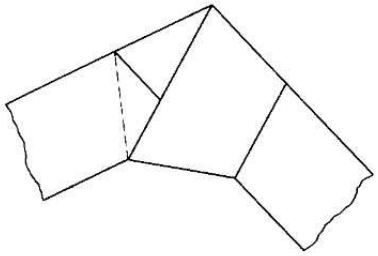
Las sumas de cada diagonal indicada del triángulo de Pascal enderezado “a la china” son los términos consecutivos de la sucesión de Fibonacci.

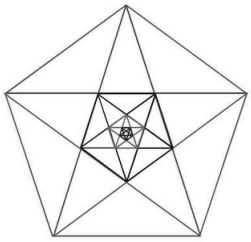
Hay una relación entre $\varphi$ y $\pi$ dada por funciones trigonométricas elementales: $\varphi = 2\cos\pi/5 \quad , \quad 1/\varphi = 2\sin\pi/10 \quad , \quad \varphi = 1 + 1/\varphi = 1 + 2\sin\pi/10$. Por lo tanto hay relaciones entre $\varphi$, el pentágono regular, el decágono y poliedros. El cociente entre la longitud de una diagonal y el lado de un pentágono regular es $\varphi$. Luego un pentágono regular es obtenido a partir de un “triángulo áureo" $ACE$ (ver figura) y cuyos ángulos son $\widehat{ACE} = \pi/5, \widehat{CAE} = \widehat{CEA} = 2\pi/5$.




*(El pentagrama o estrella pentagonal formada por las 5 diagonales del pentágono era usado por los Pitagóricos como símbolo de reconocimiento y fraternidad. )*

A causa de los ángulos del pentágono éste puede ser obtenido como un “nudo áureo" a partir de una tira rectangular de papel (ver figura).





Las intersecciones interiores de las diagonales son los cinco vértices de un nuevo pentágono y así indefinidamente hacia el interior o el exterior.[2]
Un decágono regular es también fácilmente obtenido pues el cociente entre el radio y el lado es $\varphi$.
Conociendo $\varphi$ se obtiene también un dodecaedro regular, con 12 pentágonos regulares como caras, 3 en cada vértice (símbolo de Schläfli $(5, 3)$ ), pues

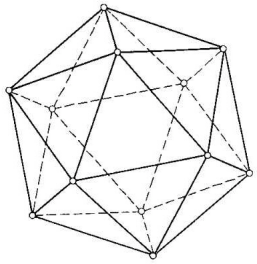
$$(\pm 1, \pm 1, \pm 1), (0, \pm\varphi, \pm 1/\varphi), (\pm 1/\varphi, 0, \pm\varphi), (\pm\varphi, \pm 1/\varphi, 0)$$

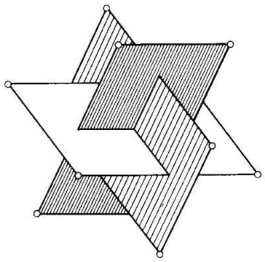
[2]*Nota.* La estrella pentagonal fue usada como logo comercial de una empresa de combustibles y también como logo político de desafío en un pequeño país sudamericano.

son las coordenads de los 20 vértices de un dodecaedro regular.

Los 12 centros de las caras del dodecaedro son los vértices de un icosaedro, dual del dodecaedro, con 20 triángulos equiláteros como caras, 5 en cada vértice (símbolo de Schläfli $(3, 5)$ ).

Cada par de aristas opuestas paralelas del icosaedro limitan un *rectángulo áureo* con $\varphi$ como cociente de lados. Hay 30 aristas, 15 rectángulos y 5 tríos de tales rectángulos que son ortogonales.





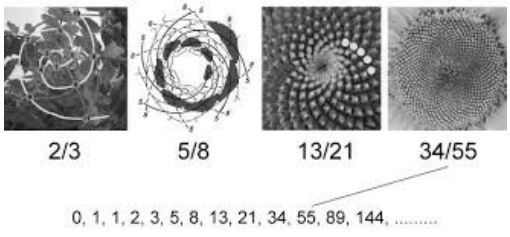
Con coordenadas con respecto a los planos de un tal trío ortogonal de rectángulos los puntos $(0, \pm\varphi, \pm 1)$, $(\pm 1, 0, \pm\varphi)$, $(\pm\varphi, \pm 1, 0)$ son los vértices del icosaedro.

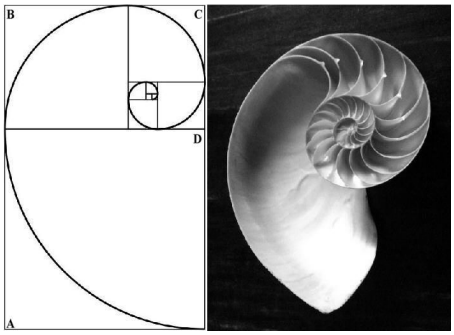
Estos puntos dividen en la proporción $\varphi : 1$ las 12 aristas del octaedro regular con vértices $(\pm\varphi^2, 0, 0), (0, \pm\varphi^2, 0), (0, 0, \pm\varphi^2)$ .

Los Elementos de Euclides culminan en el Libro XIII con la construcción de los cinco tipos de poliedros regulares o platónicos, tetraedro, cubo, octaedro, dodecaedro e icosaedro. Los poliedros regulares son conocidos desde tiempos inmemoriales, pues se han encontrado dados neolíticos de piedra en Escocia (datados aproximadamente del año -2000) con las formas de los cinco poliedros regulares, que se encuentran en el Museo de Oxford. Pero el primer estudio matemático formal conocido de estos poliedros es debido a los matemáticos griegos.

El número $\varphi$ y la sucesión de Fibonacci aparecen sin fin en la naturaleza (e.g en filotaxis, la distribución de hojas a lo largo de tallos de plantas, o una espiral logarítmica en el molusco Nautilus), en arte (e.g. en pinturas y dibujos de Botticelli y Leonardo da Vinci, en esculturas de Fidias), y en música (e.g. en el ciclo de quintas de Pitágoras, en las proporciones de violines de Stradivarius).

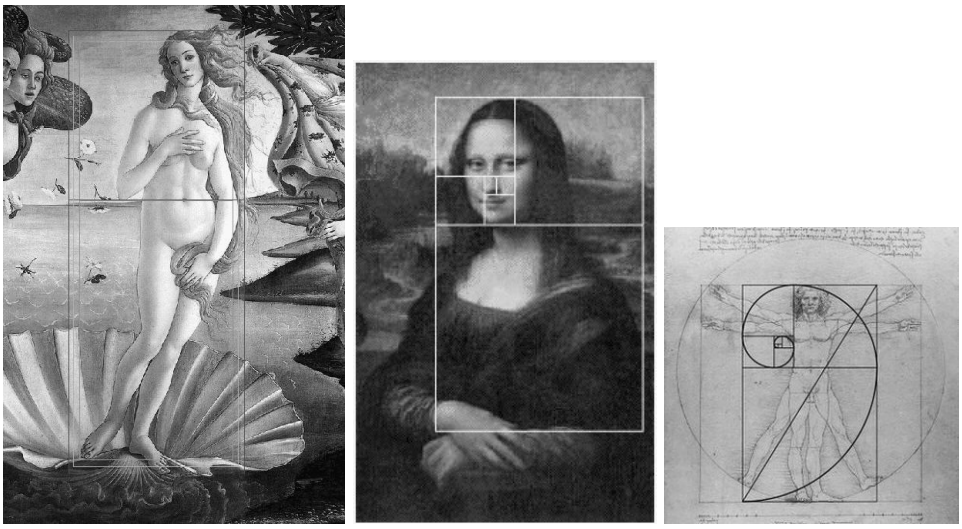
Filotaxis: recorriendo el tallo en espiral, se representa el número de vueltas sobre el número de hojas encontradas por una fracción cacterística de las diferentes especies. Se encuentran cocientes de la sucesión de Fibonacci en girasoles, piñas de pinos, ananás, ...






Espiral logarítmica de ecuación polar $\boldsymbol{r = \varphi^{2\theta/\pi}}$ en $(r, \theta)$ con polo $\{AC \cap BD\}$ y nautilus

Dividiendo un rectángulo áureo en un cuadrado y otro rectángulo, éste último será también áureo y se puede iterar la división. Inscribiendo un cuarto de círculo en cada cuadrado así construído se obtiene una curva que aproxima una espiral logarítmica (ver figura).

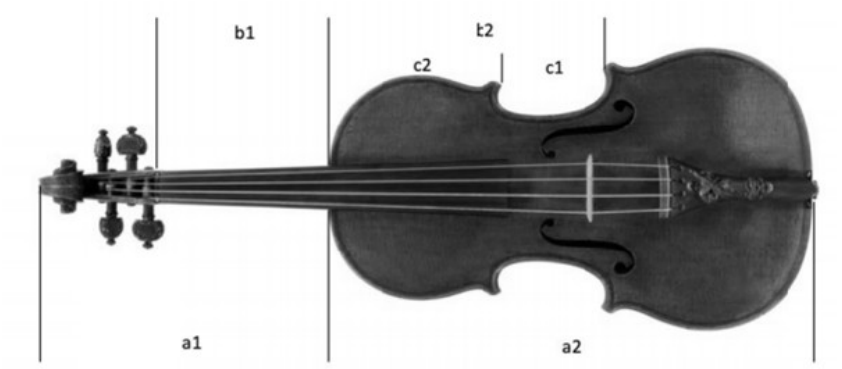


Nacimiento de Venus de Boticelli, La Gioconda y hombre de Leonardo da Vinci.

Friso del Partenón y cariátides en la Acrópolis de Fidias.

Proporciones áureas en violín de Stradivarius.

$$\frac{a1+a2}{a2}=\frac{a2}{a1}=\frac{b2}{b1}=\frac{b2}{c2}=\frac{c2}{c1}=\phi$$




## 2. Fracciones continuas de números racionales e irracionales

Las fracciones continuas de números racionales resultan del algoritmo clásico de Euclides. Dado $a/b$, con $a$ y $b$ enteros (sin factor común), $b>0$, la división entera $a = bc_0 + r_0$, con $c_0$ el cociente y $r_0$ el resto, $0 \leq r_0 < b$, puede ser iterada (finitamente) hasta que el último resto obtenido sea cero : $a/b = c_0 + 1/(b/r_0) = c_0 + 1/(c_1 + 1/(r_0/r_1) = \cdots$

(*Nota.* El último resto no nulo es $mcd(a,b)$. Esta es la presentación original de Euclides, como la medida común entre dos segmentos conmensurables, Libro 7,2.)

*Un número es racional si y sólo si su fracción continua es finita.*

La expresión decimal puede ser infinita, como $1/7 = 0,\overline{142857}$ con período supraraya-do. La fracción continua de un racional es única (con último término $>1$ si tiene al menos 2 términos), pero su expresión decimal no es única, e.g. $1,2 = 1,1\overline{9}$.

Se demuestra que *un número es racional si y sólo si su expresión decimal es periódica.*

Este algoritmo puede ser generalizado, notando que el cociente entero es la *parte entera* del racional, i.e. el mayor entero menor o igual al número dado $r$, notación : $\lfloor r \rfloor$ ("piso" de $r$).

Si $r$ es un número real irracional, sea $\{r\} = r - \lfloor r \rfloor$ su parte fraccionaria. Entonces también por iteración la fracción continua es obtenida:

$$r = \lfloor r \rfloor + \frac{1}{1/\{r\}} = \lfloor r \rfloor + \frac{1}{\lfloor 1/\{r\} \rfloor + \frac{1}{1/\{1/\{r\}\}}} = \cdots = c_0 + \cfrac{1}{c_1 + \cfrac{1}{\ddots + \cfrac{1}{c_n + \cfrac{1}{\ddots}}}}$$

$$= c_0 + \underline{1 \Big/ c_1} + \cdots + \underline{1 \Big/ c_n} + \cdots \qquad \text{Notación: } [c_0, c_1, \cdots, c_n, \cdots]$$

Es una fracción continua infinita pues si no sería un racional. Inversamente *cualquier sucesión infinita de enteros, positivos a partir del segundo, define una fracción continua que converge hacia un número irracional.*

Hemos ya visto el ejemplo notable: $\varphi = (1+\sqrt{5})/2 = [1, 1, \cdots, 1, \cdots]$

Es un caso particular de la raíz positiva de $x^2 = ax + 1$ para $a = 1$. Para cualquier entero $a > 0$ su raíz positiva tiene la fracción continua $[a, a, \cdots, a, \cdots]$.
Por ejemplo si $a = 2$, la raíz positiva es $1 + \sqrt{2} = [2, 2, \cdots, 2, \cdots]$. Resulta entonces

$$\sqrt{2} = [1, 2, \cdots, 2, \cdots].$$

(*Nota*: Este número es usado para los formatos estándar pues si la proporción entre los lados de una hoja rectangular de papel es $\sqrt{2}$, entonces al plegarla la proporción es la misma. A partir de una hoja de 1 $m^2$, $A_0 = 84,1 \times 118,9$ $cm$, al plegarla cuatro veces se obtiene el formato usual $A_4 = 21 \times 29,7$ $cm$.)

Hay también resultados importantes de fracciones continuas de irracionales *trascendentes* e.g. $e = [2, 1, 2, 1, 1, 4, 1, 1, 6, 1, 1, 8, 1, 1, 10, 1, \cdots]$ $(Euler)$
o ejemplos de fracciones continuas *generalizadas* (i.e. con numeradores $\neq 1$).

$$e = 2 + \underline{1 \,\big|\, 1} + \underline{1 \,\big|\, 2} + \underline{2 \,\big|\, 3} + \underline{3 \,\big|\, 4} + \cdots \qquad (Euler, Cesàro)$$

$$\pi = 3 + \underline{1^2 \,\big|\, 6} + \underline{3^2 \,\big|\, 6} + \underline{5^2 \,\big|\, 6} + \underline{7^2 \,\big|\, 6} + \cdots \qquad (Wallis)$$

Una fracción continua infinita $[c_0, c_1, \cdots, c_n, \cdots]$ es llamada *periódica* si se reproduce periódicamente a partir de un cierto rango, i.e. $[c_0, \cdots, c_{k-1}, \overline{c_k, \cdots, c_{k+n}}]$.
Es *periódica pura* si $k = 0$, i.e. $[\overline{c_0, \cdots, c_n}]$.

Un teorema notable de Joseph-Louis Lagrange prueba que *todo irracional cuadrático tiene una fracción continua periódica.* La proposición inversa, más fácil de probar, era conocida por Leonhard Euler, por lo que hay una equivalencia.
Evariste Galois probó en particular que *una fracción continua es periódica pura si y solo si es un irracional cuadrático* $> 1$ *cuyo conjugado está entre* $0$ *y* $1$.

## 3. Interpretación geométrica de las fracciones continuas.

Sea $r$ un número real positivo y $L$ la semirrecta con equación $y = rx$, de pendiente $r$, en el primer cuadrante del plano. Sea la fracción continua de $r$ con convergentes $p_n/q_n = [c_0, \cdots, c_n]$.

Felix Klein propuso la interpretación geométrica siguiente. *Si un hilo a lo largo de L se aleja de L en ambos lados entonces los primeros puntos con coordenadas enteras que encuentra tienen coordenadas* $(q_n, p_n)$ *dadas por los convergentes. Estos puntos son los vértices de dos líneas poligonales que aproximan L, una por arriba y la otra por debajo.*

Algunas pruebas y corolarios:

Las relaciones de recurrencia siguientes son probadas por inducción:

$$(1) \quad p_k = c_k p_{k-1} + p_{k-2} \quad , \quad (2) \quad q_k = c_k q_{k-1} + q_{k-2} \quad , \quad (\text{para} \quad k \geq 2)$$

Si se pone $p_{-1} := 1$ y $q_{-1} := 0$, entonces son válidas también para $k = 1$.
La diferencia $[(2)p_{k-1} - (1)q_{k-1}]$ elimina $c_k$ y entonces
$q_k p_{k-1} - p_k q_{k-1} = -(q_{k-1} p_{k-2} - p_{k-1} q_{k-2})$. como $q_0 p_{-1} - p_0 q_{-1} = 1$ se obtiene

$$(3) \quad q_k p_{k-1} - p_k q_{k-1} = (-1)^k$$

Reulta de (3) que el representante $p_n/q_n$ de cada convergente is irreducible, pues cualquier divisor de $p_n$ y de $q_n$ es divisor of $(-1)^k$.
Con la diferencia $[(2)p_{k-2} - (1)q_{k-2}]$ se obtiene, por (3),
$q_k p_{k-2} - p_k q_{k-2} = c_k(q_{k-1} p_{k-2} - p_{k-1} q_{k-2}) = (-1)^{k-1} c_k$ , luego

$$p_{k-2}/q_{k-2} - p_k/q_k = (-1)^{k-1} c_k / q_k q_{k-2}.$$

En consecuencia *si $r$ es irracional entonces los convergentes de orden par (resp. impar) forman una sucesión creciente (resp. decreciente)*

$$\frac{p_0}{q_0} < \frac{p_2}{q_2} < \frac{p_4}{q_4} < \frac{p_6}{q_6} < \cdots < r < \cdots < \frac{p_7}{q_7} < \frac{p_5}{q_5} < \frac{p_3}{q_3} < \frac{p_1}{q_1}$$

Los puntos $V_k$ de coordenadas $(q_k, p_k)$ en el plano están por debajo de la recta $L$ si $k$ es par, y por arriba si $k$ es impar.
Como $q_k = c_k q_{k-1} + q_{k-2} \geq q_{k-1} + q_{k-2} \geq 2q_{k-2}$ sigue $q_k \geq 2^{(k-1)/2}$, y por (3) se obtiene:

*Las dos sucesiones convergen hacia el mismo límite, $r$. Si $r$ is racional, ambas son finitas y una de ellas llega a $r$. Los puntos con coordenadas dadas por los convergentes tienden monótonamente por debajo y por arriba hacia la recta $L$.*

La igualdad (3) $q_k p_{k-1} - p_k q_{k-1} = (-1)^k$ puede verse como el determinante de la matriz con vectores columna $V_k = \begin{pmatrix} q_k \\ p_k \end{pmatrix}$ y $V_{k-1} = \begin{pmatrix} q_{k-1} \\ p_{k-1} \end{pmatrix}$

Entonces $|\det(V_k, V_{k-1})| = 1$ y se deduce que $(V_k, V_{k-1})$ es una base de $\mathbb{Z}^2$. El área del triángulo $(0, V_k, V_{k-1})$ es $1/2$, luego sus vértices son sus únicos puntos enteros.
*Esto prueba que $V_k$ y $V_{k-1}$ son los puntos enteros más próximos de la semirrecta $L$ en este triángulo, y se ha demostrado la interpretación geométrica de los convergentes de la fracción continua.*

La relaciones (1) y (2) pueden ser interpretadas geometricamente: $V_k - V_{k-2} = c_k V_{k-1}$, que significa que el segmento $\overline{(V_{k-2}, V_k)}$ es paralelo a $V_{k-1}$ y lo contiene $c_k$ veces. Luego el número de puntos enteros en $\overline{(V_{k-2}, V_k)}$ es $c_k + 1$.

## 4. Fracción continua negativa de Jung-Hirzebruch

Una fracción continua con términos negativos $\leq -2$ a partir del segundo es obtenida considerando para $r$ racional la división euclídea por exceso, o de manera equivalente para cualquier real $r$ el menor entero $\geq r$ , notación: $\lceil r \rceil$ ("techo" de $r$).
Si $r$ no es entero entonces $\lceil r \rceil = \lfloor r \rfloor + 1$ , $r = \lceil r \rceil - (1 - \{r\}) =$.

$$\lceil r \rceil - 1/(1/(1 - \{r\})) = \lceil r \rceil - 1/(\lceil 1/(1 - \{r\}) \rceil - (1 - \{1/(1 - \{r\})\})) = \cdots$$

$$= d_1 - \underline{1 \left| d_2 \right.} - \cdots - \underline{1 \left| d_n \right.} - \cdots \quad \text{Notación: } [d_1, d_2 \cdots, d_n, \cdots]^-$$

$$d_k \geq 2 \text{ si } k > 1$$

Si $0 < a < b$ y $b/a$ es un racional irreducible, se obtienen con la fracción continua de Jung-Hirzebruch todos los puntos enteros próximos por debajo de la semirrecta $L$ en el primer cuadrante de ecuación $y = (b/a)x$,

$$Q_0 = (1, 0), Q_1 = (1, 1), \quad Q_{i+1} = d_i Q_i - Q_{i-1}, i \geq 1$$

Los puntos $Q_{i-1}, Q_i, Q_{i+1}$ están alineados si $d_i = 2$. Hay una relación con la fracción continua ordinaria con la que se obtienen primero los vértices de la poligonal por debajo de $L$ con las reducidas pares y luego todos los puntos enteros en cada segmento de esta poligonal con las reducidas impares.
La fracción continua de Jung-Hirzebruch de $\varphi$ es más difícil de encontrar:

$$\varphi = [2, \overline{3}]^- = 2 - \underline{1 \left| 3 \right.} - \cdots - \underline{1 \left| 3 \right.} - \cdots$$

*Nota histórica*: Jung usa este tipo de fracción continua en su método de resolución de singularidades de superficies. La singularidad en el origen en la superficie de ecuación $x^n = yz^p$ después de ser normalizada puede ser descrita como una singularidad tórica, definida por el álgebra asociada al semigrupo de puntos enteros debajo de la semirecta $L$ de ecuación $y = (n/p)x$. Su resolución es obtenida como un abanico de conos limitados por las semirrectas por todos los puntos enteros debajo de $L$. De hecho estos puntos

enteros forman una base (llamada base de Hilbert) del semigrupo. Estas singularidades son esenciales en el método de resolución de Jung. La resolución minimal posee una cadena de curvas racionales en su divisor excepcional (que se contrae en el punto singular), con auto intersecciones dadas por la fracción continua negativa de $n/p$.

Las fracciones continuas negativas periódicas de irracionales cuadráticos han sido usadas por Hirzebruch para resolver las singularidades cuspidales de Hilbert-Blumenthal. Se obtienen ciclos de divisores excepcionales para estas singularidades elípticas ($h_1 = 1$).

## 5. Fracciones continuas en dimensión superior

Sea $C$ un cono simplicial en el espacio euclídeo $\mathbb{R}^n$ y $(e_1, \cdots, e_n)$ la base canónica de $\mathbb{R}^n$. La adherencia convexa de los puntos no nulos de $C \cap \mathbb{Z}^n$ es llamada el *poliedro de Klein de* $C$.

El borde del poliedro de Klein es una variedad poliédrica llamada *"vela"*. Si $r$ es un número real positivo , por la interpretación geométrica su fracción continua es equivalente a dar las velas de los conos $\langle e_1, (1, r)\rangle$ y $\langle (1, r), e_2)\rangle$, cuyos vértices o puntos angulares corresponden a los convergentes.

Esta equivalencia permite una generalización de fracciones continuas en dimensión superior. Si $L$ es una semirrecta en el octante positivo de $\mathbb{R}^3$, los puntos con coordenadas enteras más próximos a $L$ son los que se encuentran en las velas de los conos $C_1 = \langle L, e_2, e_3\rangle, C_2 = \langle e_1, L, e_3\rangle, C_3 = \langle e_1, e_2, L\rangle$.

También puede darse una descripción matricial de la fracción continua que es generalizable en dimensión superior. Las relaciones de recurrencia (1) y (2) son expresadas como un producto en $GL(2, \mathbb{Z})$:

$$\begin{pmatrix} p_k & p_{k-1} \\ q_k & q_{k-1} \end{pmatrix} = \begin{pmatrix} c_0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} c_1 & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} c_k & 1 \\ 1 & 0 \end{pmatrix}$$

Algunos resultados en dimensión 3 (artículo en preparación)

Sea $L_1 = (x_1, y_1, z_1)$ un punto en $\mathbb{R}^3$ con coordenadas positivas que define una semirrecta $OL_1$ en el octante positivo de $\mathbb{R}^3$. Sea $P_1$ el primer punto entero más próximo a la semirrecta $OL_1$. Si $P_1$ pertenece al cono $C_i$ remplazar $e_i$ por $P_1$ in la base canónica y denotar por $T_1$ la matriz de cambio de base. Sea $L_2 = T_1^{-1} L_1$, se itera el procedimiento con respecto a la semirrecta $OL_2$ y la nueva base y así sucesivamente. La sucesión $T_k$ de matrices de cambio de base pertenece de hecho a $GL(3, \mathbb{Z})$ y es el equivalente matricial de la fracción continua. Las matrices $M_k := T_1 T_2 \cdots T_k$ juegan el rol de los convergentes y se tiene $M_{k-1} L_k = L_1$ para $k \geq 2$.

Ejemplo: la raíz cúbica de 2

$\sqrt[3]{2} = 1.25992104989\cdots$ es un número cúbico irracional bastante misterioso. Su fracción continua (en dimensión 2) no es periódica por el teorema de Lagrange pues no es un irracional cuadrático y no hay ninguna regularidad

$$\sqrt[3]{2} = [1, 3, 1, 5, 1, 1, 4, 1, 1, 8, 1, 14, 1, 10, 2, 1, 4, 12, 2, 3, 2, 1, 3, 4, 1, 1, 2, 14, 3, 12, 1, 15, 3, 1, 4, 534, 1, 1, 5, 1, 1, 121, 1, 2, 2, 4, 10, \cdots]$$

*Resultado: con el procedimiento anterior se obtiene una fracción continua periódica para* $r = \sqrt[3]{2}$*, en el sentido que la sucesión de matrices* $T_k$ *tiene un período de longitud 3 a partir del tercer término, para la semirrecta dada por* $L_1 = (1, r, r^2)$ *( base natural de* $\mathbb{Q}[r]$ *sobre* $\mathbb{Q}$*).*

$$\left[ \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 3 & 0 \\ 0 & 1 & 0 \\ 0 & 2 & 1 \end{pmatrix}, \overline{\begin{pmatrix} 1 & 0 & 3 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 3 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 3 & 0 \\ 0 & 1 & 0 \\ 0 & 3 & 1 \end{pmatrix}} \right]$$

Los puntos enteros más próximos a la semirrecta, en la base canónica de partida, son columnas de las matrices $M_k$. Tienen coordenadas que crecen muy rápidamente, como se puede observar con los primeros nueve puntos:

$$\begin{pmatrix}1\\1\\1\end{pmatrix}, \begin{pmatrix}3\\4\\5\end{pmatrix}, \begin{pmatrix}12\\15\\19\end{pmatrix}, \begin{pmatrix}46\\58\\73\end{pmatrix}, \begin{pmatrix}177\\223\\281\end{pmatrix}, \begin{pmatrix}681\\858\\1081\end{pmatrix}, \begin{pmatrix}2620\\3301\\4159\end{pmatrix}, \begin{pmatrix}10080\\12700\\16001\end{pmatrix}, \begin{pmatrix}38781\\48861\\61561\end{pmatrix}$$

El cociente de la segunda sobre la primera (o de la tercera sobre la segunda) coordenada converge rápidamente hacia $r$.

El cono dado por tres puntos consecutivos contiene la semirrecta y tiene determinante 1, por lo que el tetraedro con vértices estos puntos y el origen contiene solamente los vértices como puntos enteros.

## References

[1] N. Beskin. *Fracciones maravillosas.* Ediciones MIR, Moscú, 1987
[2] E. R. Gentile. *Aritmética elemental* Secretaría de la OEA, Publicaciones Matemáticas no. 25, Buenos Aires, 1985
[3] H. S. M. Coxeter. *Regular polytopes* Dover Publications, New York, 1973.
[4] G. Gonzalez Sprinberg. *Éventails en dimension deux et transformé de Nash.* Secrétariat Mathémathique de l'E.N.S., Paris, 1977.
[5] O. Perron.*Die Lehre von den Kettenbrüchen.* Teubner, Germany, 1954

CMat, Facultad de Ciencias, Universidad de la República. Montevideo, Uruguay.
Université Grenoble Alpes, France
*Email address*: gerardogs@cmat.edu.uy

PONENCIA

# TEORÍAS Y MODELOS: UNA INTRODUCCIÓN A LA LÓGICA DE PRIMER ORDEN

ALEXANDRE MIQUEL

Publicaciones Matemáticas del Uruguay
Volumen 17, Julio 2019, Páginas 195–224
S 0797-1443

# TEORÍAS Y MODELOS: UNA INTRODUCCIÓN A LA LÓGICA DE PRIMER ORDEN

ALEXANDRE MIQUEL

## Introducción

El objetivo de este artículo[1] es presentar los conceptos y los resultados básicos de la lógica de primer orden, con un énfasis sobre la articulación entre la noción de *teoría de primer orden* (el punto de vista sintáctico) y la noción de *modelo* (el punto de vista semántico). Este artículo se restringe al estudio de los formalismos de primer orden, en los cuales las variables sólo representan «objetos de base» —como los enteros naturales, los elementos de un grupo, o los elementos de un universo conjuntista. Por otro lado, los formalismos de segundo o de alto orden introducen otros tipos de variables para representar (y para cuantificar sobre) las proposiciones y las relaciones. Sin embargo, cabe destacar que los formalismos de alto orden siempre se pueden presentar como teorías de primer orden (mediante traducciones adecuadas), lo que explica por qué la lógica de primer orden constituye el marco fundamental de la lógica.

Así, en las páginas que siguen, intentaremos contestar a las siguientes preguntas:

¿Cómo se construyen las expresiones del lenguaje matemático?
¿Qué es una teoría, un axioma, una demostración, un teorema?
¿Se puede demostrar que una teoría es consistente? y ¿cómo?
¿A qué objetos refieren los símbolos del lenguaje matemático?
¿Por qué los «universos» descritos por los formalismos matemáticos son ambiguos?
¿En qué sentido los formalismos matemáticos son intrínsecamente incompletos?

**Problema de metodología.** El propio objeto de la lógica —que se propone razonar sobre el razonamiento— hace inevitablemente surgir dudas en cuanto a su viabilidad. ¿Cómo se puede razonar sobre las demostraciones matemáticas si todavía no tenemos una noción de demostración? ¿En qué lenguaje se puede definir el lenguaje matemático? Y ¿el razonamiento sobre el razonamiento no introduce un círculo vicioso en la matemática? Por suerte, este problema de circularidad no es específico a la lógica, pero también aparece en la lingüística, cuyo objeto es estudiar los mecanismos del lenguaje... con el propio lenguaje. En lingüística, se resuelve tradicionalmente el problema distinguiendo dos niveles de lenguaje:

- el nivel del *discurso lingüístico*, que se escribe del modo usual;
- el nivel de los *objetos lingüísticos*, que se destacan en el discurso lingüístico mediante comillas, las cuales sirven para transformar partes del discurso en *objetos* del discurso.

[1]Este artículo está basado en una charla dada en el 6$^{to}$ coloquio uruguayo de matemática, que tuvo lugar en Montevideo en diciembre de 2017.

©2019 PMU

Por ejemplo, el enunciado lingüístico

La palabra «contiene» contiene 8 letras

tiene dos ocurrencias de la palabra "contiene"[2]. La segunda ocurrencia (sin comillas) es parte del discurso, mientras la primera ocurrencia (entre comillas) es el objeto lingüístico sobre el cual trata el discurso. También se observa que la parte del discurso entre comillas se usa como un sustantivo (es decir: como un objeto), aunque ésta esté formada a partir de un verbo.

En lógica, se resuelve el problema de circularidad de modo análogo.

De mismo modo que la lingüística no intenta definir el lenguaje a partir de nada, pero sólo propone estudiar los mecanismos del lenguaje con las herramientas del lenguaje, el objetivo de la lógica no es definir la matemática a partir de nada, pero sólo estudiar los mecanismos de razonamiento de la matemática con las herramientas de la matemática, presuponiendo que los lectores ya tienen un conocimiento básico de éstas. Y como en lingüística, se necesita distinguir dos niveles de discurso, con el fin de evitar las ambigüedades... y los círculos viciosos. Para ello, se modelizan los enunciados, las demostraciones y las teorías que aparecen en el discurso matemático como objetos matemáticos particulares, que se llaman *fórmulas*, *teorías de primer orden* y *derivaciones*. En la práctica, estos objetos matemáticos que representan «el discurso matemático entre comillas» están definidos con técnicas estándar, y se pueden combinar con otros objetos matemáticos, como grupos, grafos o espacios topológicos.

Así, podremos formular enunciados sobre las fórmulas (los «enunciados entre comillas»), demostrar propiedades sobre las derivaciones (las «demostraciones entre comillas») y —sobre todo— desarrollar una teoría de las teorías de primer orden (las «teorías entre comillas»), esto sin temer los círculos viciosos, en la medida en que el vínculo entre el discurso matemático y su modelación sólo se ubica en el nivel filosófico, afuera de la matemática.

Sin embargo, la experiencia[3] muestra que las fórmulas y las derivaciones constituyen una representación fiel de los enunciados matemáticos y de sus demostraciones (a menos de cambios de notaciones). Y en la práctica, esto nos permitirá a veces justificar que cierta fórmula tiene una derivación en cierta teoría de primer orden, usando el hecho que el enunciado correspondiente es un teorema en la teoría correspondiente.

**Plan del artículo.** Este artículo está dividido en tres secciones. En la primera, definimos los *lenguajes de primer orden* y la noción de *teoría de primer orden*, y mostramos cómo las teorías matemáticas usuales se pueden modelizar en este marco. En la segunda sección, introducimos la noción de *modelo* (en el sentido de Tarski), así como los teoremas fundamentales que describen los vínculos entre las teorías de primer orden y sus modelos. En la última sección, presentamos (con un enfoque más histórico-filosófico) los teoremas de incompletitud de Gödel, que describen los límites intrínsecos de los formalismos matemáticos.

[2]Este enunciado es un ejemplo de discurso metalingüístico sobre un discurso lingüístico, lo que explica por qué necesita dos niveles de comillas (la sangría y las comillas tipográficas).

[3]Así como el reciente desarrollo de los asistentes a la prueba, que permitieron transformar miles de páginas de matemática en derivaciones formales almacenadas en bases de datos y verificadas por computadoras.

## 1. Teorías de primer orden

**1.1. Lenguajes de primer orden.** Cada teoría de primer orden está definida a partir de un lenguaje de primer orden, que distingue dos tipos de expresiones:

- los *términos*, que sirven para representar los objetos de la teoría;
- las *fórmulas*, que sirven para representar los enunciados de la teoría.

Formalmente, un lenguaje de primer orden está definido a partir de un vocabulario:

**Definición 1.1** (Vocabulario)**.** Un *vocabulario* consta de:

- un conjunto de *símbolos de función* (notación: $f$, $g$, $h$, etc.) dados con sus aridades;
- un conjunto de *símbolos de predicado* (notación: $p$, $q$, $r$, etc.) dados con sus aridades.

Se recuerda que la *aridad* de un símbolo (de función o de predicado) es un entero positivo o nulo que indica (de modo convencional) la cantidad de argumentos esperada por dicho símbolo. En particular, se llama *símbolo de constante* a todo símbolo de función de aridad 0.

**Ejemplos 1.2** (Vocabularios usuales)**.** En este artículo, consideraremos ejemplos de teorías de primer orden basadas en los siguientes vocabularios:

(1) El vocabulario de la *teoría de grupos*, que consta de:
   - un símbolo de constante «$e$» (elemento neutro), un símbolo de función «$I$» (inverso) de aridad 1, y un símbolo de función «$*$» (composición) de aridad 2;
   - un único símbolo de predicado «$=$» (igualdad) de aridad 2.

(2) El vocabulario de la *teoría de anillos*, que consta de:
   - dos símbolos de constante «0» (cero) y «1» (uno), un símbolo de función «$-$» (opuesto) de aridad 1, y dos símbolos de función «$+$» (suma) y «$\times$» (producto) de aridad 2;
   - un único símbolo de predicado «$=$» (igualdad) de aridad 2.

(3) El vocabulario de la *teoría del orden*, que consta de:
   - ningún símbolo de función o de constante;
   - dos símbolos de predicado «$=$» (igualdad) y «$\leq$» (orden) de aridad 2.

(4) El vocabulario de la *aritmética de Peano*, que consta de:
   - un símbolo de constante «0» (cero), un símbolo de función «$s$» (sucesor) de aridad 1, y dos símbolos de función «$+$» (suma) y «$\times$» (producto) de aridad 2;
   - un único símbolo de predicado «$=$» (igualdad) de aridad 2.

(5) El lenguaje de la *teoría de conjuntos*, que consta de:
   - ningún símbolo de función o de constante;
   - dos símbolos de predicado «$=$» (igualdad) y «$\in$» (pertenencia) de aridad 2.

**Observaciones 1.3.** (1) Los vocabularios anteriores contienen todos un símbolo de igualdad, que es un símbolo de predicado binario (escrito «$=$») con un sentido particular en lógica. En lo siguiente, sólo consideraremos vocabularios que contienen un símbolo de igualdad.

(2) En lo siguiente, usaremos la notación $e_1 \equiv e_2$ (en lugar de $e_1 = e_2$) para expresar que dos expresiones sintácticas $e_1$ y $e_2$ (dos términos o dos fórmulas) son

idénticas, con el fin de evitar confusiones con expresiones sintácticas construidas a partir del símbolo «=».

Sea $\mathcal{V}$ un vocabulario. Se construyen los términos y las fórmulas del correspondiente lenguaje de primer orden mediante un conjunto infinito numerable de símbolos cuyos elementos se llaman *variables* (notación: $x$, $y$, $z$, etc.) Se supone que el conjunto de las variables es infinito numerable, y disjunto del vocabulario considerado.

**Definición 1.4** (Términos)**.** Se llama *término* (notación: $t$, $u$, $v$, etc.) a toda expresión construida por aplicación finita de las siguientes reglas:

(1) Si $x$ es una variable, entonces $x$ es un término.
(2) Si $f$ es un símbolo de función de aridad $k \geq 0$ y si $t_1, \ldots, t_k$ son $k$ términos, entonces la expresión $f(t_1, \ldots, t_k)$ es un término.

Si $c$ es un símbolo de constante —es decir un símbolo de función de aridad 0—, se escribe $c$ antes de $c()$ el término obtenido aplicando el símbolo de función $c$ a la sucesión vacía de términos. Algunos símbolos de función binarios tales que $+$ o $\times$ se usan con una notación de infijo; así se escribe $t + u$ antes de $+(t, u)$, y $t \times u$ o $tu$ antes de $\times(t, u)$.

**Ejemplos 1.5.** (1) En el lenguaje de la aritmética de Peano, se usan las abreviaturas $0 :\equiv s(0)$, $2 :\equiv s(1)$, $3 :\equiv s(0)$, etc. Así el término $+(\times(s(s(0)), x), s(0))$ se escribe $2x + 1$. Más generalmente, se llama *entero de Peano* $n$ y (todavía) se escribe $n$ al término

$$s^n(0) \;\equiv\; \underbrace{s(\cdots s}_{n}(0)\cdots)$$

que representa el entero natural[4] $n \in \mathbb{N}$ en el lenguaje de la aritmética de Peano.

(2) En el lenguaje de la teoría de conjuntos así como en el lenguaje de la teoría del orden, los únicos términos son las variables.

**Definición 1.6** (Fórmulas)**.** Se llama *fórmula* (notación: $\phi$, $\psi$, $\chi$, etc.) a toda expresión construida por aplicación finita de las siguientes reglas:

(1) Si $p$ es un símbolo de predicado de aridad $k \geq 0$ y si $t_1, \ldots, t_k$ son $k$ términos, entonces la expresión $p(t_1, \ldots, t_k)$ es una fórmula.
(2) Las expresiones $\top$ («obviedad») y $\bot$ («absurdidad») son fórmulas.
(3) Si $\phi$ y $\psi$ son fórmulas, entonces las expresiones $\phi \wedge \psi$ («$\phi$ y $\psi$»), $\phi \vee \psi$ («$\phi$ o $\psi$») y $\phi \Rightarrow \psi$ («$\phi$ implica $\psi$») son fórmulas.
(4) Si $x$ es una variable y si $\phi$ es una fórmula, entonces las expresiones $\forall x\, \phi$ («para todo $x$, tenemos que $\phi$») y $\exists x\, \phi$ («existe $x$ tal que $\phi$») son fórmulas.

**Observaciones 1.7** (Construcciones definidas)**.** (1) En la definición anterior, no introdujimos ni la negación $\neg\phi$ ni la equivalencia lógica $\phi \Leftrightarrow \psi$ como construcciones primitivas, pues éstas siempre se pueden definir[5] mediante las abreviaturas

$$\neg\phi \;:\equiv\; \phi \Rightarrow \bot \qquad \text{y} \qquad \phi \Leftrightarrow \psi \;:\equiv\; (\phi \Rightarrow \psi) \wedge (\psi \Rightarrow \phi)\,.$$

También se usan las abreviaturas $x \neq y \;:\equiv\; \neg(x = y)$, $x \not\leq y \;:\equiv\; \neg(x \leq y)$, $x \notin y \;:\equiv\; \neg(x \in y)$ etc. y $\exists! x\, \phi(x) \;:\equiv\; \exists x\, (\phi(x) \wedge \forall y\, (\phi(y) \Rightarrow y = x))$.

