POLINOMIOS ENUMERADORES DE PRIMOS

Ana Carrasco Martín

POLINOMIOS ENUMERADORES DE PRIMOS

Ana Carrasco Martín

Memoria presentada como parte de los requisitos para la obtención del título de Grado en Matemáticas por la Universidad de Sevilla.

Tutorizada por

Manuel Jesús Soto Prieto

Abstract

Primes are known to be the building blocks of numbers, and their beauty has fascinated mathematicians since the very beginning of time. While numerous formulas defining primes are found in literature, can they be elegantly expressed by a polynomial? This, and other related questions are the core of this dissertation. First, the fact that no polynomial can represent only primes is proved. However, we present an astonishing and seemingly unlikely result: the existence of multivariable polynomials whose represented set of positive values is precisely the entire set of primes. What is more: this result solves Hilbert's Tenth Problem in the negative. We named this polynomials "prime enumerating polynomials". We not only construct such a polynomial but also attempt practical computations and explore the minimum degree and number of variables required for this type of construction.

Resumen

Los primos son los átomos que conforman los números, y su belleza ha cautivado a los matemáticos desde tiempos inmemoriales. Existen numerosas fórmulas que definen los primos, pero ¿pueden ser descritos mediante una expresión tan elegante como un polinomio? La respuesta a esta y otras cuestiones afines constituyen esta memoria. En primer lugar, probamos que no existe un polinomio que represente únicamente números primos. Sin embargo, presentamos un resultado sorprendente y aparentemente inverosímil: la existencia de polinomios multivariables para los que el conjunto de enteros positivos que representa coincide, exactamente, con el conjunto de los números primos. Más aún: este resultado demuestra el carácter indecidible del décimo problema de Hilbert. A estos polinomios los hemos denominado polinomios enumeradores de primos. No solo construimos un polinomio enumerador de primos, sino que también tratamos de realizar cálculos prácticos e indagar en el menor grado y número de variables necesarios para este tipo de construcción.

Índice general

1 Introducción: ¿Existen funciones que definan los primos?

A modo de introducción, trataremos de dar respuesta en este primer capítulo a una de las muchas preguntas que surgen al sumergirnos en el mundo de los números primos: ¿existen funciones que los definan?, ¿es posible dar una fórmula que genere números primos? Como el término de función es quizá demasiado amplio, esta pregunta sugiere considerar distintas aproximaciones. Realizaremos un breve recorrido por los casos más estudiados apoyándonos de ejemplos notables, dando respuesta a las clásicas preguntas propuestas en [HW79]. Para ello, seguiremos la clasificación de [Rib12]. Concluiremos el capítulo presentando el caso que nos ocupará en esta memoria.

1.1 $f(n) = p_n$

Como es habitual, denotamos por p_n al $n-$ésimo primo. Es bien conocido que, por el Teorema de los Números Primos (PNT),

$$p_n \sim n \log n.$$

Existen muchas otras expresiones asintóticas que mejoran sustancialmente esta aproximación. Una infinidad de libros y artículos tratan este tema; puede consultarse, por ejemplo, [HW79] para el tratamiento clásico del PNT, y [AdRJ12] para un estudio reciente que aporta una nueva expresión asintótica de p_n, y una cota precisa del error que se comete, suponiendo cierta la Hipótesis de Riemann. La pregunta que pretendemos contestar aquí es: ¿existe una expresión cerrada para el $n-$ésimo primo, en términos de n? Son muchas las fórmulas que se han presentado para dar respuesta. El problema de estas reside en que son, esencialmente, una redefinición tautológica de lo que es un número primo. La construcción de estas fórmulas suele basarse en el conocimiento previo del número primo que queremos calcular así como sus pre-

cedentes, de funciones poco tratables en la práctica o de funciones clásicas en teoría de números como la función contadora de primos $\pi(x)$ o la función de Möbius $\mu(n)$. Uno de los resultados que permiten la construcción de gran cantidad de ejemplos es el Teorema de Wilson, cuyo enunciado y demostración están recogidos en el Teorema 3.3, y será una herramienta fundamental en el resto del trabajo. Veamos un ejemplo:

Ejemplo 1.1. En 1975, James P. Jones proporcionó una fórmula para el $n-$ésimo primo basada en el Teorema de Wilson y la función $\pi(x)$ (véase [Jon75]):

$$p_n = \sum_{i=0}^{n^2} \left(1 \dot{-} \left(\left(\sum_{j=0}^{i} \operatorname{rem}((j \dot{-} 1)!^2, j)\right) \dot{-} n\right)\right).$$

Aquí, hemos denotado por $\operatorname{rem}(a, b)$ al resto de la división euclídea de a entre b, y $x \dot{-} y$ es la operación que efectúa la resta si $y \leq x$, y devuelve 0 en caso contrario.

Como consecuencia del teorema principal del segundo capítulo, veremos que podemos representar el $n-$ésimo primo mediante un polinomio, aunque no en una variable y con ciertas inexactitudes que estudiaremos posteriormente. En particular, estaremos en las condiciones de probar el siguiente teorema:

Teorema 1.1. *Existe un polinomio $Q(n, x_1, \dots, x_k)$ tal que, para cualesquiera n y m enteros positivos, se tiene que*

$$p_n = m \iff (\exists x_1, \dots, x_k)\{Q(n, x_1, \dots, x_k) = m\}.$$

Un resultado muy reciente, dado en [FGG+19], conecta esta sección con la siguiente: existe una constante $f_1 = 2.920050977316\dots$ y una sucesión recursiva

$$f_n = \lfloor f_{n-1} \rfloor (f_{n-1} - \lfloor f_{n-1} \rfloor + 1)$$

tal que $\lfloor f_n \rfloor = p_n$, para todo n.

1.2 Constantes que representan primos

¿Existe una función $f(n)$ que devuelva únicamente valores primos y que sean distintos para todo valor de n? El resultado de [FGG+19] lo confirma. El ejemplo más conocido se debe a Mills, quien demostró en 1947 que existe una constante $A \in \mathbb{R}$ tal que $f(n) = \lfloor A^{3^n} \rfloor$ representa únicamente valores primos (aunque no todos) para todo

$n \in \mathbb{N} \setminus \{0\}$ (véase [Mil47]). Mills no aporta información alguna sobre la constante A, posteriormente conocida como constante de Mills. En [CC10] se calcula el mínimo valor de A verificando el resultado Mills, que como tantos otros resultados, está sujeto a la veracidad de la Hipótesis de Riemann. Así,

$$A \approx 1.306377883863\ldots$$

Los primos representados por esta constante son conocidos como primos de Mills, y pueden consultarse en [OEI, A051254].

El hecho de que el valor de la constante quede sujeto a la Hipótesis de Riemann implica que el número de decimales que conocemos de la misma depende de los primos que conocemos, por lo que, en realidad, podríamos decir que son los primos los que estarían generando la constante y no al contrario. Por este motivo, no es de utilidad en la práctica para encontrar nuevos primos y su interés es puramente teórico.

1.3 No existe un polinomio que represente únicamente primos

Mostramos en esta sección que, salvo polinomios constantes o casos triviales como $P(x) = x$, donde x recorre el conjunto de los números primos, no existen polinomios que puedan representar únicamente valores primos.

Definición 1.1. *Diremos que un polinomio $P(x_1, \ldots, x_n)$ representa al entero a si existen enteros $b_1, \ldots, b_n$ tales que $a = P(b_1, \ldots, b_n)$.*

El resultado en una variable y con coeficientes enteros puede encontrarse en [HW79], pág. 18. Damos aquí la prueba para un polinomio con un número arbitrario de variables y coeficientes complejos.

Teorema 1.2. *Cualquier polinomio $P(x_1, \ldots, x_k)$ con coeficientes complejos que represente únicamente valores primos al evaluarse en los enteros no negativos es constante.*

Demostración. Notemos en primer lugar que los coeficientes de un polinomio

$$\begin{aligned} P : \mathbb{Z} &\longrightarrow \mathbb{Z} \\ x &\mapsto P(x) = a_0 + a_1 x^1 + \cdots + a_n x^n \end{aligned}$$

son números racionales. Esto se debe a que los coeficientes del polinomio están unívocamente determinados por $P(0), P(1), \dots, P(n)$, pues se verifica el siguiente sistema de ecuaciones en forma matricial

$$\begin{pmatrix} 1 & 0 & \dots & 0 \\ 1 & 1^1 & \dots & 1^1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & n^1 & \dots & n^n \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} P(0) \\ P(1) \\ \vdots \\ P(n) \end{pmatrix}.$$

La matriz $n \times n$ es invertible (tipo Vandermonde), y podemos resolver el sistema de manera única, obteniendo los valores de los coeficientes $a_0, \dots, a_n$ como resultado de efectuar la multiplicación entre una matriz con coeficientes racionales y un vector columna con coeficientes enteros, por lo que serán, en efecto, números racionales. Para el caso multivariable, $P : \mathbb{Z}^n \longrightarrow \mathbb{Z}$, esta afirmación también es cierta, y es consecuencia de la interpolación de Lagrange en varias variables. La idea de fondo es la misma: suficientes puntos enteros determinan los coeficientes, que serán racionales. En [San08] puede consultarse la expresión del polinomio interpolador.