[4]En este artículo, se adopta la convención que $\mathbb{N} = \{0, 1, 2, 3, \ldots\}$ [1, 11].

[5]Estas definiciones son válidas en lógica clásica como en lógica intuicionista [7] (que no presentaremos aquí).

(2) En el lenguaje de la aritmética de Peano, se define el orden por

$$t \leq u \ :\equiv \ \exists x\,(t + x = u)\,.$$

(3) En el lenguaje de la teoría de conjuntos, se definen la relación de inclusión y las cuantificaciones restringidas mediante las abreviaturas

$$x \subseteq y \ :\equiv \ \forall z\,(z \in x \Rightarrow z \in y) \qquad \begin{array}{l}(\forall x \in y)\,\phi(x) \ :\equiv \ \forall x\,(x \in y \Rightarrow \phi(x)) \\ (\exists x \in y)\,\phi(x) \ :\equiv \ \exists x\,(x \in y \wedge \phi(x))\end{array}$$

*Variables libres y variables ligadas.* Los cuantificadores $\forall$ y $\exists$ son símbolos particulares, que ligan todas las ocurrencias (libres) de la variable cuantificada en el alcance del cuantificador, de tal modo que dicha variable no sea más «visible» afuera de la fórmula cuantificada. Así, se dice que una ocurrencia de una variable $x$ está *ligada* en una fórmula $\phi$ cuando figura en el alcance de una cuantificación de la forma $\forall x$ o $\exists x$ (con el mismo nombre de variable). Todas las otras ocurrencias de la variable $x$ son dichas *libres*. Se observa que:

- todas las ocurrencias de una variable $x$ en un término $t$ son libres, así como todas sus ocurrencias en una fórmula atómica de la forma $p(t_1, \ldots, t_k)$;
- toda ocurrencia libre (resp. ligada) de una variable $x$ en una de las dos fórmulas $\phi$ o $\psi$ sigue siendo libre (resp. ligada) en las fórmulas $\phi \wedge \psi$, $\phi \vee \psi$ y $\phi \Rightarrow \psi$;
- toda ocurrencia libre (resp. ligada) de una variable $y \not\equiv x$ en la fórmula $\phi$ sigue siendo libre (resp. ligada) en las fórmulas $\forall x\,\phi$ y $\exists x\,\phi$, pero toda ocurrencia libre de la variable $x$ en la fórmula $\phi$ se vuelve ligada en las fórmulas $\forall x\,\phi$ y $\exists x\,\phi$.

Se dice que una variable $x$ es *libre* en un término $t$ (resp. en una fórmula $\phi$) cuando tiene al menos una ocurrencia libre en dicho término (resp. en dicha fórmula), y se escribe $FV(t)$ (resp. $FV(\phi)$) al conjunto finito de las variables libres de $t$ (resp. de $\phi$). Se dice que un término $t$ está *cerrado* cuando $FV(t) = \varnothing$; si no, se dice que $t$ está *abierto*. De mismo modo, se dice que una fórmula $\phi$ está *cerrada* cuando $FV(\phi) = \varnothing$; si no, se dice que $\phi$ está *abierta*. (Así, las expresiones cerradas no dependen de ningún contexto.)

*Alfa-equivalencia.* En una fórmula de la forma $\forall x\,\phi$ o $\exists x\,\phi$, la variable cuantificada es *muda*, y se puede cambiar su nombre sin cambiar el sentido de la fórmula. Así, la fórmula $\exists x\,(x > 3)$ tiene el mismo sentido que la fórmula $\exists y\,(y > 3)$, aunque las fórmulas $x > 3$ e $y > 3$ tengan sentidos distintos (suponiendo que las variables $x$ e $y$ son distintas).

Se llama *$\alpha$-equivalencia* a la relación que expresa que dos fórmulas son idénticas a menos de los nombres de sus variables ligadas. (La definición formal [13] de esta relación es delicada, pues un cambio de nombre desafortunado puede cambiar completamente el sentido de la fórmula.) En lógica, siempre se trabaja a menos de $\alpha$-equivalencia, considerando que dos fórmulas que sólo difieren por los nombres de sus variables ligadas son idénticas.

*Sustitución.* Dados términos $t$, $u$ y una variable $x$, se escribe $t[x := u]$ al término obtenido sustituyendo en $t$ todas las ocurrencias (libres) de la variable $x$ por el término $u$. (Esta operación se define directamente por recurrencia sobre la estructura del término $t$.)

De mismo modo, dada una fórmula $\phi$, se escribe $\phi[x := u]$ a la fórmula obtenida sustituyendo en $\phi$ todas las ocurrencias libres de la variable $x$ por el término $u$.

¡Cuidado! Antes de efectuar esta operación, puede ser necesario cambiar algunos nombres de variables ligadas en $\phi$, para impedir que una de las variables libres de $u$ esté capturada por una cuantificación que usa el mismo nombre de variable en la fórmula $\phi$. (De hecho, la operación de sustitución en las fórmulas sólo está definida a menos de $\alpha$-equivalencia.)

**Ejemplo 1.8.** Si $\phi$ es la fórmula aritmética $\exists y\,(x = 2y)$ («$x$ es par»), entonces $\phi[x := y + 3]$ es la fórmula $\exists z\,(y + 3 = 2z)$ («$y + 3$ es par») —o cualquier fórmula $\alpha$-equivalente—, pero no puede ser la fórmula $\exists y\,(y+3 = 2y)$, que no tiene el sentido deseado.

## 1.2. Noción de teoría de primer orden.

**Definición 1.9** (Teoría de primer orden)**.** Una *teoría de primer orden* $\mathscr{T}$ está definida por:

(1) un lenguaje de primer orden $\mathscr{L}$ (que contiene un símbolo de igualdad);
(2) un conjunto de fórmulas cerradas del lenguaje $\mathscr{L}$, escrito $Ax(\mathscr{T})$, cuyos elementos se llaman los *axiomas* de la teoría $\mathscr{T}$.

Se dice que una fórmula cerrada $\phi \in \mathscr{L}$ es *demostrable* en $\mathscr{T}$ —o que $\phi$ es un *teorema* de $\mathscr{T}$— y se escribe $\mathscr{T} \vdash \phi$ cuando existe una lista finita de axiomas $\phi_1, \ldots, \phi_n \in Ax(\mathscr{T})$ $(n \geq 0)$ tal que la fórmula $\phi_1 \wedge \cdots \wedge \phi_n \Rightarrow \phi$ (o el secuente $\phi_1, \ldots, \phi_n \vdash \phi$) tiene una derivación en lógica clásica. El conjunto de los teoremas de la teoría $\mathscr{T}$ se escribe $Th(\mathscr{T})$.

La definición anterior depende de la noción de *derivación*, que definiremos formalmente en la Sección 1.4 más abajo. Por el momento, basta precisar que una derivación es un objeto combinatorio finito, que sólo puede usar una cantidad finita de axiomas de la teoría $\mathscr{T}$.

**Ejemplos 1.10** (Teorías de primer orden usuales)**.**

(1) La *teoría de grupos*[6] es la teoría de primer orden cuyo lenguaje está construido sobre el vocabulario del Ejemplo 1.2 (1), y cuyos axiomas son:

$$\forall x\,\forall y\,\forall z\,((x * y) * z = x * (y * z))$$
$$\forall x\,(x * e = x \wedge e * x = x)$$
$$\forall x\,(x * I(x) = e \wedge I(x) * x = e)$$

(2) La *teoría de anillos* es la teoría de primer orden cuyo lenguaje está construido sobre el vocabulario del Ejemplo 1.2 (2), y cuyos axiomas son:

$$\forall x\,\forall y\,\forall z\,((x + y) + z = x + (y + z))$$
$$\forall x\,\forall y\,(x + y = y + x)$$
$$\forall x\,(x + 0 = x)$$
$$\forall x\,(x + (-x) = 0)$$
$$\forall x\,\forall y\,\forall z\,(x \times (y \times z) = x \times (y \times z))$$
$$\forall x\,(x \times 1 = x \wedge 1 \times x = x)$$
$$\forall x\,\forall y\,\forall z\,((x + y) \times z = x \times z + y \times z)$$
$$\forall x\,\forall y\,\forall z\,(z \times (x + y) = z \times x + z \times y)$$

(3) La *teoría del orden* es la teoría de primer orden cuyo lenguaje está construido sobre el vocabulario del Ejemplo 1.2 (3), y cuyos axiomas son:

$$\forall x\,(x \leq x)$$
$$\forall x\,\forall y\,\forall z\,(x \leq y \wedge y \leq z \Rightarrow x \leq z)$$
$$\forall x\,\forall y\,(x \leq y \wedge y \leq x \Rightarrow x = y)$$

[6]La terminología de «teoría de grupos» es abusiva, pues dicha teoría sólo describe *un* grupo —pero un grupo genérico. La misma observación vale para la teoría de anillos.

(4) La *aritmética de Peano*[7] (notación: PA) es la teoría de primer orden cuyo lenguaje está construido sobre el vocabulario del Ejemplo 1.2 (4), y cuyos axiomas son:

$$\begin{array}{ll} \forall x\,(x+0=x) & \forall x\,\forall y\,(x+s(y)=s(x+y)) \\ \forall x\,(x\times 0=0) & \forall x\,\forall y\,(x\times s(y)=x\times y+x) \\ \forall x\,(s(x)\neq 0) & \forall x\,\forall y\,(s(x)=s(y)\Rightarrow x=y) \end{array}$$

$$\forall \vec{z}\,[\phi(0,\vec{z})\ \wedge\ \forall x\,(\phi(x,\vec{z})\Rightarrow\phi(s(x),\vec{z}))\ \Rightarrow\ \forall x\,\phi(x,\vec{z})]$$

donde $\phi(x,\vec{z})$ es cualquier fórmula cuyas variables libres ocurren en el conjunto $\{x,\vec{z}\}$[8]. Se observa que la última fórmula —el principio de inducción— no es un axioma, pero un esquema de axiomas, que define un axioma para cada fórmula $\phi\equiv\phi(x,\vec{z})$ tal que $FV(\phi)\subseteq\{x,\vec{z}\}$. Por lo tanto, el conjunto de axiomas de la aritmética de Peano es infinito.

(5) La *teoría de conjuntos de Zermelo-Fraenkel*[9] (notación: ZF) es la teoría de primer orden cuyo vocabulario es el del Ejemplo 1.2 (5), y cuyos axiomas son:

$$\begin{array}{ll} \text{(Extensión)} & \forall x\,\forall y\,[\forall z\,(z\in x\Leftrightarrow z\in y)\ \Rightarrow\ x=y] \\ \text{(Par)} & \forall a\,\forall b\,\exists c\,\forall x\,(x\in c\ \Leftrightarrow\ x=a\vee x=b) \\ \text{(Comprensión)} & \forall\vec{z}\,\forall a\,\exists b\,\forall x\,(x\in b\ \Leftrightarrow\ x\in a\wedge\phi(x,\vec{z})) \\ \text{(Unión)} & \forall a\,\exists b\,\forall x\,(x\in b\ \Leftrightarrow\ (\exists y\in a)(x\in y)) \\ \text{(Potencia)} & \forall a\,\exists b\,\forall x\,(x\in b\ \Leftrightarrow\ x\subseteq a) \\ \text{(Infinito)} & \exists a\,[(\exists x\in a)\,\forall z\,(z\notin x)\wedge(\forall x\in a)(\exists y\in a)\,\forall z\,(z\in y\Leftrightarrow z\in x\vee z=x)] \\ \text{(Remplazo)} & \forall\vec{z}\,\forall a\,[(\forall x\in a)\exists!y\,\psi(x,y,\vec{z})\ \Rightarrow\ \exists b\,(\forall x\in a)(\exists y\in b)\,\psi(x,y,\vec{z})] \\ \text{(Fundación)} & \forall\vec{a}\,[\exists x\,(x\in a)\ \Rightarrow\ (\exists x\in a)(\forall y\in a)(y\notin x)] \end{array}$$

donde $\phi(x,\vec{z})$ y $\psi(x,y,\vec{z})$ son fórmulas cualesquiera cuyas variables libres ocurren en los conjuntos $\{x,\vec{z}\}$ y $\{x,y,\vec{z}\}$, respectivamente. Como para la aritmética de Peano, el conjunto de axiomas de la teoría de conjuntos de Zermelo-Fraenkel es infinito.

**Observaciones 1.11.** (1) A pesar de su lenguaje minimalista (que sólo permite hablar de enteros naturales) y de su sistema axiomático muy sencillo, la aritmética de Peano (PA) es un formalismo muy expresivo, en donde se puede expresar y demostrar gran parte de los resultados de la aritmética y de la combinatoria. Por ejemplo, usando las abreviaturas

$$\begin{array}{rcl} x\geq y & :\equiv & \exists z\,(x=y+z) \\ \mathrm{Prim}(x) & :\equiv & x\neq 1\wedge\forall x_1\,\forall x_2\,(x=x_1\times x_2\Rightarrow x_1=1\vee x_2=1)\,, \end{array}$$

se puede expresar y demostrar el teorema de Euclides en PA:

$$\mathrm{PA}\ \vdash\ \forall x\,\exists y\,(y\geq x\wedge\mathrm{Prim}(y))\,.$$

Además, aunque los términos de PA sólo puedan representar expresiones polinomiales, el lenguaje de fórmulas de PA es bastante expresivo para representar todas las

[7] Al honor de Giuseppe Peano (1858–1932), que introdujo la primera axiomatización [15] de la aritmética.

[8] Aquí, $x$ es la variable de inducción, mientras $\vec{z}\equiv z_1,\ldots,z_n$ indica cualquier lista finita (posiblemente vacía) de variables (distintas a pares), que son los parámetros de inducción.

[9] La primera axiomatización de la teoría de conjuntos moderna fue propuesta en 1908 por Ernst Zermelo (1871–1953) —que también introdujo el axioma de elección— y completada en 1922 por Thoralf Skolem (1887–1963) y Abraham Fraenkel (1891–1965), que introdujeron independientemente el esquema de remplazo.

funciones computables como relaciones. Esta propiedad fundamental de PA es un ingrediente central de la prueba del primer teorema de incompletitud de Gödel (véase Sección 3.3 más abajo).

(2) La teoría de conjuntos de Zermelo-Fraenkel (ZF) es una de las teorías de primer orden más potentes[10], pues permite expresar y demostrar virtualmente todos los resultados de la matemática —a veces con la ayuda del *axioma de elección* (AC). Cabe destacar que algunos autores [12] consideran que el sistema ZF no contiene el axioma de Fundación[11], que sólo tiene interés en teoría de conjuntos, pero nunca se usa en la matemática.

*El problema de la consistencia.* Se dice que una teoría de primer orden $\mathscr{T}$ es *consistente* cuando la fórmula absurda $\bot$ no es demostrable en $\mathscr{T}$, es decir: cuando $\mathscr{T} \nvdash \bot$. En caso contrario, se dice que $\mathscr{T}$ es *inconsistente*.

En lógica, la fórmula absurda $\bot$ es la fórmula más fuerte, en el sentido en que implica cualquier otra fórmula (véase Ejemplos 1.20 (2) más abajo). Por lo tanto, una teoría de primer orden $\mathscr{T}$ es inconsistente si y sólo si todas la fórmulas cerradas (de su lenguaje) son demostrables en $\mathscr{T}$. Y al contrario, una teoría de primer orden $\mathscr{T}$ es consistente si y sólo si existe al menos una fórmula cerrada no demostrable en $\mathscr{T}$.

Dadas una teoría de primer orden $\mathscr{T}$ y una fórmula cerrada $\phi$ del lenguaje de $\mathscr{T}$, se escribe $\mathscr{T} + \phi$ la teoría obtenida añadiendo la fórmula $\phi$ a los axiomas de $\mathscr{T}$ (sin cambiar el lenguaje). En lo siguiente, usaremos frecuentemente el siguiente lema:

**Lema 1.12.** *Sean $\mathscr{T}$ una teoría de primer orden y $\phi$ una fórmula cerrada de su lenguaje.*

(1) *La teoría $\mathscr{T} + \phi$ es consistente si y sólo si $\mathscr{T} \nvdash \neg\phi$ ($\neg\phi$ no es demostrable en $\mathscr{T}$).*
(2) *La teoría $\mathscr{T} + \neg\phi$ es consistente si y sólo si $\mathscr{T} \nvdash \phi$ ($\phi$ no es demostrable en $\mathscr{T}$).*

**1.3. Subteorías, extensiones y extensiones conservativas.** Sean $\mathscr{T}$ y $\mathscr{T}'$ teorías de primer orden sobre lenguajes $\mathscr{L}$ y $\mathscr{L}'$, respectivamente. Se dice que $\mathscr{T}$ es una *subteoría* de $\mathscr{T}'$ o que $\mathscr{T}'$ es una *extensión* de $\mathscr{T}$ y se escribe $\mathscr{T} \subseteq \mathscr{T}'$ cuando:

(1) el lenguaje de $\mathscr{T}$ está incluido en el lenguaje de $\mathscr{T}'$: $\mathscr{L} \subseteq \mathscr{L}'$;
(2) todo teorema de $\mathscr{T}$ también es un teorema de $\mathscr{T}'$: $Th(\mathscr{T}) \subseteq Th(\mathscr{T}')$.

(Para (2), basta verificar que $Ax(\mathscr{T}) \subseteq Th(\mathscr{T}')$.)

Es claro que toda subteoría de una teoría consistente es consistente, mientras toda extensión de una teoría inconsistente es inconsistente. (En particular, se puede mostrar que todas las teorías sin axiomas son consistentes.) Por otro lado, una extensión de una teoría consistente puede ser inconsistente, lo que justifica la siguiente restricción de la noción de extensión:

[10]Sin embargo, existen teorías de primer orden más potentes, que se obtienen añadiendo a ZF axiomas de existencia de ciertos cardinales grandes (es decir cardinales más grandes que todos los que se pueden construir efectivamente en ZF). El estudio de estas extensiones de ZF es el objeto de la *teoría de los cardinales grandes*.

[11]El axioma de fundación expresa que la relación de pertenencia está bien fundada, es decir: que no existe ninguna sucesión infinita de conjuntos tales que $x_0 \ni x_1 \ni x_2 \ni x_3 \ni \cdots$. En teoría de conjuntos, el axioma de Fundación sirve para «regularizar» la estructura del universo conjuntista (prohibiendo la existencia de conjuntos mal fundados), pero no se necesita en la matemática, pues los conjuntos construidos por los matemáticos son naturalmente bien fundados. El axioma de Fundación no cambia la fuerza teórica de ZF, en el sentido en que el sistema sin Fundación es equiconsistente al sistema con Fundación.

**Definición 1.13** (Extensión conservativa)**.** Sean $\mathscr{T}$ y $\mathscr{T}'$ teorías de primer orden sobre lenguajes $\mathscr{L}$ y $\mathscr{L}'$, respectivamente. Se dice que $\mathscr{T}'$ es una *extensión conservativa* de $\mathscr{T}$ (notación: $\mathscr{T} \subseteq_{\text{cons}} \mathscr{T}'$) cuando $\mathscr{L} \subseteq \mathscr{L}'$ y $Th(\mathscr{T}) = Th(\mathscr{T}') \cap \mathscr{L}$.

Dicho de otro modo, una extensión conservativa de una teoría $\mathscr{T}$ es una extensión de $\mathscr{T}$ que no afecta el conjunto de los teoremas expresables en el lenguaje de $\mathscr{T}$. Y como la fórmula absurda $\bot$ pertenece a todos los lenguajes de primer orden, es claro que:

**Proposición 1.14.** *Si $\mathscr{T}'$ es una extensión conservativa de $\mathscr{T}$, entonces las teorías $\mathscr{T}$ y $\mathscr{T}'$ son equiconsistentes, en el sentido en que $\mathscr{T}' \vdash \bot$ si y sólo si $\mathscr{T} \vdash \bot$.*

Como el señor Jourdain [14] practicaba la prosa sin saberlo, los matemáticos construyen todos los días extensiones conservativas sin darse cuenta, cada vez que introducen una nueva notación. Aquí están dos ejemplos muy útiles de extensiones conservativas:

*Extensiones de Henkin.* Sean $\mathscr{T}$ una teoría de primer orden y $\phi$ un teorema existencial de $\mathscr{T}$, de la forma $\phi \equiv \exists x\, \psi(x)$. Se llama *extensión de Henkin* de la teoría $\mathscr{T}$ respecto al teorema $\phi$ a la teoría $\mathscr{T}'$ obtenida añadiendo a $\mathscr{T}$ un nuevo símbolo de constante $c$ (llamado *constante de Henkin*) así como el nuevo axioma $\psi(c)$. Se demuestra sin dificultad que:

**Teorema 1.15.** *Toda extensión de Henkin es conservativa.*

Por ejemplo, se puede construir en ZF una fórmula $\psi(x)$ (muy larga) que expresa que «$x$ es un cuerpo totalmente ordenado completo», y también se puede demostrar (en ZF) la fórmula cerrada $\exists x\, \psi(x)$ («existe un cuerpo totalmente ordenado completo»). En general, justo después de haber demostrado el teorema anterior, los matemáticos asignan un nombre a un cuerpo totalmente ordenado completo arbitrario[12], diciendo:

*Ahora, se escribe $\mathbb{R}$ a un cuerpo totalmente ordenado completo.*

Diciendo esto, construyen una extensión de Henkin de ZF, introduciendo un nuevo símbolo de constante «$\mathbb{R}$» y el nuevo axioma $\psi(\mathbb{R})$ («$\mathbb{R}$ es un cuerpo totalmente ordenado completo»). Así, cada vez que usan el símbolo «$\mathbb{R}$», los matemáticos no trabajan más en ZF, pero en una extensión particular de ZF —con un símbolo más y un axioma más. Por suerte, dicha extensión es conservativa, lo que significa que no cambia la clase de los teoremas expresables en el lenguaje anterior (sin el símbolo $\mathbb{R}$). Además, como la relación $\mathscr{T} \subseteq_{\text{cons}} \mathscr{T}'$ es transitiva, se puede iterar el proceso para introducir sucesivamente todos los símbolos de constante usuales en matemática: «$\varnothing$», «$\mathbb{N}$», «$\mathbb{Z}$», «$\mathbb{Q}$», «$\mathbb{C}$», «0», «1», «2», «3», «$\pi$», «$e$», «$i$», etc.

*Extensiones definicionales.* Consideremos ahora un teorema $\phi$ de $\mathscr{T}$ de la forma

$$\phi \;\equiv\; \forall x_1 \cdots \forall x_k\, \exists! y\, \psi(x_1, \ldots, x_k, y)\,.$$

Se llama *extensión definicional* de la teoría $\mathscr{T}$ respecto al teorema $\phi$ a la teoría $\mathscr{T}''$ obtenida añadiendo a $\mathscr{T}$ un nuevo símbolo de función $f$ de aridad $k$ (llamado *función*

[12]Recordemos que «el» cuerpo totalmente ordenado completo de ZF sólo es único *a menos de isomorfismo (único)*, pero no es único en el sentido de la igualdad conjuntista. En efecto, existen múltiples implementaciones de dicho cuerpo: por las sucesiones de Cauchy, por las cortaduras de Dedekind, etc.

*de Skolem*) así como el nuevo axioma

$$\forall x_1 \cdots \forall x_k \, \psi(x_1, \ldots, x_k, f(x_1, \ldots, x_k)) \, .$$

Intuitivamente, el símbolo $f$ representa la función que asocia a cada $k$-upla $x_1, \ldots, x_k$ de objetos el único objeto $y$ tal que $\psi(x_1, \ldots, x_k, y)$. De vuelta, se demuestra que:

**Teorema 1.16.** *Toda extensión definicional es conservativa.*

Por ejemplo, la fórmula

$$\psi(x, y, z) \ :\equiv \ \forall u \, (u \in z \Leftrightarrow u \in x \vee u \in y)$$

expresa en el lenguaje de ZF que «$z$ es la unión de $x$ e $y$». Y como

$$\mathrm{ZF} \ \vdash \ \forall x \, \forall y \, \exists! z \, \psi(x, y, z) \, ,$$

se puede considerar la correspondiente extensión definicional, que añade a ZF un nuevo símbolo de función «$_ \cup _$» de aridad 2 y el nuevo axioma $\forall x \, \forall y \, \psi(x, y, x \cup y)$, es decir:

$$\forall x \, \forall y \, \forall u \, (u \in x \cup y \Leftrightarrow u \in x \vee u \in y) \, .$$

(Por el teorema anterior, dicha extensión de ZF es conservativa.)

De mismo modo se introducen los símbolos de función usuales en teoría de conjuntos: «$\{_\}$» (conjunto unitario), «$\{_, _\}$» (par no ordenado), «$(_, _)$» (par ordenado), «$_ \cap _$» (intersección binaria), «$\mathfrak{P}(_)$» (conjunto potencia), «$\mathrm{Card}(_)$» (cardinal de un conjunto), etc.

En la práctica, las extensiones de Henkin y las extensiones definicionales sirven esencialmente para acercar los lenguajes formales del lenguaje semiformal usado en la matemática. (Esta facilidad es especialmente importante en teoría de conjuntos, cuyo lenguaje oficial no contiene ningún símbolo de constante o de función.) Lo importante es que tales extensiones no afectan ni el poder expresivo de la teoría (respecto al lenguaje de base) ni su consistencia. Es la razón por qué en lógica como en matemática, se introducen los símbolos de Henkin y de Skolem de modo implícito (sin mencionar las correspondientes extensiones), como si dichos símbolos ya estuvieran en el lenguaje de base.

**1.4. Un sistema de deducción.** Antes de concluir esta presentación de las teorías de primer orden, se necesita decir algunas palabras sobre la noción de deducción en lógica clásica. Ésta se puede definir mediante múltiples sistemas formales (sistemas de Hilbert [10], deducción natural [6, 16], cálculo de secuentes de Gentzen [6, 7], etc.) que últimamente definen la misma noción de consecuencia lógica. Por razones pedagógicas, adoptamos aquí el sistema de *deducción natural clásica* (notación: NK) con secuentes explícitos. Formalmente:

**Definición 1.17** (Secuentes)**.** Se llama *secuente* a todo par $(\Gamma, \phi)$ escrito

$$\Gamma \vdash \phi \qquad (\text{“}\Gamma \text{ lleva a } \phi\text{”})$$

donde $\Gamma$ es una lista finita de fórmulas (posiblemente vacía) y $\phi$ una fórmula.

Intuitivamente, un secuente $\Gamma \vdash \phi$ representa una etapa particular de razonamiento, en la cual se trata de establecer $\phi$ (el *consecuente*) a partir de las hipótesis en $\Gamma$ (el *antecedente*).

En lo siguiente, las listas finitas de fórmulas —también llamadas *contextos*— se escriben $\Gamma \equiv \phi_1, \ldots, \phi_n$ (fórmulas separadas por comas), y la concatenación de dos listas $\Gamma$ y $\Delta$ se escribe $\Gamma, \Delta$. Un secuente con un antecedente vacío se escribe

| | | | |
|---|---|---|---|
| (Axioma) | $\dfrac{}{\phi \vdash \phi}$ | (Debilitamiento) | $\dfrac{\Gamma, \Gamma' \vdash \phi}{\Gamma, \psi, \Gamma' \vdash \phi}$ |
| ($\top$-intro) | $\dfrac{}{\vdash \top}$ | (Absurdo) | $\dfrac{\Gamma, \neg\phi \vdash \bot}{\Gamma \vdash \phi}$ |
| ($\Rightarrow$-intro) | $\dfrac{\Gamma, \phi \vdash \psi}{\Gamma \vdash \phi \Rightarrow \psi}$ | ($\Rightarrow$-elim) | $\dfrac{\Gamma \vdash \phi \Rightarrow \psi \quad \Gamma \vdash \phi}{\Gamma \vdash \psi}$ |
| ($\wedge$-intro) | $\dfrac{\Gamma \vdash \phi \quad \Gamma \vdash \psi}{\Gamma \vdash \phi \wedge \psi}$ | ($\wedge$-elim$_{1,2}$) | $\dfrac{\Gamma \vdash \phi \wedge \psi}{\Gamma \vdash \phi} \qquad \dfrac{\Gamma \vdash \phi \wedge \psi}{\Gamma \vdash \psi}$ |
| ($\vee$-intro$_{1,2}$) | $\dfrac{\Gamma \vdash \phi}{\Gamma \vdash \phi \vee \psi} \qquad \dfrac{\Gamma \vdash \psi}{\Gamma \vdash \phi \vee \psi}$ | ($\vee$-elim) | $\dfrac{\Gamma \vdash \phi \vee \psi \quad \Gamma, \phi \vdash \chi \quad \Gamma, \psi \vdash \chi}{\Gamma \vdash \chi}$ |
| ($\forall$-intro) | $\dfrac{\Gamma \vdash \phi}{\Gamma \vdash \forall x\, \phi}\ (x \notin FV(\Gamma))$ | ($\forall$-elim) | $\dfrac{\Gamma \vdash \forall x\, \phi}{\Gamma \vdash \phi[x := t]}$ |
| ($\exists$-intro) | $\dfrac{\Gamma \vdash \phi[x := t]}{\Gamma \vdash \exists x\, \phi}$ | ($\exists$-elim) | $\dfrac{\Gamma \vdash \exists x\, \phi \quad \Gamma, \phi \vdash \psi}{\Gamma \vdash \psi}\ (x \notin FV(\Gamma, \psi))$ |
| (=-intro) | $\dfrac{}{\vdash t = t}$ | (=-elim) | $\dfrac{\Gamma \vdash t = u \quad \Gamma \vdash \phi[x := t]}{\Gamma \vdash \phi[x := u]}$ |

CUADRO 1. Reglas de inferencia de la deducción natural clásica (NK)

simplemente $\vdash \phi$ (antes de $\emptyset \vdash \phi$). Las notaciones $FV(\Gamma)$ (conjunto de variables libres) y $\Gamma[x := u]$ (sustitución) se extienden a las listas finitas de fórmulas del modo obvio: si $\Gamma \equiv \phi_1, \ldots, \phi_n$, entonces

$$\begin{array}{rcl} FV(\Gamma) & := & FV(\phi_1) \cup \cdots \cup FV(\phi_n) \\ \Gamma[x := u] & :\equiv & \phi_1[x := u], \ldots, \phi_n[x := u] \end{array}$$

La deducción natural clásica está definida a partir de las 18 *reglas de deducción* (o *reglas de inferencia*) dadas en el Cuadro 1. Cada (instancia de una) regla de deducción es de la forma

$$\frac{\Gamma_1 \vdash \phi_1 \quad \cdots \quad \Gamma_p \vdash \phi_p}{\Gamma \vdash \phi}$$

donde los secuentes $\Gamma_1 \vdash \phi_1, \ldots, \Gamma_p \vdash \phi_p$ se llaman las *premisas* de la regla, y el secuente $\Gamma \vdash \phi$ la *conclusión* de la regla. Intuitivamente, tal regla expresa un paso de deducción que permite establecer la validez de la conclusión a partir de la validez de las premisas, es decir:

*como:* $\phi_1$ *se cumple (bajo las hipótesis* $\Gamma_1$*), y*
$\vdots$
$\phi_p$ *se cumple (bajo las hipótesis* $\Gamma_p$*);*
*luego:* $\phi$ *se cumple (bajo las hipótesis* $\Gamma$*)*

Algunas reglas tienen una *condición de borde* (por la derecha), que restringe el uso de la regla a elementos sintácticos con una forma particular. Por ejemplo, la regla de introducción de $\forall$ (generalización) restringe el uso de la regla a los contextos $\Gamma$ en los cuales la variable $x$ no tiene ocurrencia libre. Intuitivamente, esta condición expresa que la variable $x$ representa un objeto "cualquiera", sobre el cual no se plantea ninguna hipótesis.

Las reglas de deducción son los ladrillos elementales del razonamiento formal, y se pueden agregar hasta formar *árboles de derivación*, o *derivaciones*. Formalmente:

**Definición 1.18** (Derivaciones)**.** — Se llama *derivación* a todo árbol finito de secuentes construido por aplicación finita de la siguiente regla de agregación:

- Si
$$\begin{array}{ccc} \vdots\, d_1 & & \vdots\, d_p \\ \Gamma_1 \vdash \phi_1 & \cdots & \Gamma_p \vdash \phi_p \end{array}$$
son derivaciones de los secuentes $\Gamma_1 \vdash \phi_1, \ldots, \Gamma_p \vdash \phi_p$ (respectivamente), y si
$$\frac{\Gamma_1 \vdash \phi_1 \quad \cdots \quad \Gamma_p \vdash \phi_p}{\Gamma \vdash \phi}$$
es (una instancia de) una de las 18 reglas del Cuadro 1, entonces el árbol
$$d \equiv \left\{ \begin{array}{c} \begin{array}{ccc} \vdots\, d_1 & & \vdots\, d_p \\ \Gamma_1 \vdash \phi_1 & \cdots & \Gamma_p \vdash \phi_p \end{array} \\ \hline \Gamma \vdash \phi \end{array} \right.$$
es una derivación del secuente $\Gamma \vdash \phi$.

Cuando tal derivación existe, se dice que el secuente $\Gamma \vdash \phi$ es *derivable*. Y cuando el secuente $\vdash \phi$ (sin hipótesis) es derivable, se dice que la fórmula $\phi$ es *derivable*.

**Analogía 1.19.** Se puede ver un sistema de deducción como un juego de construcción en el cual cada regla de deducción define una familia (infinita) de ladrillos. En este juego, cada ladrillo tiene un único enchufe por debajo (la conclusión) y cero, una o múltiples tomas por encima (las premisas). Además, las "formas" de los enchufes y de las tomas están determinadas por las correspondientes fórmulas. Según esta analogía, una derivación es un montaje de ladrillos en el cual sólo se queda un enchufe libre (por debajo): la conclusión de la derivación.

**Ejemplos 1.20.** (1) El siguiente árbol
$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{}{x=y \vdash x=y}\,(\text{axioma}) \quad \dfrac{\dfrac{}{\vdash x=x}\,(=\text{-intro})}{x=y \vdash x=x}\,(\text{debil.})}{x=y \vdash y=x}\,(=\text{-elim, con } \phi(z) \equiv z{=}x)}{\vdash x=y \Rightarrow y=x}\,(\Rightarrow\text{-intro})}{\vdash \forall y\,(x=y \Rightarrow y=x)}\,(\forall\text{-intro})}{\vdash \forall x\,\forall y\,(x=y \Rightarrow y=x)}\,(\forall\text{-intro})$$
es una derivación de la fórmula $\forall x\,\forall y\,(x=y \Rightarrow y=x)$ (sin hipótesis) en el sistema NK.

(2) El siguiente árbol

$$\dfrac{\dfrac{\dfrac{\dfrac{}{\bot \vdash \bot}\,(\text{axioma})}{\bot, \neg\phi \vdash \bot}\,(\text{debil.})}{\bot \vdash \phi}\,(\text{absurdo})}{\vdash \bot \Rightarrow \phi}\,(\Rightarrow\text{-intro})$$

es una derivación de la implicación $\bot \Rightarrow \phi$ en el sistema NK (donde $\phi$ es cualquier fórmula).

En fin, se puede asociar a cada secuente $\phi_1, \ldots, \phi_n \vdash \phi$ la fórmula $\phi_1 \wedge \cdots \wedge \phi_n \Rightarrow \phi$. Ésta tiene el mismo sentido lógico que el secuente $\phi_1, \ldots, \phi_n \vdash \phi$, en la medida en que:

**Proposición 1.21** (Equivalencia)**.** *El secuente $\phi_1, \ldots, \phi_n \vdash \phi$ es derivable en el sistema NK si y sólo si la fórmula asociada $\phi_1 \wedge \cdots \wedge \phi_n \Rightarrow \phi$ es derivable (sin hipótesis) en NK.*

Así, los secuentes no añaden nada al poder expresivo del lenguaje de fórmulas, y sólo constituyen una estructura de datos cómoda para describir el proceso de descomposición de las fórmulas durante la construcción de una derivación («búsqueda de demostración»). Es la razón por qué en el estudio de las teorías de primer orden, siempre se pueden remplazar los secuentes por las fórmulas asociadas, así como lo haremos en lo siguiente.

## 2. Modelos de Tarski

En la Sección 1, dimos una presentación puramente sintáctica de la noción de teoría de primer orden, donde sólo hablamos de los mecanismos que permiten formar las expresiones del lenguaje (los términos y las fórmulas) así como de las reglas de deducción que permiten construir las derivaciones. Sin embargo, los *símbolos* de función y los *símbolos* de predicado de una teoría de primer orden refieren implícitamente a *funciones* y a *predicados* (o *relaciones*) cuyo comportamiento está definido —o al menos delimitado— por los axiomas de dicha teoría.

El primer objetivo de la *teoría de modelos* —que fue iniciada en los años 1920 por los trabajos de Leopold Löwenheim (1878-1957), Thoralf Skolem (1887–1963), Alfred Tarski (1901–1983) y Kurt Gödel (1906–1978)— es explicitar, estudiar y clasificar las estructuras matemáticas a las cuales se refieren las teorías de primer orden.

### 2.1. Interpretación de un lenguaje de primer orden.

**Definición 2.1** (Interpretación de un lenguaje de primer orden)**.** Una *interpretación $\mathscr{M}$ de un lenguaje de primer orden $\mathscr{L}$* está definida por:

- un conjunto de base $\mathscr{M} \neq \varnothing$, llamado el *dominio* de la interpretación;
- una función $f^{\mathscr{M}} : \mathscr{M}^k \to \mathscr{M}$ para cada símbolo de función $f \in \mathscr{L}$ de aridad $k \geq 0$;
- una relación $p^{\mathscr{M}} \subseteq \mathscr{M}^k$ para cada símbolo de predicado $p \in \mathscr{L}$ de aridad $k \geq 0$.

Por convención, se supone que el predicado de igualdad «=» del lenguaje $\mathscr{L}$ está interpretado por la relación de igualdad en el conjunto $\mathscr{M}$, es decir: $(=)^{\mathscr{M}} = \{(a, a) : a \in \mathscr{M}\}$.

**Observación 2.2.** En lógica de primer orden, el «universo del discurso» (es decir: el dominio de las cuantificaciones) nunca es vacío, pues las reglas de deducción de la lógica clásica (véase Cuadro 1) permiten demostrar la fórmula $\exists x\,(x = x)$, que expresa la existencia de un objeto. Es la razón por qué el dominio de una interpretación $\mathscr{M}$ nunca puede ser vacío.

Más generalmente, una interpretación $\mathscr{M}$ de un lenguaje de primer orden $\mathscr{L}$ permite interpretar cualquier término cerrado de $\mathscr{L}$ por cierto elemento del conjunto $\mathscr{M}$, y cualquier fórmula cerrada $\phi$ de $\mathscr{L}$ por cierto valor de verdad («verdadero» o «falso»). Para ello, se necesita introducir las nociones de término y de fórmula con parámetros en $\mathscr{M}$:

**Definición 2.3** (Términos y fórmulas con parámetros en $\mathscr{M}$)**.** Sean $\mathscr{L}$ un lenguaje de primer orden y $\mathscr{M}$ una interpretación de $\mathscr{L}$. Se llama *término con parámetros en* $\mathscr{M}$ a todo término construido en el vocabulario del lenguaje $\mathscr{L}$ extendido añadiendo un símbolo de constante para cada elemento $a \in \mathscr{M}$. (Este símbolo de constante todavía se escribe $a$.) Las *fórmulas con parámetros en* $\mathscr{M}$ están definidas en la misma extensión del vocabulario de $\mathscr{L}$.

En la práctica, los parámetros sirven para cerrar los términos abiertos y la fórmulas abiertas del lenguaje $\mathscr{L}$, sustituyendo un parámetro tomado en $\mathscr{M}$ a cada variable libre de la expresión considerada. Así, dados $k$ parámetros $a_1, \ldots, a_k \in \mathscr{M}$:

- se asocia a cada término $t \equiv t(x_1, \ldots, x_k)$ del lenguaje $\mathscr{L}$ que sólo depende de las variables $x_1, \ldots, x_k$ el término cerrado con parámetros

$$t(a_1, \ldots, a_k) \;:\equiv\; t[x_1 := a_1, \ldots, x_k := a_k]\,;$$

- se asocia a cada fórmula $\phi \equiv \phi(x_1, \ldots, x_k)$ del lenguaje $\mathscr{L}$ que sólo depende de las variables $x_1, \ldots, x_k$ la fórmula cerrada con parámetros

$$\phi(a_1, \ldots, a_k) \;:\equiv\; \phi[x_1 := a_1, \ldots, x_k := a_k]\,.$$

**Definición 2.4** (Interpretación de un término cerrado con parámetros)**.** Sean $\mathscr{L}$ un lenguaje de primer orden y $\mathscr{M}$ una interpretación de $\mathscr{L}$. A cada término cerrado $t$ con parámetros en $\mathscr{M}$, se asocia un elemento $t^{\mathscr{M}} \in \mathscr{M}$, que se llama la *interpretación del término* $t$ *en* $\mathscr{M}$. Formalmente, el elemento $t^{\mathscr{M}} \in \mathscr{M}$ está definido por recursión sobre el número de símbolos en el término $t$, mediante las siguientes ecuaciones:

$$\begin{aligned} a^{\mathscr{M}} &:= a && (a \in \mathscr{M}) \\ \big(f(t_1, \ldots, t_k)\big)^{\mathscr{M}} &:= f^{\mathscr{M}}(t_1^{\mathscr{M}}, \ldots, t_k^{\mathscr{M}}) && (\text{aridad}(f) = k) \end{aligned}$$

**Definición 2.5** (Interpretación de una fórmula cerrada con parámetros)**.** Sean $\mathscr{L}$ un lenguaje de primer orden y $\mathscr{M}$ una interpretación de $\mathscr{L}$. A cada fórmula cerrada $\phi$ con parámetros en $\mathscr{M}$, se asocia un enunciado $\mathscr{M} \models \phi$ (verdadero o falso), que se lee «$\mathscr{M}$ satisface $\phi$», «$\phi$ se cumple en $\mathscr{M}$» o «$\phi$ es verdadera en $\mathscr{M}$». Formalmente, el enunciado $\mathscr{M} \models \phi$ está definido por recursión sobre el número de conectivas y de cuantificadores en la fórmula cerrada $\phi$, mediante las siguientes

equivalencias:

$$\begin{array}{rcll}
\mathscr{M} \models p(t_1,\dots,t_k) & \text{sii} & (t_1^{\mathscr{M}},\dots,t_k^{\mathscr{M}}) \in p^{\mathscr{M}} & (\text{aridad}(p)=k)\\
\mathscr{M} \models \top & \text{es} & \text{verdadero} & \\
\mathscr{M} \models \bot & \text{es} & \text{falso} & \\
\mathscr{M} \models \phi \wedge \psi & \text{sii} & \mathscr{M} \models \phi \ \text{ y } \ \mathscr{M} \models \psi & \\
\mathscr{M} \models \phi \vee \psi & \text{sii} & \mathscr{M} \models \phi \ \text{ o } \ \mathscr{M} \models \psi & \\
\mathscr{M} \models \phi \Rightarrow \psi & \text{sii} & \mathscr{M} \models \phi \ \text{ implica } \ \mathscr{M} \models \psi & \\
\mathscr{M} \models \forall x\, \phi(x) & \text{sii} & \mathscr{M} \models \phi(a) \ \text{ para todo } a \in \mathscr{M} & \\
\mathscr{M} \models \exists x\, \phi(x) & \text{sii} & \mathscr{M} \models \phi(a) \ \text{ para algún } a \in \mathscr{M} &
\end{array}$$

*Equivalencia elemental e isomorfismo.* Sean $\mathscr{M}_1$ y $\mathscr{M}_2$ dos interpretaciones de un mismo lenguaje de primer orden $\mathscr{L}$. Se dice que:

- $\mathscr{M}_1$ y $\mathscr{M}_2$ son *elementalmente equivalentes* cuando $\mathscr{M}_1$ y $\mathscr{M}_2$ satisfacen exactamente las mismas fórmulas cerradas (sin parámetros) del lenguaje $\mathscr{L}$, es decir: cuando $\mathscr{M}_1 \models \phi$ sii $\mathscr{M}_2 \models \phi$ para toda fórmula cerrada $\phi \in \mathscr{L}$;
- $\mathscr{M}_1$ y $\mathscr{M}_2$ son *isomorfas* cuando existe una biyección entre $\mathscr{M}_1$ y $\mathscr{M}_2$ que transforma todas las funciones $f^{\mathscr{M}_1} : \mathscr{M}_1^k \to \mathscr{M}_1$ y todas las relaciones $p^{\mathscr{M}_1} \subseteq \mathscr{M}_1^k$ asociadas a los símbolos de $\mathscr{L}$ (según la interpretación $\mathscr{M}_1$) en las funciones $f^{\mathscr{M}_2} : \mathscr{M}_2^k \to \mathscr{M}_2$ y las relaciones $p^{\mathscr{M}_2} \subseteq \mathscr{M}_2^k$ asociadas a los mismos símbolos (según la interpretación $\mathscr{M}_2$).

Se verifica fácilmente que dos interpretaciones isomorfas son elementalmente equivalentes, pero el recíproco no se cumple en general.

## 2.2. La noción de modelo.

**Definición 2.6** (Modelo)**.** Sea $\mathscr{T}$ una teoría de primer orden. Se llama *modelo* (o *modelo de Tarski*) de la teoría $\mathscr{T}$ a toda interpretación $\mathscr{M}$ del lenguaje de $\mathscr{T}$ tal que $\mathscr{M} \models \phi$ para todo axioma de $\mathscr{T}$. El enunciado «$\mathscr{M}$ es un modelo de $\mathscr{T}$» se escribe $\mathscr{M} \models \mathscr{T}$.

Por definición, un modelo de una teoría $\mathscr{T}$ satisface todos los axiomas de $\mathscr{T}$, y se verifica fácilmente que esta propiedad se extiende a todos los teoremas de $\mathscr{T}$. Luego:

**Proposición 2.7** (Corrección)**.** *Sean $\mathscr{T}$ una teoría de primer orden y $\phi$ una fórmula cerrada del lenguaje de $\mathscr{T}$. Si $\mathscr{T} \vdash \phi$, entonces $\mathscr{M} \models \phi$ para todo modelo $\mathscr{M} \models \mathscr{T}$.*

**Ejemplos 2.8.** (1) Por lo anterior, una interpretación del lenguaje de la teoría de grupos (Ejemplo 1.2 (1)) es un conjunto $\mathscr{M} \neq \varnothing$ equipado con un elemento $e^{\mathscr{M}} \in \mathscr{M}$, una operación binaria $*^{\mathscr{M}} : \mathscr{M}^2 \to \mathscr{M}$ y una función $I^{\mathscr{M}} : \mathscr{M} \to \mathscr{M}$. (Recordemos que por convención, el símbolo de igualdad siempre está interpretado por la igualdad en $\mathscr{M}$.)

Es claro que tal estructura constituye un modelo de la teoría de grupos (Ejemplo 1.10 (1)) si y sólo si $(\mathscr{M}, *^{\mathscr{M}})$ es un grupo, $e^{\mathscr{M}}$ su elemento neutro, e $I^{\mathscr{M}} : \mathscr{M} \to \mathscr{M}$ la función que asocia a cada elemento de $\mathscr{M}$ su inverso. Además, dos modelos de la teoría de grupos son isomorfos (como interpretaciones del lenguaje de la teoría de grupos) si y sólo si los grupos subyacentes son isomorfos (como grupos).

(2) De mismo modo, los modelos de la teoría de anillos (Ejemplo 1.10 (2)) son los anillos, y dos modelos de la teoría de anillos son isomorfos (como interpretaciones del lenguaje de la teoría de anillos) si y sólo si los anillos subyacentes son isomorfos (como anillos).

(3) Los modelos de la teoría del orden (Ejemplo 1.10 (3)) son los conjuntos ordenados no vacíos, y dos modelos de la teoría del orden son isomorfos (como interpretaciones) si y sólo si los conjuntos ordenados subyacentes son isomorfos (como conjuntos ordenados).

(4) El conjunto (no vacío) $\mathbb{N}$ equipado con el número 0, la función sucesor $n \mapsto n+1$, la suma $(n,m) \mapsto n+m$ y el producto $(n,m) \mapsto nm$ en $\mathbb{N}$ constituye un modelo de la aritmética de Peano (Ejemplo 1.10 (4)), que se llama el *modelo estándar* de PA. Veremos más adelante que este modelo no es único, ni a menos de isomorfismo, ni siquiera a menos de equivalencia elemental. (De hecho, existen modelos de PA con cardinales infinitos arbitrariamente grandes.) Sin embargo, el modelo estándar constituye el modelo más pequeño de PA, en el sentido en que todo modelo de PA contiene un (único) submodelo isomorfo al modelo estándar.

(5) Un modelo de la teoría de conjuntos de Zermelo-Fraenkel (Ejemplo 1.10 (5)) es esencialmente un grafo dirigido $(\mathscr{M}, \in^{\mathscr{M}})$ cuyas aristas representan la relación de pertenencia, y en el cual todos los axiomas de Zermelo-Fraenkel se cumplen (interpretando $\in$ por $\in^{\mathscr{M}}$). Por ejemplo, la validez del axioma del par en el grafo dirigido $(\mathscr{M}, \in^{\mathscr{M}})$ significa que para todo par de vértices $v_1, v_2 \in \mathscr{M}$, existe un (único) vértice $v_3 \in \mathscr{M}$ cuyos solos hijos son $v_1$ y $v_2$. No se sabe construir explícitamente un modelo de ZF, pero el teorema de completitud (Sección 2.3) implica que si ZF es consistente, entonces tal grafo dirigido existe.

**2.3. Completitud semántica.** La Prop. 2.7 expresa que toda fórmula demostrable en una teoría $\mathscr{T}$ se cumple en todos los modelos de $\mathscr{T}$. Por contrarrecíproco, si una fórmula cerrada no se cumple en un modelo de $\mathscr{T}$, entonces no puede ser demostrable en $\mathscr{T}$. Luego:

**Corolario 2.9.** *Toda teoría de primer orden que tiene un modelo es consistente.*

*Demostración.* Tenemos que $\mathscr{M} \not\models \bot$ en cualquier modelo $\mathscr{M} \models \mathscr{T}$, entonces $\mathscr{T} \not\vdash \bot$. □

En la práctica, la construcción de un modelo constituye el método natural para mostrar que una teoría de primer orden es consistente. Por ejemplo, la existencia del modelo estándar de PA (Ejemplos 2.8 (4)) implica que la aritmética de Peano es consistente[13].

De manera más sorprendente, Gödel mostró en su tesis de doctorado [8] el *teorema de completitud*, que expresa el recíproco del Corolario 2.9:

**Teorema 2.10** (Completitud)**.** *Toda teoría de primer orden consistente tiene un modelo.*

De tal modo que una teoría de primer orden es consistente si y sólo si tiene un modelo.

*Arquitectura de la demostración del teorema de completitud.* La demostración del teorema de completitud es un ejemplo notable de transformación de un objeto sintáctico (una teoría de primer orden consistente) en un objeto semántico (un modelo de dicha teoría).

Se efectúa la construcción en tres etapas:

[13]¡Cuidado! La construcción del modelo estándar de PA necesita trabajar en una teoría de conjuntos (ZF por ejemplo) que ya supone la existencia del conjunto $\mathbb{N}$. Aunque tal demostración de la consistencia de PA sea matemáticamente correcta, no puede ser filosóficamente convincente.

1. *Completación de Henkin.* Dada una teoría $\mathscr{T}$ consistente sobre un lenguaje $\mathscr{L}$, se construye una extensión $\mathscr{T}' \supseteq \mathscr{T}$ sobre un lenguaje $\mathscr{L}' \supseteq \mathscr{L}$ tal que:
   *a*) $\mathscr{T}'$ es una teoría *Henkin-completa*, es decir:
   para toda fórmula $\phi(x) \in \mathscr{L}'$ que sólo depende de la variable $x$, si $\mathscr{T}' \vdash \exists x\, \phi(x)$, entonces existe un símbolo de constante $c_\phi \in \mathscr{L}'$ (un «testigo») tal que $\mathscr{T}' \vdash \phi(c_\phi)$.
   *b*) $\mathscr{T}'$ es una *extensión conservativa* de la teoría $\mathscr{T}$, es decir:
   para toda fórmula cerrada $\phi \in \mathscr{L}$ ($\subseteq \mathscr{L}'$), tenemos que $\mathscr{T}' \vdash \phi$ si y sólo si $\mathscr{T} \vdash \phi$.

   Técnicamente, se construye la extensión conservativa $\mathscr{T}' \supseteq_{\text{cons}} \mathscr{T}$ añadiendo símbolos de Henkin (con los correspondientes axiomas) a la teoría $\mathscr{T}$, de tal modo que cada teorema existencial de la teoría $\mathscr{T}'$ tenga un testigo. Además, la propiedad *b*) (conservatividad) nos asegura que la nueva teoría $\mathscr{T}'$ es consistente.

2. *Completación de Gödel.* A partir de la teoría $\mathscr{T}' \supseteq_{\text{cons}} \mathscr{T}$ (sobre el lenguaje $\mathscr{L}' \supseteq \mathscr{L}$), se construye otra extensión $\mathscr{T}'' \supseteq \mathscr{T}'$ (no conservativa) sobre el lenguaje $\mathscr{L}'$ tal que:
   *a*) $\mathscr{T}''$ es una teoría *completa*, es decir:
   para toda fórmula cerrada $\phi \in \mathscr{L}'$: $\quad \mathscr{T}'' \vdash \phi \quad$ o $\quad \mathscr{T}'' \vdash \neg\phi$.
   *b*) $\mathscr{T}''$ es consistente, es decir: $\quad \mathscr{T}'' \nvdash \bot$.

   Para ello, se parte de la teoría $\mathscr{T}'' := \mathscr{T}'$, y se enumera todas las fórmulas cerradas del lenguaje $\mathscr{L}'$. Cada vez que se encuentra una fórmula $\phi$ que no es ni demostrable ni refutable en la teoría $\mathscr{T}''$ en construcción, se añade la fórmula $\phi$ a los axiomas de dicha teoría, antes de pasar a la siguiente fórmula.

3. *Construcción del modelo $\mathscr{M}$.* A partir de la teoría (consistente y completa) $\mathscr{T}''$, se construye el conjunto cociente $\mathscr{M} := \mathscr{U}/\!\sim$, donde $\mathscr{U}$ es el conjunto de los términos cerrados del lenguaje $\mathscr{L}'$ y $\sim$ la relación de equivalencia definida por

$$t \sim u \quad \text{sii} \quad \mathscr{T}'' \vdash t = u \qquad \text{(para todos } t, u \in \mathscr{U}\text{)}$$

   En el conjunto cociente $\mathscr{M} := \mathscr{U}/\!\sim$, se interpreta cada símbolo de función $f \in \mathscr{L}'$ (de aridad $k \geq 0$) por la función $f^{\mathscr{M}} : \mathscr{M}^k \to \mathscr{M}$ definida por

$$f^{\mathscr{M}}([t_1], \ldots, [t_k]) \;:=\; [f(t_1, \ldots, t_k)] \qquad \text{(para todos } t_1, \ldots, t_k \in \mathscr{U}\text{)}$$

   y cada símbolo de predicado $p \in \mathscr{L}'$ (aridad $k \geq 0$) por la relación $p^{\mathscr{M}} \subseteq \mathscr{M}^k$ dada por

$$([t_1], \ldots, [t_k]) \in p^{\mathscr{M}} \quad \text{sii} \quad \mathscr{T}'' \vdash p(t_1, \ldots, t_k) \qquad \text{(para todos } t_1, \ldots, t_k \in \mathscr{U}\text{)}$$

   A partir de las propiedades anteriores, se verifica fácilmente que el conjunto $\mathscr{M}$ equipado con las funciones $f^{\mathscr{M}}$ y las relaciones $p^{\mathscr{M}}$ es una interpretación del lenguaje $\mathscr{L}'$ que captura los teoremas de la teoría (consistente y completa) $\mathscr{T}''$, en el sentido en que

$$\mathscr{M} \models \phi \qquad \text{sii} \qquad \mathscr{T}'' \vdash \phi \qquad \text{(para toda fórmula cerrada } \phi \in \mathscr{L}'\text{)}$$

   Luego, $\mathscr{M} \models \mathscr{T}''$, y por lo tanto $\mathscr{M} \models \mathscr{T}$ (pues $\mathscr{T} \subseteq \mathscr{T}''$).

*Completitud del sistema de deducción.* El teorema de completitud (Teorema 2.10) implica el recíproco de la propiedad de corrección (Prop. 2.7 p. 209):

**Corolario 2.11** (Completitud de la deducción)**.** *Sean $\mathscr{T}$ una teoría de primer orden y $\phi$ una fórmula cerrada del lenguaje de $\mathscr{T}$. Si $\mathscr{M} \models \phi$ para todo modelo $\mathscr{M} \models \mathscr{T}$, entonces $\mathscr{T} \vdash \phi$.*

*Demostración.* Razonando por contrarrecíproco, supongamos que $\mathscr{T} \nvdash \phi$. Por el Lema 1.12, la teoría $\mathscr{T} + \neg\phi$ es consistente y tiene un modelo $\mathscr{M}$ por el teorema de completitud. Pero este modelo $\mathscr{M}$ también es un modelo de $\mathscr{T}$ en el cual $\mathscr{M} \models \neg\phi$, es decir: $\mathscr{M} \not\models \phi$. □

Dicho de otro modo, el resultado anterior expresa que si una fórmula cerrada $\phi$ se cumple en todos los modelos de $\mathscr{T}$ —es decir: en todas las interpretaciones del lenguaje de $\mathscr{T}$ que cumplen los axiomas de $\mathscr{T}$—, entonces la fórmula $\phi$ tiene una derivación formal a partir de los axiomas de $\mathscr{T}$. Intuitivamente, este resultado expresa la completitud del sistema de deducción en lógica clásica (Sección 1.4), es decir: que las reglas del Cuadro 1 p. 205 son suficientes para expresar todas las deducciones lógicas que se necesitan en la matemática.

**2.4. El teorema de compacidad.** Se dice que una teoría de primer orden $\mathscr{T}$ es *finita* cuando su vocabulario es finito, así como su conjunto de axiomas. Por ejemplo, la teoría de grupos, la teorías de anillos y la teoría del orden son finitas, pero la aritmética de Peano y la teoría de conjuntos de Zermelo-Fraenkel no lo son, pues tienen conjuntos de axiomas infinitos.

Ya vimos en la Sección 1.4 que una derivación de una fórmula $\phi$ en la teoría $\mathscr{T}$ es un objeto finito, que sólo involucra una cantidad finita de símbolos del lenguaje de $\mathscr{T}$ y una cantidad finita de axiomas de $\mathscr{T}$. En particular, toda inconsistencia de la teoría $\mathscr{T}$ es una inconsistencia de una subteoría finita de $\mathscr{T}$, de tal modo que:

**Proposición 2.12** (Compacidad sintáctica)**.** *Una teoría de primer orden es consistente si y sólo si todas sus subteorías finitas lo son.*

Pasando por la equivalencia entre las teorías consistentes y las teorías que tienen un modelo (Corolario 2.9 y Teorema 2.10), se deduce que:

**Teorema 2.13** (Compacidad)**.** *Una teoría de primer orden tiene un modelo si y sólo si todas sus subteorías finitas tienen un modelo.*

En la práctica, el teorema de compacidad es una herramienta muy importante para construir modelos. Un ejemplo típico es el siguiente:

*Construcción de un modelo no estándar de PA.* Sea $\mathscr{T}$ la teoría de primer orden cuyo vocabulario es el vocabulario de la aritmética de Peano (Ejemplos 1.2 (4)) extendido con un nuevo símbolo de constante $c$, y cuyos axiomas son los axiomas de la aritmética de Peano (Ejemplos 1.10 (4)), más los siguientes nuevos axiomas:

$$c \neq 0, \qquad c \neq 1, \quad c \neq 2, \quad c \neq 3, \quad \ldots, \quad c \neq s^n(0) \qquad \text{(para todo } n \in \mathbb{N}\text{)}$$

Por construcción, $\mathscr{T}$ es una extensión de PA con una constante que indica un *entero no estándar*, es decir: un entero que no es de la forma $s^n(0)$ para ningún $n \in \mathbb{N}$.

El teorema de compacidad permite demostrar que:

**Proposición 2.14.** *La teoría $\mathscr{T} \supseteq PA$ tiene un modelo.*

*Demostración.* Sea $\mathscr{T}_0$ una subteoría finita cualquiera de $\mathscr{T}$. Como $\mathscr{T}_0$ sólo contiene una cantidad finita de axiomas de la forma $c \neq n$ ($n \in \mathbb{N}$), existe un entero $n_0 \in \mathbb{N}$ tal que $n_0 \neq n$ para todo axioma $(c \neq n) \in Ax(\mathscr{T}_0)$. Consideremos ahora la interpretación $\mathscr{M}_0$ del lenguaje de $\mathscr{T}$ cuyo dominio es $\mathbb{N}$ (el dominio del modelo estándar), y en la cual:

- los símbolos 0, $s$, + y $\times$ están interpretados del modo usual en $\mathbb{N}$;
- el símbolo de constante $c$ está interpretado por el entero $c^{\mathscr{M}} := n_0$.

Por construcción, tenemos que $\mathscr{M}_0 \models \phi$ para todo axioma $\phi \in Ax(\mathscr{T}_0)$, luego $\mathscr{M}_0$ es un modelo de $\mathscr{T}_0$. Por el teorema de compacidad, se concluye que la teoría $\mathscr{T}$ tiene un modelo. □

Por construcción, un modelo $\mathscr{M} \models \mathscr{T}$ es un modelo de la aritmética de Peano que contiene un objeto particular $c^{\mathscr{M}} \in \mathscr{M}$ distinto de todo entero estándar, en el sentido en que

$$\mathscr{M} \;\models\; c \neq s^n(0) \qquad \text{(para todo } n \in \mathbb{N}\text{)}$$

Por lo tanto, se deduce que:

**Proposición 2.15** (Existencia de modelos no estándar)**.** *Cada modelo de la teoría $\mathscr{T}$ constituye un modelo de* PA *no isomorfo al modelo estándar.*

Aquí, es importante entender que desde el punto de vista de la teoría $\mathscr{T} \supseteq$ PA y de cualquier modelo $\mathscr{M} \models \mathscr{T}$, la constante $c$ sigue siendo un entero natural, pues tenemos que

$$\mathscr{T} \;\vdash\; \phi(0) \wedge \forall x\,(\phi(x) \Rightarrow \phi(s(x))) \;\Rightarrow\; \phi(c)$$

para toda fórmula aritmética $\phi(x)$. (Es decir: el principio de inducción alcanza la constante $c$, como cualquier entero natural.) En particular, en cualquier modelo $\mathscr{M} \models \mathscr{T}$:

- se pueden calcular $c^2$, $2^c$, $c!$ y $c^c$;
- se puede determinar si $c$ es par o impar;
- se puede determinar si $c$ es primo o no;
- se puede descomponer $c$ en sus factores primos, etc.

(En efecto, todas las operaciones anteriores son definibles en PA, y luego en $\mathscr{M}$.)

**2.5. La paradoja de Löwenheim-Skolem.** Consideremos una teoría de primer orden consistente $\mathscr{T}$ sobre un lenguaje $\mathscr{L}$ de cardinal $\kappa$ (infinito)[14]. Una observación atenta de la demostración del teorema de completitud (cuya arquitectura ha sido presentada en la Sección 2.3) revela que el lenguaje intermedio $\mathscr{L}'$ usado en la construcción del modelo «sintáctico» $\mathscr{M} := \mathscr{U}/\!\sim$ tiene el mismo cardinal que el lenguaje $\mathscr{L}$ de la teoría inicial. Por lo tanto, el conjunto $\mathscr{U}$ de los términos cerrados del lenguaje $\mathscr{L}'$ tiene un cardinal menor o igual a $\kappa$, así como el conjunto cociente $\mathscr{M} := \mathscr{U}/\!\sim$. Se obtiene así el siguiente refinamiento del teorema de completitud:

**Teorema 2.16** (Löwenheim-Skolem)**.** *Toda teoría de primer orden consistente tiene un modelo cuyo cardinal es menor o igual al cardinal de su lenguaje.*

En particular:

[14]Observemos que el cardinal de un lenguaje de primer orden $\mathscr{L}$ siempre es infinito. Más precisamente, si el lenguaje $\mathscr{L}$ está construido sobre un vocabulario $\mathcal{V}$ (finito o infinito), se verifica fácilmente que $\mathrm{Card}(\mathscr{L}) = \mathrm{máx}(\mathrm{Card}(\mathcal{V}), \aleph_0)$, donde $\aleph_0$ indica el primer cardinal infinito —es decir: el cardinal de $\mathbb{N}$.

**Corolario 2.17** (Paradoja de Löwenheim-Skolem)**.** *Si la teoría de conjuntos de Zermelo-Fraenkel (ZF) es consistente, entonces tiene un modelo infinito numerable.*

*Demostración.* En efecto, como el lenguaje de ZF es numerable, el teorema anterior implica que si ZF es consistente, entonces tiene un modelo (finito o infinito) numerable. Pero se verifica fácilmente que tal modelo no puede ser finito, entonces es infinito numerable. □

*Interpretación de la paradoja.* Es importante entender por qué el resultado anterior —a pesar de su carácter paradójico— no lleva a ninguna contradicción. Para ello, recordemos que un modelo de ZF es un grafo dirigido $(\mathscr{M}, \in_{\mathscr{M}})$ cuyas aristas representan la relación de pertenencia, y en el cual se cumplen todos los axiomas de ZF (interpretando $\in$ por $\in^{\mathscr{M}}$).

Ahora, supongamos que $\mathscr{M} = (\mathscr{M}, \in^{\mathscr{M}})$ es un modelo numerable de ZF.

En el grafo dirigido $(\mathscr{M}, \in^{\mathscr{M}})$, existe un (único) vértice $\mathbb{N}^{\mathscr{M}} \in \mathscr{M}$ que representa el conjunto $\mathbb{N}$ de los enteros naturales, y cuyos hijos $x_0, x_1, x_2, \ldots \in^{\mathscr{M}} \mathbb{N}^{\mathscr{M}}$ representan los enteros naturales en el sentido del modelo $\mathscr{M}$. Por otro lado, existe otro vértice $\mathbb{R}^{\mathscr{M}} \in \mathscr{M}$ que representa el conjunto $\mathbb{R}$ de los números reales, y cuyos hijos $y_0, y_1, y_2, \ldots \in^{\mathscr{M}} \mathbb{R}^{\mathscr{M}}$ representan los números reales en el sentido del modelo $\mathscr{M}$. Pero como el modelo $\mathscr{M}$ es infinito numerable, ambos conjuntos $\mathbb{N}^{\mathscr{M}} = \{x_0, x_1, x_2, \ldots\}$ y $\mathbb{R}^{\mathscr{M}} = \{y_0, y_1, y_2, \ldots\}$ son infinitos numerables, y por lo tanto, se puede construir una biyección $\tilde{f} : \mathbb{N}^{\mathscr{M}} \tilde{\to} \mathbb{R}^{\mathscr{M}}$.

Ahora, se trata de ver por qué la existencia de una biyección $\tilde{f} : \mathbb{N}^{\mathscr{M}} \tilde{\to} \mathbb{R}^{\mathscr{M}}$ no contradice el teorema de Cantor, que expresa que no existe ninguna biyección entre $\mathbb{N}$ y $\mathbb{R}$.

En efecto, desde el punto de vista de la lógica de primer orden, el teorema de Cantor es una fórmula cerrada del lenguaje de la teoría de conjuntos que es demostrable en ZF, lo que implica que dicha fórmula se cumple en el modelo $\mathscr{M} = (\mathscr{M}, \in^{\mathscr{M}})$. Pero esto sólo significa que *en el modelo* $\mathscr{M}$, no existe ningún vértice que representa una biyección entre $\mathbb{N}^{\mathscr{M}}$ y $\mathbb{R}^{\mathscr{M}}$. Y lo único que podemos concluir es que la biyección $\tilde{f} : \mathbb{N}^{\mathscr{M}} \tilde{\to} \mathbb{R}^{\mathscr{M}}$ que construimos anteriormente no corresponde a ningún vértice del grafo $(\mathscr{M}, \in^{\mathscr{M}})$; sólo existe afuera de $\mathscr{M}$.

Dicho de otro modo:

- Afuera del modelo $\mathscr{M}$, se observa que ambos conjuntos $\mathbb{N}^{\mathscr{M}}$ y $\mathbb{R}^{\mathscr{M}}$ son infinitos numerables, de tal modo que existe una biyección (externa) $f : \mathbb{N}^{\mathscr{M}} \tilde{\to} \mathbb{R}^{\mathscr{M}}$.
- Adentro del modelo $\mathscr{M}$, se observa que el conjunto $\mathbb{R}$ $(= \mathbb{R}^{\mathscr{M}})$ no es numerable, pues no existe ninguna biyección (interna) entre $\mathbb{R}$ $(= \mathbb{R}^{\mathscr{M}})$ y $\mathbb{N}$ $(= \mathbb{N}^{\mathscr{M}})$.

Esto nos lleva a una idea fundamental de la teoría de modelos, a saber que la noción de cardinal (en la teoría de conjuntos) depende del modelo considerado, y que cada modelo de la teoría de conjuntos define su noción interna de cardinal, distinta de la noción externa de cardinal[15].

En fin, cabe destacar que el Teorema 2.16 se puede generalizar del siguiente modo:

**Teorema 2.18** (Löwenheim-Skolem, versión descendente y ascendente)**.** *Si $\mathscr{T}$ es una teoría de primer orden sobre un lenguaje $\mathscr{L}$ que tiene un modelo infinito $\mathscr{M}$, entonces para todo cardinal $\kappa \geq \mathrm{Card}(\mathscr{L})$, la teoría $\mathscr{T}$ tiene un modelo $\mathscr{M}_\kappa$ de cardinal $\kappa$. Además, se puede construir el modelo $\mathscr{M}_\kappa \models \mathscr{T}$ de tal modo que:*

[15]Es decir, la noción de cardinal de la metateoría.

(1) $\mathscr{M}_\kappa$ *es elementalmente equivalente a* $\mathscr{M}$
(2) *Si* $\kappa \leq \mathrm{Card}(\mathscr{M})$*, entonces* $\mathscr{M}_\kappa$ *es un submodelo de* $\mathscr{M}$ *(versión «descendente»)*
(3) *Si* $\kappa \geq \mathrm{Card}(\mathscr{M})$*, entonces* $\mathscr{M}$ *es un submodelo de* $\mathscr{M}_\kappa$ *(versión «ascendente»)*

Ya vimos que la versión «descendente» del teorema de Löwenheim-Skolem implica que ZF tiene modelos numerables (ojalá ZF sea consistente), pero la versión «ascendente» también implica que la aritmética de Peano (PA) tiene modelos con cardinales arbitrariamente grandes:

**Corolario 2.19** (Cardinales de los modelos de PA)**.** *Para todo cardinal infinito* $\kappa$*, la aritmética de Peano tiene un modelo* $\mathscr{M}_\kappa$ *de cardinal* $\kappa$*. Además, se puede construir el modelo* $\mathscr{M}_\kappa$ *de tal modo que sea elementalmente equivalente al modelo estándar* $\mathbb{N}$ *(numerable).*

## 3. Los teoremas de incompletitud de Gödel

**3.1. El programa de Hilbert.** Al comienzo del siglo XX, la matemática se encontraba en una situación paradójica. Por un lado, Georg Cantor (1845–1918) había mostrado en 1878 que todas las ramas de la matemática se podían unificar en la teoría de conjuntos [2], mientras Gottlob Frege (1848–1925) había definido en 1879 un lenguaje universal (la *conceptografía* [3], el antepasado del cálculo de predicados) que permitía formalizar todos los razonamientos matemáticos. Por otro lado, la teoría de conjuntos de Cantor (cuyas bases formales no eran claras) todavía contenía múltiples paradojas, y el primer intento de formalización [4, 5] por Frege en 1893 había sido un fracaso, después de que Bertrand Russell (1872–1970) hubo demostrado su inconsistencia en 1903. En 1908, Ernst Zermelo (1871–1953) había introducido una nueva axiomatización de la teoría de conjuntos que permitía eliminar la inconsistencia descubierta por Russell, pero sin lograr demostrar la consistencia de la nueva axiomatización.

En este contexto de «crisis de los fundamentos», David Hilbert (1862–1943) formuló en los años 1920 un programa de investigación muy ambicioso para asegurar los fundamentos de la matemática. El programa de Hilbert se puede resumir en cuatro puntos:

1. Definir una teoría axiomática «universal» suficientemente expresiva para formalizar toda la matemática, sobre el modelo de la teoría de conjuntos de Cantor[16].
2. Demostrar por medios puramente aritméticos que la teoría universal es consistente.
3. Demostrar por medios puramente aritméticos que la teoría universal es completa, en el sentido en que toda fórmula cerrada es demostrable o refutable en dicha teoría.
4. Definir algoritmos que permiten resolver cualquier pregunta matemática[17].

Por supuesto, Hilbert no pensaba demostrar la consistencia y la completitud de la teoría universal a partir de nada. Sin embargo, había observado que las demostraciones matemáticas son objetos combinatorios finitos similares a enteros

[16]Según Hilbert: *„Aus dem Paradies, das Cantor uns geschaffen, soll uns niemand vertreiben können.“* («Del paraíso que Cantor nos creó, nadie nos debe poder expulsar.»)

[17]Según Hilbert: *„Wir müssen wissen, wir werden wissen.“* («Debemos saber, sabremos»).

naturales, que se pueden estudiar adentro de la aritmética de Peano, vía una codificación adecuada. Y como Hilbert estaba convencido que la aritmética era «segura», proponía utilizar la aritmética de Peano (o un formalismo similar) como marco metamatemático para realizar su programa.

En 1931, Kurt Gödel (1906–1978) publicó sus famosos dos *teoremas de incompletitud* [9], que arruinaron el programa de Hilbert, mostrando que:

1. Toda teoría consistente y «suficientemente expresiva» es necesariamente incompleta.
2. Una teoría consistente y «suficientemente expresiva» siempre puede expresar su propia consistencia, pero no puede demostrarla. En particular: no se puede demostrar la consistencia de ZF (ni siquiera la de PA) adentro de PA, salvo si PA es inconsistente.

Antes de presentar los dos teoremas de incompletitud de Gödel, se necesita precisar lo que queremos decir cuando hablamos de una teoría «suficientemente expresiva».

### 3.2. Teorías recursivas y aritméticas, teorías completas.

*3.2.1. Teorías recursivas.* Cuando un lenguaje de primer orden $\mathscr{L}$ es numerable (es decir: cuando su vocabulario es finito o numerable), siempre se puede definir una *función de codificación*[18] $e \mapsto \ulcorner e \urcorner$ (inyectiva) que asocia a cada expresión sintáctica del lenguaje $\mathscr{L}$ (un término, una fórmula o una derivación) un *código* $\ulcorner e \urcorner \in \mathbb{N}$, de tal modo que todas las operaciones sintácticas usuales (aplicación de un símbolo de función o de predicado, construcción de fórmula con una conectiva o con un cuantificador, sustitución de variable por un término, aplicación de una regla de deducción, etc.) correspondan a funciones computables en $\mathbb{N}$.

Formalmente, se dice que una teoría de primer orden $\mathscr{T}$ es *recursiva* cuando su lenguaje es numerable, y cuando el conjunto de (los códigos de) sus axiomas es computable, vía una función de codificación dada. (Recordemos que un conjunto $X \subseteq \mathbb{N}$ es computable cuando su función característica $\mathbb{1}_X : \mathbb{N} \to \{0, 1\}$ es computable.)

Más generalmente, se demuestra que:

**Proposición 3.1.** *Si $\mathscr{T}$ es una teoría recursiva, entonces:*

(1) *Existe un algoritmo que permite decidir si un árbol finito $A$ de secuentes constituye una demostración (bien formada) de la teoría $\mathscr{T}$, en el sentido en que:*
   a) *El árbol $A$ es una derivación bien formada en el sistema NK (Sección 1.4).*
   b) *La conclusión de la derivación $A$ es un secuente de la forma $\phi_1, \ldots, \phi_n \vdash \phi$, donde $\phi_1, \ldots, \phi_n$ son axiomas de $\mathscr{T}$, y donde $\phi$ es una fórmula cerrada del lenguaje de $\mathscr{T}$ (lo que implica que $\phi$ es un teorema de $\mathscr{T}$).*

(2) *Existe un algoritmo que enumera todas las demostraciones de $\mathscr{T}$.*

(3) *Existe un algoritmo que enumera todos los teoremas de $\mathscr{T}$.*

[18] Hoy sería más adecuado hablar de *función de digitalización*, desde que el desarrollo de la informática nos acostumbró a la idea que todos los datos que manipulamos en nuestras computadoras (los textos, los programas, las bases de datos, los vídeos y los sonidos) están representados en las entrañas de la máquina por sucesiones finitas de 0 y 1 —es decir: por enteros naturales especialmente grandes.

Intuitivamente, una teoría recursiva es una teoría cuyas demostraciones se pueden *reconocer* mediante un algoritmo adecuado. Es claro que todas las teorías matemáticas usuales son recursivas —y en particular, todas las que presentamos en los ejemplos anteriores.

*3.2.2. Teorías aritméticas.* Se dice que una teoría $\mathscr{T}$ es *aritmética* cuando se pueden expresar todas la fórmulas (y demostrar todos los teoremas) de la aritmética de Peano adentro de la teoría $\mathscr{T}$, vía una representación adecuada de los enteros naturales en $\mathscr{T}$.

Formalmente, tal representación se define mediante una *traducción lógica* de PA en $\mathscr{T}$, es decir: una función computable $\phi \mapsto \phi^*$ que asocia a cada fórmula $\phi$ del lenguaje de PA una fórmula $\phi^*$ del lenguaje de $\mathscr{T}$, de tal modo que:

(1) La fórmula $\phi^*$ tiene las mismas variables libres que $\phi$: $FV(\phi^*) = FV(\phi)$.
(2) La función $\phi \mapsto \phi^*$ conmuta con todas las construcciones de la lógica, es decir:

$$\begin{array}{ll} \mathscr{T} \vdash (\phi \Rightarrow \psi)^* \Leftrightarrow \phi^* \Rightarrow \psi^* & \mathscr{T} \vdash \bot^* \Leftrightarrow \bot \\ \mathscr{T} \vdash (\phi \wedge \psi)^* \Leftrightarrow \phi^* \wedge \psi^* & \mathscr{T} \vdash (\forall x\, \phi(x))^* \Leftrightarrow \forall x\, (N(x) \Rightarrow \phi(x)^*) \\ \mathscr{T} \vdash (\phi \vee \psi)^* \Leftrightarrow \phi^* \vee \psi^* & \mathscr{T} \vdash (\exists x\, \phi(x))^* \Leftrightarrow \exists x\, (N(x) \wedge \phi(x)^*) \end{array}$$

donde $N(x)$ es una fórmula del lenguaje de $\mathscr{T}$ que expresa que «$x$ es un entero natural».

(3) La función $\phi \mapsto \phi^*$ transforma todos los teoremas de PA en teoremas de $\mathscr{T}$: si $\mathrm{PA} \vdash \phi$, entonces $\mathscr{T} \vdash \phi^*$.

**Ejemplos 3.2.** La aritmética de Peano (PA) es obviamente una teoría aritmética (vía la traducción identidad), pero la teoría de conjuntos de Zermelo-Fraenkel (ZF) también lo es, vía la representación de los enteros naturales por los ordinales de von Neumann:

$$\begin{array}{rcccl} 0 & = & \varnothing & & \\ 1 & = & 0 \cup \{0\} & = & \{0\} \\ 2 & = & 1 \cup \{1\} & = & \{0, 1\} \\ \vdots & & \vdots & & \vdots \\ n+1 & = & n \cup \{n\} & = & \{0, \ldots, n\} \end{array}$$

(En este caso, la fórmula $N(x)$ que expresa en el lenguaje de ZF que «$x$ es un entero natural» está definida por $N(x) :\equiv x \in \omega$, donde $\omega$ es el primer ordinal límite.)

*3.2.3. Teorías completas.* En fin, se dice que una teoría de primer orden $\mathscr{T}$ es *completa* cuando toda fórmula cerrada $\phi$ del lenguaje de $\mathscr{T}$ es decidible en $\mathscr{T}$, es decir: cuando $\mathscr{T} \vdash \phi$ («$\phi$ es demostrable en $\mathscr{T}$») o $\mathscr{T} \vdash \neg\phi$ («$\phi$ es refutable en $\mathscr{T}$»).

Cabe destacar que en la definición anterior, el «o» es inclusivo, de tal modo que toda teoría inconsistente es trivialmente completa (ya que $\mathscr{T} \vdash \phi$ y $\mathscr{T} \vdash \neg\phi$ para toda fórmula cerrada $\phi$). Por otro lado, cuando una teoría $\mathscr{T}$ es a la vez consistente y completa, tenemos que $\mathscr{T} \vdash \phi$ ó $\mathscr{T} \vdash \neg\phi$ («o» exclusivo) para toda fórmula cerrada $\phi$ del lenguaje de $\mathscr{T}$.