Sea entonces $P(x_1, \dots, x_k)$ un polinomio con coeficientes racionales que representa únicamente números primos para todo $\mathbf{x} \in \mathbb{N}^k$. Consideremos p un número primo cualquiera, y supongamos que $P(1, \dots, 1) = p$. Sean $n_1, \dots, n_k$ enteros no negativos, y l un múltiplo común de los denominadores de los coeficientes de P. Entonces,

$$P(1 + n_1 lp, 1 + n_2 lp, \dots, 1 + n_k lp) \equiv P(1, \dots, 1) \mod p.$$

Es decir, se tiene que $P(1 + n_1 lp, 1 + n_2 lp, \dots, 1 + n_k lp)$ es divisible por p para cualesquiera enteros no negativos $n_1, \dots, n_k$. Además, sabemos que debe resultar un primo por hipótesis, luego, necesariamente, $P(1 + n_1 lp, 1 + n_2 lp, \dots, 1 + n_k lp) = p$, para cualesquiera $n_1, \dots, n_k$. Por tanto, $P(x_1, \dots, x_k)$ es un polinomio constante, pues como consecuencia del Teorema Fundamental del Álgebra, un polinomio no constante no puede representar infinitas veces el mismo valor.

En 1943, Irving Reiner probó que también se tiene este resultado negativo para funciones exponenciales en una variable (ref. [Rei43]). Veamos un resultado multivariable de este tipo, extraído de [JSWW76]:

Teorema 1.3. *Sean $P_i(x_1, \dots, x_n)$ y $Q_i(x_1, \dots, x_n)$ polinomios con coeficientes enteros tales que $P_i(x_1, \dots, x_n) \geq 0$ y $Q_i(x_1, \dots, x_n) \geq 0$ para todo $\mathbf{x} \in \mathbb{N}^n$, con $i = 1, \dots, m$. Consideremos además $a_1, \dots, a_m \in \mathbb{Z}_+$. Entonces, si la función*

$$F(x_1, \dots, x_n) = \sum_{i=1}^{m} P_i(x_1, \dots, x_n) {a_i}^{Q_i(x_1, \dots, x_n)}$$

representa únicamente valores primos al evaluarse en los enteros no negativos, es constante.

Demostración. Basta demostrar el resultado en una variable, pues si F es constante en cada variable, también lo será $F(x_1, \dots, x_n)$. Sea entonces $F(x) = \sum_{i=1}^{m} P_i(x) a_i^{Q_i(x)}$, que representa únicamente números primos para todo $x \in \mathbb{N}$. Contemplamos dos escenarios posibles: que represente finitos o infinitos primos distintos.

Supongamos que $F(x)$ representa infinitos valores primos distintos p para $x \in \mathbb{N}$. Consideremos $x_1 \in \mathbb{N}$ tal que $F(x_1) = p$, con p coprimo con a_i para todo $i = 1, \dots, m$, y sea $k \in \mathbb{Z}_+$. Si evaluamos F en $x_1 + kp(p-1) \in \mathbb{N}$, como $P_i(x_1 + kp(p-1)) \equiv P_i(x_1)$ mód p y $Q_i(x_1 + kp(p-1)) \equiv Q_i(x_1) \mod (p-1)$, tomando módulo p se tiene que

$$F(x_1 + kp(p-1)) \equiv \sum_{i=1}^{m} P_i(x_1) a_i^{Q_i(x_1)+M(p-1)} \mod p.$$

Ahora, teniendo en cuenta que, para todo $i = 1, \dots, m$, se tiene que p es coprimo con $a_i \neq 0$, por el Pequeño Teorema de Fermat se deduce que $a_i^{M(p-1)} \equiv 1 \mod p$. Luego

$$p = F(x_1) \equiv F(x_1 + kp(p-1)) \equiv \sum_{i=1}^{m} P_i(x_1) a_i^{Q_i(x_1)} \mod p.$$

Esto es, $F(x_1 + kp(p-1))$ es divisible por p, pero por hipótesis el resultado debe ser un primo, por lo que es exactamente p para todo $k \in \mathbb{Z}_+$.

Ahora, si F representa solo una cantidad finita de números primos distintos, quiere decir que al menos uno de estos primos se toma infinitas veces.

En ambos casos, para algún primo p se tiene que $F(x) = p$ para infinitos valores de $x \in \mathbb{N}$. Y esto concluye la prueba, pues como $P_i(x), Q_i(x) \geq 0$ para todo $x \in \mathbb{N}$ e $i = 1, \dots, m$, también lo será F, de lo que se deduce que su coeficiente líder es positivo. Así, $\lim_{x\to\infty} F(x) = \infty$, salvo que F sea constante e igual a p. ∎

En 1946, Robert C. Buck demostró que las funciones racionales tampoco pueden representar únicamente valores primos (véase [Buc46]). Es más, en el apartado 4 de [JSWW76] se demuestra que este resultado puede generalizarse a funciones algebraicas cualesquiera, como consecuencia del Teorema 1.2 y el hecho de que una función algebraica de $\mathbb{Z}^n \longrightarrow \mathbb{Z}$ es un polinomio.

1.4 Polinomios que representan cadenas de primos

Acabamos de probar que no existe un polinomio no constante que represente únicamente números primos. Sin embargo, existen polinomios que representan un conjunto finito de números primos distintos para valores consecutivos de la variable. Es decir, existen polinomios tales que $f(x)$ es primo para todo $0 \leq m \leq x \leq n$, siendo la diferencia $n - m$ no demasiado pequeña para que el cardinal del conjunto de números primos representados de manera consecutiva no se reduzca a un número insignificante.

El ejemplo por excelencia de un polinomio de este tipo se debe a Euler, quien en 1772 intercambió en una carta con Bernoulli lo siguiente: el polinomio $f(x) = x^2 + x + 41$ devuelve números primos para $x = 0, 1, \dots, 39$. Detrás de este conocido ejemplo, resultaron esconderse las equivalencias entre:

1. $p = 2, 3, 5, 11, 17$ o 41,
2. $f(x) = x^2 + x + p$ asume valores primos cuando $x = 0, 1, \dots, p - 2$,
3. El número de clases del cuerpo $\mathbb{Q}(\sqrt{1 - 4p})$ es $h = 1$.

La implicación (1) $\Rightarrow$ (2) fue el contenido de la carta de Euler a Bernoulli, aunque en nuestra memoria haya quedado únicamente el ejemplo más significativo. Posteriormente, en 1912, Frobenius y Rabinovitch probaron la equivalencia entre (2) y (3), y sabemos gracias a Gauss que (3) implica (1). En 1983, Le Lionnais bautizó a los números p que verifican esto como los números de la suerte de Euler.

Existen otros muchos polinomios que representan un conjunto finito de primos para valores consecutivos de la variable, pero no de mayor relevancia: muchos de ellos representan el mismo conjunto de primos que el polinomio de Euler salvo orden y repetición. A principios del siglo XXI se aportaron algunos ejemplos que representaban un número ligeramente mayor de primos, como el polinomio $f(x) = \frac{1}{4}(x^5 - 133x^4 + 6729x^3 - 158379x^2 + 1720294x - 6823316)$ dado por F. Dress y Landau, que devuelve primos distintos para $x = 0, \dots, 56$.

1.5 Polinomios enumeradores de primos

Damos en esta sección una introducción al tipo de polinomios que dan nombre al trabajo. El resultado que pretendemos desarrollar es el siguiente: el conjunto de los números primos coincide con el conjunto de los valores positivos representados por ciertos polinomios. En la literatura existente pueden encontrarse estos polinomios co-

mo *prime-representing polynomials* o *formula for primes*, pero estos conceptos engloban, generalmente, polinomios y fórmulas como las tratadas en las secciones previas. Por este motivo, hemos optado por distinguirlos, denominándolos polinomios enumeradores de primos; el sentido de esta elección se muestra en el segundo capítulo.

A continuación, presentamos los principales registros de polinomios enumeradores de primos:

- En 1971, Yuri Matijasevič indicó la existencia de un polinomio enumerador de primos de grado 37 en 24 variables, haciendo uso de la sucesión de Fibonacci (véase [Mat71]). En el apéndice de la traducción al inglés del artículo, se proporciona una mejora del mismo, quedando reducido a grado 21 en 21 variables.
- El primer polinomio enumerador de primos escrito explícitamente se encuentra en [JSWW76], artículo publicado en 1976 por James P. Jones, Daihachiro Sato, Hideo Wada y Douglas Wiens. Se trata de un polinomio de grado 25 en 26 variables, que desarrollaremos detenidamente en el capítulo 3. En este mismo artículo se demuestra además la existencia de un polinomio enumerador de primos en 12 variables, y afirman que puede obtenerse un polinomio de grado 5 en 42 variables; daremos más detalles sobre esto en el capítulo 5.
- En 1981, Yuri Matijasevič demostró en [Mat81] la existencia de un polinomio enumerador de primos en 10 variables.

A partir de estos polinomios pueden obtenerse algunos otros con una ligera reducción en el número de variables mediante sustituciones en el sistema de ecuaciones que determinarán estos polinomios, o una reducción en el grado gracias a la posibilidad de reemplazar algunos términos o expresiones usadas en su construcción por otras.

Existen también polinomios de esta índole para ciertos subconjuntos de primos, como los primos de Mersenne o los de Fermat; véase [Jon79]. La idea es la misma que presentamos en esta memoria para el conjunto completo de números primos.

En el próximo capítulo veremos en qué contexto emergieron los polinomios enumeradores de primos y el porqué de su existencia, cuya demostración proporciona además el paso final en la construcción de los mismos.

2 | En el principio era Hilbert

Hilbert desafió a la comunidad matemática con una lista de 23 problemas que establecerían el curso de las matemáticas durante el siglo XX. Ya en el primer capítulo hemos hecho referencia en varias ocasiones al octavo problema de esta lista, que a día de hoy sigue en pie: la Hipótesis de Riemann. Sin embargo, no es este el problema que nos ocupa en este trabajo, aunque a primera vista es el que tiene una consecuencia directa sobre la distribución de los números primos. Una lectura detallada sobre la descripción de estos problemas tal y como los formuló Hilbert se encuentra en [Hil02].

2.1 El décimo problema de Hilbert

Centrémonos en el décimo problema de esta lista, y veamos qué tiene que ver con los números primos y por qué fue el inicio de toda la teoría desarrollada en esta memoria para los polinomios enumeradores de primos. Dice así:

> 10. DETERMINATION OF THE SOLVABILITY OF A DIOPHANTINE EQUATION.
>
> Given a Diophantine equation with any number of unknown quantities and with rational integral numerical coefficients: To devise a process according to which it can be determined by a finite number of operations whether the equation is solvable in rational integers.

En otras palabras, el problema consiste en encontrar un algoritmo que, dada una ecuación diofántica cualquiera, determine si existe solución de dicha ecuación. Comencemos la discusión del problema estableciendo qué es una ecuación diofántica:

Definición 2.1 (Ecuación diofántica). *Una ecuación diofántica es una ecuación* $P(x_1, \dots, x_n) = 0$*, donde* $P(x_1, \dots, x_n)$ *es un polinomio con coeficientes enteros y varia-*

bles enteras.

El problema habitual cuando se trata con ecuaciones diofánticas consiste en: dada una ecuación de este tipo, encontrar el conjunto de sus soluciones. Es necesario para nuestros propósitos lidiar aquí con el problema inverso: dado el conjunto de sus "soluciones", encontrar la ecuación diofántica correspondiente. Más precisamente, vamos a añadirle un parámetro a la ecuación diofántica, y el conjunto con el que tratamos será el conjunto de los valores positivos de ese parámetro para los que dicha ecuación tiene solución.

Definición 2.2 (Conjunto diofántico). *Un conjunto D de enteros positivos se dice que es diofántico si existe un polinomio $P(x_1, \dots, x_n, y)$ tal que la ecuación diofántica $P(x_1, \dots, x_n, y) = 0$ tiene solución en $x_1, \dots, x_n \in \mathbb{Z}_+$ si y solo si $y \in D$, es decir,*

$$D = \{y \mid \exists(x_1, \dots, x_n)(P(x_1, \dots, x_n, y) = 0)\}. \tag{2.1}$$

La restricción de los valores del parámetro y y de las variables $x_1, \dots, x_n$ a los enteros positivos puede eliminarse con una simple reescritura, como veremos más adelante.

Veamos un par de ejemplos que nos servirán de motivación:

Ejemplos 2.1.

1. Los números que no son potencias de dos.
   Consideremos el polinomio $P(x_1, x_2, y) = x_1(2x_2 + 1) - y$. Para un valor fijo del parámetro $y \in \mathbb{Z}_+$, la ecuación $y = x_1(2x_2 + 1)$ tiene solución si y solo si y no es una potencia de dos. Los números que no son potencias de dos forman entonces un conjunto diofántico
   $$D = \{y \mid (\exists x_1, x_2)(x_1(2x_2 + 1) - y = 0)\}.$$

2. Los números compuestos.
   Consideremos ahora el polinomio $P(x_1, x_2, y) = (x_1 + 1)(x_2 + 1) - y$. Para un valor fijo del parámetro $y \in \mathbb{Z}_+$, la ecuación $y = (x_1+1)(x_2+1)$ tiene solución si y solo si y es un número compuesto. Los números compuestos forman entonces un conjunto diofántico
   $$D = \{y \mid (\exists x_1, x_2)((x_1 + 1)(x_2 + 1) - y = 0)\}.$$

A la vista de estos dos ejemplos, parece natural preguntarse si las potencias de dos o los números primos también son conjuntos diofánticos. La respuesta en ambos casos es afirmativa, aunque no resulta tan sencillo encontrar un polinomio en las condiciones de la definición como en los dos ejemplos anteriores. Es el motivo de este trabajo demostrar el carácter diofántico del conjunto de los números primos. En la siguiente sección nos detendremos a construir explícitamente un polinomio que verifica (2.1), para el que D será precisamente el conjunto de los números primos.

Ahora bien, ¿qué relación tiene nuestro propósito de demostrar el carácter diofántico del conjunto de los números primos con el décimo problema de Hilbert? Vamos a recorrer brevemente la historia del problema, mencionando los resultados principales que nos atañen. Para ello, necesitamos conceptos básicos de teoría de la computación, que expondremos conforme nos resulten necesarios. Puede consultarse el tercer capítulo del manual autocontenido [DSW94], que recoge en profundidad estos conceptos.

En 1953, Davis probó varios resultados (véase [Dav53]) que le llevaron a conjeturar lo siguiente:

«Todo conjunto recursivamente enumerable es diofántico».

Para tratar de comprender este enunciado, necesitamos establecer qué es un conjunto recursivamente enumerable:

Definición 2.3 (Conjunto recursivamente enumerable). *Un conjunto $S \subset \mathbb{N}^n$ se dice recursivamente enumerable (r.e.) si existe un algoritmo tal que, para cualquier $\mathbf{x} \in \mathbb{N}^n$, devuelva el valor 0 en tiempo finito si y solo si $\mathbf{x} \in S$. A estos conjuntos también se les conoce como semidecidibles.*

Dicho de otro modo que motiva el término, un conjunto es r.e. si existe una función recursiva que lo enumera, esto es, el conjunto coincide con el rango de dicha función. Un subconjunto propio de estos son los llamados conjuntos recursivos.

Definición 2.4 (Conjunto recursivo). *Un conjunto $S \subset \mathbb{N}^n$ se dice recursivo si existe un algoritmo que, para cualquier $\mathbf{x} \in \mathbb{N}^n$ devuelva, en tiempo finito, el valor 0 si $\mathbf{x} \in S$ y 1 si $\mathbf{x} \notin S$. A estos conjuntos también se les conoce como decidibles.*

Ahora que disponemos de esta definición, notemos que si el décimo problema de Hilbert tuviera solución, todo conjunto diofántico sería recursivo. Pero existen conjuntos recursivamente enumerables que no son recursivos (consúltese, por ejemplo, el notorio problema de la parada), luego la conjetura de Davis implicaría que existen conjun-

tos diofánticos que no son recursivos, lo que conduce a la indecibilidad del problema de Hilbert.

En [G̈31], Kurt Gödel trata la noción de recursividad, y demuestra que una amplia lista de funciones son recursivas, construyéndolas a partir de otros ejemplos más elementales. Entre ellos, encontramos la definición recursiva de que un número sea primo, que no es más que una reescritura en términos lógicos de la definición habitual (y errónea, aunque equivalente en $\mathbb{Z}$) de un número primo:

$$p \in \mathbb{N} \text{ es primo si } p > 1 \text{ y sus únicos divisores naturales son 1 y } p.$$

Este enunciado está compuesto de otros predicados más elementales que son, a su vez, recursivos. Así, como el conjunto de los números primos es r.e., de verificarse la conjetura se tendría que dicho conjunto es diofántico. Esto es, podríamos afirmar que existe una ecuación diofántica

$$P(p, z_1, \dots, z_n) = 0 \tag{2.2}$$

que tiene solución en $z_1, \dots, z_n$ si y solo si p es un número primo.

En 1960, Hilary Putnam consiguió acercar este problema a un lenguaje más algebraico, invitando así a expertos en teoría de números a adentrarse en él. Putnam demostró en [Put60] que la existencia de solución de la ecuación (2.2) era equivalente a la de la ecuación

$$p = z_0(1 - P'^2(z_0, \dots, z_n)). \tag{2.3}$$

Trataremos este resultado en la siguiente sección de este capítulo. Con esto, se sigue de la conjetura de Davis que el conjunto de los números primos coincide con el conjunto de enteros positivos representados por un polinomio. Es a estos polinomios a los que hemos nombrado polinomios enumeradores de primos, en consonancia con el carácter recursivamente enumerable del conjunto.

En 1969, Julia Robinson observó en [Rob69] que la demostración del caso particular de la conjetura de Davis para los números primos implicaría la veracidad de la conjetura en su totalidad. Esto es, el problema de probar que todo conjunto r.e. es diofántico quedó reducido a la existencia de un polinomio enumerador de primos.

Martin Davis, Julia Robinson y Hilary Putnam siguieron trabajando en esta línea, desarrollando nuevos resultados (véase [DPR61]) que ayudaron a la definitiva resolución del problema. Fue Yuri Matijasevič quien, en 1970, reunió todos estos resultados de esfuerzo colectivo y demostró la conjetura de Davis en [Mat70], hoy día conocida como

el Teorema de Matijasevič o Teorema DPRM, por las iniciales de los principales contribuidores. Un resumen detallado del devenir del resultado final puede encontrarse en el apéndice histórico de [Dav73].