En la teoría de modelos, las teorías completas se caracterizan del modo siguiente:

**Proposición 3.3** (Caracterización de las teorías completas)**.** *Una teoría de primer orden es completa si y sólo si todos sus modelos son elementalmente equivalentes.*

*Demostración.* Implicación directa: Supongamos que $\mathscr{T}$ es completa, y consideremos dos modelos $\mathscr{M}_1, \mathscr{M}_2 \models \mathscr{T}$. Dada una fórmula cerrada $\phi$ del lenguaje de $\mathscr{T}$, se observa que:

- o bien $\mathscr{T} \vdash \phi$, lo que implica que $\mathscr{M}_1 \models \phi$ y $\mathscr{M}_2 \models \phi$;
- o bien $\mathscr{T} \vdash \neg\phi$, lo que implica que $\mathscr{M}_1 \models \neg\phi$ y $\mathscr{M}_2 \models \neg\phi$, es decir $\mathscr{M}_1 \not\models \phi$ y $\mathscr{M}_2 \not\models \phi$.

En ambos casos, tenemos que $\mathscr{M}_1 \models \phi$ sii $\mathscr{M}_2 \models \phi$. Implicación recíproca: Ahora se supone que $\mathscr{T}$ es incompleta, y se considera una fórmula cerrada $\phi$ tal que $\mathscr{T} \nvdash \phi$ y $\mathscr{T} \nvdash \neg\phi$.

- Como $\mathscr{T} \nvdash \phi$, la teoría $\mathscr{T} + \neg\phi$ es consistente, y tiene un modelo $\mathscr{M}_1$ (por el teorema de completitud). Por construcción, $\mathscr{M}_1$ es un modelo de $\mathscr{T}$ tal que $\mathscr{M}_1 \models \neg\phi$.
- Como $\mathscr{T} \nvdash \neg\phi$, la teoría $\mathscr{T} + \phi$ es consistente, y tiene un modelo $\mathscr{M}_2$ (por el teorema de completitud). Por construcción, $\mathscr{M}_2$ es un modelo de $\mathscr{T}$ tal que $\mathscr{M}_2 \models \phi$.

Así construimos dos modelos $\mathscr{M}_1$ y $\mathscr{M}_2$ de la teoría $\mathscr{T}$ que no son elementalmente equivalentes, pues $\mathscr{M}_1 \not\models \phi$ y $\mathscr{M}_2 \models \phi$. □

**3.3. El primer teorema de incompletitud.** Ahora tenemos todos los ingredientes para enunciar el primer teorema de incompletitud de Gödel:

**Teorema 3.4** (Primer teorema de incompletitud)**.** *Toda teoría de primer orden recursiva, aritmética y consistente es incompleta.*

La demostración del teorema se basa en las siguientes observaciones:

1. Como la teoría $\mathscr{T}$ es recursiva, se puede definir una función de codificación $e \mapsto \ulcorner e \urcorner$ que asocia un código $\ulcorner e \urcorner \in \mathbb{N}$ a cada expresión sintáctica del lenguaje de $\mathscr{T}$, de tal modo que todas las operaciones sintácticas usuales correspondan a funciones recursivas en $\mathbb{N}$. Además, el conjunto de (los códigos de) los axiomas de $\mathscr{T}$ es computable, así como el conjunto de (los códigos de) las demostraciones de $\mathscr{T}$.
2. Como todas las funciones computables son «representables» en PA (en un sentido que no precisaremos aquí), se puede construir en el lenguaje de PA una fórmula $Dem_{\mathscr{T}}(x)$ que expresa que «$x$ es el código de una fórmula cerrada demostrable en $\mathscr{T}$».
   (Cabe destacar que $Dem_{\mathscr{T}}(x)$ es una fórmula del lenguaje de PA que expresa la noción de demostrabilidad de la teoría $\mathscr{T}$, vía la función de codificación $\phi \mapsto \ulcorner \phi \urcorner$.)
3. Ahora, como $\mathscr{T}$ es una teoría aritmética, existe una traducción lógica $\phi \mapsto \phi^*$ que transforma toda fórmula (resp. todo teorema) $\phi$ de PA en una fórmula (resp. en un teorema) $\phi^*$ de $\mathscr{T}$. En particular, la fórmula aritmética $Dem_{\mathscr{T}}(x)$ se traduce en el lenguaje de la teoría $\mathscr{T}$ en una fórmula $Dem^*_{\mathscr{T}}(x)$, que expresa la misma propiedad sobre los códigos de las fórmulas de $\mathscr{T}$ (pero ahora en el lenguaje de $\mathscr{T}$).
4. Usando una técnica similar a la diagonal de Cantor, se construye una fórmula aritmética cerrada $\Phi$ (¡muy larga!) tal que

$$\mathscr{T} \vdash \Phi^* \Leftrightarrow \neg Dem^*_{\mathscr{T}}(\ulcorner \Phi^* \urcorner)\,.$$

Intuitivamente, la fórmula aritmética cerrada $\Phi$ dice: «no soy demostrable en $\mathscr{T}$» (vía la codificación $\phi \mapsto \ulcorner\phi\urcorner$ y la traducción lógica $\phi \mapsto \phi^*$ de PA en $\mathscr{T}$).

5. Por lo anterior, se deduce fácilmente que si la teoría $\mathscr{T}$ es consistente, entonces la fórmula $\Phi^*$ no es demostrable en $\mathscr{T}$, es decir: $\mathscr{T} \nvdash \Phi^*$. Y con un poco más trabajo[19], también se demuestra que la fórmula $\Phi^*$ no es refutable en $\mathscr{T}$, es decir: $\mathscr{T} \nvdash \neg\Phi^*$.

Al final, se obtiene una fórmula cerrada $\Phi^*$ del lenguaje de $\mathscr{T}$ que es indecidible en $\mathscr{T}$.

**Observaciones 3.5.** (1) Cabe destacar que la fórmula indecidible $\Phi^*$ no es cualquier fórmula cerrada del lenguaje de $\mathscr{T}$, pero la traducción de una fórmula aritmética en el lenguaje de $\mathscr{T}$. Dicho de otro modo, el primer teorema de incompletitud de Gödel no sólo expresa la incompletitud de los formalismos matemáticos aritméticos, pero expresa su incompletitud *respecto a las fórmulas aritméticas* —un resultado cuyo alcance filosófico es mucho más grande.

Por ejemplo, si la teoría de conjuntos de Zermelo-Fraenkel (ZF) es consistente, a pesar de toda la fuerza teórica brindada por la jerarquía de conjuntos que se pueden construir por encima de $\mathbb{N}$, todavía existen fórmulas *aritméticas* indecidibles en ZF. (Esta observación no aminora el interés de ZF, en qué se pueden demostrar mucho más fórmulas aritméticas que en PA.)

(2) Dada una fórmula indecidible $\phi$ en la teoría $\mathscr{T}$, es muy tentador añadir la fórmula $\phi$ (o su negación) a los axiomas de $\mathscr{T}$, con el fin de «completar» la teoría $\mathscr{T}$. Desgraciadamente, el primer teorema de incompletitud todavía se aplica a la teoría (consistente) $\mathscr{T} + \phi$, y nos da otra fórmula cerrada $\phi'$ que es indecidible en la nueva teoría $\mathscr{T} + \phi$. Esta observación muestra que el primer teorema de incompletitud de Gödel no viene de un «bug» de algún formalismo matemático, pero que expresa una incompletitud fundamental de todos los sistemas formales (recursivos) que contienen la aritmética de Peano.

(3) En el caso donde $\mathscr{T} = \mathrm{PA}$, se verifica fácilmente que la fórmula de Gödel $\Phi^*$ (indecidible en PA) se cumple en el modelo estándar (Ejemplos 2.8 (4)), es decir: $\mathbb{N} \models \Phi^*$. Así, la fórmula de Gödel $\Phi^*$ asociada a la aritmética de Peano también constituye un ejemplo de fórmula verdadera en el modelo estándar que no es demostrable en PA: $\mathbb{N} \models \Phi^*$ pero $\mathrm{PA} \nvdash \Phi^*$.

*Consecuencias del primer teorema de incompletitud en la teoría de modelos.* Desde el punto de vista de la teoría de modelos, ya vimos (Prop. 3.3) que las teorías incompletas son las teorías (consistentes) que tienen al menos dos modelos no elementalmente equivalentes. Por el primer teorema de incompletitud de Gödel, se deduce que:

**Corolario 3.6.** *Toda teoría $\mathscr{T}$ de primer orden recursiva, aritmética y consistente tiene dos modelos no elementalmente equivalentes.*

Y en particular:

[19]En su demostración original de 1931, Gödel suponía que la teoría $\mathscr{T}$ era $\omega$-consistente —una hipótesis mucho más fuerte que la consistencia de $\mathscr{T}$— para demostrar que $\mathscr{T} \nvdash \neg\Phi^*$ («$\Phi$ no es refutable en $\mathscr{T}$»). Sin embargo, Rosser introdujo en 1936 un truco (el «truco de Rosser») para demostrar que $\mathscr{T} \nvdash \neg\Phi^*$ sólo a partir de la hipótesis de consistencia, usando una versión modificada de la fórmula $Dem_{\mathscr{T}}(x)$.

**Corolario 3.7.** *La aritmética de Peano tiene modelos no elementalmente equivalentes (y tampoco isomorfos) al modelo estándar.*

**3.4. Ejemplos de teorías completas.** El primer teorema de incompletitud sólo se aplica a las teorías recursivas, aritméticas y consistentes, y es fácil ver que no se puede eliminar ninguna de estas tres hipótesis. Por ejemplo, es obvio que toda teoría inconsistente $\mathscr{T}$ es completa, ya que $\mathscr{T} \vdash \phi$ y $\mathscr{T} \vdash \neg\phi$ para cada fórmula cerrada $\phi$ del lenguaje de $\mathscr{T}$.

Un ejemplo obvio de teoría consistente y completa es la *teoría del modelo estándar*, a saber la teoría $\mathscr{T}_{\mathbb{N}}$ cuyo lenguaje es el lenguaje de PA, y cuyos axiomas son (por definición) todas las fórmulas aritméticas cerradas que se cumplen en el modelo estándar $\mathbb{N}$ (Ejemplos 2.8 (4)). Más generalmente, se verifica fácilmente que los teoremas de la teoría $\mathscr{T}_{\mathbb{N}}$ son las fórmulas aritméticas cerradas que se cumplen en $\mathbb{N}$[20], de tal modo que

$$\mathscr{T}_{\mathbb{N}} \vdash \phi \qquad \text{sii} \qquad \mathbb{N} \models \phi$$

para toda fórmula aritmética cerrada $\phi$.

Por la equivalencia anterior, es obvio que la teoría $\mathscr{T}_{\mathbb{N}}$ es consistente y completa. Además, la teoría $\mathscr{T}_{\mathbb{N}}$ es aritmética, pues contiene obviamente PA como subteoría. Por lo tanto, se deduce por el primer teorema de incompletitud que:

**Teorema 3.8.** *La teoría del modelo estándar $\mathscr{T}_{\mathbb{N}}$ no es recursiva.*

Es decir: no existe ningún algoritmo que permita enumerar todas la fórmulas aritméticas verdaderas en $\mathbb{N}$. Y luego, tampoco existe un algoritmo que permita decidir si una fórmula aritmética cerrada es verdadera o falsa en $\mathbb{N}$.

Ahora, queda mostrar que existen teorías consistentes, completas y recursivas, pero no aritméticas. Un ejemplo muy sencillo es la teoría $\mathscr{T}_1$ del «objeto único», cuyo vocabulario sólo contiene el símbolo de igualdad «=», y cuyo único axioma

$$\forall x \, \forall y \, (x = y)$$

expresa que «sólo existe un objeto en el universo». Es claro que la teoría $\mathscr{T}_1$ tiene un único modelo (a menos de isomorfismo), a saber el modelo con un único elemento. Por lo tanto, todos los modelos de la teoría $\mathscr{T}_1$ son elementalmente equivalentes, y se deduce por la Prop. 3.3 que $\mathscr{T}_1$ es una teoría completa. (Por supuesto, la teoría $\mathscr{T}_1$ no es aritmética.)

Otros ejemplos mucho más interesantes son los siguientes:

*3.4.1. La aritmética de Presburger.* Se llama *aritmética de Presburger* y se escribe $\mathscr{P}$ a la teoría de primer orden cuyo lenguaje es el lenguaje de PA sin símbolo «×» (producto), y cuyos axiomas son los axiomas de PA que no involucran el símbolo «×». Por construcción, la aritmética de Presburger tiene un lenguaje poco expresivo, en qué sólo se pueden expresar desigualdades lineales tales que

$$3x + 2y \leq 4z + 7$$

(es decir, formalmente: $\exists u \, (x + x + x + y + y + u = z + z + z + z + 7)$).

Es claro que $\mathscr{P}$ es una teoría consistente (pues $\mathbb{N} \models \mathscr{P}$) y recursiva. Además, Mojżesz Presburger (1904–1943) demostró [17] en 1929 que $\mathscr{P}$ es completa, lo que implica que:

(1) Todos los modelos de $\mathscr{P}$ son elementalmente equivalentes a $\mathbb{N}$ (sin producto).

[20]De hecho, la teoría $\mathscr{T}_{\mathbb{N}}$ no hace ninguna distinción entre las nociones de axioma y de teorema.

(2) Para toda fórmula cerrada del lenguaje de $\mathscr{P}$, tenemos que

$$\mathscr{P} \vdash \phi \qquad \text{sii} \qquad \mathrm{PA} \vdash \phi \qquad \text{sii} \qquad \mathbb{N} \models \phi\,.$$

(3) Existe un algoritmo que permite decidir si una fórmula cerrada del lenguaje de $\mathscr{P}$ es demostrable en $\mathscr{P}$ / demostrable en PA / verdadera en $\mathbb{N}$.

Por supuesto, la aritmética de Presburger no es una teoría aritmética (en el sentido de Peano), pues no es suficientemente expresiva para expresar la relación $z = x \times y$.

*3.4.2. La teoría de los cuerpos reales cerrados.* Se llama *teoría de los cuerpos reales cerrados* y se escribe $\mathscr{R}$ a la teoría de primer orden cuyo lenguaje está definido a partir de los símbolos «0» (cero), «1» (uno), «+» (suma), «−» (opuesto), «×» (producto), «=» (igualdad) y «≤» (orden); y cuyos axiomas son

- los axiomas de los cuerpos totalmente ordenados:

$$\begin{array}{ll}
\forall x\,(0 + x = x) & \forall x\,(1x = x) \\
\forall x\,\forall y\,(x + y = y + x) & \forall x\,\forall y\,(xy = yx) \\
\forall x\,\forall y\,\forall z\,((x + y) + z = x + (y + z)) & \forall x\,\forall y\,\forall z\,((xy)z = x(yz)) \\
\forall x\,(x + (-x) = 0) & \forall x\,(x \neq 0 \Rightarrow \exists y\,(xy = 1)) \\
0 \neq 1 & \forall x\,\forall y\,\forall z\,((x + y)z = (xz) + (yz)) \\
\forall x\,(x \leq x) & \\
\forall x\,\forall y\,(x \leq y \wedge y \leq x \Rightarrow x = y) & \forall x\,\forall y\,\forall z\,(x \leq y \wedge y \leq z \Rightarrow x \leq z) \\
\forall x\,\forall y\,\forall z\,(x \leq y \Rightarrow x + z \leq y + z) & \forall x\,\forall y\,\forall z\,(x \leq y \wedge 0 \leq z \Rightarrow xz \leq yz);
\end{array}$$

- el axioma que expresa que todo número mayor o igual a 0 tiene raíz cuadrada:

$$\forall x\,(0 \leq x \Rightarrow \exists y\,(x = yy))\,;$$

- para cada entero natural impar $n$ $(= 2k + 1)$, el axioma que expresa que todos los polinomios de grado $n$ tienen una raíz:

$$\forall a_0\,\forall a_1 \cdots \forall a_n\,(a_n \neq 0 \Rightarrow \exists x\,(a_0 + a_1 x + \cdots + a_n x^n = 0))$$

(donde $a_p x^p$ es una abreviatura para $a_p \times x \times \cdots \times x$, con $p$ productos).

Es claro que $\mathscr{R}$ es una teoría consistente, pues $\mathbb{A} \models \mathscr{R}$, donde $\mathbb{A}$ es el subcuerpo numerable de $\mathbb{R}$ formado por los *números reales algebraicos*. (Por supuesto, el cuerpo $\mathbb{R}$ de los números reales también es un modelo —pero no numerable— de la teoría $\mathscr{R}$.) Además, la teoría $\mathscr{R}$ es claramente recursiva. Sin embargo, Alfred Tarski (1901–1983) demostró [18] en 1951 que $\mathscr{R}$ es una teoría completa, lo que implica de vuelta que:

(1) Todos los modelos de la teoría $\mathscr{R}$ son elementalmente equivalentes a $\mathbb{A}$.
(2) Para toda fórmula cerrada del lenguaje de $\mathscr{R}$, tenemos que

$$\mathscr{R} \vdash \phi \qquad \text{sii} \qquad \mathbb{A} \models \phi \qquad \text{sii} \qquad \mathbb{R} \models \phi\,.$$

(3) Existe un algoritmo que permite decidir si una fórmula cerrada del lenguaje de $\mathscr{R}$ es demostrable en $\mathscr{R}$ / verdadera en $\mathbb{A}$ / verdadera en $\mathbb{R}$[21].

[21] En particular, este algoritmo permite resolver automáticamente todos los problemas de la geometría elemental (en cualquier dimensión finita). Al día de hoy, existen múltiples implementaciones eficientes de este algoritmo, que están usadas en la industria robótica para la detección de colisión.

Aquí, cabe destacar que, aunque el lenguaje de la teoría $\mathscr{R}$ permita expresar todos los enteros naturales ($0$, $1$, $2 :\equiv 1+1$, $3 :\equiv 2+1$, etc.), no permite construir ninguna fórmula $N(x)$ que exprese que «$x$ es un entero natural». Por lo tanto, es imposible definir una traducción lógica $\phi \mapsto \phi^*$ de PA en $\mathscr{R}$, lo que explica por qué la teoría $\mathscr{R}$ no es aritmética.

**3.5. El segundo teorema de incompletitud.** El primer teorema de incompletitud viene de que toda teoría $\mathscr{T}$ recursiva y aritmética es suficientemente expresiva para formalizar su propia teoría de la demostración, mediante un predicado $Dem^*_{\mathscr{T}}(x)$ que expresa en el lenguaje de $\mathscr{T}$ que «$x$ es el código de una fórmula cerrada demostrable en $\mathscr{T}$». En particular, este predicado permite definir la fórmula

$$Cons\mathscr{T} \;:\equiv\; \neg Dem^*_{\mathscr{T}}(\ulcorner \bot \urcorner)\,,$$

que expresa la consistencia de la teoría $\mathscr{T}$ en el propio lenguaje de $\mathscr{T}$.

Pero, ¿se puede demostrar la fórmula $Cons\mathscr{T}$ en $\mathscr{T}$?

Para contestar a esta pregunta, Gödel observa que la prueba de la primera implicación del primer teorema de incompletitud

*Si $\mathscr{T}$ es consistente, entonces $\Phi^*$ no es demostrable en $\mathscr{T}$*

(donde $\Phi^*$ es la fórmula indecidible construida anteriormente) sólo se basa en argumentos combinatorios que se pueden formalizar adentro de la aritmética de Peano (mediante la función de codificación $e \mapsto \ulcorner e \urcorner$), y luego adentro de la teoría $\mathscr{T}$ (vía la traducción lógica $\phi \mapsto \phi^*$).

Transformando el razonamiento anterior en una derivación formal adentro de la teoría $\mathscr{T}$, Gödel deduce que

$$\mathscr{T} \;\vdash\; Cons\mathscr{T} \;\Rightarrow\; \neg Dem^*_{\mathscr{T}}(\ulcorner \Phi^* \urcorner)\,.$$

Pero como las fórmulas cerradas $\Phi^*$ y $\neg Dem^*_{\mathscr{T}}(\ulcorner \Phi^* \urcorner)$ son equivalentes en la teoría $\mathscr{T}$ (por construcción de la fórmula $\Phi^*$), se deduce que:

$$\mathscr{T} \;\vdash\; Cons\mathscr{T} \;\Rightarrow\; \Phi^*$$

Y como ya vimos que la fórmula $\Phi^*$ no es demostrable en $\mathscr{T}$ (si $\mathscr{T}$ es consistente), es claro que la fórmula $Cons\mathscr{T}$ tampoco es demostrable en $\mathscr{T}$. Por lo tanto:

**Teorema 3.9** (Segundo teorema de incompletitud)**.** *Si $\mathscr{T}$ es una teoría de primer orden recursiva, aritmética y consistente, entonces la fórmula cerrada $Cons\mathscr{T}$ que expresa la consistencia de la teoría $\mathscr{T}$ en el lenguaje de $\mathscr{T}$ no es demostrable en $\mathscr{T}$: $\mathscr{T} \nvdash Cons\mathscr{T}$.*

**Observación 3.10.** Cabe destacar que el segundo teorema de incompletitud sólo dice que la fórmula $Cons\mathscr{T}$ no es demostrable en $\mathscr{T}$, pero no dice nada sobre su negación. De hecho, se puede construir un ejemplo (muy artificial) de una teoría de primer orden recursiva, aritmética y consistente que demuestra su propia inconsistencia —aunque sea consistente[22].

[22]Intuitivamente, tal teoría expresa la existencia de un entero natural (el código de una demostración de la fórmula absurda) que no tiene ninguna existencia concreta. Se dice de tal teoría que es $\omega$-inconsistente.

**3.6. La noción de fuerza teórica.** El segundo teorema de incompletitud no impide que una teoría $\mathscr{T}$ (recursiva, aritmética y consistente) demuestre la consistencia de otra teoría $\mathscr{T}'$. Cuando es el caso, esto significa que la teoría $\mathscr{T}$ es estrictamente más fuerte que la teoría $\mathscr{T}'$. Por ejemplo, ya vimos que la existencia del modelo estándar implica la consistencia de la aritmética de Peano. Pero como dicho modelo se construye fácilmente en ZF, se deduce que:

**Proposición 3.11.** $\mathrm{ZF} \vdash \mathit{Cons}\,\mathrm{PA}$.

En lógica de primer orden, se dice que una teoría $\mathscr{T}$ es *más fuerte* que otra teoría $\mathscr{T}'$ y se escribe $\mathscr{T}' < \mathscr{T}$ cuando la consistencia de $\mathscr{T}'$ es demostrable en $\mathscr{T}$, es decir: $\mathscr{T} \vdash \mathit{Cons}\,\mathscr{T}'$. (Esta relación sólo tiene sentido cuando $\mathscr{T}$ es aritmética y cuando $\mathscr{T}'$ es recursiva.)

Es fácil ver que la relación $\mathscr{T}' < \mathscr{T}$ es transitiva, mientras el segundo teorema de incompletitud expresa que $\mathscr{T} \not< \mathscr{T}$ para toda teoría $\mathscr{T}$ recursiva, aritmética y consistente.

En la práctica, la relación $\mathscr{T} < \mathscr{T}'$ permite clasificar las teorías recursivas y aritméticas en términos de fuerza teórica, y de modo sorprendente, todas las teorías usuales están totalmente ordenadas, aunque la relación $\mathscr{T} < \mathscr{T}'$ no sea total:

$$\mathrm{PA} \;<\; \mathrm{PA2} \;<\; \mathrm{PA3} \;<\; \cdots \;<\; \mathrm{PA}\omega \;<\; \mathrm{Z} \;<\; \cdots \;<\; \mathrm{ZF} \;<\; \cdots$$

Así, en el diagrama anterior:

- PA es la aritmética de Peano, también llamada *aritmética de primer orden* (notación: PA1), cuyos solos objetos son los enteros naturales.
- PA2 es la *aritmética de segundo orden*, cuyos solos objetos son los enteros naturales y los conjuntos de enteros naturales.
- PA3 es la *aritmética de tercer orden*, cuyos solos objetos son los enteros naturales, los conjuntos de enteros naturales y los conjuntos de conjuntos de enteros naturales.
- De modo similar se define PA$n$ para todo $n \geq 2$. Por construcción, tenemos la inclusión $\mathrm{PA}n \subseteq \mathrm{PA}(n+1)$ para todo $n \geq 1$.
- PA$\omega$ es la *aritmética de alto orden*, definida por $\mathrm{PA}\omega := \bigcup_{n\geq 1} \mathrm{PA}n$.
- Z es la *teoría de conjuntos de Zermelo*, es decir: la teoría de Zermelo-Fraenkel sin el esquema de remplazo (véase Ejemplos 1.10 (5) p. 200).
- ZF es la teoría de conjuntos de Zermelo-Fraenkel.

(También se pueden definir muchas teorías intermedias entre Z y ZF, así como muchas teorías más fuertes que ZF, añadiendo a ésta axiomas sobre los cardinales grandes.)

## Referencias


[1] N. Bourbarki. *Théorie des ensembles*. Springer, 1939.

[2] G. Cantor. Ein Beitrag zur Mannigfaltigkeitslehre. *Journal für die Reine und Angewandte Mathematik*, 84:242–258, 1878.

[3] G. Frege. *Begriffsschrift: eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*. Halle, 1879.

[4] G. Frege. *Grundgesetze der Arithmetik, Band I*. Verlag Hermann Pohle, 1893.

[5] G. Frege. *Grundgesetze der Arithmetik, Band II*. Verlag Hermann Pohle, 1903.

[6] G. K. E. Gentzen. Untersuchungen über das logische Schließen. I & II. *Mathematische Zeitschrift*, 39(1):176–210,405–431, 1935.

[7] J.-Y. Girard, Y. Lafont, and P. Taylor. *Proofs and Types*. Cambridge Tracts in Theoretical Computer Science. Cambridge University Press, 1990.

[8] K. Gödel. *Über die Vollständigkeit des Logikkalküls*. PhD thesis, University of Vienna, 1929.
[9] K. Gödel. Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme, I. *Monatshefte für Mathematik und Physik*, 38:173–198, 1931.
[10] D. Hilbert and P. Bernays. *Grundlagen der Mathematik I*. Springer Verlag, 1934. (2nd ed., 1968).
[11] International Organization for Standardization. ISO 80000-2:2009 Quantities and units – Part 2: Mathematical signs and symbols to be used in the natural sciences and technology. Technical report, ISO, 2009.
[12] J.-L. Krivine. *An introduction to axiomatic set theory*. Springer, 1971.
[13] E. Mendelson. *Introduction to Mathematical Logic (5th ed.)*. Chapman & Hall, 2009.
[14] Molière. *Le Bourgeois gentilhomme*. 1670.
[15] G. Peano. *Arithmetices principia, nova methodo exposita*. Bocca, Torino, 1889.
[16] D. Prawitz. Natural deduction: A proof-theoretical study. *Acta Universitatis Stockholmiensis, Stockholm studies in philosophy*, 3, 1965.
[17] M. Presburger. Über die Vollständigkeit eines gewissen Systems der Arithmetik ganzer Zahlen, in welchem die Addition als einzige Operation hervortritt. In *Comptes Rendus du Ier congrès de Mathématiciens des Pays Slaves*, pages 92–101, 1929.
[18] A. Tarski. *A Decision Method for Elementary Algebra and Geometry*. Univ. of California Press, 1951.

IMERL, Facultad de Ingeniería, Universidad de la República. Montevideo, Uruguay.
*Email address*: amiquel@fing.edu.uy

PONENCIA

# CONJUNTOS MINIMALES PARA HOMEOMORFISMOS DEL CíRCULO

ALDO PORTELA

Publicaciones Matemáticas del Uruguay
Volumen 17, Julio 2019, Páginas 227–228
S 0797-1443

# CONJUNTOS MINIMALES PARA HOMEOMORFISMOS DEL CíRCULO

ALDO PORTELA

En el estudio de la dinámica de homeomorfismos del círculo es importante conocer los conjuntos minimales.

Sea $f : S^1 \to S^1$ un homeomorfismo. Decimos que un conjunto no vacío de $S^1$ es minimal si es compacto invariante y minimal respecto a la inclusión. Ejemplos simples de conjuntos minimales son los puntos fijos o las órbitas periódicas de un homeomorfismo. Un punto $x$ de $S^1$ se llama periódico para $f$ si existe un número natural $n$ mayor o igual a 1 tal que $f^n(x) = x$.

Cuando $f$ no conserva el sentido es fácil probar que $f$ tiene por lo menos dos puntos fijos, y en este caso es simple determinar los conjuntos minimales. En el caso en que $f$ conserva el sentido y tiene puntos periódicos también es simple determinar los conjuntos minimales. Siempre son órbitas periódicas.

Llamamos órbita de $x$ al conjunto $o(x) = \{f^n(x) : n \in Z\}$. También se definen la órbita futura y pasada de $x$ como $o^+(x) = \{f^n(x) : n \in N\}$ y $o^-(x) = \{f^{-n}(x) : n \in N\}$ respectivamente.

En el caso en que $f$ no tiene puntos periódicos la situación es más compleja. De ahora en más vamos a suponer que $f : S^1 \to S^1$ es un homeomorfismo sin puntos periódicos. A continuación vamos a enunciar un conjunto de definiciones y un conjunto de propiedades que no son difíciles de probar. Dado $x \in S^1$ se definen los conjuntos alfa límite y omega límite como,

$$\alpha(x) = \{z : z \text{ es punto de acumulación de } o^-(x)\},$$
$$\omega(x) = \{z : z \text{ es punto de acumulación de } o^+(x)\}.$$

Propiedades:

1. Los conjuntos límites son cerrados y no vacíos.
2. Los conjuntos límites son totalmente invariantes o sea $f(\alpha(x)) = \alpha(x)$ y $f(\omega(x)) = \omega(x)$.
3. Para todo $x$, $\alpha(x) = \omega(x)$.
4. El conjunto $\alpha(x)$ no depende de $x$. A dicho conjunto lo llamaremos $K$.
5. El conjunto $K$ es minimal para $f$, luego $f$ tiene un solo conjunto minimal.

**Teorema 0.1.** *Las posibilidades para K son $S^1$ o un conjunto de Cantor.*

Un ejemplo en el que $K$ es $S^1$ es cuando el homeomorfismo es una rotación irracional en $\pi$. Para un ejemplo en el que el conjunto minimal es un conjunto de Cantor ver los ejemplos de Denjoy. Los ejemplos que construyó Denjoy son minimales para difeomorfismos de clase $C^1$, [D]. Este trabajo es de 1932. Como en $S^1$ todo par de conjuntos de Cantor son homeomorfos, conjugando adecuadamente un ejemplo de Denjoy se puede probar que cualquier conjunto de Cantor es minimal para algún homeomorfismo (en este caso decimos que el conjunto de Cantor es $C^0$- minimal).

©2019 PMU

O sea todo conjunto de Cantor es $C^0$- minimal. También en [D] Denjoy prueba que si $f$ es un difeomorfismos de clase $C^2$ entonces su conjunto minimal es $S^1$. O sea, ningún conjunto de Cantor es $C^2$-minimal ( minimal para un difeomorfismo de clase $C^2$). Entonces una pregunta posible, que aún no tiene respuesta, es que conjuntos de Cantor son $C^1$-minimales. O sea, que conjuntos de Cantor son minimales para algún difeomorfismos de clase $C^1$.

Sí existen respuestas parciales al respecto:

- En 1981, McDuff prueba que el conjunto de Cantor triádico usual no es $C^1$-minimal. También da una condición que implica la no $C^1$ minimalidad [MD].
- En 2002, Alec Norton prueba que los conjuntos afines de Cantor generados por dos ramos crecientes no son $C^1$ minimales [N]. En este mismo trabajo A. Norton conjetura que los conjuntos hiperbólicos de Cantor no son $C^1$-minimales.
- En [BIP] se generaliza el resultado de Norton a todo los conjuntos afines de Cantor.

También se conocen otras familias de conjuntos de Cantor que no son $C^1$-minimales. Ver por ejemplo [P]. En resumen, lo que se conoce hasta ahora es que ciertas familias de conjuntos de Cantor no son $C^1$ minimales y los únicos conjuntos de Cantor $C^1$ minimales conocidos son los dados por Denjoy o conjugados a ellos u otros obtenidos dinámicamente. Pero no se conocen, por ejemplo, condiciones geométricas sobre un conjunto de Cantor que impliquen que este sea, o no, $C^1$-minimal.

En [MD] McDuff enuncia una conjetura que aún está sin resolver.

Sea $K$ un conjunto de Cantor de $S^1$ y se considera $K^c = \bigcup I_j$ donde $I_j$ son las componentes conexas de $K^c$. Entonces se llama espectro de $K$ al conjunto $\{\lambda_i\}$ ordenado de forma decreciente estricta donde los $\lambda_i$ son las medidas de los intervalos $I_j$. Notar que dado un elemento del espectro de K pueden existir más de una componente conexa de $K^c$ con medida dicho número.

La conjetura de McDuff afirma que si $\frac{\lambda_n}{\lambda_{n+1}} \nrightarrow 1$ entonces el conjunto $K$ no es $C^1$-minimal. Se pueden ver que en los ejemplos de Denjoy $\frac{\lambda_n}{\lambda_{n+1}} \to 1$ y además es fácil construir conjuntos de Cantor no $C^1$-minimales tales que $\frac{\lambda_n}{\lambda_{n+1}} \to 1$. También se puede construir conjuntos de Cantor afines (que no son $C^1$-minimales) que satisfagan que $\frac{\lambda_n}{\lambda_{n+1}} \nrightarrow 1$ y otros en que se satisfaga que $\frac{\lambda_n}{\lambda_{n+1}} \to 1$. En [IP] se puede ver algún avance en la dirección afirmativa de la conjetura.

## Referencias

[BIP] L. Bordignon, J. Iglesias, A. Portela. About $C^1$-minimality of the hyperbolic Cantor sets. Bulletin of the Brazilian Mathematical Society, New Series, 45 3 45 (2014), 525 - 542.

[D] A. Denjoy. Sur les courbes définies par les équations différentielles á la surface du tore. J. de Math Pure et Appl., 9(11)(1932),333-375.

[IP] J. Iglesias, A.Portela. Dynamically defined Cantor sets under the conditions of Mc Duff's conjecture. Colloquium Mathematicum, 2010.

[MD] D. McDuff. $C^1$-minimal subsets of the circle. Ann. Inst. Fourier, Grenoble, 31 (1981), 177-193.

[N] A. Norton. Denjoy minimal sets are far from affine. Ergod. Th. & Dynam. Sys., 22 (2002), 1803-1812.

[P] A. Portela. New examples of Cantor sets in $S^1$ that are not $C^1$-minimal. Bull Braz Math Soc, New Series 38(4) (2007), 623-633.

IMERL, Facultad de Ingeniería, Universidad de la República. Montevideo, Uruguay.
*Email address*: aldo@fing.edu.uy

PONENCIA

# WAGNER AND PETERSEN ARE UNIFORMLY MOST-RELIABLE GRAPHS

PABLO ROMERO

Publicaciones Matemáticas del Uruguay
Volumen 17, Julio 2019, Páginas 231–244
S 0797-1443

# WAGNER AND PETERSEN ARE UNIFORMLY MOST-RELIABLE GRAPHS

PABLO ROMERO

Abstract. If we are given $n$ nodes and $m$ links, what is the most reliable network topology? Partial answers are offered in the literature. Here, we show that Wagner and Petersen graphs are uniformly most-reliable graphs, and a conjecture on the existence of new such graphs is provided.

## 1. Motivation

Extremal graph theory is inspirational for network design [9]. In the second book ever written in graph theory, Berge challenges the readers to find the graph with maximum connectivity among all graphs with a fixed number of nodes and links. Frank Harary provided not only a full answer, but also found connected graphs with minimum and maximum diameter [16]. Gustav Kirchhoff solved linear time-invariant resistive circuits, and as corollary he introduced the Matrix-Tree theorem, where he counts the number of spanning trees of a graph (i.e., the tree-number) using the determinant of a matrix [18]. This breakthrough in electrical systems launched the theory of trees, which provides the building blocks in communication design [16]. However, the corresponding extremal problem is not well understood: find the graph with a fixed number of nodes and links that maximizes the tree-number [24].

The previous problems are deterministic. In network reliability analysis, the goal is to determine the probability of correct operation of a system [15, 4]. In its most elementary setting, we are given a simple graph $G$ with perfect nodes but random link failures with identical and independent probability $\rho$.

Even though network reliability is probabilistic in nature, there is a strong interplay with the previous deterministic problems. The motivation of this paper is to have a better understanding of the interplay between network reliability analysis and deterministic problems. The main contributions are the following:

(1) The interplay between uniformly most-reliable graphs and easy graphs is considered.
(2) Optimal augmentations of a cycle are produced, which lead us to Möbius graphs and Wagner graph in a special case.
(3) Wagner and Petersen graphs are formally proved to be uniformly most-reliable.
(4) A conjecture on uniformly most reliable graphs is here posed, inspired by Wagner graph and related works on the field.

©2019 PMU

This paper is organized as follows. Section 2 presents a formal definition of reliability polynomial, easy graphs and uniformly most-reliable graphs. Section 3 covers the body of related works on uniformly most-reliable graphs.

The practical value and potential application of uniformly most-reliable graphs in communication systems is discussed in Section 4. Augmentations arise as a natural approach once the network is already deployed. Assuming that most communication systems are 2-node connected, the analysis is focused on iterative augmentations of a cycle.

The result of these augmentations is Möbius graphs $M_n$. In Section 5 it is formally proved that Wagner graph $M_4$ is uniformly most-reliable. However, $M_5$ does not belong in the category, since Petersen graph is uniformly most-reliable.

Inspired by prior works in the field, it is conjectured that $(n, n+4)$ uniformly most-reliable graphs are obtained by successive elementary subdivisions of Wagner graph. Finally, Section 6 presents concluding remarks and trends for future work.

## 2. Uniformly Most-Reliable Graphs

We are given a simple graph $G = (V, E)$, with perfect nodes and unreliable links with failure probability $\rho$. The all-terminal reliability $R_G(\rho)$ measures the probability that the resulting random graph remains connected, and it is a polynomial in $\rho \in [0, 1]$. For convenience, in this paper we work with the unreliability polynomial $U_G(\rho) = 1 - R_G(\rho)$. Let us denote $p = |V|$ and $q = |E|$ the respective order and size of the graph $G$. Further, denote by $m_k(G)$, or simply $m_k$, the number of link-disconnecting sets with cardinality $k$, this is, the number of subsets $E' \subseteq E$ such that $|E'| = k$ and $G' = G - E'$ is disconnected. By sum-rule, the unreliability polynomial can be expressed as follows:

$$U_G(\rho) = \sum_{k=0}^{q} m_k \rho^k (1-\rho)^{q-k}. \tag{1}$$

Let us denote $(p, q)$-graph to the family of graphs with $p$ nodes and $q$ links. Clearly, if we consider a fixed $\rho$, there is at least one graph $H$ that attains the minimum unreliability, i.e., $U_H(\rho) \leq U_G(\rho)$ for all $(p, q)$ graph $G$. Further, if the previous condition holds for all $\rho \in [0, 1]$ and all $(p, q)$-graphs $G$, the graph $H$ is uniformly most-reliable.

The determination of the unreliability polynomial is equivalent to finding the coefficients $m_k$ for all $k$. Ball and Provan showed that this problem belongs to the $\mathcal{NP}$-Hard class [1]. However, they also proved that the determination of $m_\lambda$ is feasible in polynomial time. On the other hand, Kirchhoff found the tree-number $\tau(G)$ efficiently, so the number $m_{q-p+1} = \binom{q}{q-p+1} - \tau(G)$ can be obtained in polynomial time [18]. Observe that $m_k = 0$ whenever $k < \lambda$, and $m_k = \binom{q}{k}$ if $k > q - p + 1$. Under these considerations, the unreliability polynomial can be

rewritten [25]:

$$
\begin{aligned}
U_G(\rho) = {} & m_\lambda \rho^\lambda (1-\rho)^{q-\lambda} + \sum_{k=\lambda+1}^{q-p} m_k \rho^k (1-\rho)^{q-k} \\
& + \left( \binom{q}{q-p+1} - \tau(G) \right) \rho^{q-p+1} (1-\rho)^{p-1} \\
& + \sum_{k=q-p+2}^{q} \binom{q}{k} \rho^k (1-\rho)^{q-k}.
\end{aligned} \tag{2}
$$

Expression (2) holds whenever $\lambda + 1 \leq q - p$. Observe that in the remaining cases all the coefficients $m_k$ are fully known beforehand, and the unreliability polynomial accepts a straightforward calculation [12].

**Definition 1** (Level of difficulty)**.** *The level of difficulty of a graph $G$, denoted by $d(G)$, is the number of unknown coefficients from the unreliability polynomial:*

$$d(G) = (q-p) - (\lambda + 1) + 1 = (q - p + 1) - \lambda - 1. \tag{3}$$

If we are given a connected $(p, q)$-graph, then $q - p + 1$ represents the difference between its size and the size of a spanning tree. The number $s = q - p + 1$ is called the *co-rank* of a graph for algebraic reasons [5].

The invariant $d(G)$ measures *how difficult* is the problem of finding the unreliability polynomial. A valuable case occurs when $d(G) \leq 0$:

**Definition 2** (Easy graph)**.** *A graph $G$ is easy if $d(G) \leq 0$.*

The following result is just a corollary of the definition of easy graphs:

**Corollary 1.** *The coefficients $m_k$ are fully known finding only $\tau(G)$ and $m_\lambda$, if and only if $G$ is an easy graph.*

If we delete more than $s = q-p+1$ edges of a $(p, q)$-graph, the resulting subgraph is not connected. Therefore, $\lambda \leq s + 1$. However, the reader can observe that the equality is achieved in trees and elementary cycles. In [12], it is proved that they are the only graphs where the equality $\lambda = s + 1$ holds. In other terms, they belong to the class of easy graphs with the lowest level of difficulty $d = -2$.

**Corollary 2.** *All $(n, n + i)$ uniformly most-reliable graphs with $i < n/2$ have level of difficulty $d = i - 2$.*

*Proof.* The maximum connectivity among $(n, n + i)$-graphs is $\lambda_{max} = \lfloor 2 + \frac{2i}{n} \rfloor = 2$. The co-rank is $s = (n + i) - n + 1 = i + 1$. Therefore, the level of difficulty of all $(n, n+i)$ uniformly most-reliable graphs with $i < n$ is $d = (i+1) - 2 - 1 = i - 2$. □

A full characterization of easy graphs is already presented [12]. An exhaustive search among easy graphs reveals a simple way to prove that trees, cycles, Monma graphs and elementary subdivisions of $K_4$ and $K_{(3,3)}$ are uniformly most-reliable graphs for $i \in \{-1, 0, 1, 2, 3\}$ respectively.

## 3. Related Work

From inspection of Expression (1), we can see that if there exists some $(p, q)$-graph $H$ such that $m_k(H) \leq m_k(G)$ for all $k$ and all $(p, q)$-graph $G$, then $H$ is uniformly most-reliable. The level of difficulty is a natural way of understanding

the coefficient-based approach of uniformly most-reliable graphs. Curiously enough, this sufficient criterion is not known to be necessary. However, to the best of our knowledge, the search of uniformly most-reliable graphs rests on the minimization of all the coefficients $m_k$. This approach is promoted by the following result, which can be proved using elementary calculus [2]:

**Proposition 1.**

(i) *If there exists some integer $k$ such that $m_i(H) = m_i(G)$ for all $i < k$ but $m_k(H) < m_k(G)$, then there exists $\rho_0 > 0$ such that $U_H(\rho) < U_G(\rho)$ for all $\rho \in (0, \rho_0)$.*

(ii) *If there exists some integer $k$ such that $m_i(H) = m_i(G)$ for all $i > k$ but $m_k(H) < m_k(G)$, then there exists $\rho_1 < 1$ such that $U_H(\rho) < U_G(\rho)$ for all $\rho \in (\rho_1, 1)$.*

By definition, there are no disconnecting sets with lower cardinality than the link connectivity. Therefore, $m_i(G) = 0$ for all $i < \lambda$, and by Proposition 1-(i) uniformly most-reliably graphs must have the maximum link-connectivity $\lambda$. Furthermore, the number of disconnecting sets $m_\lambda$ must be minimized. On the other hand, $m_i(G) = \binom{q}{i}$ for all $i > q - p + 1$, since trees are minimally connected with $q = p - 1$ links. The number of connected sets with $q - p + 1$ links is precisely the tree-number $\tau(G)$, so $m_{q-p+1}(G) = \binom{q}{q-p+1} - \tau(G)$. Using Proposition 1-(ii), the tree-number should be maximized. Prior observations directly connect this network design problem with extremal graph theory:

**Corollary 3.** *A uniformly most-reliable $(p, q)$-graph $H$ must have the maximum tree-number $\tau(H)$, maximum connectivity $\lambda(H)$, and the minimum number of disconnecting sets $m_\lambda(H)$ among all $(p, q)$-graphs with maximum connectivity.*

For convenience we say that a $(p, q)$-graph, $H$, is $t$-optimal if $\tau(H) \geq \tau(G)$ for every $(p, q)$ graph $G$. Briefly, Corollary 3 claims that uniformly most-reliable graphs must be $t$-optimal and max-$\lambda$ min-$m_\lambda$, where $\lambda$ denotes the edge connectivity.

Frank Harary found the maximum connectivity of a $(p, q)$ graph. By handshaking, the average-degree of every $(p, q)$-graph is $\frac{2q}{p}$. If we denote $\delta(G)$ and $\lambda(G)$ the minimum degree and link-connectivity respectively, we immediately get that $\lambda(G) \leq \delta(G) \leq \lfloor \frac{2q}{p} \rfloor$. The candidate connectivity is $\lambda_{max} = \lfloor \frac{2q}{p} \rfloor$. It suffices to find a $(p, q)$-graph with connectivity $\lambda_{max}$ whenever $p \geq q - 1$ (otherwise, the graph is not connected). The evidence is the following family of graphs [16]:

**Definition 3** (Harary Graphs $H_{(n,k)}$)**.** *Let $n$ and $k$ be positive integers. Harary graph $H_{(n,k)}$ consists of $n$ nodes $\{v_0, \ldots, v_{n-1}\}$ equally spaced around a circle, and the following links:*

- *If $k$ is even, each vertex is adjacent to the $k/2$ nearest nodes in each direction.*
- *If $k$ is odd and $n$ is even, $H_{(n,k)}$ is $H_{(n,k-1)}$ with additional links $\{v_i, v_{i+\frac{n}{2}}\}$ for each $i = 0, \ldots, \frac{n}{2}$.*
- *If $k$ and $n$ are both odd, $H_{(n,k)}$ is $H_{(n,k-1)}$ with additional links $\{v_i, v_{i+\frac{n-1}{2}}\}$ for each $i = 0, \ldots, n - 1$.*

We immediately check that Harary graphs have maximum connectivity $\lambda_{\max} = \lfloor \frac{2q}{p} \rfloor$, so, they are max-$\lambda$. The number of disconnecting sets should be minimized as well; max-$\lambda$ graphs that minimize the disconnecting sets with $\lambda$ nodes are called max-$\lambda$ min-$m_\lambda$ graphs. Prior works from Bauer et. al. fully characterize max-$\lambda$

min-$m_\lambda$ graphs [3]. A key idea is to observe that in max-$\lambda$ graph, the number of disconnecting sets $m_\lambda$ is at least the number of nodes with degree $\lambda$. If this bound is achieved, a max-$\lambda$ min-$m_\lambda$ graph is retrieved. For that purpose, they define generalized Harary graphs, which are just an augmentation of the original Harary graphs with random matchings (this is, edges with non-adjacent nodes). In that way, the number of nodes with degree $\lambda$ is minimized, and the authors show that no other disconnecting sets with that size exists.

By Corollary 3, Bauer et. al. provide a family of graphs that contain all uniformly most-reliable graphs. Later works try to find uniformly $(p, p+i)$-most-reliable graphs for $i$ small, by a simultaneous minimization of all the coefficients $m_k$. The cases $i = -1$ and $i = 0$ are trivial. Indeed, when $q = p - 1$ all the trees have the same reliability polynomial $\rho^q$, so they are uniformly most-reliable (the reliability is zero if the graph is not connected). When $i = 0$ we have $q = p$, and the elementary cycle $C_p$ is $t$-optimal. All the other graphs with $p = q$ are not 2-connected, and by direct inspection we can see that $C_p$ achieves the minimum coefficients $m_k$.

Perhaps the first non-trivial uniformly most-reliable graphs were found by Boesch et. al. in 1991 [7]. A new reading of Bauer et. al. construction lead them to find that Monma graphs are $(n, n + 1)$ uniformly most-reliable graphs, whenever the number of nodes in each path differ by at most one. Interestingly enough, Clyde Monma et. al. used these graphs for the design of minimum cost two-node connected metric networks [22]. Figure 1 depicts Monma graphs. The reader is invited to find a combinatorial proof of Monma's $t$-optimality when the length of the paths differ by at most one in [13].

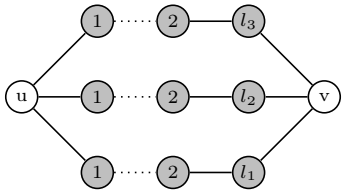



FIGURE 1. Monma graph $M_{(l_1+1,l_2+1,l_3+1)}$.

A more challenging problem is to find $(n, n + 2)$ uniformly most-reliable graphs. Boesch et. al. minimize the four effective terms $m_0$, $m_1$, $m_2$ and $m_3$ from Expression (1). An $(n, n+2)$ max-$\lambda$ min-$m_\lambda$ graph already minimizes the first three terms. If in addition the tree-number is minimized all the coefficients are simultaneously minimized, and the result must be a uniformly most-reliable graph. The merit of the paper [7] is to adequately select the feasible graphs from Bauer et. al. that minimizes the tree-number. Observe that $K_4$ can be partitioned into three perfect matchings, $PM_1$, $PM_2$ and $PM_3$. The result is that we should insert $n - 4$ points in the six links of $K_4$ in such a way that:

i the number of inserted nodes in all the links differ by at most one, and
ii if we insert the same number of nodes in two different matchings $PM_i \neq PM_j$, then the number of nodes in the four links from $PM_i \cup PM_j$ are identical.

The resulting $(n, n + 2)$-graph defines, for every $n \geq 4$, a single graph up to isomorphism. The authors formally prove that the resulting graph is uniformly

most-reliable $(n, n+2)$-graph. Furthermore, inspired by a previous research on $t$-optimality in multipartite graphs authored by Cheng [14], they conjecture that the shape of uniformly most-reliable $(n, n+3)$-graphs are elementary subdivisions of $K_{(3,3)}$. This conjecture is correct, and it was proved by Wang [26]. To the best of our knowledge, a full characterization of $(n, n+4)$-graphs is still open.

It is worth to remark that there are $(p, q)$-pairs where a uniformly most-reliable graphs does not exist [23]. The reader can consult [8] for a valuable survey on uniformly most-reliable graphs.

A full determination of $t$-optimal graphs for every $(p, q)$-pair is a related open problem. Indeed, a historical result credited by Leggett and Bedrosian asserts that $t$-optimal graphs must be almost regular, this is, the degrees do not differ by most than one [20]. Even though closed formulas are available for the tree-number of specific graphs, the progress on $t$-optimality is effective on special regularity conditions [14], almost-complete graphs or other special graphs with few links [24].

## 4. Practice and Augmentation

In this section we highlight potential application of uniformly most-reliable graphs, as an inspirational design tool. Real-life fiber-optics communication is already deployed in most countries, and a cost-effective solution is to add a single link in order to maximize the reliability. This is an augmentation problem: given a simple graph $G$, add a single link $e$ such that the unreliability $U_{G\cup e}(\rho)$ is minimized.

We know from previous result on uniformly most-reliable graphs that an unreliability minimization over the whole compact set $p \in [0, 1]$ is not always feasible for all graphs. Since fiber-optics are highly-reliable systems the elementary unreliability $\rho$ is small, and by Proposition 1, the unreliability is $U_G(\rho) \approx m_\lambda \rho^\lambda (1-\rho)^{q-\lambda}$, where $\lambda$ is the link-connectivity. Furthermore, we know that real-life physical implementation of fiber-optics are 2-node connected, so, every pair of nodes are included in a ring.

As a point of departure, we assume we are given a ring $G = C_n$ with $n$ even, and we study the augmentation problem in a step-by-step fashion. Specifically, we want to find the sequence of graphs $\{G^{(i)}\}_{i=0,\ldots,\lfloor \frac{n}{2} \rfloor}$ with $G^{(0)} = C_n$, such that the graph $G^{(i+1)} = G^{(i)} \cup \{e_{i+1}\}$ represents the best augmentation. In words, we consider iterative augmentations of the cycle. In network planning, this means that the operator decides to add a single link greedily, for instance, at different dates. In the following paragraphs we find the sequence $\{G^{(i)}\}_{i=0,\ldots,\lfloor \frac{n}{2} \rfloor}$.

By Handshaking Lemma, 3-regular graphs with $n$ nodes must have $m = 3n/2$ links. Thus, if we add less than $\lfloor \frac{n}{2} \rfloor$ links to the cycle $C_n$ there is at least one node $v_i$ with degree $deg(v_i) = 2$, and $\lambda(G^{(i)}) = 2$ for all $i < \lfloor \frac{n}{2} \rfloor$. This means that we should minimize the coefficient $m_2$ in a step-by-step fashion.

The key is to observe that a selected class of disconnecting sets is subsequently partitioned into *two classes*, whenever we iteratively add a single link to the cycle. Specifically, sort the nodes of the ring $\{0, 1, \ldots, 2k-1\}$ in clockwise. Without loss of generality, we choose $G^{(1)} = C_n \cup e_1$, where $e_1 = (0, x)$ for some node $x : 2 \leq x \leq 2k-2$. The minimally disconnecting sets are divided into two classes: pair of links that belong either to the elementary path $P_1 = 0, 1, \ldots, x$ or

$P_2 = x, x+1, \ldots, 0$. Since $C_n$ is already 2-node connected, there is no hope to delete some of extra links $e_i$ and obtain a minimally disconnecting set during the process for some graph $G^{(i)}$.

Mathematically, the number of minimally disconnecting sets in $G^{(1)} = C_n \cup \{(0, x)\}$ is:

$$m_2(x) = \binom{x}{2} + \binom{n-x}{2} = x^2 - nx + \frac{n^2 - n}{2} \tag{4}$$

The expression for $m_2(x)$ is minimized when $x = \frac{n}{2}$. The result is Monma graph $C_n + (0, \frac{n}{2}) = M_{(\frac{n}{2}, \frac{n}{2}, 1)}$. This is not uniformly most-reliable, since the three elementary paths are not balanced. However, it represents the best single augmentation to the ring. The following steps are straightforward, using the fact that each additional link divides the classes of the minimally disconnecting sets. The first class is two pair of links from the elementary path $P_1 = 0, 1, \ldots, y_1$, where $y_1$ is the first node adjacent to some of the additional links $e_i$. The following class is $P_2 = y_1, \ldots, y_2$ is the following node adjacent to some $e_i$, and so on. We define the sequence $G^{(i)}$ using an analogy of *fair cake-cutting*.

Suppose that you have a cake, but the number of guests is unknown. You are the host, and guests come before midnight. Guests only require to have one piece of cake, and the only rule you have at hand is to cut the cake as minimum as possible. A cut is a division of the cake by the diameter, and it must be performed whenever a guest has no piece of cake. As soon as the first guest arrives, the cake should be cut into two identical parts (this is the first link addition). When guest number 2 arrives, the following cut is performed such that we get four identical pieces of cake (one is yours). If guest number 3 arrives, he/she already has a piece of cake, and no cut is needed. At midnight, you and your guests eat their corresponding pieces of cake. In the analogy, every cut is a link addition (that connects the farthest nodes), and the process is finite. By the previous combinatorial argument, we get the following

**Theorem 1** (Fair Cake-Cutting)**.** *The best iterative augmentation of the cycle $C_n$ must be performed choosing the links $e_i$ as in a finite fair cake-cutting process.*

Observe that the Cake-Cutting process is finished after $\frac{n}{2}$ steps, and the result is a special cubic graph:

**Definition 4.** *For every even natural $n$, Möbius graph $M_n$ is constructed from the cycle $C_{2n}$ adding $n$ new links joining every pair of opposite nodes.*

**Corollary 4.** *Fare Cake-Cutting builds $M_{n/2}$.*

## 5. Finding Uniformly Most-Reliable Graphs

In this section we give two additional steps, studying $(n, n+i)$-graphs for $i \in \{4, 5\}$. Clearly, $K_5 - e$ is the only $(n, n+4)$-graph with 5 nodes, so it is uniformly most-reliable. The cases $n \in \{6, 7\}$ can be studied using exhaustive search [21]. By Proposition 1, if there exists a uniformly most-reliable graph it must be cubic (three-regular and connected). There are precisely 5 cubic graphs with 8 nodes [11]; see Figs 2-6. They were classified in [11] and systematically generated in [10]. Among

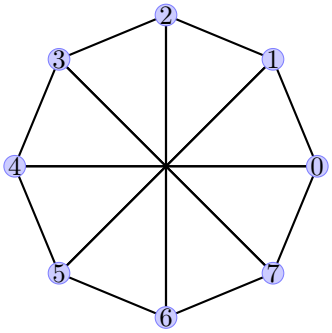



FIGURE 2. Wagner Graph $M_4$

$(8, 12)$-graphs, the cubic graph with the highest tree-number is precisely a special Harary graph $H_{8,3}$, named Wagner graph $M_4$ in the memory of the author for his research on Möbius ladders. Furthermore, since $t$-optimal graphs must be almost regular, in this case all nodes must have the same degree; so Wagner is $t$-optimal.

5.1. **Wagner graph.** The tree-number of Möbius graphs is [5]:

$$\tau(M_n) = \frac{n}{2}[(2+\sqrt{3})^n + (2-\sqrt{3})^n + 2] \tag{5}$$

The tree-number in $(8, 12)$-graphs is maximized in $\tau(M_4) = 392$. Therefore, the only candidate of uniformly most-reliable graph with 8 nodes is Wagner graph $M_4$ depicted in Figure 2. Incidentally, this graph is the product of Fair Cake-Cutting process after four augmentations. The following result can be obtained by an exhaustive search of disconnecting sets. Here we choose a combinatorial argument, since it is preparatory for the main theorem.

**Lemma 1.** *Wagner graph $M_4$ is minimizer of the coefficient $m_4$ among all cubic $(8, 12)$-graphs.*

*Proof.* After the removal of 4 links, the minimum degree of the resulting graph must be $\delta \in \{0, 1, 2\}$. Let us denote $u_i$ to the number of those disconnected subgraphs whose minimum degree is $\delta = i$. Therefore $m_4 = u_0 + u_1 + u_2$. Let us count and sum the three disjoint cases in order:

(1) Type-0: three links incident to a fixed node and another link, $u_0 = 8 \times 9 = 72$ cases ($\delta = 0$).
(2) Type-1: if we remove the four links adjacent to a fixed link we disconnect the graph. Thus, $u_1 \geq 12$.
(3) Type-2: the resulting graph is 2-regular. Therefore, $u_2$ counts all disconnecting perfect matchings.

By inspection we see that $u_1(M_4) = 12$ and $u_2(M_4) = 2$, so $m_4(M_4) = 72 + 12 + 2 = 86$. Recall that $u_0 = 72$ and $u_1 \geq 12 = u_1(M_4)$ in all $(8, 12)$-cubic graphs. In the cube $u_2(Q_3) = 3 > 2$, so $m_4$ is greater than in Wagner graph.

Graphs $G_1$, $G_2$ and $G_3$ present disconnecting sets of three non-adjacent links. There are 9 ways to pick another link. At most one of them is type-2, and clearly no-one is type-0. Therefore $u_1(G_i) \geq 12 + 8 = 20$, so $m_4(G_i) \geq 92 \geq m_4(M_4)$, for $i \in \{1, 2, 3\}$. □

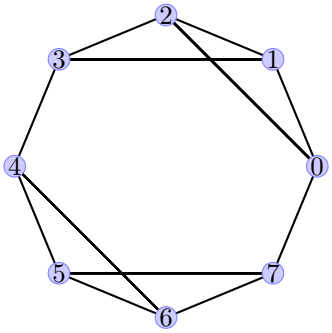



FIGURE 3. Cubic Graph $G_1$

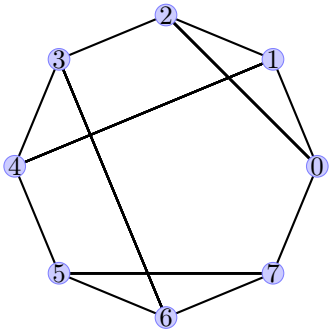



FIGURE 4. Cubic graph $G_2$

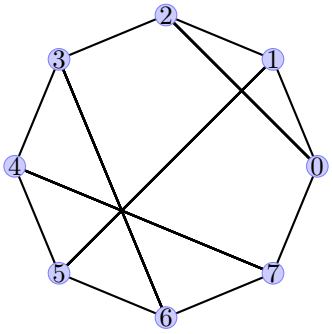



FIGURE 5. Cubic graph $G_3$

**Theorem 2.** *Wagner graph is uniformly most-reliable.*

*Proof.* Wagner graph is $t$-optimal and max-$\lambda$, with $\lambda = 3$. By inspection it is super-$\lambda$, so $m_\lambda = m_3 = 8$ is minimum. Using Expression (5) we get that $m_5 = \binom{12}{5} - \tau(M_4) = 792 - 392 = 400$ is also minimum. The level of difficulty is $d(M_4) = 1$. Using Lemma 1, we know that $m_4(M_4) = 86$. It suffices to prove that $m_4(H) \geq 86$ for all $(8, 12)$-graphs $H$.

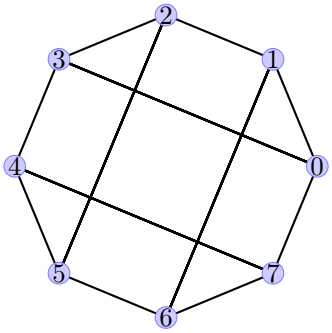



FIGURE 6. Cube $Q_3$

We will study disjoint cases and use combinatorial arguments. If $H$ has a bridge $m_4(H) \geq \binom{11}{3} > 86$. We consider bridgeless connected graphs, so in the following, $\delta(H) \geq 2$. If $\delta(H) = 2$ and we assume that $deg(v_1) = deg(v_2) = 2$ for two different nodes in $H$, there are two disjoint cases:

(i) $v_1$ and $v_2$ are adjacent.
(ii) $v_1$ and $v_2$ are not adjacent.

If (i), we know that $U = \{(v_1, v_2), (v_1, x), (v_2, y)\} \subseteq E(H)$ for some nodes $x$ and $y$. If we pick two links from $U$ and two from $E(H) - U$, we get disconnecting sets. Therefore, $m_4(H) \geq \binom{3}{2} \times \binom{9}{2} > 86$.

If (ii), $U = \{(v_1, x_1), (v_1, y_1), (v_2, x_2), (v_2, y_2)\} \subseteq E(H)$ for some $x, y \in V(H)$. We can pick $(v_1, x_1), (v_1, y_1)$ and two other links from $E(H) - U$: $\binom{8}{2} = 28$ cases; $(v_2, x_2), (v_2, y_2)$ and two other links from $E(H) - U$: $\binom{8}{2} = 28$ cases; or three links from $U$ and one from $E(H) - U$: $8 \times 4 = 32$ cases. Thus $m_4(H) \geq 28 + 28 + 32 > 86$.

The remaining cases are $\delta(H) = 2$ with a single node with degree 2, or $\delta(H) \geq 3$, which must be cubic graphs. By handshaking, in the first case the degree sequence must be $(4, 3, 3, 3, 3, 3, 3, 2)$. By Erdos-Gallai characterization theorem, this sequence is graphic. We consider two disjoint disconnecting sets:

(1) The incident links of $v$ and two additional links: $\binom{10}{2} = 45$ cases;
(2) Three incident links of a degree-3 node and another link: $6 \times 9 = 54$ cases.

Therefore, all feasible graphs with degree-sequence $(4, 3, 3, 3, 3, 3, 3, 2)$ have at least $m_4(H) \geq 45 + 54 > 86$ disconnecting sets with 4 links. The result for cubic graphs holds by Lemma 1. □

5.2. **Conjecture.** The following classes of graphs are uniformly most reliable:

- Trees among $(n, n-1)$-graphs.
- Cycles among $(n, n)$-graphs.
- Monma with balanced paths among $(n, n+1)$-graphs.
- $K_4$ and special subdivisions among $(n, n+2)$-graphs [7].
- $K_{(3,3)}$ and special subdivisions among $(n, n+3)$-graphs [26].
- Wagner graph $M_4$ (Theorem 3).

Trees, cycles and Monma graphs are easy graphs. They are all uniformly most-reliable graphs, and the analysis is elementary. The rationale behind the elementary subdivisions of $K_4$ and $K_{(3,3)}$ is identical. Indeed, the authors in both papers consider a partition into three disjoint perfect matchings, and introduce elementary

subdivisions to them. All Möbius graphs accept a partition into three disjoint perfect matchings. Furthermore, $M_2 = K_4$, $M_3 = K_{(3,3)}$ and $M_4$ is Wagner graph. Recall that all $(n, n+2)$ and $(n, n+3)$ uniformly most-reliable graphs are obtained adding links in a balanced manner to the three disjoint perfect matchings. This promotes a conjecture:

**Conjecture 1.** *All uniformly most-reliable $(n, n+4)$ graphs with $n \geq 8$ are elementary subdivisions of Wagner graph.*

An optimistic prediction from Conjecture 1 is that all Möbius ladders and special subdivisions are uniformly most-reliable $(n, n+i)$ graphs. However, this generalization is false, since Petersen graph is $t$-optimal [17]. Curiously enough, this reinforces Donald Knuth's statement that Petersen graph serves as a counterexample to several optimistic predictions in graph theory [19].

5.3. **Petersen Graph is Uniformly Most-Reliable.** Petersen graph is the complement of the line-graph of $K_5$ (the reader can find alternative definitions in the book [17]). By Sachs theorem [6], its eigenvalues are 3 (simple), 1 (multiplicity 5) and $-2$ (multiplicity 4). Therefore, its tree-number is $\tau(P) = \frac{1}{10} \times (3-1)^5 \times (3-(-2))^4 = 2000$. By prior works in the literature it is known that Petersen is $t$-optimal. Therefore, Petersen graph is the only candidate to be uniformly most-reliable $(10, 15)$-graph. It is clearly super-$\lambda$ with connectivity $\lambda = 3$, so $m_3 = 10$ is minimum among $(10, 15)$-graphs. From inspection we find that all the disconnecting sets with 4 links are either incident to a fixed node or fixed link, so $m_4 = 10 \times \binom{3}{3}\binom{12}{1} + 15 = 135$. Furthermore, all cubic $(10, 15)$-graphs possess the previous disconnecting sets. In order to count $m_5$ we observe that such disconnecting sets isolate nodes, links, 2-paths or 5-cycles, so, $m_5 = (10 \times \binom{12}{2} - 15) + 15 \times 10 + (\binom{10}{2} - 15) + 6 = 831$. From now on, we will assume that the ground graph-set is always $(10, 15)$-graphs. The following two lemmas are preparatory for the main result, and their proofs only use combinatorial arguments:

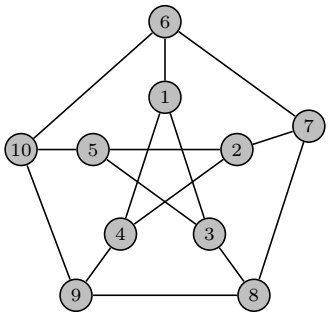



FIGURE 7. Petersen graph

**Lemma 2.** *The coefficient $m_4$ is minimized in Petersen graph.*

*Proof.* The result is trivial for cubic graphs. Consider an arbitrary $(10, 15)$-graph $H$, and denote $m_4 = 135$ the number of disconnecting sets in Petersen graph. If $H$ has a bridge, then $m_4(H) \geq \binom{13}{3} \geq m_4$. It suffices to prove the result when $\delta(H) = 2$. If there exists non-adjacent nodes $v_1$ and $v_2$ such that $deg(v_1) = deg(v_2) = 2$, then

$m_4(H) \geq 2 \times \binom{13}{2} - 1 > m_4$. If $v_1$ and $v_2$ are adjacent nodes then $m_4(H) \geq \binom{3}{2}\binom{12}{2} > m_4$. Finally, if there is a single node $v$ such that $deg(v) = 2$, by Handshaking Lemma the degree-sequence must be $(4, 3, 3, 3, 3, 3, 3, 3, 3, 2)$. In this case, counting the ways to disconnect a node we know that $m_4(H) \geq \binom{4}{4} + 8 \times \binom{3}{3}\binom{12}{1} + \binom{2}{2}\binom{13}{2} > m_4$. □

**Lemma 3.** *The coefficient $m_5$ is minimized in Petersen graph among all cubic $(10, 15)$-graphs.*

*Proof.* □

The following result is analogous to Lemma 2:

**Lemma 4.** *The coefficient $m_5$ is minimized in Petersen graph.*

*Proof.* By Lemma 3 we know that the result holds for in cubic graphs. We know that $m_5 = 831$ in Petersen graph. In the following, we remark that only trivial disconnecting sets are considered for counting, unless specified otherwise. If $H$ has a bridge, then $m_5(H) \geq \binom{14}{4} \geq m_5$. It suffices to prove the result when $\delta(H) = 2$. If $deg(v_1) = deg(v_2) = deg(v_3) = 2$ for three different nodes, we consider three disjoint and exhaustive cases:

(i) They are connected in a 2-path: $m_5(H) \geq \binom{4}{2}\binom{11}{3} + \binom{4}{3}\binom{11}{2} + \binom{4}{4}\binom{11}{1} >> m_5$
(ii) There are exactly two adjacent nodes: $m_5(H) \geq \binom{13}{3} + \binom{3}{2}\binom{12}{3} + \binom{3}{3}\binom{12}{2} - 31 \geq m_5$;
(iii) Non-adjacent nodes: $m_5(H) \geq 3 \times \binom{13}{3} - 3 \times 11 = 825$. Observe that there exists at least one node $v_4$ with degree 3. Adding trivial disconnecting sets related to $v_4$ we find $\binom{12}{2} - 3 = 63$ more sets. Therefore $m_5(H) \geq 825 + 63 = 888 > m_5$.

Assume that there are precisely two different nodes $v_1 \neq v_2$ such that $deg(v_1) = deg(v_2) = 2$. We know that $deg(v_i) = 3 + \delta_i$ for $i = 3, \ldots, 10$. By Handshaking Lemma we know that $30 = \sum_i deg(v_i)$, so $\sum_{i=3}^{10} \delta_i = 2$. Therefore, the only graphic degree-sequences with two degree-two nodes are $D_1 = (4, 4, 3, 3, 3, 3, 3, 3, 2, 2)$ and $D_2 = (5, 3, 3, 3, 3, 3, 3, 3, 2, 2)$. We consider four cases: $D_1$ or $D_2$ with adjacent or non-adjacent ($A$-$NA$) nodes $v_1$ and $v_2$:

(i) $D_1$ and $A$: $m_5(H) \geq 2 \times \binom{4}{4}\binom{11}{1} + 6 \times \binom{3}{3}\binom{12}{2} + \binom{3}{2}\binom{12}{3} + \binom{3}{3}\binom{12}{2} - 6 \times 3 = 1126 > m_5$;
(ii) $D_1$ and $NA$: $m_5(H) \geq 2 \times \binom{4}{4}\binom{11}{1} + 6 \times \binom{3}{3}\binom{12}{2} + 2 \times \binom{2}{2}\binom{13}{3} - 2 \times 5 - 11 = 969 > m_5$;
(iii) $D_2$ and $A$: $m_5(H) \geq \binom{5}{5} + 7 \times \binom{3}{3}\binom{12}{2} + \binom{3}{2}\binom{12}{3} + \binom{3}{3}\binom{12}{2} - 7 \times 3 > m_5$;
(iv) $D_2$ and $NA$: $m_5(H) \geq \binom{5}{5} + 7 \times \binom{3}{3}\binom{12}{2} + 2 \times \binom{2}{2}\binom{13}{3} - 2 \times 7 - 11 = 1010 > m_5$.

Finally, we consider the case where there exists only one degree-2 node. In this case, the degree-sequence must be $(4, 3, 3, 3, 3, 3, 3, 3, 3, 2)$. Counting trivial disconnecting sets we get that $m_5(H) \geq \binom{4}{4}\binom{11}{1} + 8 \times \binom{3}{3}\binom{12}{2} + \binom{2}{2}\binom{13}{3} - 8 = 825$. However, it does not suffice to close the proof. We must find at least 6 non-trivial disconnecting sets. Observe that there exists at least 11 links whose extremes are nodes with degree 2 or 3, so $m_5(H) \geq 825 + 11 = 836 > m_5$. □

**Theorem 3.** *Petersen is uniformly most-reliable.*

*Proof.* Recall that Petersen is super-$\lambda$, so $m_3$ is minimized. Clearly, $m_i = 0$ for $i \in \{0, 1, 2\}$, and $m_i = \binom{15}{i}$ for all $(10, 15)$-graphs, when $i \geq 7$. Petersen is $t$-optimal,

thus, it minimized the coefficient $m_6$. Combining Lemmas 2 and 4 we know that Petersen graph minimizes simultaneously all the coefficients $m_i$. □

## 6. Conclusions and Trends for Future Work

Uniformly most-reliable graphs represent a synthesis in network reliability analysis. Finding them is a hard task not well understood. Prior works in the field try to globally minimize the coefficients of disconnecting sets. This methodology provides uniformly most-reliable $(n, n+i)$ graphs for every $i \in \{-1, 0, 1, 2, 3\}$. Here, we show the interplay between easy graphs and uniformly most-reliable graphs. We observe that Wagner graph $M_4$ is uniformly most-reliable. The result is suggested by an iterative augmentation of a cycle, here called Fair Cake-Cutting theorem. The paper is closed with the conjecture that $(n, n+4)$ uniformly most-reliable graphs are elementary subdivisions of Wagner graph.

There are several trends for future work. A powerful methodology to find uniformly most-reliable graphs is not known. A full characterization of $t$-optimal graphs is an open problem. Conjecture 1 could be studied with a parallel reasoning of Boesch [7] and Wang [26].

## 7. Acknowledgements

This work is partially supported by Project 395 CSIC I+D *Sistemas Binarios Estocásticos Dinámicos.*

## References

[1] Michael O. Ball and J. Scott Provan. The complexity of counting cuts and of computing the probability that a graph is connected. *SIAM J. Computing*, 12:777–788, 1983.

[2] D. Bauer, F. Boesch, C. Suffel, and R. Van Slyke. On the validity of a reduction of reliable network design to a graph extremal problem. *IEEE Transactions on Circuits and Systems*, 34(12):1579–1581, Dec 1987.

[3] D. Bauer, F. Boesch, C. Suffel, and R. Tindell. Combinatorial optimization problems in the analysis and design of probabilistic networks. *Networks*, 15(2):257–271, 1985.

[4] L.W. Beineke, R.J. Wilson, and O.R. Oellermann. *Topics in Structural Graph Theory.* Encyclopedia of Mathematics and its Applications. Cambridge University Press, 2012.

[5] N. Biggs. *Algebraic Graph Theory.* Cambridge Mathematical Library. Cambridge University Press, 1993.

[6] N. Biggs. *Algebraic Graph Theory.* Cambridge Mathematical Library. Cambridge University Press, 1993.

[7] F. T. Boesch, X. Li, and C. Suffel. On the existence of uniformly optimally reliable networks. *Networks*, 21(2):181–194, 1991.

[8] F.T. Boesch, A. Satyanarayana, and C.L. Suffel. A survey of some network reliability analysis and synthesis results. *Networks*, 54(2):99–107, 2009.

[9] B. Bollobás. *Extremal Graph Theory.* Dover Books on Mathematics. Dover Publications, 2004.

[10] Gunnar Brinkmann, Jan Goedgebeur, and Brendan D. Mckay. Generation of Cubic graphs. *Discrete Mathematics and Theoretical Computer Science*, Vol. 13 no. 2, January 2011.

[11] F.C Bussemake, S Cobeljic, D.M Cvetkovic, and J.J Seidel. Cubic graphs on $\leq$14 vertices. *Journal of Combinatorial Theory, Series B*, 23(2):234 – 235, 1977.

[12] E. Canale, P. Romero, G. Rubino, and X. Warnes. Network utility problem and easy reliability polynomials. In *2016 8th International Workshop on Resilient Networks Design and Modeling (RNDM)*, pages 79–84, Sept 2016.

[13] Eduardo Canale, Juan Piccini, Franco Robledo, and Pablo Romero. Diameter-constrained reliability: Complexity, factorization and exact computation in weak graphs. In *Proceedings of the Latin America Networking Conference on LANC 2014*, LANC '14, pages 1–7, New York, NY, USA, 2014. ACM.

[14] Ching-Shui Cheng. Maximizing the total number of spanning trees in a graph: Two related problems in graph theory and optimum design theory. *Journal of Combinatorial Theory, Series B*, 31(2):240 – 248, 1981.
[15] Charles J. Colbourn. Reliability issues in telecommunications network planning. In *Telecommunications network planning, chapter 9*, pages 135–146. Kluwer Academic Publishers, 1999.
[16] Frank Harary. The maximum connectivity of a graph. *Proceedings of the National Academy of Sciences of the United States of America*, 48(7):1142–1146, 1962.
[17] D.A. Holton and J. Sheehan. *The Petersen Graph*. Australian Mathematical Society Lecture Series. Cambridge University Press, 1993.
[18] G. Kirchoff. Über die auflösung der gleichungen, auf welche man bei der untersuchung der linearen verteilung galvanischer ströme geführt wird. *Ann. Phys. Chem.*, 72:497–508, 1847.
[19] D.E. Knuth. *The art of computer programming: Introduction to combinatiorial algorithms and Boolean functions*. Addison-Wesley series in Computer Science and Information Proceedings. Addison-Wesley, 2008.
[20] J. D. Leggett and S. D. Bedrosian. On networks with the maximum numbers of trees. In *Proc. of Eighth Midwest Symposium on Circuit Theory*, pages 1–8, June 1965.
[21] Behrang Mahjani. *Exploring connectivity of random subgraphs of a graph*. PhD thesis, Chalmers University of Technology, Göteburg, Sweden, 2010.
[22] ClydeL. Monma, BethSpellman Munson, and WilliamR. Pulleyblank. Minimum-weight two-connected spanning networks. *Mathematical Programming*, 46(1-3):153–171, 1990.
[23] Wendy Myrvold, Kim H. Cheung, Lavon B. Page, and Jo Ellen Perry. Uniformly-most reliable networks do not always exist. *Networks*, 21(4):417–419, 1991.
[24] L. Petingi, F. Boesch, and C. Suffel. On the characterization of graphs with maximum number of spanning trees. *Discrete Mathematics*, 179(1):155 – 166, 1998.
[25] F. Robledo, P. Romero, and P. Sartor. A novel interpolation technique to address the edge-reliability problem. In *2013 5th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, pages 187–192, Sept 2013.
[26] Guifang Wang. A proof of Boesch's conjecture. *Networks*, 24(5):277–284, 1994.

IMERL, Facultad de Ingeniería, Universidad de la República. Montevideo, Uruguay.
*Email address*: promero@fing.edu.uy

PONENCIA

# UNA INTRODUCCIÓN A MODELOS MATEMÁTICOS PARA COMPOSICIÓN AUTOMÁTICA DE MÚSICA TONAL

VERÓNICA RUMBO

Publicaciones Matemáticas del Uruguay
Volumen 17, Julio 2019, Páginas 247–259
S 0797-1443

# UNA INTRODUCCIÓN A MODELOS MATEMÁTICOS PARA COMPOSICIÓN AUTOMÁTICA DE MÚSICA TONAL

VERÓNICA RUMBO

## 1. Introducción

En este trabajo se procura hacer una breve presentación de algunos modelos explorados en busca de responder la siguiente pregunta: ¿Hasta que punto es posible generar aleatoriamente música con cierto "sentido"? ¿Cómo se puede hacer para aprender los rasgos carácterísticos de un estilo? Para ello será necesario presentar algunas nociones muy básicas de teoría musical que serán necesarias para los ejemplos que se propondrán a continuación.

Consideremos un modelo –bastante simplificado– en el cual un sonido está determinado por dos parámetros:

- **Altura:** depende la frecuencia del sonido percibido y se asocia con la condición de *grave* o *agudo* de éste.
- **Duración:** es la longitud del sonido en el tiempo.

Supondremos que los valores que estos parámetros toman son discretos. Se le llama *nota* a un par (altura, duración). Cuando trabajemos sólo con alturas la palabra "nota" podrá aparecer también en referencia a alturas.

En este contexto, llamaremos melodía a una secuencia de notas (alturas) y supondremos que cada nota depende –a lo sumo– de la anterior.

En el tipo de música con el que trabajaremos (conocida como *música tonal*) las notas están fuertemente jerarquizadas, distribuyendo sus roles en torno a una nota principal denominada tónica. Para que nuestros ejemplos respeten esta jerarquía elegiremos manualmente el subconjunto de notas posibles a utilizar.[1]

Otros conceptos importantes son los de *armonía* y *acorde*. Un acorde suele definirse como un conjunto de notas tocadas simultáneamente, sin embargo, incluso en las melodías en las que no hay superposición de voces suele haber siempre acordes subyacentes, cada uno de los cuales tiene una función y se combinan de acuerdo a ciertas reglas. Armonía refiere justamente a esa combinación de acordes, así como a la disciplina que trata cómo combinarlos.

No nos detendremos en la clasificación de acordes ni en los criterios para determinar un acorde que no siempre es explícito. A los efectos de comprender el trabajo basta entender el acorde como conjunto de (habitualmente 3 o 4) notas, por lo general sin distinción de octava. Mientras rige un acorde, las notas de uso prioritario en la melodía son precisamente las notas que lo integran y el uso de las demás notas queda supeditado a éstas (es decir, se usan como ornamentos o puentes entre notas del acorde).