2.2 Método de Putnam

Exponemos y demostramos aquí el resultado de Putnam sobre la equivalencia entre (2.2) y (2.3). Damos previamente un resultado que necesitaremos en la demostración del teorema que recoge el método. Este se conoce como la conjetura de Bachet o el Teorema de los cuatro cuadrados, probado en 1770 por Lagrange. Es además un caso particular (caso cuadrado) del Teorema del número poligonal de Fermat:

Teorema 2.1 (Teorema de los cuatro cuadrados). *Todo entero no negativo n puede escribirse como suma de cuatro cuadrados, esto es, existen enteros no negativos a, b, c, d tales que*

$$n = a^2 + b^2 + c^2 + d^2.$$

La demostración de Lagrange se encuentra en el tercer tomo de sus reconocidas *Oeuvres*, págs. 189-201. Una prueba alternativa puede verse en [HW79], pág. 302.

Teorema 2.2 (Putnam, 1960). *Todo conjunto diofántico de enteros positivos coincide con el conjunto de enteros positivos representados por un polinomio.*

Demostración. Sea D un conjunto diofántico cualquiera, definido por el polinomio $P(x_1, \dots, x_n, y)$ como en (2.1). La restricción de las variables $x_1, \dots, x_n$ sobre los enteros positivos puede ser eliminada con una simple reescritura usando el Teorema 2.1, pues como todo entero no negativo es la suma de cuatro cuadrados, todo entero positivo puede ser escrito como $x_1 = x_{11}^2 + x_{12}^2 + x_{13}^2 + x_{14}^2 + 1$. Se tiene entonces que la expresión (2.1) es equivalente a esta otra

$$D = \{y > 0 \mid (\exists x_{11}, x_{12}, x_{13}, x_{14}, \dots, x_{n1}, x_{n2}, x_{n3}, x_{n4})(P(x_{11}^2 + x_{12}^2 + x_{13}^2 + x_{14}^2 + 1, \dots, x_{n1}^2 + x_{n2}^2 + x_{n3}^2 + x_{n4}^2 + 1) = 0)\}.$$

Esto es, podemos escribirlo como $D = \{y > 0 \mid (\exists z_1, \dots, z_m)(P'(z_1, \dots, z_m, y) = 0)\}$, donde $m = 4n$ y $z_1, \dots, z_m \in \mathbb{Z}$. Ahora, como $y > 0$, también podemos reescribirla del mismo modo:

$$\begin{cases} y = u_1^2 + u_2^2 + u_3^2 + u_4^2 + 1, \\ P'(z_1, \dots, z_m, y) = P'(z_1, \dots, z_m, u_1^2 + u_2^2 + u_3^2 + u_4^2 + 1) = 0. \end{cases} \tag{2.4}$$

Este sistema es equivalente a la siguiente ecuación:

$$y = (1 - P'^2(z_1, \dots, z_m, u_1^2 + u_2^2 + u_3^2 + u_4^2 + 1))(u_1^2 + u_2^2 + u_3^2 + u_4^2 + 1),\ y \in \mathbb{Z}_+. \quad (2.5)$$

La equivalencia entre (2.4) y (2.5) es sencilla:

$\boxed{(2.4) \Rightarrow (2.5)}$ $P' = 0 \Rightarrow P'^2 = 0$. Reescribiendo la expresión de y, tenemos

$$y = (1 - 0)(u_1^2 + u_2^2 + u_3^2 + u_4^2 + 1) = (1 - P'^2)(u_1^2 + u_2^2 + u_3^2 + u_4^2 + 1) \in \mathbb{Z}_+.$$

$\boxed{(2.5) \Rightarrow (2.4)}$ Recíprocamente, como y es producto de dos factores (y el segundo es siempre positivo), y será positivo si y solo si $1 - P'^2 > 0$, y esto se tiene si y solo si $P'^2 = 0$. Entonces, la expresión de y en (2.5) resulta

$$y = (1 - 0)(u_1^2 + u_2^2 + u_3^2 + u_4^2 + 1) = u_1^2 + u_2^2 + u_3^2 + u_4^2 + 1,$$

y como $P'^2 = 0$, también tendremos $P' = 0$, luego se da (2.4).

Por último, si consideramos

$$Q := (1 - P'^2)(u_1^2 + u_2^2 + u_3^2 + u_4^2 + 1),$$

podemos escribir nuestro conjunto diofántico D como sigue:

$$D = \{y > 0 \mid (\exists z_1, \dots, z_m, u_1, u_2, u_3, u_4)(Q(z_1, \dots, z_m, u_1, u_2, u_3, u_4) = y)\}.$$

Luego el conjunto D de enteros positivos coincide precisamente con el conjunto de enteros positivos representados por el polinomio Q, como queríamos demostrar. ∎

Ejemplo 2.1. Veamos una aplicación de este resultado. Demostramos el Teorema 1.1, es decir, que existe un polinomio Q que representa al primo n–ésimo. Una definición recursiva del n–ésimo primo se encuentra en la lista de [Gö31], compuesta de nuevo por otras funciones recursivas más elementales y por la de que un número sea primo, que ya vimos que también es recursiva. Por tanto, podremos expresarlo en términos diofánticos: sabemos que existe un polinomio P tal que

$$p_n = m \iff (\exists x_1, \dots, x_l)(P(n, m, x_1, \dots, x_l) = 0).$$

Siguiendo la demostración del resultado anterior, basta tomar

$$Q = x_{l+1}(1 - P^2(n, x_{l+1}, x_1, \dots, x_l)).$$

En el capítulo 5 vemos que puede tomarse $l = 13$ y que, en definitiva, el n–ésimo primo puede definirse como un polinomio en 14 variables (en el sentido de la definición diofántica).

3 Construcción explícita de un polinomio enumerador de primos

El propósito de este capítulo reside en la construcción explícita de un polinomio enumerador de primos. En particular, desarrollaremos con detalle todo el material necesario para obtener el polinomio enumerador de primos de grado 25 en 26 variables dado por Jones et al. en [JSWW76]. Más concretamente, el capítulo convergerá en el siguiente teorema:

Teorema 3.1. *El conjunto de los números primos coincide con los valores positivos representados por el polinomio*

$$\begin{aligned}P = (k+2)\{&1-(wz+h+j-q)^2-[(gk+2g+k+1)(h+j)+h-z]^2-(2n+p+q+z-e)^2\\&-[16(k+1)^3(k+2)(n+1)^2+1-f^2]^2-[e^3(e+2)(a+1)^2+1-o^2]^2-[(a^2-1)y^2+1-x^2]^2\\&-[16r^2y^4(a^2-1)+1-u^2]^2-[((a+u^2(u^2-a))^2-1)(n+4dy)^2+1-(x+cu)^2]^2-(n+l+v-y)^2\\&-[(a^2-1)l^2+1-m^2]^2-(ai+k+1-l-i)^2-[p+l(a-n-1)+b(2an+2a-n^2-2n-2)-m]^2\\&-[q+y(a-p-1)+s(2ap+2a-p^2-2p-2)-x]^2-[z+pl(a-p)+t(2ap-p^2-1)-pm]^2\},\end{aligned}$$

donde las variables recorren los enteros no negativos.

Si observamos el polinomio P, es fácil ver que viene dado como producto de dos factores. Escribámoslo de la siguiente forma:

$$P = (k+2)\{1-M\},$$

donde M es una suma de cuadrados (y, por tanto, no negativo). Comparando esto con la expresión (2.3), nos damos cuenta de que esto no es más que una construcción siguiendo el método de Putnam. Dado M, es inmediato construir el polinomio, pero es precisamente en este punto donde radica la dificultad, pues la construcción de M

requiere de una gran parte de las técnicas inventadas para la resolución del décimo problema de Hilbert, y ya hemos visto en el segundo capítulo que esto no es mera casualidad: el Teorema 3.1 implica, de hecho, el carácter indecidible del problema.

En lo sucesivo, la tarea que nos ocupará será construir un polinomio M que verifique lo siguiente:

Proposición 3.1. Para cada entero no negativo k, se tiene que

$$k+2 \text{ es primo si y solo si } M = 0 \text{ tiene solución en los enteros no negativos.}$$

Un tal M que verifique esta proposición será precisamente una suma de cuadrados, y obtenemos el polinomio enumerador de primos P del Teorema 3.1 mediante el método de Putnam, estableciendo $P = (k+2)\{1 - M\}$, como hicimos notar previamente en la estructura del polinomio. Dicho esto, la aparente paradoja de que el polinomio factorice se diluye, pues cuando $k+2$ sea primo se tendrá que uno de los factores es exactamente 1.