[1]También pueden utilizarse todas las notas disponibles, estimando sus probabilidades de aparición a partir de un conjunto de melodías. Un ejemplo con este enfoque puede verse en [6]

©2019 PMU

## 2. Modelando alturas: Cadenas de Markov con restricciones

**2.1. Algunos conceptos básicos.** En primer lugar nos enfocaremos en modelar las alturas, para lo cual utilizaremos cadenas de Markov, es decir, sucesiones de variables aleatorias que dependen (a lo sumo) de la variable inmediatamente anterior. Se presentan a continuación algunas definiciones y propiedades básicas que necesitaremos, para una introducción más detallada a cadenas de Markov en espacios finitos se sugiere consultar [2].

**Definición 2.1.** Sean $E = \{e_i\}_{i\in I} \subset \mathbb{R}$ un conjunto finito o numerable con una numeración $I$ [2], $\mu = (\mu_i)_{i\in I}$ un vector o familia numerable de reales positivos con $\sum_{i\in I} \mu_i = 1$ y $\{X_k\}_{k\in\mathbb{N}}$ una sucesión de variables aleatorias que toman valores en $E$.
Diremos que $\{X_k\}$ es una cadena de Markov con espacio de estados $E$ y distribución inicial $\mu$ si:

- $X_0$ tiene distribución $\mu$, es decir, si $P(X_0 = e_i) = \mu_i$, $\forall i \in I$.
- Para todo $n \in \mathbb{N}$ y para toda $n$-upla $e_{i_0}, e_{i_1}, ..., e_{i_n}$ de elementos de $E$ se cumple que

$$P(X_n = e_{i_n} | X_{n-1} = e_{i_{n-1}}, X_{n-2} = e_{i_{n-2}}, ..., X_0 = e_{i_0}) = \\ = P(X_n = e_{i_n} | X_{n-1} = e_{i_{n-1}}).$$

Al conjunto $E$ lo denominamos *espacio de estados* mientras que a $\mu$ lo llamamos *distribución inicial*

Si en una cadena de Markov se tiene que para todo par de estados $e_i, e_j$ las probabilidades $P(X_n = e_j | X_{n-1} = e_i)$ no dependen de $n$ se dice que la cadena es *homogénea* (en el tiempo). En ese caso tiene sentido denominar a dicha probabilidad como $p_{ij}$ (la probabilidad de ir de i a j en un paso). Si la cadena no es homogénea será necesario también indicar el instante considerado.

Para simplificar la notación, asumiremos en general que el espacio de estados es de la forma $\{1, \ldots, n\}$ o bien $\mathbb{N}$ (eventualmente $\mathbb{Z}$) si el espacio es infinito). No se pierde generalidad, ya que simplemente identificamos el espacio de estados $E$ con la numeración $I$. Asimismo, notaremos $p_{ij} := P(X_1 = j | X_0 = i)$.

**Definición 2.2.** Sea $\{X_k\}_{k\in\mathbb{N}}$ una cadena de Markov homogénea con espacio de estados $E$ como en la definición 2.1. Definimos su matriz de transición $P$ como la matriz $(p_{ij})_{i,j\in E}$ tal que $p_{ij} = P(X_n = j | X_{n-1} = i)$.

*Observación* 2.3.

- Las entradas de una matriz de transición $p_{ij}$ verifican que $\sum_{j\in I} p_{ij} = 1$ para todo $i \in E$. Decimos que tales matrices son *estocásticas*.
- Para el caso de las cadenas no homogéneas puede definirse una familia de matrices $\{P^{(n)}\}$ con entradas $p_{ij}^{(n)} = P(X_n = j | X_{n-1} = i)$.

*Definición* 2.4. Dada $\{X_k\}_{k\in\mathbb{N}}$ cadena de Markov con espacio de estados $E$, matriz de transición $P$ y distribución inicial $\mu$ definimos su *matriz de transición de orden* $n$ como[3]

$$P^n = (p_{ij}^n)_{i,j\in E}, \quad \text{siendo} \quad p_{ij}^n = P(X_n = j | X_0 = i),$$

[2]Se puede pensar por ejemplo en que el espacio de estados es $\{1, \ldots N\}$ o $\mathbb{N}$ según corresponda

[3]No confundir con $P^{(n)}$, la matriz de transición del tiempo $n-1$ al $n$ en cadenas no homogéneas definida antes.

donde las probabilidades $p_{ij}^n$ se denominan *probabilidades de transición de orden n*. Asimismo, llamaremos *distribución de probabilidad de orden n* al vector

$$\mu^n = (\mu_i^n)_{i\in E}, \quad \text{siendo} \quad \mu_i^n = P(X_n = i).$$

Cabe notar que $P^n$ es por el momento sólo una notación y no refiere a la potencia la matriz $P$. Sin embargo se puede probar que ambas coinciden.

*Proposición* 2.5. *Sea $\{X_k\}_{k\in\mathbb{N}}$ una cadena de Markov con espacio de estados $E$, matriz de transición $P = (p_{ij})$ y distribución inicial $\mu$. Para todo $n \in \mathbb{N}$ se cumple:*

1. *$P^n$ definida como antes es también la enésima potencia de la matriz $P$.*
2. *$\mu^n = \mu \times P^n, \forall n \in \mathbb{N}$.*

*Además la probabilidad de una trayectoria $s_0, s_1 \ldots, s_n$ dada es*

$$\begin{aligned} P(X_n = s_n, X_{n-1} = s_{n-1}, \ldots, X_0 = s_0) = \\ = P(X_n = s_n | X_{n-1} = s_{n-1}) P(X_{n-1} = s_{n-1} | X_{n-2} = s_{n-2}) \ldots \\ \ldots P(X_1 = s_1 | X_0 = s_0) P(X_0 = s_0) = \\ = p_{s_{n-1}s_n} p_{s_{n-2}s_{n-1}} \ldots p_{01} \mu_{s_0}. \end{aligned}$$

De ese modo una cadena de Markov (homogénea) queda determinada por su espacio de estados, su distribución inicial y una matriz estocástica que será su matriz de transición.

*Ejemplo* 2.6 (Paseos al azar). Si $\{X_k\}_{k\in\mathbb{N}}$ una sucesión de variables aleatorias i.i.d. que toman valores en $\mathbb{Z}$ se tiene que $S_k = \sum_{i=1}^k X_i$ es una cadena de Markov. En efecto, si $i_1, \ldots, i_n$ son números enteros,

$$\begin{aligned} P(S_n = i_n | S_{n-1} = i_{n-1}, \ldots, S_0 = i_0) = \\ = P(\sum_{i=1}^{n} X_k = i_n | \sum_{i=1}^{n-1} X_k = i_{n-1}, \ldots, X_0 = i_0) \\ = P(X_n + i_{n-1} = i_n) = P(X_n = i_n - i_{n-1}). \end{aligned}$$

Lo mismo ocurre al calcular $P(S_n = i_n | S_{n-1} = i_{n-1})$ ya que por la independencia de $\{X_k\}$ lo único que se necesita es conocer el valor de la suma parcial en el instante $n-1$.

Además, dados $i, j \in \mathbb{Z}$ la probabilidad de transición de $i$ a $j$ es $d_{j-i}$, notando $d_k := P(X = k)$. Así, considerando los propios estados como índices la matriz de transición resulta ser $P = (p_{ij})_{i,j} = (d_{j-i})_{i,j}$.

Este ejemplo nos da una estrategia (bastante ingenua) para generar secuencias de alturas: Se simula $\{X_0, \ldots, X_n\}$ una trayectoria de un paseo al azar en $\mathbb{Z}$, asociando a cada estado una altura (considerando por ejemplo notas de cierta escala o acorde).

*¿Cómo simular el paseo al azar?* Si consideramos, por ejemplo, un estado inicial $X_0$ fijo y trayectorias de largo finito, el paseo al azar se comporta como una cadena de Markov con espacio de estados finitos (ya que el conjunto de los enteros a los que se llega con probabilidad no nula es finito). Para tales cadenas tenemos el siguiente esquema de simulación:

- Simular $X_0$ con distribución $\mu$.
- Para $k \geq 1$, si $X_{k-1} = i$ simular $X_k$ con distribución $(p_{i1}, p_{i2}, \ldots, p_{in})$ (es decir la $i$-ésima fila de la matriz de transición).

Para una descripción más detallada del procedimiento puede consultarse [2].

**2.2. Estimación de las probabilidades.** Para simular paseos con el procedimiento anterior es necesario disponer de la distribución inicial y la matriz de transición. Una posible estrategia para elegirlas consiste en "aprenderlas" a partir de una familia de trayectorias[4] (que suponemos se comportan de acuerdo a la cadena).

Supongamos que observamos un conjunto finito $S = \{s^n\}_{n\in\{1,\dots,N\}}$ de trayectorias finitas e independientes entre sí de una cadena de Markov homogénea con espacio de estados finito. Podemos estimar su distribución inicial y matriz de transición estimando sus entradas por máxima verosimilitud, es decir,

$$\hat{\mu_i} = \frac{1}{N}\sum_{n=1}^{n=N} \mathbb{1}_{\{s_0^n=i\}},$$

$$\hat{p}_{ij} = \begin{cases} \frac{1}{N_i}\sum_{n=1}^{n=N}\sum_{s_k:s_{k-1}=i} \mathbb{1}_{\{s_k=j\}}, & \text{si } N_i \neq 0, \\ 0, & \text{si } N_i = 0, \end{cases}$$

con $N_i = \sum_{n\in\{1,\dots,N\}} \#\{s_k : s_{k-1} = i\}$ la cantidad total de observaciones precedidas por $i$. En otras palabras, $\hat{\mu_i}$ es la proporción de trayectorias que comienza en $i$, y para hallar $\hat{p_{ij}}$ se consideran todas las transiciones en $S$ que comienzan en $i$, calculando la proporción de éstas que resuelve en $j$.

*Observación* 2.7. El procedimiento anterior se utiliza para la estimación de cadenas homogéneas en general, no únicamente paseos al azar.

A la hora de modelar música, elegir bien el conjunto de trayectorias (melodías) sobre los cuales estimar es importante. Por ejemplo al generar música tonal consideramos para la estimación melodías en la misma tonalidad. De este modo las notas y transiciones "extrañas" aparecen con baja probabilidad. Otra posibilidad es establecer a priori que notas pueden utilizarse (y luego estimar las probabilidades restantes).

**2.3. Agregando restricciones.** El modelo anterior tiene varias falencias para simular melodías, siendo una muy importante que no permite controlar el comportamiento futuro de las trayectorias. Por ejemplo, modificando la distribución inicial es posible crear trayectorias con estado inicial $X_0$ fijo, pero no podemos fijar el estado $X_k$ para $k$ mayores (su comportamiento está dado por el vector $\mu^k$, distribución de probabilidad de orden $k$). Es decir, no podemos elegir en que nota terminarán las melodías que simulamos[5].

Para resolver este problema utilizaremos ciertas cadenas de Markov no homogéneas que nos permitirán imponer cierto tipo de *restricciones* sobre las trayectorias. Estas cadenas (que llamaremos cadenas de Markov con restricciones) y su uso en la generación de melodías son presentadas en detalle por Pachet, Roy y Barbieri en [4].

Comencemos enunciando una generalización de la propiedad 2.5 al caso de cadenas no necesariamente homogéneas:

**Proposición 2.8.** *Sean $\{X_k\}_{k\in\mathbb{N}}$ una cadena de Markov con espacio de estados $E$, matrices de transición $P^{(k)} = (p_{ij}^{(k)})$ y distribución inicial $\mu$, y $n, m$ naturales con $n > m$. Se cumple:*

[4]Es decir, secuencias finitas de variables, no necesariamente del mismo largo

[5]En caso de hacerlo, la trayectoria resultante podría tener probabilidad nula de ocurrir.

1. *Si llamamos* $P_{nm} = (p_{ij}(n,m))_{i,j\in E}$ *a la matriz cuyas entradas son las probabilidades de transición de* $i$ *a* $j$ *en* $n-m$ *pasos partiendo de* $i$ *en el instante* $m$ (i.e, $p_{ij}(n,m) = P(X_n = j|X_m = i)$), *entonces se verifica que* $P_{nm} = P^{(m+1)}P^{(m+2)}\dots P^{(n)}$.
2. *Se cumple que* $\mu^n = \mu \times P_{n0} = \mu P^{(1)}P^{(2)}\dots P^{(n)}, \forall n \in \mathbb{N}$, *siendo* $\mu^n$ *la distribución en el instante* $n$.
3. *Las probabilidades de la forma* $P(X_n = i_n, X_{n-1} = i_{n-1}, \dots, X_m = i_m)$ *pueden calcularse como*

$$P(X_n = i_n, X_{n-1} = i_{n-1}, \dots, X_m = i_m) = p^{(n)}_{i_{n-1}i_n} p^{(n-1)}_{i_{n-2}i_{n-1}} \dots p^{(m+1)}_{i_m i_{m+1}} \mu^{(m)}_{i_m}.$$

Necesitamos ahora definir que entenderemos por *restricciones*. Informalmente, supongamos que queremos obtener trayectorias finitas $x_0, x_1, \dots, x_N$ a partir de una cadena de Markov, de forma tal que podamos imponer algunas condiciones para el comportamiento de la cadena en los distintos instantes $0, 1, \dots, N$. Dichas condiciones podemos clasificarlas como:

- **Restricciones unitarias:** son condiciones que refieren a un único instante $k \in \{0, 1, \dots N\}$ (siendo $N \in \mathbb{Z}^+$ el largo de la trayectoria a considerar) y consisten en indicar cuáles son los estados que puede tomar la variable $X_k$.
- **Restricciones binarias:** son condiciones que refieren a dos instantes consecutivos $k$ y $k+1$ ($k \in \mathbb{N}$) e indican cuáles son los pares de estados posibles en $X_k$ y $X_{k+1}$.

Veremos que es posible considerar este tipo de restricciones sin perder la condición de Markov. Sin embargo no será posible considerar restricciones que involucren a más de 2 estados consecutivos ya que la "pérdida de memoria" de las cadenas de Markov lo impide[6].

Definamos más rigurosamente las restricciones.

**Definición 2.9.** Sea $\{X_k\}_{k\in\mathbb{N}}$ una cadena de Markov con espacio de estados $E$. Para todo $n \in \mathbb{N}$ definimos $U_n \subset E$ *restricciones (unitarias) sobre el instante* $n$ como el conjunto de los estados en los que la variable $X_n$ tiene probabilidad positiva, es decir,

$$U_n := \{i \in E : P(X_n = i) > 0\}.$$

Asimismo para $n \geq 1$ definimos $B_n \subset E \times E$ conjunto de *restricciones (binarias) sobre la transición de* $n-1$ *a* $n$ como los pares de estados en los que el vector $(X_{n-1}, X_n)$ tiene probabilidad positiva, es decir:

$$B_n = \{(i,j) \in E \times E : P(X_{n-1} = i, X_n = j) > 0\}.$$

Nos interesará, dada de una cadena de Markov homogénea, modificarla imponiendo algunas restricciones adicionales. De esto resultará en una cadena no homogénea, cuyas restricciones estarán contenidas en las de la original. Un ejemplo de aplicación de restricciones sobre el paseo al azar puede verse en [6].

Observemos que al imponer restricciones es de esperarse que se reduzca el espacio de trayectorias posibles (i.e. que tienen probabilidad positiva). Se definirán pues las nuevas probabilidades de transición de modo que las trayectorias prohibidas por las

[6]En general, para cadenas de orden $N$ pueden considerarse restricciones que involucren hasta $N$ estados consecutivos.

restricciones tengan probabilidad nula y las demás tendrán la probabilidad condicionada al nuevo conjunto de trayectorias posibles. Es decir, dada una trayectoria $s$ definiremos su nueva probabilidad $\tilde{P}(s)$ del siguiente modo:

$$\tilde{P}(s) = \begin{cases} 0 & \text{si } s \notin S', \\ P(s|s \in S') & \text{si } s \in S', \end{cases} \tag{1}$$

donde $S'$ es el conjunto de las trayectorias que satisfacen las restricciones y $P$ la probabilidad bajo la cadena homogénea. Más adelante determinaremos las probabilidades de transición bajo estas condiciones.

Notemos que es necesario asegurar cierta consistencia en las restricciones si queremos que el nuevo conjunto de trayectorias posibles sea no vacío. Para formalizar esto consideremos la siguiente definición.

**Definición 2.10.** Sean $\{X_k\}_{k\in\mathbb{N}}$ una cadena de Markov con espacio de estados $E$, $N \geq 2$ un número entero, $\{U_n\}_{n\in\{0,\dots,N\}}$ una familia de subconjuntos de $E$ y $\{B_n\}_{n\in\{1,\dots,N\}}$ una familia de subconjuntos de $E \times E$. Diremos que $\{U_n\}$ y $\{B_n\}$ son *consistentes por caminos* como restricciones de la cadena si para todo $n \in \{0,\dots,N-1\}$ se verifica que

$$\forall i \in U_n \ \exists j \in U_{n+1} \ / \ (i,j) \in B_{n+1}.$$

Además, diremos que una secuencia de estados $i_0, i_1, \dots, i_l$ con $l \leq N$ es *consistente* si puede realizarse verificando todas las restricciones, es decir si cumple:

- $i_k \in U_k \ \forall k \in \{0,\dots,l\}$,
- $(i_{k-1}, i_k) \in B_k \ \forall k \in \{1,\dots,l\}$.

En palabras, que las restricciones sean consistentes por caminos asegura que toda trayectoria que comience verificando las restricciones podrá terminar haciéndolo, como se muestra en la siguiente proposición.

**Proposición 2.11.** *Sean $E$ un espacio de estados, $N \geq 2$ entero y $\{U_n\}_{n\in 0,\dots,N}$ y $\{B_n\}_{n\in 1,\dots N}$ familias de restricciones unitarias y binarias respectivamente. Si dichas restricciones son consistentes por caminos para toda trayectoria parcial consistente $i_0, i_1, \dots, i_l$ con $l < N$ (es decir, que puede realizarse verificando las restricciones para los instantes 0 a $l$) se cumple:*

1. *Existe $i_{l+1} \in U_{l+1}$ tal que $i_0, i_1, \dots, i_l, i_{l+i}$ es consistente.*
2. *La trayectoria $i_0, i_1, \dots, i_l, i_{l+1}$ es consistente si y sólo si $i_l, i_{l+1}$ lo es.*

*Demostración.*

1. Como la trayectoria $i_0, i_1, ..., i_l$ es consistente sabemos que $i_l \in U_l$ y como las restricciones son consistentes por caminos, para $i_l \in U_l$ tenemos $i_{l+1} \in U_{l+1}$ tal que $(i_l, i_{l+1}) \in B_{l+1}$. Luego la trayectoria $i_0, \dots i_{l+1}$ resulta consistente.
2. Veamos que si $i_l, i_{l+1}$ es consistente, la secuencia entera lo es (la otra implicancia es inmediata). Para ello basta notar que como la secuencia $i_0, i_1, \dots, i_l$ es consistente, y $i_l, i_{l+1}$ también, se tiene que $i_k \in U_k \ \forall k \in \{0, 1, \dots l+1\}$ y $(i_{k-1}, i_k) \in B_k \ \forall k \in \{1, \dots l+1\}$ lo cual concluye la demostración.

□

Con esta proposición podemos dar un algoritmo para determinar las nuevas restricciones de modo tal que sean consistentes por caminos, **siempre y cuando en la cadena original exista alguna trayectoria con probabilidad positiva que verifique las condiciones que queremos imponer**. Para ello pondremos

el foco en la propagación de las restricciones unitarias, mientras que las restricciones binarias que de ello se desprendan quedarán de hecho impuestas en las matrices de transición.

- **Fijando un estado:** si se quiere imponer una condición de la forma $U_k = \{a\}$, hay que eliminar de $U_{k+1}$ todos los estados a los que no se puede acceder desde $a$, es decir, los $b \in E \ / P(X_{k+1} = b | X_k = a) = 0$. De modo similar, hay que quitar de $U_{k-1}$ los estados que no pueden ir hacia $a$, que son los $b \in E$ tales que $P(X_k = a | X_{k-1} = b) = 0$. Una vez removidos estos estados hay que seguir propagando las restricciones que generó la remoción de más estados, de acuerdo a lo siguiente.
- **Removiendo estados:** si se quiere quitar un estado $a$ del conjunto $U_k$, habra que quitar de $U_{k+1}$ todos los estados a los que sólo se puede acceder desde $a$, esto es, quitar los $b \in E$ tales que $P(U_{k+1} = b | U_k = c) = 0 \ \forall c \neq a$. De $U_{k-1}$ quitaremos los estados que sólo podían ir hacia $a$, es decir los $b \in E$ tales que $P(X_k = c | X_{k-1} = b) = 0 \ \forall c \neq a$.

El proceso termina cuando las restricciones son consistentes por caminos, esto es, cuando en los pasos anteriores no hay más nada por hacer. Nótese que si hay al menos una trayectoria con probabilidad no nula en la cadena original que satisface las restricciones, el procedimiento anterior preserva los estados y trayectorias involucrados y por lo tanto el espacio de restricciones resultante no tiene conjuntos vacíos.

Veamos ahora cómo construir las matrices de transición $\{\tilde{P}^{(n)}\}_{\{i \in 1,\dots,N\}}$ de una cadena con restricciones. Para ello tenemos inicialmente una cadena homogénea $\{X_k\}_{k \in \mathbb{N}}$ con distribución inicial $\mu$ y matriz de transición $P$, $N$ un entero positivo y $\{U_n\}_{n \in \{0,\dots,N\}}$, $\{B_n\}_{n \in \{1,\dots,N\}}$ conjuntos de restricciones unitarias y binarias **consistentes por caminos** que queremos imponer. El procedimiento general consiste en

- Construir una familia auxiliar de matrices $\{Z^{(n)}\}_{\{i \in 1,\dots,N\}}$ que indica que estados y transiciones están permitidos. Estas matrices generalmente no son estocásticas
- Renormalizar las matrices obtenidas de modo que sean estocásticas[7] y respeten la condición (1).

Construimos la familia $\{Z^{(n)}\}_{n \in \{0,\dots N\}}$ del siguiente modo:

- **Inicialización.** Definimos $Z^{(0)} = \mu$ y $Z^{(n)} = P \ \forall n \in \{1, \dots, N\}$.
- **Remoción de estados.** Llamemos $z_{ij}^{(n)}$ a la entrada $i, j$ de la matriz $Z^{(n)}$ Para cada $j \in E$ removido de $U_n$, se establece $z_{ij}^{(n)} = 0 \ \forall i \in E$ (es decir, se lleva la j-ésima columna de la matriz a 0).
- **Remoción de transiciones.** Las transiciones prohibidas imponen ceros en las matrices del siguiente modo: $\forall i, j \in E, \ n \in \{1, \dots, N\}$ tales que $(i, j) \notin B_n$ se establece $z_{ij}^{(n)} = 0$.

Llamemos $S$ al conjunto de trayectorias de largo $N$ posibles para la cadena homogénea original. Al imponer restricciones queda determinado un subconjunto $S' \subset S$ que contiene únicamente a las trayectorias que verifican las restricciones. Para cada trayectoria $i = i_0 i_1 \dots i_N \in S'$ queremos definir $\{\tilde{P}^{(n)}\}$ de modo que la nueva probabilidad sea $\tilde{P}(i) = P(i | i \in S')$ (siendo $\tilde{P}(i) = 0$ si $i \notin S'$). Cabe notar

[7] Abusando de la nomenclatura, admitiremos por estocásticas matrices que tengan filas de ceros.

que este modo de definir las probabilidades **no** es equivalente a renormalizar las matrices $Z^{(n)}$ por filas.

El siguiente resultado nos permite construir las nuevas matrices de transición.

**Teorema 2.12.** *Consideremos $\{X_k\}_{k\in\mathbb{N}}$ una cadena de Markov homogénea en $E$ finito con distribución inicial $\mu$ y matriz de transición $P$, $N \geq 2$ un entero, $\{U_n\}_{n\in\{0,\dots,N\}}$, $\{B_n\}_{n\in\{1,\dots,N\}}$ familias de restricciones –unitarias y binarias respectivamente– consistentes por caminos, y $\{Z^{(n)}\}_{n\in\{1,\dots,N\}}$ matrices construidas según el procedimiento anterior. Definiremos las matrices de transición $\{\tilde{P}^{(n)}\}$, $n \in \{1,\dots,N\}$, con $\tilde{P}^{(n)} = (\tilde{p}_{ij}^{(n)})_{i,j}$ de modo que,*

$$\tilde{p}_{ij}^{(N)} = \frac{z_{ij}^{(N)}}{\alpha_i^{(N)}}, \qquad \text{siendo} \quad \alpha_i^{(N)} = \sum_{k\in E} z_{ik}^{(N)},$$
$$\tilde{p}_{ij}^{(n)} = \frac{\alpha_j^{(n+1)} z_{ij}^{(n)}}{\alpha_i^{(n)}}, \quad \text{siendo} \quad \alpha_i^{(n)} = \sum_{k\in E} \alpha_k^{(n+1)} z_{ik}^{(n)} \ \forall n \in \{1,\dots,N-1\}.$$

*Y la distribución inicial $\tilde{\mu}$ de modo que*

$$\tilde{\mu}_i = \frac{\alpha_i^{(1)} z_i^{(0)}}{\alpha^{(0)}}, \quad \text{con} \quad \alpha^{(0)} = \sum_{k\in E} \alpha_k^{(1)} z_k^{(0)}.$$

*En caso de que $\alpha_i^{(n)} = 0$ las expresiones de los $p_{ij}^{(n)}$ son de la forma $\frac{0}{0}$ e impondremos $p_{ij}^{(n)} = 0$. Entonces la cadena (no homogénea) de trayectorias finitas, con matrices de transición $\{\tilde{P}^{(n)}\}_{n\in\{1,\dots,N\}}$ y distribución inicial $\tilde{\mu}$ definidas como antes verifica la condición* (1).

En palabras, el procedimiento para hallar las matrices de transición consiste en normalizar individualmente la matriz correspondiente a la última transición, para luego propagar la normalización hacia atrás de modo que cumpla la propiedad buscada, como se verá en la demostración.

*Demostración.* La demostración de que las matrices $\tilde{P}^{(n)}$ son estocásticas y que $\tilde{\mu}^{(n)}$ es distribución puede verse en [6]. Verificaremos directamente que se cumple la condición (1), para lo cual consideramos una trayectoria $s = s_0, s_1, \dots, s_N$ que verifica las restriciones (en caso contrario es inmediato que $\tilde{P}(s) = 0$). Tenemos entonces,

$$\begin{aligned}
\tilde{P}(s) &= \tilde{\mu}_{s_0} p_{s_0 s_1}^{(1)} p_{s_1 s_2}^{(2)} \dots p_{s_{N-1} s_N}^{(N)} \\
&= \frac{\cancel{\alpha_{s_0}^{(1)}} z_{s_0}^{(0)}}{\alpha^{(0)}} \frac{\cancel{\alpha_{s_1}^{(2)}} z_{s_0 s_1}^{(1)}}{\cancel{\alpha_{s_0}^{(1)}}} \cdots \frac{\cancel{\alpha_{s_{N-1}}^{(N)}} z_{s_{N-2} s_{N-1}}^{(N-1)}}{\cancel{\alpha_{s_{N-2}}^{(N-1)}}} \frac{z_{s_{N-1} s_N}^{(N)}}{\cancel{\alpha_{s_{N-1}}^{(N)}}} \\
&= \frac{1}{\alpha^{(0)}} z_{s_0}^{(0)} z_{s_0 s_1}^{(1)} \dots z_{s_{N-2} s_{N-1}}^{(N-1)} z_{s_{N-1} s_N}^{(N)},
\end{aligned}$$

donde el producto resultante $z_{s_0}^{(0)} z_{s_0 s_1}^{(1)} \dots z_{s_{N-2} s_{N-1}}^{(N-1)} z_{s_{N-1} s_N}^{(N)}$ es no nulo ya que $s$ verifica las restricciones. Más aún, por construcción, se tiene que si $z_{ij}^{(n)} \neq 0$,

$z_{ij}^{(n)} = p_{ij}$ y $z_i^{(0)} = \mu_i$ con lo cual,

$$\begin{aligned}\tilde{P}(s) &= \frac{1}{\alpha^{(0)}} z_{s_0}^{(0)} z_{s_0 s_1}^{(1)} \dots z_{s_{N-2} s_{N-1}}^{(N)} z_{s_{N-1} s_N}^{(N)} \\ &= \frac{1}{\alpha^{(0)}} \mu_{s_0} p_{s_0 s_1} \dots p_{s_{N-1} s_N} \\ &= \frac{1}{\alpha^{(0)}} P(s),\end{aligned}$$

siendo $P$ la probabilidad bajo la cadena homogénea. Es decir, toda trayectoria $s$ que cumpla las restricciones verifica $\tilde{P}(s) = cP(s)$ donde $c = \frac{1}{\alpha^{(0)}}$ es una constante que no depende de $s$. Luego si $S'$ es el conjunto de trayectorias que verifican las restricciones se tiene que $\tilde{P}(S') = 1$. Por lo tanto $P(S') = \alpha^{(0)}$ y

$$\tilde{P}(s) = \frac{P(s)}{P(S')} = P(s|s \in S') \ \forall s \in S'.$$

Lo cual verifica la condición (1) y concluye la demostración. □

Tenemos así un algoritmo que nos permite, dada una cadena de Markov homogénea y ciertas restricciones, determinar todos los parámetros de la cadena inducida por las mismas.

**2.4. Una aplicación a simulación de melodías.** En el siguiente ejemplo consideramos una melodía preexistente ("Arroz con leche") e intentamos generar otras que puedan percibirse como "variaciones" del tema original. Para ello preservamos

- La estructura rítmica, es decir, la cantidad y duraciones de las notas originales.
- Algunas notas –convenientemente elegidas– de la melodía.
- La estructura armónica, es decir, los acordes subyacentes.

Las notas preservadas oficiarán como restricciones unitarias de una cadena de Markov cuya matriz de transición "inicial" (es decir, previo a la implementación de las restricciones) es $P = (p_{ij})$ con

$$p_{ii+1} = p, \quad p_{ii} = q \quad , p_{ii-1} = r,$$

con $p + q + r = 1$. En los demás casos $p_{ij} = 0$. Los valores de $p$, $q$ y $r$ empleados se estimaron de la melodía original[8]. Se conservan las notas que aparecen en cada cambio de acorde, y se elige como diccionario a utilizar el conjunto de las notas del acorde.

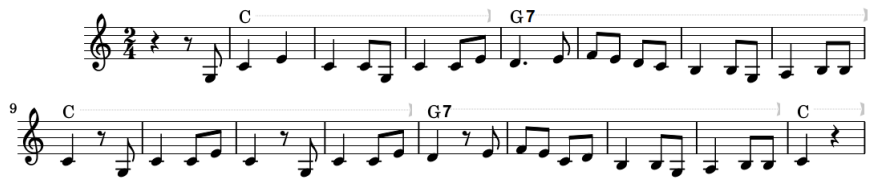



FIGURA 1. Partitura de la melodía de Arroz con leche, con sus acordes cifrados arriba.

[8]Notar que dicha estimación es anecdótica, en tanto se realizó sobre una conjunto muy pequeño de transiciones.

Observemos ahora la partitura de uno de los ejemplos simulados.

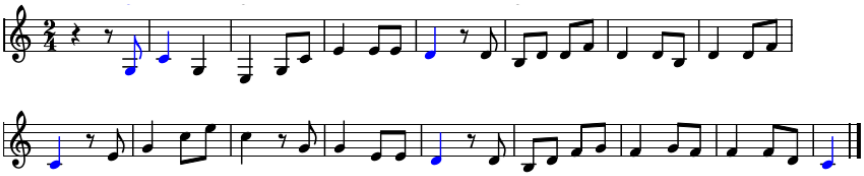



FIGURA 2. Una de las variaciones simuladas.

El audio, así como otros ejemplos y el código para generarlos, se encuentran en [7].

Nótese que la elección de las notas asociadas a cada estado se hizo de forma manual, al igual que la determinación de los acordes y la región asociada a éstos. Una vez determinado esto se pueden generar automáticamente tantas variaciones como se quiera, sin embargo no es una estrategia viable cuando las piezas con las que se trabaja son muy extensas. Asimismo, resultaría deseable utilizar notas ajenas al acorde sin perder la estructura armónica. Estas son algunas de las limitaciones que el ejemplo presenta.

Para resolver estas situaciones será necesario explorar otros modelos. En particular una breve revisión de modelos Markovianos orientados a la composición musical puede verse en [1].

## 3. Modelando duraciones

**3.1. Esquema basado en compases.** Abordaremos a continuación el problema de la simulación de duraciones de modo que tenga cierta estructura métrica. Para ello consideraremos un modelo simplificado en el que la música se organiza en *compases*. Entendemos por compases secciones de igual duración, subdivididas en cierta cantidad de *beats* o golpes. Identificaremos la duración de una nota con el golpe en que esta ocurre, es decir, toda nota ocurre al mismo tiempo que un golpe y sólo nos interesan los instantes de estas ocurrencias (ataques).

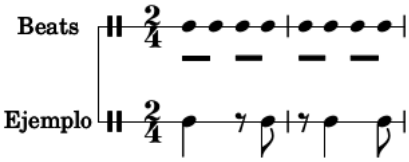



FIGURA 3. Ejemplo en 2/4 con golpes de corchea. Se observan dos compases con golpes en 1,4 y 2,4 respecivamente.

Así, la métrica de un compás (dividido en $n$ golpes) puede verse como un subconjunto de $\{1, \ldots, n\}$ que indica en que golpes ocurrió una nota. El objetivo es, entonces, construir secuencias de compases, eligiendo adecuadamente los instantes de ataque de las notas. Para ello Temperley [8] propone algunos modelos, de creciente complejidad:

- Modelo de posición uniforme: En cada golpe decide (con probabilidad $p$ constante) si hay o no un ataque. Las decisiones son independientes y en consecuencia todos los golpes tienen la misma probabilidad de contener un ataque.
- Modelo de orden cero: Se sortea la cantidad de golpes a ocurrir entre ataques. Cada posible valor tiene una probabilidad fija.
- Modelo de posición métrica: En cada golpe se decide si hay o no un ataque, pero la probabilidad de que lo haya depende de la *fuerza* del golpe implicado.
- Modelo de posición refinado: En cada golpe se decide si hay o no un ataque. La probabilidad de que lo haya depende de la posición del golpe en el compás.
- Modelo jerárquico: En cada golpe se decide si hay o no un ataque. La probabilidad de que lo haya depende del *nivel* del golpe y de si hay o no ataques en el golpe de nivel inmediato superior.
- Modelo de orden uno: Se sortea la posición de los ataques, condicionada a la posición del ataque anterior.

En todos los casos hay probabilidades a determinar, las cuales estimaremos por máxima verosimilitud a partir de un *corpus* dado. Tanto en [8] como en el ejemplo que presentaremos se considera parte del Essen Folksong Collection. Temperley describe los 6 modelos utilizando ejemplos en 4/4. Aquí consideraremos únicamente el modelo jerárquico aplicado a compases de 2/4.

Dos conceptos que fueron mencionados en el listado de modelos y nos serán de utilidad son el de *fuerza* y *nivel* de un golpe. En ellos, asumimos que los golpes dentro de cada compás no son equivalentes, sino que hay algunos más fuertes que otros. Se clasifican pues los golpes en *niveles*, correspondiendo niveles más altos a golpes fuertes y niveles más bajos a golpes débiles. Cabe notar que dicha clasificación depende del tipo de compás empleado y se conoce de antemano. Por ejemplo para el compás 2/4 la clasificación en niveles es

FIGURA 4. Niveles propuestos por Temperley para el compás de 2/4. La cantidad de puntos representa el nivel

En [5] puede verse una explicación más detallada de los modelos rítmicos, así como los niveles propuestos para otros compases.

**3.2. El modelo jerárquico.** Describiremos aquí con un poco más de detalle el modelo jerárquico propuesto por Temperley y lo implementaremos sobre el ejemplo de "Arroz con leche". Como se mencionó en la sección anterior, en este modelo decidiremos si en un golpe hay ataque o no dependiendo de lo que suceda con los golpes de nivel superior más cercanos. Por ejemplo, en los compases de 2/4:

- En el único golpe de nivel 3 (el primero) hay un ataque con probabilidad $p$ fija e independiente de otros compases. Por simplicidad asumimos $p = 1$.
- En el golpe de nivel 2 hay un ataque con probabilidad condicionada a lo que haya ocurrido en los golpes de nivel 3 adyacentes (esto incluye al primer golpe del siguiente compás). Dado que supusimos que siempre hay ataques

en el primer compás el modelo se simplifica y la probabilidad siempre será la misma.
- En los dos golpes de nivel 1 (posiciones 2 y 4) la probabilidad de ataque está condicionada a lo que ocurrió en los golpes 1 y 3 (para el golpe 2) y el 3 y 1 del compás siguiente (para el golpe 4).

Denominamos a estas probabilidades condicionales bi-anclada (si hay ataque en ambos vecinos de nivel superior), pre-anclada (si hay ataque únicamente en el anterior), post-anclada (si hay ataque únicamente en el posterior) o inanclada (si no hay ataque en ninguno de sus dos vecinos). Nótese que en el caso que estamos describiendo no hay probabilidades inancladas.

Estimamos de las obras en 2/4 del corpus Essen las probabilidades condicionales obteniendo

| | Nivel 1 | Nivel 2 | Nivel 3 |
|---|---|---|---|
| bi-ancladas | 0.573 | 0.783 | 1 |
| post-ancladas | 0.973 | – | – |
| pre-ancladas | 0.0805 | – | – |
| inancladas | – | – | – |

CUADRO 1. Probabilidades de ataque en cada golpe. No se incluyen las probabilidades que no se utilizarán en el ejemplo

Se utiliza el modelo para generar la métrica de 16 compases de 2/4, a partir de los cuales luego se simula una variación de "Arroz con leche", es decir, una melodía con las mismas restricciones y armonía que la original. A continuación se muestra uno de los resultados obtenidos.

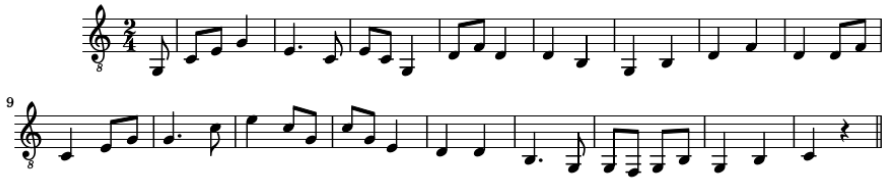



FIGURA 5. Partitura de una variación de "Arroz con leche" con melodía y ritmo aleatorios.

**3.3. Un tema pendiente: La comparación de modelos.** Se obtuvo un primer procedimiento de generación de melodías con alturas y duraciones aleatorias. Resulta bastante limitado en cuanto a la información que requiere (por ejemplo la estructura armónica y/o un diccionario adecuado de notas a usar) y algo conservador en su comportamiento. La exploración de otros modelos (como por ejemplo los propuestos en [8]) introduce una nueva interrogante ¿Qué criterio utilizar para comparar modelos? En ese sentido referimos a Mavromatis [3], quien propone un criterio de selección de modelos basado en el largo de descripción. Así, un buen modelo para un conjunto de datos (por ejemplo de obras musicales) es aquel que permite representarlas más comprimidas (considerando el costo de representar los datos **y** el modelo).

## Referencias

[1] Ames, Charles: *The Markov Process as a Compositional Model: a Survey and Tutorial*. Leonardo, 22(2):175–187, 1989.

[2] Häggström, Olle: *Finite Markov Chains and algorithmic applications*. Cambridge University Press, 2002.

[3] Mavromatis, Panayotis: *Minimum description length modelling of musical structure*. Journal of Mathematics and Music, 2009.

[4] Pachet, François, Pierre Roy y Gabriele Barbieri: *Finite-Length Markov Processes with Constraints*. Proceedings of the Twenty-Second International Joint Conference on Artificial Itelligence, páginas 638–642, 2011.

[5] Pearce, Marcus, Daniel Müllensiefen, David Lewis y Christophe Rhodes: *David Temperley, Music and Probability*. 2, Octubre 2007.