Antes de pasar a la construcción de M, es importante recordar que, como vimos en el primer capítulo del trabajo, no existen polinomios que representen únicamente números primos. En particular, el polinomio P toma también, además de números primos (que no aparecen en orden y que pueden repetirse), valores negativos. Para subsanar esta inexactitud en la representación polinomial, Julia Robinson observó, en un contexto más general pero perfectamente aplicable a nuestro caso (véase [DPR61], Corolario 7), que se podía hacer uso de la función 0^x, estableciendo previamente que $0^0 = 1$. Así, tomando M en las condiciones de la Proposición 3.1, se puede probar que

Teorema 3.2. *El conjunto de los números primos coincide exactamente con el rango de una función de la forma*

$$2 + k \cdot 0^{M(k,x_1,\ldots,x_n)}.$$

Ahora sí, vamos a dotarnos de la artillería necesaria para construir el polinomio M.

3.1 Ecuación de Pell

En el sistema de ecuaciones diofánticas que definirán el conjunto diofántico que buscamos encontraremos varias ecuaciones con una estructura idéntica y características comunes, y que reciben el nombre de ecuaciones de Pell.

Definición 3.1 (Ecuación de Pell). *Una ecuación de Pell es una ecuación diofántica de la forma*

$$x^2 - dy^2 = 1,$$

donde $d \in \mathbb{Z}$ es un parámetro previamente determinado y x e y son variables enteras.

En esta memoria, supondremos que tanto el parámetro d como las variables x e y son números naturales. El estudio de las soluciones de esta ecuación depende en su totalidad del parámetro d. Distinguimos varios casos:

- Si $d = 0$, las soluciones vienen dadas por $(1, y)$, donde $y \in \mathbb{N}$.
- Supongamos que $d > 0$ y es un cuadrado perfecto. Veamos que la única solución es la solución trivial $(1, 0)$. Consideremos $d = c^2$, con $c \in \mathbb{Z} \setminus \{0\}$. Así, la ecuación resulta $x^2 - dy^2 = x^2 - (cy)^2 = 1$, y los únicos cuadrados que se diferencian en una unidad son el 0 y el 1. Por tanto, $x^2 = 1$ y $c^2y^2 = dy^2 = 0$. El resultado se sigue de que $x, y \in \mathbb{N}$ y $d \neq 0$.
- Por el contrario, supongamos ahora que $d > 0$, no cuadrado. Lagrange demostró, en 1768, que la ecuación $x^2 - dy^2 = 1$ tiene una solución no trivial. Veremos en las próximas páginas que esto implica que, cuando d no es un cuadrado perfecto, esta ecuación posee infinitas soluciones.

Nos centraremos en estudiar las soluciones de la siguiente ecuación de Pell:

$$\begin{cases} x^2 - dy^2 = 1, & x, y \geq 0. \\ d = a^2 - 1, & a > 1. \end{cases} \tag{3.1}$$

Como $d = a^2 - 1$ donde $a > 1$, d no es un cuadrado, por lo que según la discusión del párrafo anterior esta ecuación tendrá infinitas soluciones. Notemos que

$$x^2 - dy^2 = (x + \sqrt{d}y)(x - \sqrt{d}y) = 1.$$

Podemos ver $x + \sqrt{d}y$ como un elemento del cuerpo cuadrático $\mathbb{Q}(\sqrt{d})$. Así, las soluciones no negativas (x, y) de (3.1) se corresponden con las unidades de norma $N(x + \sqrt{d}y) = x^2 - dy^2 = 1$ de su anillo de enteros, que como $d = a^2 - 1 \not\equiv 1$ mód 4, es $\mathbb{Z}[\sqrt{d}]$. Existe una amplia literatura sobre las ecuaciones de Pell y las diferentes formas de obtener sus soluciones; nos detendremos primero en demostrar que las soluciones vienen generadas por ciertas ecuaciones recursivas, denominadas sucesiones de Lucas.

Definición 3.2 (Sucesión de Lucas). *Se denomina sucesión de Lucas a cualquier sucesión de enteros que satisface la siguiente relación de recurrencia*

$$X_0 = A, \quad X_1 = B, \quad X_{n+2} = CX_{n+1} + DX_n, \quad A, B, C, D \in \mathbb{Z}.$$

Si denotamos por $x = \chi_a(n)$ e $y = \psi_a(n)$ a las soluciones de la ecuación (3.1), ordenadas según la magnitud de y, estas resultan venir dadas por las siguientes sucesiones crecientes:

$$\begin{aligned} &\chi_a(0) = 1, \quad \chi_a(1) = a, \quad \chi_a(n+2) = 2a\chi_a(n+1) - \chi_a(n), \\ &\psi_a(0) = 0, \quad \psi_a(1) = 1, \quad \psi_a(n+2) = 2a\psi_a(n+1) - \psi_a(n). \end{aligned} \tag{3.2}$$

Notemos que estas sucesiones son también válidas para $a = 1$, en cuyo caso $d = 0$ y el conjunto de soluciones venía dado por $(1, n) = (\chi_1(n), \psi_1(n))$, para todo $n \in \mathbb{N}$. Veamos que, en efecto, que este par de sucesiones caracterizan las soluciones de la ecuación de Pell (3.1). Llegaremos a ellas a través de una serie de lemas, todos ellos extraídos de [Dav73].

En primer lugar, es inmediato ver que $(1, 0)$ y $(a, 1)$ son soluciones de la ecuación (3.1), las cuales son correspondientes con $1 + \sqrt{d} \cdot 0$ y $a + \sqrt{d} \cdot 1$, respectivamente. Obsérvese que estas soluciones están en orden creciente de y, y veamos que no existe otro par (x, y) solución entre ellas.

Lema 3.1. No existe un par de enteros (x, y) tales que $1 < x + \sqrt{d}y < a + \sqrt{d}$ que satisfagan la ecuación de Pell (3.1).

Demostración. Sean $x, y \in \mathbb{Z}$ verificando (3.1). Entonces,

$$\begin{aligned} x^2 - dy^2 &= (x + \sqrt{d}y)(x - \sqrt{d}y) = 1, \\ d = a^2 - 1 &\Rightarrow (a + \sqrt{d})(a - \sqrt{d}) = 1. \end{aligned}$$

Multiplicando por -1 en ambas ecuaciones, obtenemos

$$\begin{aligned} (x + \sqrt{d}y)(-x + \sqrt{d}y) &= -1, \\ (a + \sqrt{d})(-a + \sqrt{d}) &= -1. \end{aligned}$$

Supongamos ahora que se verifica la desigualdad del enunciado. Entonces, también debe verificarse $-1 < -x + \sqrt{d}y < -a + \sqrt{d}$. Sumando ambas desigualdades, se tiene

que $0 < 2\sqrt{d}y < 2\sqrt{d}$, lo cual implicaría que $0 < y < 1$, pero esto es una contradicción: hemos supuesto que y es solución de la ecuación de Pell (esto es, $y \in \mathbb{Z}, y \geq 0$). Luego no existen soluciones de la ecuación (3.1) que satisfagan la desigualdad del enunciado.

Lema 3.2. Sean (x, y) y (x', y') dos soluciones de la ecuación de Pell (3.1), y consideremos x'' e y'' tales que

$$x'' + \sqrt{d}y'' = (x + \sqrt{d}y)(x' + \sqrt{d}y'). \tag{3.3}$$

Entonces, el par (x'', y'') es también solución de la ecuación.

Demostración. Tomando conjugado en la ecuación (3.3), tenemos que

$$x'' - \sqrt{d}y'' = (x - \sqrt{d}y)(x' - \sqrt{d}y').$$

Multiplicando ambas ecuaciones, concluimos que (x'', y'') es solución de la ecuación de Pell, pues del hecho de que (x, y) y (x', y') sean soluciones se deduce que

$$(x'')^2 - d(y'')^2 = (x^2 - dy^2)((x')^2 - d(y')^2) = 1.$$

Este lema establece que podemos obtener nuevas soluciones de la ecuación de Pell a partir de otras, lo que motiva la siguiente definición-proposición:

Definición 3.3. *Denotemos por $\chi_a(n)$ y $\psi_a(n)$ para $n \geq 0$ al par que verifica*

$$\chi_a(n) + \psi_a(n)\sqrt{d} = (a + \sqrt{d})^n. \tag{3.4}$$

El par $(\chi_a(n), \psi_a(n))$ es solución de la ecuación de Pell (3.1)*, y toda solución es de esta forma.*

Por tanto, la ecuación (3.4) genera, de manera algebraica, todas las soluciones de la ecuación de Pell considerada. Veamos primero que el par $(\chi_a(n), \psi_a(n))$ así definido es solución de la ecuación para todo $n \geq 0$:

- Para $n = 0$ se obtiene la solución trivial: $(\chi_a(0), \psi_a(0)) = (1, 0)$.
- Para $n = 1$, se tiene $(\chi_a(1), \psi_a(1)) = (a, 1)$. Basta sustituir en la ecuación de Pell (3.1) para comprobar que el par verifica la ecuación. A esta solución se le conoce como solución fundamental.

Además, sabemos que no existen soluciones a la ecuación que se encuentren entre las dadas para $n = 0$ y $n = 1$ gracias al Lema 3.1.

Con objeto de no sobrecargar la notación, omitiremos la dependencia de a cuando no sea necesario. Denotaremos $(x_n, y_n) := (\chi_a(n), \psi_a(n))$.

Lema 3.3. El par (x_n, y_n) es solución de la ecuación de Pell (3.1) para todo $n \geq 0$.