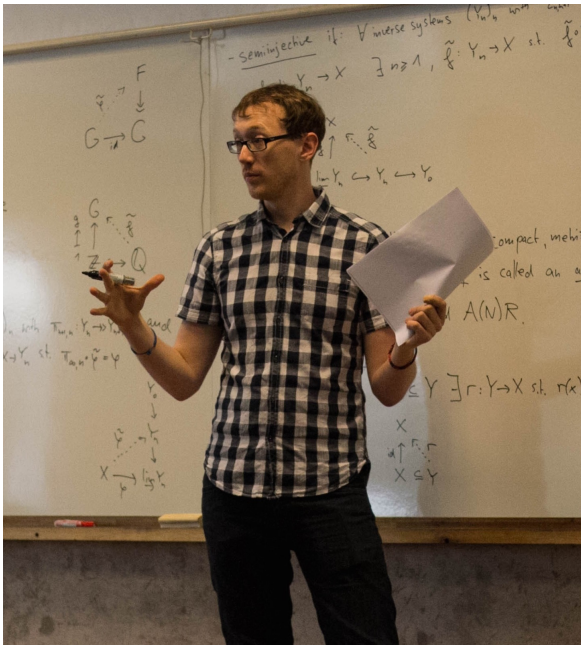
[6] Rumbo, Verónica: *Cadenas de Markov con restricciones y aplicación a la composición automática de música tonal*. `http://www.cmat.edu.uy/~vrumbo/monografia`. Trabajo monográfico.

[7] Rumbo, Verónica: *Repositorio online del trabajo monográfico*. `http://www.cmat.edu.uy/~vrumbo/repositorio`.

[8] Temperley, David: *Modeling Common-Practice Rhythm*. Music Perception, 27(5):355–376, 2010. Available at https://openmusiclibrary.org/article/37125/.

CMat, Facultad de Ciencias, Universidad de la República. Montevideo, Uruguay.
*Email address*: `vrumbo@cmat.edu.uy`

PONENCIA




# SEMIPROJECTIVITY AND SEMIINJECTIVITY IN DIFFERENT CATEGORIES

HANNES THIEL

PONENCIA

# RIGIDEZ Y GEOMETRICIDAD DE ACCIONES DE GRUPOS DE SUPERFICIES EN EL CÍRCULO

MAXIME WOLFF

Publicaciones Matemáticas del Uruguay
Volumen 17, Julio 2019, Páginas 277–279
S 0797-1443

# RIGIDEZ Y GEOMETRICIDAD DE ACCIONES DE GRUPOS DE SUPERFICIES EN EL CÍRCULO

MAXIME WOLFF

Resumen. Esto es el acta de mi charla en el sexto coloquio uruguayo de matemáticas. Consideramos representaciones desde un grupo de superficie cerrada a $\mathrm{Homeo}^+(S^1)$. Kathryn Mann ha probado que las representaciones geométricas (ie, que levantan a una representacion fiel y discreta en $\mathrm{PSL}(2,\mathbb{R})$) son rígidas (ie, todas sus deformaciones son semi-conjugadas). Junto con ella, probamos la recíproca: todas las representaciones rígidas son geométricas.

## 1. Semi-conjugación

El estudio de acciones de grupos sobre el círculo es un tema bastante clásico, empezando desde Poincaré, con la teoría del número de rotación, para resolver problemas de sistemas dinámicos. Más recientemente, la teoría de cohomologia acotada empezó praticamente con el trabajo de Ghys [4], donde caracterizó la dinámica rotacional de una acción en términos de una clase de cohomologia acotada (la clase acotada de Euler). La teoría de las variedades de dimensión tres, con el teorema de geometrización de Thurston, también utiliza acciones de grupos de superficies o de variedades de dimensión tres sobre el círculo, con la caracterización de los fibrados de Seifert a travéz de los grupos de convergencia [11, 3, 2], o el estudio de foliaciones tensas sobre una variedad de dimension tres (ver [1] por ejemplo). Se pueden encontrar muchas más motivaciones y observaciones en el libro [10].

Aquí nos interesa estudiar las *deformaciones* de estas acciones. Entonces, si $\Gamma$ es un grupo finitamente generado, consideramos el "espacio de representaciones",

$$R_\Gamma = \mathrm{Hom}(\Gamma, \mathrm{Homeo}^+(S^1)),$$

con la topología producto. Hay una manera bastante obvia de deformar una representación $\rho \in R_\Gamma$, por conjugación: si $g_t \in \mathrm{Homeo}^+(S^1)$ es un camino que empieza en la identidad, entonces $g_t^{-1}\rho g_t$ es una deformación, trivial en el sentido de que no cambia la dinámica de $\rho$. Una observación de Denjoy es que siempre es posible deformar acciones de otra manera: por ejemplo si $\rho(\Gamma)$ tiene una órbita infinita, digamos, $(x_n)_{n\geq 1}$, se puede construir un nuevo círculo insertando un intervalo de tamaño $\frac{t}{2^n}$ en lugar de $x_n$, y luego, extender $\rho$ a una acción $\rho_t$ sobre el nuevo círculo. Entre $t = 0$ y $t \neq 0$, estas acciónes tipicamente no son conjugadas (por ejemplo si $\rho$ es minimal), pero tienen la misma dinámica rotacional. Además, son "casi-conjugadas", en el sentido que $\rho_t$ tiene conjugados arbitrariamente cerca de $\rho$.

**Definición 1** (Ghys [4])**.** Decimos que dos acciones $\rho_1$ y $\rho_2$ son *semi-conjugadas* si existe un mapa creciente $h\colon \mathbb{R} \to \mathbb{R}$, tal que $h(x+1) = h(x) + 1$, y, para cada $\gamma \in \Gamma$, existen levantamientos $\widetilde{\rho_i(\gamma)}\colon \mathbb{R} \to \mathbb{R}$ tales que $h \circ \widetilde{\rho_1(\gamma)} = \widetilde{\rho_2(\gamma)} \circ h$.

©2019 PMU

Esta definición cumple que $\rho_1$ y $\rho_2$ son semi-conjugadas si y solamente si *las clausuras* de sus clases de conjugación se intersectan. Así que definimos el "espacio de caracteres",

$$X_\Gamma = R_\Gamma / \sim,$$

donde $\sim$ es la semi-conjugación.

Por el trabajo de Ghys y Matsumoto [4, 8], es fácil ver que $X_\Gamma$ es el cociente Hausdorff más grande del espacio $R_\Gamma / \mathrm{Homeo}^+(S^1)$, cociente de $R_\Gamma$ por conjugación. Esta observación permite generalizar la definición de $X_\Gamma$ a representaciones en otros grupos que $\mathrm{Homeo}^+(S^1)$, y también, de verla como una generalización de los cocientes de la teoría geométrica de los invariantes (GIT), en el caso de representaciones en grupos de Lie.

## 2. Rígidez y geometricidad

Decimos que una (clase de) representación $[\rho] \in X_\Gamma$ es *rígida*, simplemente si es un punto aislado de $X_\Gamma$. El primer ejemplo de representaciones rígidas fue dado por Matsumoto. Si $\Sigma_g$ es una superficie cerrada de género $g \geq 2$, entonces podemos elegir una estructura hiperbólica sobre $\Sigma_g$. El cubrimiento universal de $\Sigma_g$, con la métrica levantada, es isométrico al plano $\mathbb{H}^2$, así que la acción natural del grupo fundamental $\pi_1\Sigma_g$ sobre $\mathbb{H}^2$, por isometrias, es un morfismo $\pi_1\Sigma_g \to \mathrm{PSL}(2,\mathbb{R})$. Ahora la acción de $\mathrm{PSL}(2,\mathbb{R}) \simeq \mathrm{Isom}^+(\mathbb{H}^2)$ sobre el círculo a infinito nos da una representación, $\pi_1\Sigma_g \to \mathrm{Homeo}^+(S^1)$, que llamamos *Fuchsiana*. Diferentes estructuras dan representaciónes Fuchsianas conjugadas en $\mathrm{Homeo}^+(S^1)$, y Matsumoto probó que estas representaciones son rígidas [9]. (Probó un teorema más fuerte, diciendo que todas las representaciónes de clase de Euler maximal son semi-conjugadas a una representación Fuchsiana.)

Mann generalizó este teorema a los levantados de representaciones Fuchsianas. Si $p\colon S^1 \to S^1$ es un cubrimiento de grado $k$, entonces, se puede verificar que una representación Fuchsiana $\rho_F$ se puede levantar a una representación $\rho_F^k$, que cumple $\rho_F(\gamma) \circ p = p \circ \rho_F^k(\gamma)$ para cada $\gamma \in \Gamma$, si y solamente si $k$ divide $2g-2$. Estas nuevas representaciónes son geométricas, en el sentido siguiente.

**Definición 2** (Mann,[6])**.** Si $G < \mathrm{Homeo}(M)$ es un grupo actuando sobre una variedad $M$ transitivamente, y si $\Gamma$ es un grupo discreto, decimos que una representación $\rho\colon \Gamma \to G$ es *geométrica* si es fiel y si su imágen es un lattice de un grupo de Lie $H < G$ que actúa transitivamente sobre $M$.

Aquí, diremos también que una representación es geométrica si es semi-conjugada a una representación geométrica, porque estamos más interesados en clases de semi-conjugación de representaciónes.

Ahora, el teorema de Mann evocado en el abstract es el siguiente.

**Teorema 3** (Mann,[5])**.** *Sea $\Gamma$ un grupo finitamente generado, y sea $\rho\colon \Gamma \to \mathrm{Homeo}^+(S^1)$ una representación geométrica. Entonces $\rho$ es rígida.*

Se puede verificar fácilmente que si $\rho\colon \Gamma \to \mathrm{Homeo}^+(S^1)$ es geométrica entonces $\Gamma$ puede ser solamente $\mathbb{Z}$ o un grupo de superficie cerrada. Así que el teorema siguiente, que logramos junto con ella, es un recíproco.

**Teorema 4** ([7])**.** *Sea $\rho\colon \pi_1\Sigma_g \to \mathrm{Homeo}^+(S^1)$ una representación rígida. Entonces es geométrica.*

Entonces, ahora tenemos la lista completa de los puntos aislados del espacio $X_{\pi_1\Sigma_g}$. Pero muchas preguntas se quedan abiertas. Por ejemplo, no se sabe si tiene un número finito de componentes conexas.

## Referencias


[1] Danny Calegari and Nathan M. Dunfield. Laminations and groups of homeomorphisms of the circle. *Invent. Math.*, 152(1):149–204, 2003.

[2] Andrew Casson and Douglas Jungreis. Convergence groups and Seifert fibered 3-manifolds. *Invent. Math.*, 118(3):441–456, 1994.

[3] David Gabai. Convergence groups are Fuchsian groups. *Ann. of Math.*, 136(3):447–510, 1992.

[4] Étienne Ghys. Groupes d'homéomorphismes du cercle et cohomologie bornée. In *The Lefschetz centennial conference, Part III (Mexico City, 1984)*, volume **58** of *Contemp. Math.*, pages 81–106. Amer. Math. Soc., 1987.

[5] Kathryn Mann. Spaces of surface group representations. *Invent. Math.*, 201(2):669–710, 2015.

[6] Kathryn Mann. Rigidity and flexibility of group actions on $S^1$. In *Handbook of group actions*. 2017.

[7] Kathryn Mann and Maxime Wolff. Rigidity and geometricity for surface group actions on the circle. arXiv:1710.04902.

[8] Shigenori Matsumoto. Numerical invariants for semiconjugacy of homeomorphisms of the circle. *Proc. AMS*, 98:163–168, 1986.

[9] Shigenori Matsumoto. Some remarks on foliated $S^1$ bundles. *Invent. Math.*, **90**:343–358, 1987.

[10] Andres Navas. *Groups of Circle Diffeomorphisms*. Chicago Lectures in Mathematics. University of Chicago Press, 2011.

[11] Pekka Tukia. Homeomorphic conjugates of Fuchsian groups. *J. Reine Angew. Math.*, 391:1–54, 1988.



Sorbonne Universités, UPMC Univ. Paris 06, Institut de Mathématiques de Jussieu-Paris Rive Gauche, UMR 7586, CNRS, Univ. Paris Diderot, Sorbonne Paris Cité, 75005 Paris, France

*Email address*: maxime.wolff@imj-prg.fr

# DIPLOMADOS, PONENCIAS

PONENCIA

# LEY DE BENFORD

MARÍA CAPUTI ZUNINI

Publicaciones Matemáticas del Uruguay
Volumen 17, Julio 2019, Páginas 285–290
S 0797-1443

# LEY DE BENFORD

MARÍA CAPUTI ZUNINI

## 1. Resumen

La Ley de Benford establece, contrariamente a la intuición, que, en algunos conjuntos de datos numéricos, la frecuencia de aparición del primer dígito significativo no es uniforme. La frecuencia con la que aparece cada dígito sigue una proporción particular que se explicita en la denominada por Benford [1] en 1938 como "Ley de los números anómalos". El 1 aparece como primer dígito significativo un 30,1 % de las veces, el 2 un 17,6 %, el 3 un 12,5 %, el 4 un 9,7 %, el 5 un 7,9 %, el 6 un 6,7 %, el 7 un 5,8 %, el 8 un 5,1 % y el 9 un 4,6 %, aproximadamente. Esta ley permaneció como una mera "curiosidad estadística" por varias décadas. En 1992 fue catapultada a la luz e interés público por el contador norteamericano Mark Nigrini, quien en su tesis de doctorado la utilizó para detectar fraudes en las declaraciones fiscales.

En este trabajo se presenta muy sintéticamente la Ley y se resumen algunos métodos estadísticos desarrollados en mi trabajo final del Diploma en Matemática mención Aplicaciones (ANEP-UdelaR)[2] para analizar si un conjunto de datos sigue o no la Ley de Benford. Se analiza el cumplimiento de la Ley de Benford en los resultados del censo realizado en Uruguay en 2011.

## 2. Palabras clave

Ley de Benford; Primer dígito significativo.

## 3. Antecedentes históricos

La primera publicación sobre la que posteriormente sería llamada "Ley de Benford" fue autoría del astrónomo norteamericano Simon Newcomb y data de 1881. En un artículo de dos páginas titulado "Note on the Frequency of Use of the Different Digits in Natural Numbers" publicado en American Journal of Mathematics el autor describe sus observaciones en el uso de las tablas de logaritmos. Newcomb detecta que las hojas más desgastadas eran las primeras. El artículo de Newcomb pasa desapercibido por la comunidad matemática de la época, quedando como una mera curiosidad.

Frank Benford en 1938, aparentemente sin conocer el artículo de Newcomb, realiza la misma observación en las tablas de logaritmos. Éstas eran muy utilizadas para realizar rápidamente multiplicaciones y divisiones antes de la era de las calculadoras. Benford tenía contacto frecuente con dichas tablas, ya que fue un físico que trabajó en el centro de investigaciones de General Electric en Nueva York. Para comprobar empíricamente la llamada por Benford como "Ley de los números anómalos" recolectó 20.229 números extraídos de las más diversas fuentes (entre otros: periódicos, revistas, estadísticas de béisbol, longitudes de ríos, tablas de constantes matemáticas y físicas)

©2019 PMU

y sin ayuda de calculadoras o procesadores de datos realizó los cálculos de las proporciones en las que aparecían los primeros 9 números enteros positivos. Esos datos los recogió en una tabla que fue retomada en Nigrini ([6], pág. 4) quien afirma que la misma contiene algunos errores, entre otros de redondeo. La publicación de Benford en Proceedings of the American Philosophical Society tuvo una repercusión mucho mayor que la de Newcomb, según Hill ([4], pág. 359), debido a estar previo al artículo de H. A. Bethe, M. E. Rose y L. P. Smith, titulado "The Multiple Scattering of Electrons", que luego sería muy renombrado. Habiendo pasado desapercibida la publicación de Newcomb, la bautizada por Benford como "Ley de los números anómalos" sería conocida como "Ley de Benford".

Mark Nigrini, un contador estadounidense, se encuentra por primera vez con la Ley de Benford en un curso de doctorado en 1989 en la Universidad de Cincinnati (EEUU). Según Nigrini [6] interesado por el tópico y tras leer muchos de los artículos que había publicados en esa materia decide contactarse con Ralph A Raimi, quien era el autor más citado en la temática del momento. Raimi consideraba que la ley no tenía fines prácticos. En ese momento, Nigrini [6] si bien vislumbraba que podía usarse la ley para la detección de anomalías en las declaraciones de impuestos, descarta la aplicación práctica por no tener los medios informáticos necesarios para poder procesar "gran cantidad" de datos. Sin embargo, a principios de 1991, Nigrini con ayuda de John Byant decide buscar un modelo matemático de cómo los contribuyentes engañan al fisco y de cómo la Ley de Benford puede ayudar a detectarlos. Esta idea sería la semilla de la tesis de doctorado de Nigrini publicada en 1992 y de muchas otras publicaciones del autor sobre el tópico. La tesis de Nigrini es la que lanza a la fama a la Ley de Benford. A partir de allí según Joseph T. Wells (autor del prólogo de Nigrini [6]) se elaboraron otras aplicaciones de la Ley de Benford.

Desde la publicación de Benford fueron numerosos los intentos por dar una prueba matemática rigurosa de la Ley. Pero, ésta recién fue conseguida por Hill en 1995.

## 4. Ley de Benford o ley del primer dígito significativo

**4.1. La Ley de Benford como distribución de probabilidad discreta. Una primera aproximación.** La Ley de Benford establece una distribución para los primeros dígitos significativos de un número real positivo. Siendo el primer dígito significativo de un real positivo el dígito distinto de cero que aparece más a la izquierda en su expresión decimal. Por ejemplo, el primer dígito significativo de 73,15 es 7, el de 0,045 es 4 y el de e es 2. Se observa que 0 no es considerado como primer dígito significativo. Al primer dígito significativo de un número real lo notaremos $D_1$.

Una primera descripción de la Ley de Benford como una ley de probabilidad podría tabularse como sigue.

| Dígito $d$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| $\mathrm{P}(D_1 = d)$ | 0,301 | 0,176 | 0,125 | 0,097 | 0,079 | 0,067 | 0,058 | 0,051 | 0,046 |

Con más formalidad podemos enunciar la Ley: $P(D_1 = d) = \log_{10}\left(1 + \frac{1}{d}\right)$ con $d = 1, \ldots, 9$.

Así definida la Ley es una distribución de probabilidad discreta. La Ley verifica las siguientes proposiciones: $P(D_1 = 1) = P(D_1 = 2) + P(D_1 = 3)$ y $P(D_1 = 1) = \sum_{i=5}^{9} P(D_1 = i)$.

## 5. Ley general de probabilidad para los primeros k dígitos significativos de un número en base 10

Hill [3] propone una Ley general de probabilidad para los primeros $k$ dígitos significativos de un número. Análogamente a como se definió el primer dígito significativo de un real positivo pueden definirse las funciones $D_1{:}R^+ \longrightarrow \{1,2,3,4,5,6,7,8,9\}$ y $D_i{:}R^+ \longrightarrow \{0,1,2,3,4,5,6,7,8,9\}$ para $i \geq 2$ tal que $D_i$ hace corresponder a un real positivo su $i$-ésimo dígito significativo en base 10.

**Definición 5.1.** Ley general para los primeros $k$ dígitos significativos de un número real positivo en base 10. Dado $k \in Z^+$, para todo $d_1$ con $d_1 \in \{1,2,3,4,5,6,7,8,9\}$, y para todos $d_j$ con $d_j \in \{0,1,2,3,4,5,6,7,8,9\}$, $j = 2,\ldots,k$ se define:

$$P\left(\bigcap_{i=1}^{i=k}\{D_i = d_i\}\right) = \log_{10}\left(1 + \frac{1}{\sum_{i=1}^{k} d_i 10^{k-i}}\right).$$

Se observa que esta Ley incluye como caso particular a la Ley de Benford para el primer dígito significativo. Se puede probar, a partir de esta Ley, que dado un natural $n$, $n \geq 2$ y siendo $D_n$ el $n$-ésimo dígito significativo de un real positivo, $P(\{D_n = d\}) = \sum_{i=10^{n-2}}^{10^{n-1}-1} \log_{10}\left(1 + \frac{1}{10i+d}\right)$ con $d = 0,\ldots,9$.

Otro corolario que se puede deducir de la Ley general de probabilidad para los primeros $k$ dígitos significativos es que existe una dependencia entre los dígitos significativos de un real positivo. Por ejemplo, la probabilidad de que el segundo dígito significativo sea 2 ($P(D_2 = 2) \cong 0,109$), es diferente a la probabilidad de que el segundo dígito significativo sea 2 sabiendo que el primer dígito significativo es 1 ($P(D_1 = 1, D_2 = 2/D_1 = 1) \cong 0,115$), y es a su vez diferente a la probabilidad de que el segundo dígito significativo sea 2 sabiendo que el primer dígito significativo es 2 ($P(D_1 = 2, D_2 = 2/D_1 = 2) \cong 0,110$). Esta dependencia se hace más débil a medida que crece la distancia entre los dígitos. Es decir, puede demostrarse, por ejemplo, que $(a_n) : a_n = P(\{D_1 = 1, D_n = 1\})$ es decreciente para $n \to +\infty$.

Otras implicancias son que $P(\{D_1 = i, D_n = j\})$ se aproxima a $P(\{D_n = j\})$, para $n \to +\infty$ , con $i \in \{1,2,3,4,5,6,7,8,9\}$, y $j \in \{0,1,2,3,4,5,6,7,8,9\}$. Además, se verifica que la distribución del $n$-ésimo dígito tiende a la distribución uniforme a medida que $n$ crece. Se ofrecerá al lector una prueba de la última afirmación (para el caso $j = 0$) que fuera enunciada por Hill en varias oportunidades pero cuya demostración no pudo ser encontrada en la bibliografía. Nigrini ([6], pág.6) plantea que en conjuntos de datos con tres o más cifras significativas, a partir del cuarto dígito significativo el error con la distribución uniforme es despreciable a fines prácticos.

**Teorema 5.2.** *Sea $D_n$ el n-ésimo dígito significativo de un real positivo, donde $D_n$ sigue la Ley general. Entonces la distribución del enésimo dígito tiende a la distribución uniforme para $n \to +\infty$. En particular $P(\{D_n = 0\}) \to \frac{1}{10}$, $n \to +\infty$.*

*Demostración.* $P(\{D_n = 0\}) = \sum_{i=10^{n-2}}^{10^{n-1}-1} \log_{10}\left(1 + \frac{1}{10i}\right) =$

$$= \frac{1}{\log_e 10} \sum_{i=10^{n-2}}^{10^{n-1}-1} \log_e\left(1 + \frac{1}{10i}\right).$$

No es difícil ver que $\sum_{i=10^{n-2}}^{10^{n-1}-1} \log_e\left(1 + \frac{1}{10i}\right) \sim \sum_{i=10^{n-2}}^{10^{n-1}-1} \frac{1}{10i}$ para $n \to +\infty$.

Sea $\mu = \dfrac{1}{\log_e 10}$. Entonces, $P(\{D_n = 0\}) \sim \mu \sum_{i=10^{n-2}}^{10^{n-1}-1} \dfrac{1}{10i} =$
$= \dfrac{\mu}{10} \sum_{i=10^{n-2}}^{10^{n-1}-1} \dfrac{1}{i}$ para $n \to +\infty$.
Como $\log_e m \sim \sum_{i=1}^{m} \dfrac{1}{i}$, para $m \to +\infty$.
Entonces, $P(\{D_n = 0\}) \sim \dfrac{\mu}{10} \left( \sum_{i=1}^{10^{n-1}-1} \dfrac{1}{i} - \sum_{i=1}^{10^{n-2}-1} \dfrac{1}{i} \right) \sim$
$\sim \dfrac{\mu}{10} \left[ \log_e(10^{n-1} - 1) - \log_e(10^{n-2} - 1) \right] = \dfrac{\mu}{10} \left[ \log_e \dfrac{10^{n-1} - 1}{10^{n-2} - 1} \right] \sim$
$\sim \dfrac{\mu}{10} \left[ \log_e \dfrac{10^{n-1}}{10^{n-2}} \right] = \dfrac{\mu}{10} \log_e 10 = \dfrac{1}{10}.$ □

## 6. Aplicaciones de la Ley de Benford

El uso de la Ley de Benford para detectar fraude o falsificación de datos en las declaraciones de impuestos fue abordado por el contador norteamericano Mark Nigrini en su tesis de doctorado. Nigrini ha recogido una extensa cantidad de evidencia empírica de la ocurrencia de la Ley de Benford en muchas áreas de la contabilidad y de la demografía y concluyó que en una amplia y variada cantidad de situaciones contables los datos reales siguen la Ley de Benford. Sin embargo, al falsear datos estos deben inventarse, lo que puede hacerse tomando números uniformemente generados o mediante una elección personal del falseador. Nigrini ha diseñado varios tests para medir la conformidad de un conjunto de datos con la Ley de Benford.

**6.1. Ley de Benford para los primeros dos dígitos.** Nigrini ([6], pág.15-19) bajo el título "Love at first sight" explica porqué usar solo la Ley de Benford para el primer dígito a veces no es suficiente al analizar si un conjunto de datos de dos o más dígitos sigue o no la Ley. Allí, construye un contraejemplo en el que podría concluirse que el conjunto de datos sigue la Ley de Benford para el primer dígito, pero en cambio se evidencia la no conformidad al considerar el análisis con los primeros dos dígitos.

Sin embargo, Nigrini ([6], pág. 20) sugiere no comparar una distribución con la Ley de Benford para los primeros dos dígitos en conjuntos de menos de trescientos datos. Para esos casos sugiere utilizar la Ley de Benford para el primer dígito.

**6.2. ¿Cómo comprobar que un conjunto de datos sigue la Ley de Benford?**

*6.2.1. Chi Cuadrado.* El test de bondad de ajuste de Chi Cuadrado $\chi^2$ permite comprobar si ciertos datos siguen una cierta distribución de probabilidad con un cierto error $\alpha$ .

El estadístico Chi Cuadrado depende de la cantidad de datos. Nigrini ([6], pág 154) afirma que el test de Chi Cuadrado sufre de un "exceso de poder" cuando la cantidad de datos es muy grande. Es decir, que casi siempre será mayor que su valor crítico, lo que implicaría que los datos no conforman la Ley de Benford aunque las diferencias sean mínimas. Afirma que se nota este problema cuando los conjuntos tienen más de 5000 datos y que con más de 25000 datos se requeriría casi la perfección para que pasaran dicho test.

*6.2.2. Mean Absolute Deviation Test (MAD Test).* Nigrini ([6], pág. 158) afirma que para desentenderse del problema de la cantidad de datos (que en las auditorías de fraude fiscal suelen ser muchos) se debe utilizar un estadístico que no dependa de ese parámetro. Propone utilizar el Mean Absolute Deviation Test (MAD Test), definiendo el estadístico $MAD = \frac{\sum_{j=1}^{K} |o_j - p_j|}{K}$, siendo $K$ la cantidad de categorías que toma la variable, $o_j$ la frecuencia relativa observada y $p_j$ la probabilidad esperada para la distribución de probabilidad. Cuanto mayor es MAD, mayor es la diferencia promedio entre la frecuencia relativa observada y la probabilidad. En el caso de comparar una distribución de datos con la Ley de Benford para el primer dígito, $K$ tomará el valor 9, con la Ley de Benford para el primer y segundo dígito $K$ será 90.

A diferencia del Test de Chi Cuadrado no hay una grilla con valores críticos objetivos según un cierto error u otros parámetros. En Nigrini ([6], pág. 160) aparece una tabla con ciertos valores críticos para decidir la conformidad o no conformidad de un conjunto de datos con la Ley de Benford. La misma fue desarrollada en base a las experiencias empíricas de Nigrini en el trabajo diario con este tópico en diferentes áreas.

Dicha tabla proporciona valores críticos para la comparación con la Ley de Benford para el primer dígito significativo, para el segundo dígito significativo, y para los primeros dos dígitos significativos. En algunas ocasiones las conclusiones pueden diferir según el Test MAD utilizado sea el del primer dígito, el segundo o el de los primeros dos dígitos. Allí, será necesario entonces, buscar explicaciones de acuerdo a las características particulares del conjunto de datos. Con respecto a esto, Nigrini sugiere una posición más bien pesimista en casos de resultados encontrados.

Nigrini ([6], pág.170) concluye que el test de Chi Cuadrado funcionará bien en conjuntos con una cantidad pequeña de datos. Para otros conjuntos con mayor cantidad de datos sugiere el uso del Test MAD siguiendo los parámetros dados por los valores críticos que aparecen en la tabla citada anteriormente.

*6.2.3. Últimos dos dígitos.* Como se demostró en el teorema 5.2 la frecuencia de los dígitos tiende a uniformizarse a medida que avanzamos hacia la derecha en la posición del dígito. Tomando esto como base, Nigrini ([6], pág.129) afirma que para propósitos prácticos a partir del tercer dígito significativo la probabilidad de los dígitos es uniforme. Utilizando esto crea el Test de los últimos dos dígitos que servirá para detectar invención de números en determinadas situaciones. El Test de los últimos dos dígitos consiste en comparar la distribución empírica con la distribución uniforme utilizando el Test de bondad de ajuste de Chi Cuadrado.

¿Cuáles son los últimos dos dígitos de un número? En general, esta pregunta carece de sentido, pero lo que importa en el análisis forense de resultados es qué dígitos serían los apropiados para este análisis. Según Nigrini ([6], pág.129) para cuestiones que involucren dinero los centavos serían los apropiados candidatos a últimos dos dígitos, y que, para cuestiones en donde puede darse invención o fraude con números naturales las decenas y las unidades serán los apropiados.

El Test de los dos últimos dígitos es utilizado en datos donde se buscan signos de invención de datos. Por ejemplo, censos poblacionales, resultados electorales, inventarios, números en las deducciones de impuestos.

## 7. Análisis del cumplimiento de la Ley de Benford: censo de la población uruguaya en 2011

En esta sección se propone estudiar la distribución de la población total de las localidades uruguayas con más de 1.000 habitantes según el censo realizado por el Instituto Nacional de Estadística de Uruguay [5] en 2011. Dicho censo tuvo la particularidad que, a diferencia de los anteriores, no fue realizado todo un mismo día si no que llevó varios meses en concluirse. Cabe preguntarse si los resultados del mismo serán válidos. La Ley de Benford provee una estrategia para analizar los datos recogidos.

A continuación, aparece una tabla en donde se registran las frecuencias del primer dígito significativo del total de población de las 176 diferentes localidades uruguayas de más de 1.000 habitantes.

| Dígito | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| Frecuencia | 70 | 37 | 21 | 9 | 7 | 10 | 9 | 6 | 7 |

Al aplicar el test de Bondad de ajuste de Chi-cuadrado, con 8 grados de libertad y $\alpha = 0,05$ , se obtiene $\chi^2 = 15,48 < 15,51$. Por lo tanto, es aceptada la hipótesis de que este conjunto de datos satisface la Ley de Benford para el primer dígito significativo.

Al aplicar el Test MAD para el primer dígito significativo obtenemos el valor 0,029 lo que describe una no conformidad de la cantidad de habitantes de las localidades uruguayas con más de mil habitantes con la Ley de Benford para el primer dígito significativo. Esto podría explicarse debido a que el conjunto de datos es pequeño o tal vez se deba a otras razones como invención de datos.

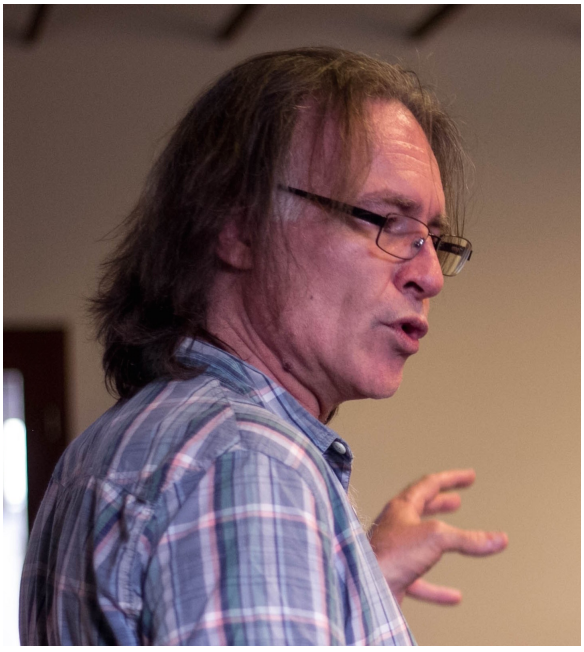
Podemos utilizar el Test de los últimos dos dígitos para analizar si hay signos de invención en los datos. Consideraremos los últimos dos dígitos como las decenas y unidades de los datos. Al aplicar el test de Bondad de ajuste de Chi-cuadrado, con 99 grados de libertad y $\alpha = 0,05$, se obtiene $\chi^2 = 98,12 < 124,34$ . Por lo tanto, es aceptada la hipótesis de que la distribución de los últimos dos dígitos satisface la distribución uniforme.

## Referencias

[1] Benford, F. (1938) The law of anomalous numbers, Proc. Amer. Phil. Soc. 78, 551-572.
[2] Caputi, M (2016) Ley de Benford (Tesis de Diploma). ANEP-UdelaR, Montevideo, Uruguay.
[3] Hill, TP (1995). A Statistical Derivation of the Significant-Digit Law. Statistical Science 10(4), 354-363.
[4] Hill, TP (1998). The First-Digit Phenomenon. American Scientist 86 (4), 358-363.
[5] Instituto Nacional de Estadística. (s.f.). Instituto Nacional de Estadística Uruguay. Recuperado el 14 de Septiembre de 2015, de `http://www.ine.gub.uy/`
[6] Nigrini, M. (2012). Benford's Law Applications for Forensic Accounting, Auditing, and Fraud Detection. Wiley.

Instituto de Formación Docente de Pando, ANEP. Canelones, Uruguay
*Email address*: `mariolinacaputi@gmail.com`

PONENCIA



# CRIPTOGRAFÍA SOBRE CURVAS ELÍPTICAS

HORACIO CASTAGNA

Publicaciones Matemáticas del Uruguay
Volumen 17, Julio 2019, Páginas 293–297
S 0797-1443

# CRIPTOGRAFÍA SOBRE CURVAS ELÍPTICAS

HORACIO CASTAGNA

## 1. Introducción

Aquí presentamos las ideas centrales desarrolladas en el trabajo final presentado para el Diploma en Matemática, cuyo título es Criptografía sobre Curvas Elípticas, tutoreado por Dr. José Vieitez.

Los temas vinculados a la seguridad en el ciberespacio y las comunicaciones son de preocupación mundial. La idea central del trabajo es presentar de manera sencilla las ideas básicas que están detrás de los métodos involucrados en la protección de las comunicaciones y los datos. Para ello trataremos los aspectos teóricos referentes a las curvas elípticas, y cómo éstas pueden usarse para implementar algoritmos de codificado, cifrado y firma digital de documentos. Luego, para ilustrar y comprender el funcionamiento de ciertos algoritmos, se construye un ejemplo de "juguete" con la implementación de los mismos en un lenguaje de alto nivel.

## 2. Curva Elíptica y estructura de Grupo Abeliano.

Dado un cuerpo $K$ definimos curva elíptica al conjunto de puntos $(x, y) \in K^2$ que verifican la ecuación,

$$y^2 + a_{11}xy + a_{01}y = x^3 + a_{20}x^2 + a_{10}x + a_{00}$$

Si la característica del cuerpo es distinta de 2 y de 3 podemos llevar la ecuación anterior a la forma,

$$y^2 = x^3 + ax + b$$

De ahora en adelante trabajaremos con esta forma denominada Ecuación de Weierstrass.

Para poder definir una operación "suma de puntos" sobre la curva elíptica, necesitamos que la curva no sea singular. Esto se logra si su discriminante no es nulo, o sea,

$$\Delta = 4a^3 + 27b^2 \neq 0$$

Tomemos el punto $O$ ("punto en el infinito" según la dirección del eje $Oy$) y el conjunto,

$$E(K) = \{(x, y) \in K^2 : y^2 = x^3 + ax + b\} \cup \{O\}$$

Definamos ahora $+ : E(K) \times E(K) \to E(K)$ de la siguiente manera,

1. Si $P, Q \in E(K)$ son distintos de $O$ y con abscisas distintas, tomamos la recta $PQ$ y la cortamos con $E(K)$, y al punto distinto de $P$ y de $Q$ así obtenido lo simetrizamos respecto de $Ox$ y obtenemos $P + Q$.
2. Si $P, Q \in E(K)$ son distintos de $O$ y con abscisas iguales, o $P = Q$ con ordenadas 0, definimos $P + Q = O$.

©2019 PMU

3. Si $P, Q \in E(K)$ con $P = Q$ y ordenadas distintas de 0, tomamos la recta tangente e $E(K)$ por $P$, la cortamos con $E(K)$, al punto distinto de $P$ así obtenido lo simetrizamos respecto de $Ox$ y obtenemos $P + Q$.
4. $P + O = P$, $\forall P \in E(K)$.

A partir de lo definido obtenemos una estructura de grupo abeliano $(E(K), +)$, considerando el conjunto $E(K)$ y la operación $+$.

## 3. Criptografía.

Desde tiempos remotos ha existido información sensible que se desea compartir entre dos personas, a la cual no tengan acceso los demás. Así nace la Criptología (del griego krypto: "oculto" y logos: "palabra"), que es la disciplina científica que se dedica al estudio de la escritura secreta.

La Criptografía, que podemos verla como parte de la Criptología, se ocupa del estudio de los algoritmos, protocolos y sistemas que se utilizan para proteger la información, y dar seguridad en las comunicaciones. Actualmente se realizan infinidad de comunicaciones y transacciones que se desean preservar de ojos y oídos ajenos (cliente-banco, compras por internet, correos electrónicos, chats, redes sociales, comercio, etc.).

Distinguimos dos grandes tipos de enfoque,

1. Criptografía de clave privada o simétrica. Sistema de cifrar que utiliza una clave solamente conocida por aquellos usuarios autorizados, que deben ponerse de acuerdo en la clave previamente a comunicarse.
2. Criptografía de clave pública o asimétrica. Sistema de cifrar que utiliza una clave pública para un usuario, de modo que solamente él pueda descifrar el mensaje.

Para el trabajo, nos enfocaremos en los algoritmos de clave pública. Los mismos intentan resolver los siguientes problemas,

1. Confidencialidad. Un documento debe poder leerlo sólo el receptor al que le fue enviado y nadie más.
2. Autenticidad. El documento proviene, verdaderamente, de quien dice provenir.
3. Integridad. El emisor y el receptor del documento deben estar seguros de que dicho documento no fue alterado.
4. No repudio. El emisor envió el documento y no puede decir que no lo envió.

Según el tiempo de ejecución en función del tamaño de la entrada, podemos clasificar los algoritmos en dos tipos,

1. Complejidad Polinomial (problema fácil).
2. Complejidad No Polinomial (problema difícil).

Para que el cifrado sea seguro (difícil de descifrar sin la clave), debemos utilizar algoritmos que se basen en complejidad no polinomial en el descifrado. Existen varios problemas para los cuales no existen, hasta el momento, algoritmos en tiempo polinomial que los resuelvan. Uno de ellos es el Problema del Logaritmo Discreto sobre Curvas Elípticas (PLDE).

El PLDE consiste en hallar $n \in N$, dados $P, Q \in E(K)$ con $Q = nP$. No existen, de momento, algoritmos en tiempo polinomial para hallar $n$.

Presentamos a continuación distintos algoritmos que utilizan curvas elípticas, dependiendo del uso,

1. Intercambio de claves.

2. Cifrado de mensajes completos.
3. Firma Digital y verificación.
4. Codificación.

**3.1. Diffie-Hellman.** Este algoritmo es utilizado para el intercambio de claves. A y B se ponen de acuerdo en utilizar cierta $E(K)$ y un punto $P \in E(K)$, en general con $K = \mathbb{Z}_p$. A elige $n_A \in \mathbb{Z}$ y B elige $n_B \in \mathbb{Z}$. A calcula $R_A = n_A P$ y se lo envía a B. Por su parte B calcula $R_B = n_B P$ y se lo envía a A. Ahora A calcula $S = n_A R_B$ y B calcula $S = n_B R_A$, teniendo ambos el mismo $S$, secreto compartido. ¿Cómo funciona? Observemos que $S = n_A R_B = n_A n_B P = n_B n_A P = n_B R_A$.

**3.2. El Gamal.** Este algoritmo se utiliza para el envío de mensajes cifrados. Al igual que en el algoritmo anterior, A y B se ponen de acuerdo en utilizar cierta $E(K)$ y un punto $P \in E(K)$. Supongamos que B desea enviar a A cierto mensaje $m$. A elige $n_A \in \mathbb{Z}$ y B elige $n_B \in \mathbb{Z}$. A calcula $R_A = n_A P$ y se lo envía a B. Ahora B calcula $R_B = n_B P$ y $S = m + n_B R_A$ y los envía a A. A calcula $m = S - n_A R_B$ recuperando así $m$. Veamos, $S - n_A R_B = m + n_B R_A - n_A R_B = m$, por lo que efectivamente A recupera el mensaje original $m$.