Demostración. Los casos $n = 0$ y $n = 1$ los hemos visto hace apenas unas líneas. Procedamos por inducción en n. Supongámoslo cierto para n y veamos que se tiene para $n + 1$:

$$x_{n+1} + \sqrt{d}\, y_{n+1} = (a + \sqrt{d})^{n+1} = (a + \sqrt{d})^n (a + \sqrt{d}) = (x_n + \sqrt{d}\, y_n)^n (x_1 + \sqrt{d}\, y_1).$$

Como hemos comprobado que $(x_1, y_1) = (a, 1)$ es solución de la ecuación de Pell, y (x_n, y_n) también lo es por hipótesis de inducción, podemos hacer uso del Lema 3.2 para concluir que (x_{n+1}, y_{n+1}) también verifica la ecuación. ∎

Veamos ahora que todas las soluciones vienen dadas de esta forma: cualquier par (x, y) solución de la ecuación coincide con (x_n, y_n) para algún n.

Lema 3.4. Sea (x, y) un par que verifica la ecuación de Pell. Entonces, para algún n se tiene que $x = x_n$ e $y = y_n$.

Demostración. Como $x + \sqrt{d}\, y \geq 1$, para algún $n \geq 0$ tendremos que

$$(a + \sqrt{d})^n \leq x + \sqrt{d}\, y < (a + \sqrt{d})^{n+1}.$$

Si se da la igualdad, hemos terminado. Por reducción al absurdo, supongamos que no. Entonces, como $(a + \sqrt{d}) = x_n + \sqrt{d}\, y_n$,

$$x_n + \sqrt{d}\, y_n < x + \sqrt{d}\, y < (x_n + \sqrt{d}\, y_n)(a + \sqrt{d}),$$

donde (x_n, y_n) es solución de la ecuación de Pell. Si multiplicamos por $x_n - \sqrt{d}\, y_n > 0$, como $(x_n + \sqrt{d}\, y_n)(x_n - \sqrt{d}\, y_n) = 1$, se tiene que

$$1 < (x + \sqrt{d}\, y)(x_n - \sqrt{d}\, y_n) < a + \sqrt{d},$$

pero esto contradice los Lemas 3.1 y 3.2. ∎

Así, cobra ahora sentido el llamar al par $(a, 1)$, solución del caso $n = 1$, solución fundamental, pues lo que tenemos en (3.4) es una ecuación que genera, a partir de esta, el resto de soluciones. Para determinar el par (x_n, y_n) para un cierto n fijo dada la solución fundamental, bastará expandir el binomio y extraer dos ecuaciones: una para la parte racional y otra para la irracional.

Observación 3.1. Notemos una vía alternativa para la obtención de la caracterización algebraica (3.4) de las soluciones a la ecuación. Si vemos las soluciones como los elementos $x + \sqrt{d}y$ invertibles de norma uno de $\mathbb{Z}[\sqrt{d}]$, como comentamos anteriormente, resulta que estos forman un grupo cíclico, generado por lo que se conoce como unidad fundamental - nótese la analogía -, que es exactamente $a + \sqrt{d}$. Y sabemos que, en un grupo cíclico, podemos obtener el resto de elementos del grupo a partir de su generador, exactamente como en (3.4). Sin embargo, esta vía requiere de teoría mucho menos elemental como la que hemos tratado aquí, con intención de que resultase autocontenido. Puede encontrarse el desarrollo de esta idea en [Zyk].

Para motivar el próximo lema, notemos la existencia de una similitud formal entre la relación $x_n + \sqrt{d}y_n = (a + \sqrt{d})^n$ y la conocida identidad de Euler:

$$\cos(\theta) + \sqrt{-1}\,\mathrm{sen}(\theta) = e^{i\theta} = (e^{i\cdot 1})^\theta = (\cos(1) + \sqrt{-1}\,\mathrm{sen}(1))^\theta.$$

Basta tomar $d = -1$ y reemplazar coseno y seno por x_n e y_n, respectivamente. Así:

$$x_n + \sqrt{d}y_n = (x_1 + \sqrt{d}y_1)^n = (a + \sqrt{d})^n.$$

Con estas consideraciones, la ecuación de Pell $x_n^2 - dy_n^2 = 1$ también tiene una estructura reconocible: la de la identidad pitagórica $\cos^2(\theta) + \mathrm{sen}^2(\theta) = 1$.

Veamos que x_n e y_n verifican propiedades aditivas análogas a las de estas funciones trigonométricas, a partir de las cuales obtendremos las soluciones de la ecuación de Pell (3.1) expresadas como sucesiones de Lucas:

Lema 3.5. Se verifica

1. $x_{m\pm n} = x_m x_n \pm d y_n y_m$,
2. $y_{m\pm n} = x_n y_m \pm x_m y_n$.

Demostración. Desarrollemos primero los cálculos para x_{m+n} e y_{m+n}:

$$\begin{aligned} x_{m+n} + \sqrt{d}y_{m+n} &= (a + \sqrt{d})^{m+n} = (a + \sqrt{d})^m (a + \sqrt{d})^n = (x_m + \sqrt{d}y_m)(x_n + \sqrt{d}y_n) \\ &= (x_m x_n + d y_n y_m) + \sqrt{d}(x_n y_m + x_m y_n), \end{aligned}$$

luego $x_{m+n} = x_m x_n + d y_n y_m$ e $y_{m+n} = x_n y_m + x_m y_n$. Obtengamos ahora las fórmulas para la resta:

$$x_{m-n} + \sqrt{d} y_{m-n} = (a + \sqrt{d})^{m-n} = (a + \sqrt{d})^m (a + \sqrt{d})^{-n} = (x_m + \sqrt{d} y_m)(x_n + \sqrt{d} y_n)^{-1}$$
$$= (x_m + \sqrt{d} y_m)(x_n - \sqrt{d} y_n) = (x_m x_n - d y_n y_m) + \sqrt{d}(x_n y_m - x_m y_n),$$

luego $x_{m-n} = x_m x_n - d y_n y_m$ e $y_{m-n} = x_n y_m - x_m y_n$.

Si tomamos $n = 1$ en el lema anterior y recuperamos la notación inicial, se obtiene

$$\chi_a(m \pm 1) = a\chi_a(m) \pm d\psi_a(m), \text{ y} \tag{3.5}$$
$$\psi_a(m \pm 1) = a\psi_a(m) \pm \chi_a(m). \tag{3.6}$$

Por último, sumando

$$\chi_a(n+1) + \chi_a(n-1) = 2a\chi_a(n), \text{ y}$$
$$\psi_a(n+1) + \psi_a(n-1) = 2a\psi_a(n).$$

Reordenando los términos llegamos a las sucesiones que buscábamos:

Lema 3.6. Las soluciones de la ecuación de Pell (3.1) expresadas como sucesiones de Lucas vienen dadas por

1. $\chi_a(n+1) = 2a\chi_a(n) - \chi_a(n-1)$,
2. $\psi_a(n+1) = 2a\psi_a(n) - \psi_a(n-1)$.

Basta considerar $n = n + 1$ para tener las expresiones tal y como las presentamos al inicio de este capítulo en (3.2).

3.1.1 Propiedades de las soluciones de la ecuación de Pell

Resulta inmediato escribir el conjunto de soluciones de $x^2 - dy^2 = 1$, fijado d, como un conjunto diofántico. Sea $D := \{y \mid (\exists x)(x^2 - dy^2 = 1)\}$. Lo que necesitaremos para demostrar el carácter diofántico de los números primos es una descripción diofántica del $n-$ésimo elemento del conjunto D, es decir, del grafo (n, y_n). Para ello, necesitaremos una gran cantidad de propiedades de las soluciones de la ecuación de Pell, que exponemos en la siguiente sucesión de lemas:

Lema 3.7 (*Congruence rule*). Sea $a \equiv b \mod c$. Para todo $n \geq 0$, se tiene que

$$\chi_a(n) \equiv \chi_b(n) \mod c, \text{ y}$$
$$\psi_a(n) \equiv \psi_b(n) \mod c.$$

Demostración. Desarrollemos la prueba para $\psi(n)$. Para $n = 0, 1$ se da la igualdad. Procediendo por inducción, supongamos que se verifica para $n-1$ y para n, y veamos que también se tiene para $n+1$:

$$\psi_a(n+1) = 2a\psi_a(n) - \psi_a(n-1) \equiv 2b\psi_b(n) - \psi_b(n-1) = \psi_b(n+1) \mod c.$$

De manera completamente análoga, se tiene el mismo resultado para $\chi(n)$, a excepción de que, para $n = 1$, no se da la igualdad sino la congruencia hipótesis, $a \equiv b \mod c$.

Un caso particular de interés se obtiene para $b = 1$ y $c = a - 1$:

Corolario 3.1. $\psi_a(n) \equiv n \mod a - 1$.

Como es habitual por sus siglas en inglés, denotaremos por $\gcd(a, b)$ al máximo común divisor entre a y b.

Lema 3.8. $\gcd(x_n, y_n) = 1$.

Demostración. Sea $t > 0$ que divide a x_n y a y_n. Entonces, también debe dividir a $x_n^2 - dy_n^2 = 1$. Luego $t = 1$.

Lema 3.9. $y_n \mid y_{nk}$.

Demostración. Para $k = 1$ es evidente. Procedamos por inducción: supongamos que se verifica para m y veamos que se tiene para $m+1$. En efecto, haciendo uso del Lema 3.5,

$$y_{n(m+1)} = y_{nm+n} = x_n y_{nm} + x_{nm} y_n.$$

Ahora, por hipótesis de inducción, como $y_n \mid y_{nm}$, dividirá a los dos sumandos de la expresión (pues también se divide a sí mismo) y por tanto se tendrá el resultado para $m+1$.

Lema 3.10. $y_n \mid y_t$ si y solo si $n \mid t$.

Demostración. Supongamos primero que $n \mid t$, es decir, $t = nk$ para cierto k. Por el lema anterior, $y_n \mid y_t$.

Recíprocamente, procedamos por reducción al absurdo y supongamos que $y_n \mid y_t$ pero $n \nmid t$, es decir, $t = nq + r$, donde $0 < r < n$. Aplicando el Lema 3.5,

$$y_t = y_{nq+r} = y_r y_{nq} + x_{nq} y_r.$$

Por el lema anterior $y_n \mid y_{nq}$, e $y_n \mid y_t$ por hipótesis, luego también $y_n \mid x_{nq} y_r$. Ahora, $\gcd(y_n, x_{nq}) = 1$ por el Corolario 3.1 y el Lema 3.8. Entonces, necesariamente, $y_n \mid y_r$, pero como $r < n$, se tiene que $y_r < y_n$. Por tanto, $n \mid t$.

Lema 3.11. $y_{nk} \equiv k x_n^{k-1} y_n \mod y_n^3$.

Demostración. Haciendo uso del binomio de Newton,

$$x_{nk} + \sqrt{d}\, y_{nk} = (a + \sqrt{d})^{nk} = (x_n + \sqrt{d}\, y_n)^k = \sum_{j=0}^{k} \binom{k}{j} x_n^{k-j} y_n^j d^{\frac{j}{2}}.$$

Tomando la parte irracional, $y_{nk} = \sum_{\substack{j=1 \\ \text{j impar}}}^{k} \binom{k}{j} x_n^{k-j} y_n^j d^{\frac{(j-1)}{2}}$. Todos los términos de esta expresión para $j > 1$ son congruentes con 0 módulo y_n^3, por lo que

$$y_{nk} \equiv \binom{k}{1} x_n^{k-1} y_n d^0 = k x_n^{k-1} y_n \mod y_n^3.$$

En particular, cuando $k = y_n$, se deduce directamente del lema anterior que

$$y_n^2 \mid y_{n y_n}. \tag{3.7}$$

El siguiente lema deriva de los anteriores, y se conoce como *first step down lemma*, pues una condición de divisibilidad en un término de la sucesión implica una condición de divisibilidad en el índice del mismo.

Lema 3.12 (*First step down lemma*). Si $y_n^2 \mid y_t$, entonces también se tiene que $y_n \mid t$.

Demostración. Por el Lema 3.10, $n \mid t$. Pongamos $t = nk$. Ahora, por el Lema 3.11, $y_n^2 \mid k x_n^{k-1} y_n \Rightarrow y_n \mid k x_n^{k-1}$. Pero según el Lema 3.8, $(x_n, y_n) = 1$, luego debe tenerse $y_n \mid k$. Así, y_n también divide a $t = nk$.

Lema 3.13. Si n es par (impar), entonces $\psi_a(n)$ es par (impar).

Demostración. Tomando módulo 2 en la expresión de la sucesión de Lucas,

$$\psi_a(n+2) = 2a\psi_a(n+1) - \psi_a(n) \equiv \psi_a(n) \mod 2.$$

De esto se deduce que, cuando n es par, $\psi_a(n) \equiv \psi_a(0) = 0 \mod 2$, y cuando n es impar, $\psi_a(n) \equiv \psi_a(1) = 1 \mod 2$.

Lema 3.14. Sea $a \geq 1$. Para cualesquiera p, n se tiene que

$$\chi_a(n) \equiv p^n + \psi_a(n)(a-p) \mod 2ap - p^2 - 1.$$

Es más, cuando $0 < p^n < a$, se tiene que $\chi_a(n) \geq p^n + \psi_a(n)(a-p)$.

Demostración. Es fácil comprobar que se verifica la igualdad para $n = 0, 1$. En efecto,

$$\begin{aligned} 1 = \chi_a(0) &\equiv 1 + \psi_a(0)(a-p) = 1 + 0 = 1 \mod 2ap - p^2 - 1, \\ a = \chi_a(1) &\equiv p + \psi_a(1)(a-p) = p + (a-p) = a \mod 2ap - p^2 - 1. \end{aligned}$$

Supongamos que se cumple para $n-1$ y n y procedamos por inducción:

$$\begin{aligned} \chi_a(n+1) &= 2a\chi_a(n) - \chi_a(n-1) \equiv 2a[p^n + \psi_a(n)(a-p)] - [p^{n-1} + \psi_a(n-1)(a-p)] \\ &= 2ap^n - p^{n-1} + (a-p)[2a\psi_a(n) - \psi_a(n-1)] = p^{n-1}(2ap-1) + (a-p)\psi_a(n+1) \\ &\equiv p^{n-1}p^2 + (a-p)\psi_a(n+1) = p^{n+1} + (a-p)\psi_a(n+1) \mod 2ap - p^2 - 1. \end{aligned}$$

Hemos usado el Lema 3.6 y que $2ap - 1 \equiv p^2 \mod 2ap - p^2 - 1$.

Supongamos ahora que $0 < p^n < a$. Al inicio de la prueba vimos que se da la igualdad $\chi_a(n) = p^n + \psi_a(n)(a-p)$ para $n = 0, 1$. Consideremos ahora $n \geq 2$. Por hipótesis, $p^n < a$, luego

$$p^n + \psi_a(n)(a-p) \leq (a-1) + \psi_a(n)(a-1). \tag{3.8}$$

Ahora, como $n \geq 2$, se tiene que $\psi_a(n) \geq 2a$, por lo que

$$a - 1 < \left[\sqrt{a^2-1} - (a-1)\right] 2a \leq \left[\sqrt{a^2-1} - (a-1)\right] \psi_a(n),$$

de donde se deduce que $(a-1) + (a-1)\psi_a(n) < \sqrt{a^2-1}\psi_a(n)$. Tomando esta cota en (3.8), obtenemos la desigualdad estricta para $n \geq 2$. Basta tener en cuenta que, como $(\chi_a(n), \psi_a(n))$ es solución de la ecuación de Pell, $\sqrt{a^2-1}\psi_a(n) < \chi_a(n)$. Así,

$$p^n + \psi_a(n)(a-p) < \sqrt{a^2-1}\psi_a(n) < \chi_a(n), \text{ para } n \geq 2. \qquad \blacksquare$$

Lema 3.15. $(2a-1)^n \leq \psi_a(n+1) \leq (2a)^n$.

Demostración. Estas cotas se deducen fácilmente de la expresión de la sucesión, teniendo en cuenta que $\psi_a(1) = 1$ y $\psi_a(2) = 2a$. Veamos primero la cota superior. Como $\psi_a(n+1)$ se obtiene de $\psi_a(n)$ multiplicando por $2a$ y restando una cantidad no negativa,

$$\psi_a(n+1) = 2a\psi_a(n) - \psi_a(n-1) \leq 2a\psi_a(n) \leq (2a)^2\psi_a(n-1) \leq \cdots \leq (2a)^n.$$

De manera similar, podemos obtener la cota inferior. Como $\psi_a(n+1)$ se obtiene de $\psi_a(n)$ multiplicando por $2a$ y restando una cantidad menor que $\psi_a(n)$,

$$\psi_a(n+1) = 2a\psi_a(n) - \psi_a(n-1) \geq 2a\psi_a(n) - \psi_a(n) = (2a-1)\psi_a(n) \geq \cdots \geq (2a-1)^n.$$

Lema 3.16. Para todo n, se tiene $\psi_a(n+1) > \psi_a(n) \geq n$ y $\chi_a(n+1) > \chi_a(n) \geq a^n$.

Demostración. Que las sucesiones son crecientes se ve fácilmente por definición, y ya hemos hecho uso de ello en demostraciones previas. Ahora, como $\psi_a(0) = 0 \geq 0$, se sigue por inducción que $\psi_a(n) \geq n$ para todo n. De manera análoga, $\chi_a(n) \geq a^n$.

Lema 3.17. $x_{2n\pm j} \equiv -x_j \mod x_n$.

Demostración. Basta desarrollar la expresión de $x_{2n\pm j} = x_{n+(n\pm j)}$ usando la fórmula aditiva del Lema 3.5, así como las de $x_{n\pm j}$ e $y_{n\pm j}$ y tomar módulo x_n:

$$x_{n+(n\pm j)} = x_n x_{n\pm j} + d\, y_n y_{n\pm j} \equiv d\, y_n^2 x_j = (x_n^2 - 1)x_j \equiv -x_j \mod x_n.$$

Lema 3.18. $x_{4n\pm j} \equiv x_j \mod x_n$.

Demostración. Aplicando dos veces el lema anterior, $x_{4n\pm j} \equiv -x_{2n\pm j} \equiv x_j \mod x_n$.