**3.3. Elliptic Curve Digital Signature Algorithm (ECDSA).** Algoritmo para firmar digitalmente. La forma en la que se puede asegurar la autenticación, el no repudio y la integridad en una comunicación, es mediante la introducción de la firma al documento o mensaje. Lo más usual es utilizar primero una Función Hash (Huella), que lo que hace es, cuando aplicamos dicha función al mensaje, devolver un bloque de longitud fija. Luego este bloque se cifra y con ello obtenemos la firma y se adjunta al mensaje original. Las propiedades que debe tener una Función Huella $h$, para que pueda cumplirse con la seguridad requerida son:

1. Dado un mensaje $m$, debe ser fácil y rápido (computacionalmente hablando) calcular $h(m)$.
2. Dada una huella digital $h_0$, debe ser muy difícil (computacionalmente no existe algoritmo en tiempo polinomial) calcular $x_0$ de modo que $h(x_0) = h_0$. Esto se denomina resistencia a hallar una preimagen.
3. Dados $x_0$ y $h(x_0) = h_0$, debe ser prácticamente imposible hallar $x_1$, de manera que $h(x_0) = h(x_1)$. Esto se denomina resistencia a hallar una segunda preimagen.
4. Debe ser prácticamente imposible calcular $x_0$ y $x_1$, con $x_0 \neq x_1$ de modo que $h(x_0) = h(x_1)$. Esto se denomina resistencia a la colisión.

Entonces, una vez aplicada la función huella elegida al documento (texto, mensaje, archivo, etc.), se cifra dicha huella. Luego de cifrar la huella (firma electrónica), esta se anexa al documento, lo que produce un documento firmado, el cual se cifra, para luego enviarlo por la red a su destinatario.

Cuando el destinatario recibe el mensaje, se descifra, se obtiene el texto en claro y la firma, y pasamos a la etapa de verificación. Dicha etapa es sumamente importante, ya que es la que garantiza las propiedades de autenticación, integridad y no repudio.

El mecanismo de verificación es el siguiente, una vez descifrado el mensaje recibido, le aplicamos al texto en claro la función huella correspondiente, obtenemos la huella del texto que nos llegó. Luego con la clave pública, de firmar, del emisor, se descifra la firma y se obtiene la función huella del documento que se envió. Ahora sólo queda verificar que la huella del texto que se envió coincida con la huella del texto recibido.

Si coincide, entonces el documento no ha sido alterado, preservando la integridad del mismo. Como solo la clave pública para firmar del emisor es la que descifra la firma, entonces se garantiza la autenticidad y el no repudio.

Ahora se presenta el algoritmo ECDSA.

A y B se ponen de acuerdo en utilizar cierta $E(K)$ de orden $n$ (por seguridad $n > 2^{160}$) y un punto $P \in E(K)$. A elige $n_A \in \mathbb{Z}$ y calcula $R_A = n_A P$, elige $k \in \mathbb{Z}$ y calcula $Q = kP$ de modo que la abscisa de $Q$, $x_Q$, sea distinta de 0. Ahora calcula $k^{-1}{}_{(mod\, n)}$ y $h_m = h(m)$, y luego $s = k^{-1}(h_m + n_A x_Q)_{(mod\, n)}$. Si $s \equiv 0_{(mod\, n)}$ elige $k$ nuevamente y se calcula todo de nuevo, en caso contrario la firma es la pareja $(x_Q, s)$. La firma se introduce en el mensaje $m$ que se cifra para enviar a B (con El Gamal por ejemplo). Veamos ahora como B puede verificar la firma. Descifra el documento y obtiene el mensaje $m$ y la firma $(x_Q, s)$. Ahora calcula $h(m) = h_m$, $V = h_m s^{-1} P + x_Q s^{-1} R_A$. Si $x_Q \equiv x_{V\ (mod\, n)}$ la firma se acepta, caso contrario se rechaza.

Veamos la justificación: $V = h_m s^{-1} P + x_Q s^{-1} R_A = (h_m (h_m + n_A x_Q)^{-1} k)P + (x_Q (h_m + n_A x_Q)^{-1} k n_A)P = ((h_m + n_A x_Q)(h_m + n_A x_Q)^{-1} k)P = kP = Q$

**3.4. Codificación.** Según lo que hemos visto hasta ahora, sabemos como cifrar y firmar un determinado documento utilizando un sistema criptográfico basado en el problema del logaritmo discreto elíptico (PLDE). Ahora queda ver cómo se procede para realizar una equivalencia entre puntos de una curva elíptica y símbolos (caracteres y otros). De esta forma podemos asociar a un conjunto de símbolos, un conjunto de puntos sobre una curva elíptica. Debemos realizar un procedimiento que garantice la existencia de un algoritmo, por el cual asociemos a un símbolo, uno y sólo un punto de la curva a utilizar. Primero elegimos un alfabeto (la asignación de un entero a un símbolo, uno a uno, como ASCII o UNICODE) de $N$ símbolos. Tomemos bloques de largo $l$ para codificar. Sea $m_i$ el entero correspondiente al $i$-ésimo símbolo del bloque $b = (m_0, ..., m_{l-1})$. Ahora elegimos $k$ (con 50 suele ser suficiente) de modo que $kN^l < q$, siendo $q$ el número de elementos del cuerpo $K$. Esto es para garantizar que dado un entero podré asociarle un punto de la elíptica con mucha probabilidad $(1 - \frac{1}{2^k})$. Calculamos $x_b = \sum_{i=0}^{i=l-1} (m_i N^{l-1-i})$, asignamos $j = 0$. Chequeamos si existe $y$ de modo que $(kx_b + j, y) \in E(K)$. Si es así, dicho punto codifica el bloque, si no, vamos incrementando $j$ de a una unidad hasta obtener el punto de la elíptica asociado al bloque.

¿Cómo recuperar el bloque $b$ teniendo el punto de la elíptica $(x, y)$? Calculamos $k^{-1}$ y asignamos $j = 0$. Ahora calculamos $k^{-1}(x - j)$ y chequeamos si pertenece al intervalo $[0, N^{l-1}]$, si esto ocurre hacemos $x_b = k^{-1}(x - j)$, si no incrementamos $j$ en una unidad y repetimos. Una vez obtenido el $x_b$ descomponemos (por divisiones sucesivas por ejemplo) para obtener los $m_i$ de modo que $x_b = \sum_{i=0}^{i=l-1} (m_i N^{l-1-i})$, recuperando el bloque $b = (m_0, ..., m_{l-1})$.

## 4. Construcción de un ejemplo.

En el trabajo diseñamos un ejemplo donde trabajamos con el cuerpo $\mathbb{Z}_{103}$, la curva elíptica $E(\mathbb{Z}_{103}) = \{(x, y) \in \mathbb{Z}_{103}^2 : y^2 = x^3 - 2x + 1\} \cup \{O\}$ y el punto $P(65, 1)$ de la misma.

Por otro lado consideramos un alfabeto propio de 29 símbolos, con la asignación correspondiente a un punto de la elíptica, según la siguiente tabla,

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| A | (0,1) | F | (20,37) | K | (30,26) | O | (36,21) | T | (41,6) | Z | (48,9) |
| B | (5,42) | G | (25,34) | L | (31,13) | P | (37,25) | U | (43,3) | | (49,21) |
| C | (13,3) | H | (26,18) | M | (33,50) | Q | (38,1) | V | (44,15) | , | (51,35) |
| D | (15,16) | I | (27,36) | N | (34,32) | R | (39,29) | X | (46,42) | . | (52,42) |
| E | (18,21) | J | (28,24) | Ñ | (35,24) | S | (40,24) | Y | (47,3) | | |

Se utilizó El Gamal para cifrar y descifrar el mensaje deseado. También diseñamos la implementación de los algoritmos requeridos en un lenguaje de alto nivel.

Para más detalles sugerimos consultar el trabajo final.

## Referencias


[1] Castagna, H. (2017). Criptografía sobre Curvas Elípticas.

[2] Blake, I., Seroussi, G., Smart, N. (2005). Advances in Elliptic Curve Cryptography. Cambridge University Press.

[3] Enge, A. (1999). Elliptic Curves and their Applications to Cryptography. Massachusetts, United States of America: Kluwer Academic Publishers.

[4] Koblitz, N. (1994). A Course in Number Theory and Cryptography. New York: Springer-Verlag.

[5] Lucena López, M. Criptografía y Seguridad en Computadores (4ª ed.). España: Universidad de Jaén.

[6] Matheu García, S. (2015). Curvas Elípticas. Universidad de Murcia, Facultad de Matemáticas.

[7] Medina Aparcana, R. (2012). Criptografía con curvas elípticas sobre cuerpos p-ádicos. Universidad Nacional de Ingeniería, Facultad de Ciencias, Lima.

[8] Silverman, J. (2008). The Arithmetic of Elliptic Curves. Providence, Rhode Island: Springer.

[9] Vieitez, J. (2015). Aritmética Modular y Criptografía. Salto, Uruguay.



IPA, CFE, Montevideo, Uruguay

*Email address*: hcastagna@gmail.com

PONENCIA

# LAS TAREAS ENFOCADAS EN SIMILITUDES Y DIFERENCIAS EN EL APRENDIZAJE DE LAS TRANSFORMACIONES LINEALES EN LA FORMACIÓN INICIAL DE PROFESORES DE MATEMÁTICA

ANA MALDONADO

Publicaciones Matemáticas del Uruguay
Volumen 17, Julio 2019, Páginas 301–303
S 0797-1443

# LAS TAREAS ENFOCADAS EN SIMILITUDES Y DIFERENCIAS EN EL APRENDIZAJE DE LAS TRANSFORMACIONES LINEALES EN LA FORMACIÓN INICIAL DE PROFESORES DE MATEMÁTICA

ANA MALDONADO

Resumen. En el marco del Diploma en Matemática mención Enseñanza (ANEP-UdelaR) se realiza un proyecto de investigación en Matemática Educativa que intenta contribuir a la mejora de la enseñanza de las transformaciones lineales en la formación inicial de profesores de matemática. Para lograr tal propósito, se diseñaron actividades enfocadas en similitudes y diferencias considerando que estas permiten generar cierto grado de incertidumbre en el estudiante. Se realizó la experimentación de una de las actividades y se analizó la producción matemática de los estudiantes centrándose en la elaboración de conjeturas y el establecimiento de conexiones entre conceptos.

## 1. Desarrollo

La idea del trabajo surgió a partir de la práctica docente realizada en un grupo de Geometría y álgebra lineal en el IPA (Instituto de Profesores Artigas), en el marco de una asignatura del Diploma en Matemática. Observé que los estudiantes reaccionaban con asombro frente a las actividades propuestas y la diversidad de respuestas correctas.

Se ha reportado que el estudiante de formación docente es propenso a enseñar del modo en que se le ha enseñado (Fernández Pérez, 1994, citado en Marcelo, 1994). Por lo tanto, las tareas que se proponen a los estudiantes de profesorado y la forma en que se les solicita que se aproximen a ellas, tienen un rol fundamental en su formación.

Se han desarrollado varios estudios referidos a la enseñanza del álgebra lineal. Algunas de las principales dificultades en el estudio del algebra lineal son el obstáculo del formalismo (Dorier, J., Robert, A., Robinet, J. y Rogalski, M., 2000, citado en Sierpinska, 2000), que refiere a la escasa competencia de los estudiantes en lógica, teoría elemental de conjuntos y manipulación de expresiones algebraicas, junto a una pérdida de conexión entre las nuevas definiciones y los conocimientos previos (Dorier et al., 2000). Esto conlleva que los estudiantes no hagan uso de los conocimientos abstractos y se limiten a trabajar con un conjunto de algoritmos y procedimientos (Dubinsky, Dauterman, Leron y Zazkis, 1994). Ortega (2002) plantea que, cuando la enseñanza se centra en adquirir habilidades de manipulación, desmotiva al estudiantado generando poca participación activa. Sobre lo reportado específicamente a transformaciones lineales, se evidencia la existencia de mecanización en la verificación de la linealidad (Oktaç y Trigueros, 2010) y dificultad en la comprensión del concepto contemplando su interpretación geométrica-funcional-matricial (Maturana

©2019 PMU

y Parraguez, 2013). Se destaca que la comprensión del concepto de transformación lineal es compleja por su gran conexión con varios conceptos tales como espacio vectorial, función, coordenadas de un vector, combinación lineal, base, entre otros (Uicab y Oktaç, 2006, Roa Fuentes y Oktaç, 2012, Maturana y Parraguez, 2014). Se recomienda el diseño de tareas que estimulen al estudiante a considerar los conceptos involucrados y las relaciones que existen entre ellos, generando experiencias que los conduzcan a construir su propio conocimiento, motivando la generación de conjeturas con sus respectivas argumentaciones (Uicab y Oktaç, 2006, Oktaç y Trigueros, 2010).

Con el objetivo de favorecer el aprendizaje de las transformaciones lineales y contemplando las metodologías sugeridas para la enseñanza de conocimientos matemáticos, diseñamos tareas enfocadas en similitudes y diferencias (Zaslavsky, 2008) que pudieran promover instancias de incertidumbre (Zaslavsky, 2005) que, a su vez, fomentan los procesos de argumentación y la elaboración de pruebas. Buscando evidenciar su potencial en el contexto planteado, testeamos una actividad diseñada en un grupo de profesorado de matemática. En esta etapa el objetivo fue explorar y analizar la producción matemática de los estudiantes, centrada en la elaboración de conjeturas y el establecimiento de conexiones entre conceptos, cuando trabajan en la tarea.

Se diseñaron tres tareas considerando el contexto en el que se trabajó el tema transformaciones lineales en la institución de formación donde se testeó una de dichas tareas.

Se realizó la puesta en escena de una tarea diseñada y el análisis de las reacciones y elaboraciones de los estudiantes frente a la resolución de esta. La actividad se trabajó de manera exploratoria con un grupo de estudiantes de segundo año de profesorado de matemática del IPA de la asignatura Geometría y álgebra lineal. La clase estuvo a cargo de la docente investigadora y la metodología de trabajo empleada para conducirla fue de corte clínico, se escuchó a los estudiantes devolviendo las preguntas que estos formularon a los otros estudiantes y planteó preguntas que promovieron el conflicto cognitivo. La clase duró sesenta minutos, los primeros treinta minutos se trabajó en subgrupos y posteriormente se realizó la puesta en común, toda la clase fue audio grabada.

En ambas etapas se integra al profesor encargado del grupo donde se testea la actividad, pues tiene gran experiencia en el dictado del curso. Además, dado su conocimiento del grupo, es quien determina el momento de aplicación en función de los contenidos trabajados en la clase.

Una diferencia manifestada por el profesor, entre la tarea diseñada testeada y las actividades tradicionales del tema, es que se logró identificar el pensamiento y la comprensión matemática de los estudiantes.

La experimentación de la actividad permitió detectar algunos aspectos de los antecedentes:

- el obstáculo del formalismo (Dorier et al, 2000);
- en algunas instancias el estudiante tendió a usar procedimientos mecánicos (Otega, 2002);
- la relevancia de los conceptos previos asociados a transformación lineal (Oktaç y Trigueros, 2010, Roa Fuentes y Oktaç, 2012);
- la importancia de la conexión entre los conceptos previos con los posteriores (Uicab y Oktaç, 2006);

- el obstáculo de hallar la imagen de un vector a partir de la matriz asociada a una transformación lineal (Maturana y Parraguez, 2014).

En la experimentación de la actividad los estudiantes:

- mantuvieron un alto nivel de atención y discusión;
- abordaron de distintas formas la actividad;
- priorizaron la comprensión sobre la obtención de resultados;
- experimentaron cierto estado de incertidumbre;
- manifestaron preferencias personales;
- conjeturaron y utilizaron métodos de prueba para sus argumentaciones, comunicando ideas matemáticas;
- utilizaron como herramienta varios conceptos y proposiciones al servicio de la exploración para comparar las transformaciones lineales.

## Referencias

[1] Dorier, J., Robert, A., Robinet, J. y Rogalski, M. (2000). The Obstacle of Formalism in Linear Algebra. En J. Dorier (Ed.), On the Teaching of Linear Algebra (pp. 85-124). Dordrecht: Kluwer Academic Publishers.

[2] Dubinsky, E., Dauterman, J., Leron, U. y Zazkis, R. (1994). On learning fundamental concepts of Group Theory. Educational studies in Mathematics, 27, 267-305. Kluwer Academic Publishers.

[3] Marcelo, C. (1994). Investigaciones sobre prácticas en los últimos años: qué nos aportan para la mejora cualitativa de las prácticas. Ponencia presentada al III Symposium Internacional sobre Prácticas Escolares, Poio.

[4] Maturana, I. y Parraguez, M. (2013). Transformaciones lineales. Una mirada desde la teoría APOE. CLAME, Acta Latinoamericana de Matemática Educativa, 26, 881-890.

[5] Maturana, I. y Parraguez, M. (2014). Una mirada cognitiva a las transformaciones lineales. Articulación entre las interpretaciones: matricial y funcional. CLAME, Acta Latinoamericana de Matemática Educativa, 27, 441-449.

[6] Oktaç, A. y Trigueros, M. (2010). ¿Cómo se aprenden los conceptos de álgebra lineal? RELIME, Revista Latinoamericana de Investigación en Matemática Educativa, 13 (4-II), 373-385.

[7] Ortega, P. (2002). La enseñanza del Álgebra Lineal mediante sistemas informáticos de cálculo algebraico. Tesis doctoral no publicada. Universidad Complutense de Madrid. España. Recuperado desde: http://biblioteca.ucm.es/tesis/edu/ucm-t25694.pdf.

[8] Roa Fuentes, S. y Oktaç, A. (2012). Validación de una descomposición genética de transformación lineal: un análisis refinado por la aplicación del ciclo de investigación de la teoría APOE. RELIME, Revista Latinoamericana de Investigación en Matemática Educativa, 15 (2), 199-232.

[9] Sierpinska, A. (2000). On some aspects of students' thinking in linear algebra. En J.-L. Dorier (Ed.), On the Teaching of Linear Algebra. Kluwer Academic Publishers, 209-246.

[10] Uicab, R. y Oktaç, A. (2006). Transformaciones lineales en un ambiente de geometría dinámica. RELIME, Revista Latinoamericana de Investigación en Matemática Educativa, 9 (3), 459-490.

[11] Zaslavsky, O. (2005). Seizing the opportunity to create uncertainty in learning mathematics. Educational Studies in Mathematics, 60, 297–321.

[12] Zaslavsky, O. (2008). Attention to similarities and differences: A fundamental principle for task design and implementation in mathematics education. Invited presentation at the Topic Study Group (TSG34) on Research and Development on Task Design and Analysis, the 11th International Congress on Mathematics Education (ICME-11), Monterrey, México. Recuperado desde http://tsg.icme11.org/document/get/290.

CES, CETP, CFE. Uruguay.
*Email address*: anamaldo@gmail.com

PONENCIA

# UN PROYECTO DE INTERVENCIÓN A PARTIR DE TAREAS DE GENERALIZAR Y PARTICULARIZAR: TRABAJO COLABORATIVO ENTRE INVESTIGADOR Y FORMADOR DE PROFESORES

VICTORIA MESA PAIS

Publicaciones Matemáticas del Uruguay
Volumen 17, Julio 2019, Páginas 307–312
S 0797-1443

# UN PROYECTO DE INTERVENCIÓN A PARTIR DE TAREAS DE GENERALIZAR Y PARTICULARIZAR: TRABAJO COLABORATIVO ENTRE INVESTIGADOR Y FORMADOR DE PROFESORES

VICTORIA MESA PAIS

## Resumen

En el presente trabajo se describirá una intervención en Matemática Educativa llevada a cabo, en el marco del Diploma en Matemática (IPES–UdelaR), en conjunto con un docente de Profundización en Geometría en torno a la creación e implementación de actividades de particularizar y generalizar sobre el concepto de baricentro. El proceso vivenciado permitió constatar que las actividades de generalizar y particularizar un problema o enunciado matemático, pensadas como actividades de final abierto en una metodología de clase que fomente la producción del conocimiento matemático por parte del estudiante, favorecen que los estudiantes de profesorado de Matemática vinculen los conocimientos que aprenden en sus clases de Matemática de la carrera con los que deberán enseñar en enseñanza media.

## Las prácticas docentes y una propuesta de trabajo colaborativo

Mellado (1996) y Marcelo (1994) reportan que la manera en que los estudiantes de profesorado aprenden los conocimientos matemáticos y el modo en que se les ha enseñado influyen en sus prácticas futuras como docentes. Por ello, proponer una metodología de clase que fomente la producción del conocimiento matemático por parte de los estudiantes y tal que el docente sea un facilitador y promotor del aprendizaje, favorece la incorporación de dichas prácticas en el futuro profesor.

Teniendo en cuenta estas consideraciones, nos propusimos realizar una intervención en cursos de Matemática del profesorado pretendiendo contribuir, desde un conjunto de instancias de trabajo colaborativo, al enriquecimiento de las prácticas docentes. Específicamente, entendemos por intervención en Matemática Educativa (ME) una propuesta de acción original y creativa, programada en el tiempo y desarrollada en forma colaborativa entre un investigador y un profesor de Matemática, desde una concepción teórica determinada, que busca contribuir a las prácticas docentes con el objetivo de mejorar los aprendizajes (Mesa, 2016; Scorza, 2016).

Como parte de la propuesta se acercan al docente herramientas metodológicas para el trabajo en el aula, producto de resultados de investigaciones en ME, poniendo el foco en el diseño y planificación de actividades, su análisis a priori, implementación y análisis a posteriori.

©2019 PMU

## Tareas de generalizar y particularizar como un camino para vincular la Matemática avanzada y la Matemática escolar en la formación de profesores de Matemática

Las tareas o actividades que proponemos a nuestros estudiantes influyen en su aprendizaje, tanto su consigna como su implementación en el aula. La elección de las tareas matemáticas para la clase y la manera en que se pide a los estudiantes que se aproximen a ellas, determinan la calidad de la Matemática en la clase (Simon, 1997; Steinbring, 1998). Es por esto que decidimos poner el foco de nuestro proyecto de trabajo colaborativo, en el diseño de tareas.

La herramienta didáctica que seleccionamos para proponer en la intervención es la de tareas de generalizar y particularizar objetos matemáticos. En Maldonado et al. (2015) se identifican estas actividades como un tipo de tareas de atención a similitudes y diferencias entre objetos matemáticos (Zaslavsky, 2008), dentro de las denominadas tareas de final abierto (Zaslavsky, 1995).

A partir de las definiciones dadas por Polya (1965) de los constructos "generalizar" y "particularizar" en la resolución de un problema matemático, en Maldonado et. al. (2015) se conceptualizan las tareas de generalizar como actividades que "invitan a atender propiedades de uno o más objetos matemáticos particulares y proponer que los alumnos investiguen si conjuntos de objetos, que incluyen al particular, conservan la o las propiedades originales" (pp. 83-84). Mientras que, en las tareas de particularizar, "partiendo de una situación o problema se recurre a un caso particular para realizar ciertas conjeturas y brindar argumentos para la misma" (p. 84). En ambos tipos de tareas se busca que el estudiante analice no solo si un cierto resultado es válido para el caso particular o el general, sino también si los argumentos dados son válidos al pasar de un contexto al otro.

Específicamente en formación docente, este tipo de actividades, además de constituir interesantes problemas matemáticos, permiten el acercamiento del estudiante, futuro profesor, a problemas que tratan un cierto saber desde distintas perspectivas, lo que brinda herramientas para pensar su abordaje en los diferentes niveles educativos. Así como el contacto con metodologías de trabajo donde el eje central es, a través de un trabajo conjunto entre estudiantes y docente, la construcción del conocimiento matemático en el aula entendida como un espacio de indagación y aprendizaje.

Por otra parte, Ticknor (2012) reporta una investigación sobre la dificultad que presentan los estudiantes de profesorado para realizar conexiones interesantes entre los conocimientos matemáticos avanzados que son abordados en un curso de su formación, concretamente en álgebra abstracta, y los conocimientos matemáticos que luego deberán enseñar a sus alumnos de educación media. Buscando atender a esta problemática, el autor plantea la necesidad de establecer, en las clases de Matemática del profesorado, conexiones explícitas con la Matemática que los estudiantes, futuros profesores, habrán de enseñar, profundizando así su comprensión de los conocimientos matemáticos.

Dos aspectos importantes para hacer explícitas dichas conexiones son: el contenido abordado, que debe ser tal que lleve al estudiante a preguntarse sobre contenidos de enseñanza media, y el rol del docente en la implementación de las actividades. Pero pensamos que el tipo de tarea propuesta también es influyente. Entendemos que las tareas de generalizar y particularizar pueden contribuir a lograr tal objetivo. Por un lado, al revisitar un tema que tienen que enseñar en enseñanza media, tanto

cuando a partir de un resultado que se maneja en ese ciclo se intenta generalizar a otro contexto como cuando se logra detectar que un determinado conocimiento a enseñar en ese ciclo es un caso particular de algo más general que están viendo en su clase de formación docente. Por otro, porque este tipo de tareas favorece que los estudiantes comprendan que muchos de los conceptos o propiedades que se enseñan en nivel medio son casos particulares de otras más generales, y, como tales, adoptan sus características. Esto posibilita una mayor comprensión de los fundamentos, argumentos y explicaciones de los contenidos abordados y las relaciones entre ellos.

## Relato de un caso: la intervención realizada

La intervención fue implementada en el año 2016 en un instituto de formación de profesores del país, con un profesor de la asignatura Profundización en Geometría del último año de la carrera de profesorado de Matemática. La misma se desarrolló en cinco etapas, que se describen a continuación.

Etapa 1: Presentación del proyecto. Se presenta en qué consiste la intervención, cuáles son sus objetivos y cuál es el rol del investigador.
El profesor acuerda con la propuesta, se muestra abierto e interesado en participar. Se reflexiona de manera conjunta sobre las prácticas docentes recomendadas para la formación docente por investigaciones en ME así como la problemática planteada por Ticknor (2012) sobre las dificultades que presentan los estudiantes de profesorado, produciéndose diferentes puntos de vista que enriquecen la reflexión.

Etapa 2: Presentación del marco teórico: caracterización y ejemplificación de las actividades de generalizar y particularizar (Polya, 1965,1966; Davis y Hersh, 1989); caracterización de tareas de atención a similitudes y diferencias (Zaslavsky, 2008) y presentación de las tareas de generalizar y particularizar como un tipo de ellas (Maldonado et al., 2015). Se reflexiona sobre la importancia de las tareas que proponemos a nuestros estudiantes y sobre el rol del docente en la implementación de estas actividades para un mejor aprovechamiento de las mismas (Swan, 2008).
En esta etapa además el docente nos expresa los objetivos y contenidos del curso, buscando identificar en conjunto en qué momentos puede ser conveniente implementar este tipo de tareas.

Etapa 3: Elaboración de tareas de generalizar y particularizar desde un trabajo colaborativo entre docente e investigador. Se analiza su adecuación a los objetivos y contenidos del curso, su implementación en la clase, las metodologías de trabajo convenientes para un mayor aprovechamiento de la propuesta y el potencial a priori de las actividades pensadas.

El profesor formador aporta desde su conocimiento de los estudiantes, de los contenidos del curso y desde su experiencia como docente de la asignatura, y el docente investigador, desde el marco teórico en que se enmarca la intervención y sobre aspectos metodológicos.

A partir del concepto de baricentro en un triángulo y su existencia (problema original), se elaboran las siguientes consignas:

1) “Particularizar la noción de baricentro de un triángulo a un segmento” (para ser trabajada en una clase previa a la visitada por el investigador). Entendemos que ?’este puede ser visto como un caso particular del problema original para el caso en que dos de los vértices del triángulo coinciden.

2) “Generalizar la noción de baricentro de un triángulo a un cuadrilátero” (para ser trabajada en la clase visitada por el investigador). La generalización del problema original consiste en definir y construir el baricentro para polígonos de cualquier cantidad de lados, pero se plantea abordar en esta primera instancia el caso de los cuadriláteros.

Se acuerda plantearlas en el aula mediante una modalidad de taller, en parejas, teniendo disponible el software GeoGebra, con tiempo suficiente para que los estudiantes logren establecer conjeturas y refutarlas o confirmarlas.

Etapa 4: Puesta en escena. El profesor implementa las actividades con sus estudiantes, el rol del investigador en esta instancia es el de acompañar para facilitar el análisis crítico posterior a la puesta en escena.
Los estudiantes desarrollaron diversas estrategias de resolución, algunas de ellas previstas en el análisis a priori de la actividad y otras nuevas.

Etapa 5: Entrevista final entre el profesor y el investigador para un análisis a posteriori de la propuesta en relación a la potencialidad de las actividades y de la metodología de trabajo utilizada, y a la posibilidad de continuidad.
El profesor valora positivamente la dinámica de la clase, que surgieran de los estudiantes algunas respuestas a las consignas y que se mostraran animados con la misma. Por otra parte, sostiene que para que los estudiantes puedan demostrar, además de conjeturar, necesitan tener una base teórica que según deja entrever se lograría a través de clases más ‘tradicionales’. Sí considera factible la construcción de conocimientos por parte de los estudiantes para una etapa posterior.
Plantea además que tiene en mente trabajar a futuro con la generalización del concepto de ortocentro y por ende la generalización de la recta de Euler. En cuanto a lo metodológico, percibe como favorable trabajar con una dinámica de clase muy similar a la dada, apostando a que con el tiempo surjan otras ideas nuevas y adquiriendo así otra dinámica de trabajo. Agrega: “yo lo pienso como una idea progresiva de ir incorporando cierto [. . . ] tipo de ideas para poder generar actividades de tipo y hacerlo más fluido”.
Respecto a la idea de generalizar y particularizar, afirma que la misma siempre aparece en las clases de Matemática de los distintos cursos, incluso en las clases más tradicionales, donde el docente puede mostrar y los estudiantes pueden observar distintas generalizaciones. Pero ello es diferente a proponer como consigna una actividad de generalizar y particularizar a los estudiantes. El docente concibe esto viable principalmente en los cursos de Geometría, no visualizando inmediatamente cómo hacerlo en otras asignaturas, pero no descartando la posibilidad.
Por otro lado, menciona que, si bien el conocimiento matemático en juego en la propuesta ya era previamente conocido por él, la intervención le permitió conocer determinados materiales que le dieron la posibilidad de mirar ese conocimiento desde otro lugar, logrando organizarlo mejor y por ende plantearlo mejor.

## Reflexiones finales

Consideramos como uno de los aportes de la intervención acercar al profesor un marco metodológico específico que permitiera plasmar un conocimiento matemático previamente investigado por él en una tarea de aula para los estudiantes de profesorado, desde una nueva metodología de trabajo, tanto en lo concerniente a la consigna de la tarea como a su implementación en la clase. Ello se evidencia en la proyección del docente de incorporar a algunas de sus clases, nuevas metodologías que enfaticen la producción de conocimientos por parte de los estudiantes.

En lo que refiere al conocimiento de los estudiantes, consideramos que la intervención pudo aportar implícitamente a la problemática detectada por Ticknor (2012). Al generalizar el concepto de baricentro a un cuadrilátero, que constituye un conocimiento avanzado, los estudiantes deben revisar sus concepciones, argumentos y justificaciones sobre el baricentro de un triángulo y otros conceptos o propiedades geométricas, que incluyen conocimientos más básicos o previamente conocidos. Además, para que el problema tuviera sentido para los estudiantes, primeramente se tuvo que recordar de manera explícita qué era el baricentro de un triángulo, cómo se construye y sus propiedades principales: métricas y físicas. Cuando los estudiantes proponen la generalización, lo hacen a partir de esas ideas, intentando mantener las propiedades abordadas, pero ahora para los cuadriláteros. Por otro lado, de manera implícita la intervención buscó aportar al conocimiento didáctico de los futuros profesores, al vivenciar una modalidad de trabajo basada en la construcción de conocimientos por parte de los estudiantes, acorde a las recomendaciones didácticas de investigaciones en ME.

Creemos, además, que el trabajo colaborativo es siempre una instancia de enriquecimiento profesional para todos, docentes e investigadores.

## Referencias

[1] Davis, P. y Hersh, R. (1989). Experiencia matemática. Sección: Generalización. pp. 105-107. Barcelona: MEC y Labor.

[2] Maldonado, A., Medina, L., Mesa, V., Molfino, V., Ochoviet, C., Pagés, D., Rivero, F.(2015). Tareas enfocadas a similitudes y diferencias como motor para el aprendizajede la matemática: nuevas categorías. En Buendía, G., Molfino, V., Ochoviet, C.Estrechando lazos entre investigación y formación en Matemática Educativa. Vol. II.Consejo de Formación en Educación. Departamento de Matemática. Uruguay.

[3] Marcelo, C. (1994). Investigaciones sobre prácticas en los últimos años: qué nos aportan para la mejora cualitativa de las prácticas. Ponencia presentada al III Symposium Internacional sobre Prácticas Escolares, Poio.

[4] Mellado, V. (1996).Concepciones y prácticas de aula de profesores de ciencias, en formación inicial de primaria y secundaria. En Investigación y experiencias didácticas. Enseñanza de las ciencias, 14 (3), 289-302.

[5] Mesa, V. (2016). Actividades de generalizar y particularizar como medio para vincular la Matemática avanzada y la Matemática escolar en la formación de profesores de Matemática(tesina de Diploma no publicada). Consejo de Formación en Educación – Universidad de la República. Montevideo, Uruguay.

[6] Polya, G. (1965): Cómo plantear y resolver problemas. Ed. Trillas. México.

[7] Polya, G. (1966). Matemáticas y razonamiento pausible. Madrid: Tecnos.

[8] Scorza, V. (2016). Las tareas de final abierto y su potencial para la enseñanza de la matemática en la formación de profesores (tesinade Diploma no publicada). Consejo de Formación en Educación – Universidad de la República. Montevideo, Uruguay.

[9] Simon, A. M. (1997). Developing new models of mathematics teaching: An imperative for research on mathematics teacher development. En E. Fennema y B. Scott-Nelson (Eds.), Mathematics teachers in transition (pp. 55–86). Mahwah, NJ: Lawrence Erlbaum Associates.

[10] Steinbring, H. (1998). Elements of epistemological knowledge for mathematics teachers. Journal of Mathematics Teacher Education, 1(2), 157–189.
[11] Swan, M. (2008). The design of multiple representation tasks to foster conceptual development. Recuperado desde http://tsg.icme11.org/document/get/289.
[12] Ticknor, C. (2012). Situated learning in an abstract algebra classroom.Educ Stud Math 81: 307-323.
Zaslavsky, O. (1995). Open-ended tasks as a trigger for mathematics teachers' professional development. For the Learning of Mathematics, 15(3), 15-20.
[13] Zaslavsky, O. (2008). Attention to similarities and differences: A fundamental principle for task design and implementation in mathematics education. En 11th International Congress on Mathematics Education (ICME-11), México. Recuperado desde http://tsg.icme11.org/document/get/290.

CES, Uruguay
*Email address*: vickymesa_16@hotmail.com

Publicaciones Matemáticas del Uruguay
Volumen 17, Julio 2019, Páginas E1–E2
S 0797-1443

# CORRECTIONS TO "HOLOMORPHIC CURVES AND CELESTIAL MECHANICS"

UMBERTO L. HRYNIEWICZ

Abstract. Here we make corrections to our discussion on Floer's boundary operator, and take the opportunity to fix a misuse of terminology concerning exact symplectic diffeomorphisms.

## 1. Corrections

In Section 2 we outlined the construction of Floer's chain complex in the symplectically aespherical case. Consider a closed symplectic manifold $(W,\omega)$ such that $\omega$ and $c_1(TW,\omega)$ vanish on $\pi_2(W)$. Let $H_t$ be a 1-periodic Hamiltonian on $W$ and $J_t$ be a 1-periodic $\omega$-compatible almost complex structure. The vector field $-J_t(c(t))[\dot{c}(t) - X_{H_t}(c(t))]$ along a contractible closed loop $c$ on $W$ corresponds to the $L^2$ **gradient** of the action functional

$$\mathcal{A}_H(c) = \int_{\mathbb{D}} v^*\omega + \int_{\mathbb{R}/\mathbb{Z}} H_t(c(t))dt$$

and not to the $L^2$ anti-gradient as we stated. Here $v : \mathbb{D} \to W$ is a capping disk for $c(t)$ and the $L^2$-inner product is defined by

$$\langle X, Y \rangle_{L^2} = \int_{\mathbb{R}/\mathbb{Z}} \omega(c(t))(X(t), J_t(c(t))Y(t))\ dt$$

for all vector fields $X, Y$ along the loop $c$. Moreover, for generic data $(H, J)$, the boundary operator is

$$\delta_{H,J}(x) = \sum_{\mathrm{CZ}(y)=\mathrm{CZ}(x)-1} (n(x,y) \mod 2)\ y$$

where $n(x,y)$ is a count of solutions $u(s,t)$ of Floer's equation satisfying

$$\lim_{s\to+\infty} u(s,t) = x(t) \qquad \lim_{s\to-\infty} u(s,t) = y(t)$$

In the published version of the paper we interchanged the positions of the asymptotic limits. This completes our set of corrections to the discussion on Floer's chain complex.

Finally, in Section 1 there is a misuse of the expression "center of mass". A diffeomorphism $\Psi$ of $\mathbb{R}^{2n}/\mathbb{Z}^{2n}$ isotopic to the identity can be lifted to a diffeomorphism $\tilde{\Psi}$ of the universal covering $\mathbb{R}^{2n}$ and the integral in equation (8)

$$\int_{[0,1]^{2n}} \Delta(z)$$

©2019 PMU

is not the center of mass, it is the **shift of the center of mass** by the lifted map $\tilde{\Psi}$. There are a few places where this erroneous use of terminology needs correction. For instance, in the definition of exact symplectic diffeomorphisms condition (H3) should read "$\Psi$ is exact: it admits a lift to $\mathbb{R}^{2n}$ that preserves center of mass". This is in accordance to the discussion in [1, appendix 9], where the expression "center of gravity" is used.

## References

[1] V.I. Arnold, *Mathematical methods of classical mechanics.* Springer-Verlag, Berlin and New York, 1978.

Las Publicaciones Matemáticas del Uruguay (PMU) tienen como objetivo reflejar parte de las actividades de investigación matemática que se lleva a cabo en Uruguay. Nuestro interés es publicar artículos de investigación, así como artículos de tipo survey, anuncios, y otros trabajos que el Consejo Editor considere adecuado.

Los volúmenes no necesariamente serán arbitrados. Esto se indicará cuidadosamente en cada volumen.

---

Los artículos de este volumen no han sido arbitrados.

---

The goal of Publicaciones Matemáticas del Uruguay (PMU) is to reflect part of the mathematics research activities taking place in Uruguay. It is our interest to publish research articles, survey-type articles, research announcements and other papers considered suitable by the Editorial Board.

The editorial process may or may not involve a revision by referees. This will be carefully indicated in each volume.

---

The articles of this volume were not peer-reviewed.

---

*(Continuación de la contratapa)*

Transformaciones en la enseñanza de la matemática para estudiantes ingresantes a la Facultad de Arquitectura, Diseño y Urbanismo
Omar Gil ..... 167

El número de oro, geometría, álgebra, aritmética
Gerardo González Sprinberg ..... 183

Teorías y modelos: una introducción a la lógica de primer orden
Alexandre Miquel ..... 195

Conjuntos minimales para homeomorfismos del círculo
Aldo Portela ..... 227

Wagner and Petersen are uniformly most-reliable graphs
Pablo Romero ..... 231

Una introducción a modelos matemáticos para composición automática de música tonal
Verónica Rumbo ..... 247

Semiprojectivity and semiinjectivity in different categories
Hannes Thiel ..... 263

Rigidez y geometricidad de acciones de grupos de superficies en el círculo
Maxime Wolff ..... 277

*Diplomados, ponencias*

Ley de Benford
María Caputi Zunini ..... 285

Criptografía sobre curvas elípticas
Horacio Castagna ..... 293

Las tareas enfocadas en similitudes y diferencias en el aprendizaje de las transformaciones lineales en la formación inicial de profesores de matemática
Ana Maldonado ..... 301

Un proyecto de intervención a partir de tareas de generalizar y particularizar: trabajo colaborativo entre investigador y formador de profesores
Victoria Mesa Pais ..... 307

Publicaciones Matemáticas del Uruguay

Volumen 17 Julio 2019

Prefacio ..... iii

*Minicursos*

Un paseo por la optimización
Marcelo Fiori ..... 1

Introducción a las Álgebras de Lie
Ana González ..... 25

La regularidad de los poliedros en la mirada moderna
Ángel Pereyra ..... 43

*Charlas Plenarias*

Holomorphic curves and celestial mechanics
Umberto L. Hryniewicz ..... 61

Solución de las ecuaciones de Hamilton-Jacobi como límite clásico de integrales con respecto al movimiento Browniano
José R. León ..... 81

Dinámica elemental de cubrimientos del anillo
Alvaro Rovella ..... 95

*Ponencias*

Complejidad en análisis numérico y el problema de valores propios
Diego Armentano ..... 109

Dendritaciones de superficies
Alfonso Artigue ..... 121

Ceros de polinomios y funciones aleatorias
Federico Dalmao ..... 127

Equivariant algebraic KK-Theory
Eugenia Ellis ..... 141

Amenabilidad, acciones de grupos, y equivalencia orbital
Eusebio Gardella ..... 153

*(Continúa en el reverso)*