Lema 3.19. Sea $x_i \equiv x_j \mod x_n$, con $i \leq j \leq 2n$ y $n > 0$. Entonces, $i = j$, salvo cuando $a = 2$, $n = 1$, $i = 0$ y $j = 2$.

Demostración. Dividiremos la prueba en dos partes: si x_n es par o impar. En ambos casos, probamos que $x_0, \ldots, x_n$ son todos distintos módulo x_n.

Supongamos que x_n es impar, y consideremos $q = (x_n - 1)/2$. Entonces,

$$\{-q, -q+1, -q+2, \ldots, -1, 0, 1, \ldots, q-1, q\}$$

es un conjunto completo de residuos mutuamente incongruentes módulo x_n. Consideremos primero $x_0, \ldots, x_{n-1}$, de los que sabemos que $1 = x_0 < x_1 < \cdots < x_{n-1}$. De la expresión (3.5) se deduce además que $x_{n-1} \leq x_n/a \leq x_n/2$, luego $x_n \leq q$. Por tanto, $x_0, \ldots, x_{n-1}$ son residuos únicos módulo x_n, menores que q. Ahora, por

el Lema 3.17, $x_{n+1}, x_{n+2}, \dots, x_{2n-1}, x_{2n}$ son respectivamente congruentes módulo x_n a $-x_{n-1}, -x_{n-2}, \dots, x_1, -x_0 = -1$. Del mismo modo, $x_{n+1}, \dots, x_{2n}$ son residuos únicos módulo x_n, mayores que $-q$. Por tanto, $x_0, \dots, x_{2n}$ forman un conjunto completo de residuos mutuamente incongruentes módulo x_n.

Supongamos ahora que x_n es par, y consideremos $q = x_n/2$. Entonces,

$$\{-q+1, -q+2, \dots, -1, 0, 1, \dots, q-1, q\}$$

es un conjunto completo de residuos mutuamente incongruentes módulo x_n (a diferencia del caso impar, no aparece $-q$, pues $-q \equiv q \mod x_n$). De manera análoga al caso impar, se tiene $x_{n-1} \leq q$ y se deduce el mismo resultado, con una excepción: si $x_{n-1} = q = x_n/2$. En este caso,

$$x_{n+1} \equiv -q \equiv q = x_{n-1} \mod x_n,$$

Pero entonces, de la expresión (3.5), $x_n = ax_{n-1} + dy_{n-1} = 2x_{n-1}$, lo que implica que $a = 2$ e $y_{n-1} = 0$, es decir, $n = 1$.

Así, en ambos casos, si tenemos que $x_i \equiv x_j \mod x_n$, con $i \leq j \leq 2n$, como $x_0, \dots, x_{2n}$ forman un conjunto completo de residuos mutuamente incongruentes módulo x_n, no queda más remedio que $x_i = x_j$, y por tanto, $i = j$. El caso excepcional del enunciado es el que hemos obtenido en el caso par, de donde $a = 2$, $n = 1$ y $x_{n+1} \equiv x_{n-1} \mod x_n$, luego $i = 0$ y $j = 2$.

Lema 3.20. Sea $x_i \equiv x_j \mod x_n$ con $0 < i \leq n$ y $0 \leq j \leq 4n$. Entonces, o bien $i = j$, o bien $j = 4n - i$.

Demostración. Supongamos primero que $j \leq 2n$. En este caso, se tiene directamente por el lema anterior que $i = j$. Supongamos ahora $j > 2n$. Sea $\bar{j} = 4n - j$, de modo que $0 < \bar{j} < 2n$. Por el Lema 3.18 y usando la hipótesis,

$$x_{4n-j} \equiv x_j \equiv x_i \mod x_n.$$

Luego del lema anterior se sigue que $i = 4n - j$, es decir, $j = 4n - i$.

El siguiente lema, así como versiones similares del mismo, suele denominarse *second step down lemma*; esto se debe a que nos da, a partir de una relación entre términos de las sucesiones de Lucas, una relación similar entre sus respectivos índices.

Lema 3.21 (*Second step down lemma*). Sea $0 < i \leq n$. Si $x_i \equiv x_j \mod x_n$, entonces

$$j \equiv \pm i \mod 4n.$$

Demostración. Consideremos $j = 4nq + \bar{j}$, con $0 \leq \bar{j} < 4n$. De la hipótesis y el Lema 3.18, $x_i \equiv x_j \equiv x_{\bar{j}} \mod x_n$. Así, por el Lema 3.20 deberá tenerse que, o bien $i = \bar{j}$, o bien $i = 4n - \bar{j}$. Tomando módulo $4n$,

$$j \equiv \bar{j} \equiv \pm i \qquad \text{mód } 4n.$$

3.1.2 Representación diofántica de la sucesión $y = \psi_a(n)$

Ya estamos listos para dar una definición diofántica de la $n-$ésima solución de la ecuación de Pell. Dado el sistema de ecuaciones, la demostración consistirá, esencialmente, en encadenar convenientemente los lemas previos. El recíproco resulta de escoger cuidadosamente las variables en función de las otras para verificar cada ecuación del sistema. Este será el método de prueba que seguiremos para el resto de resultados de este tipo.

Proposición 3.2 (Representación diofántica de $y = \psi_a(n)$). Sean $a \geq 2, n \geq 1$ e $y \in \mathbb{N}$. Se tiene que $y = \psi_a(n)$ si y solo si existen enteros no negativos $b, c, d, r, s, t, u, v, x$ tales que:

(1) $x^2 = (a^2 - 1)y^2 + 1$,
(2) $u^2 = (a^2 - 1)v^2 + 1$,
(3) $s^2 = (b^2 - 1)t^2 + 1$,
(4) $v = 4ry^2$,
(5) $b = a + u^2(u^2 - a)$,
(6) $s = x + cu$,
(7) $t = n + 4dy$,
(8) $n \leq y$.

Demostración. Consideremos dados $a \geq 2, n \geq 1$ e $y \in \mathbb{N}$, y supongamos que existen $b, c, d, r, s, t, u, v, x$ verificando las ecuaciones $(1) - (8)$. Por la ecuación (5), tenemos que $b > a > 1$. Entonces, como (1), (2) y (3) son ecuaciones de Pell, sabemos por el Lema 3.4 que existirán $i, j, k > 0$ tales que

$$\begin{aligned} x &= \chi_a(i), \quad y = \psi_a(i), \\ u &= \chi_a(k), \quad v = \psi_a(k), \\ s &= \chi_b(j), \quad t = \psi_b(j). \end{aligned}$$

Por (4) se tiene que $v = \psi_a(k) \geq \psi_a(i) = y$, y por tanto $k \geq i$. Ahora, de (5) y (6) se deduce que $b \equiv a \mod u$ y $s \equiv x \mod u$, luego sustituyendo u, s y x, tenemos

$$\begin{aligned} b &\equiv a && \text{mód } \chi_a(k), \\ \chi_b(j) &\equiv \chi_a(i) && \text{mód } \chi_a(k). \end{aligned}$$

Gracias a la primera congruencia, podemos hacer uso del Lema 3.7 para obtener

$$\chi_b(j) \equiv \chi_a(j) \qquad \text{mód } \chi_a(k),$$

y uniendo esto con la segunda congruencia anterior se tiene que $\chi_a(i) \equiv \chi_a(j) \mod \chi_a(k)$. Como antes vimos que $k \geq i > 0$, del Lema 3.21 se sigue que

$$j \equiv \pm i \qquad \text{mód } 4k. \tag{3.9}$$

La ecuación (4) con $v = \psi_a(k)$ e $y = \psi_a(i)$ implica que $\psi_a^2(i) \mid \psi_a(k)$, luego por el Lema 3.12, $\psi_a(i) \mid k$. Aplicando esto a la expresión (3.9),

$$j \equiv \pm i \qquad \text{mód } 4\psi_a(i). \tag{3.10}$$

Combinando las ecuaciones (2), (4) y (5) se deduce que $b \equiv 1 \ mod \ 4\psi_a(i)$, luego por el Corolario 3.1,

$$\psi_b(j) \equiv j \qquad \text{mód } 4\psi_a(i). \tag{3.11}$$

De la ecuación (7) y teniendo en cuenta que $t = \psi_b(j)$ e $y = \psi_a(i)$, se tiene que

$$\psi_b(j) \equiv n \qquad \text{mód } 4\psi_a(i). \tag{3.12}$$

Encadenando entonces (3.10), (3.11) y (3.12), obtenemos

$$n \equiv \pm i \qquad \text{mód } 4\psi_a(i). \tag{3.13}$$

Tomando la desigualdad en (8) con $y = \psi_a(i)$, y por el Lema 3.16, se verifican las desigualdades

$$n \leq \psi_a(i) \tag{3.14}$$
$$\psi_a(i) \geq i. \tag{3.15}$$

Por último, como

$$-2\psi_a(i) + 1, -2\psi_a(i) + 2, \ldots, -1, 0, 1, \ldots, 2\psi_a(i)$$

forman un conjunto completo de residuos mutuamente incongruentes módulo $4\psi_a(i)$, las desigualdades (3.14) y (3.15) muestran que la congruencia (3.13) implica, necesariamente, que $n = i$. Por tanto,

$$x = \chi_a(i) = \chi_a(n), \text{ e}$$
$$y = \psi_a(i) = \psi_a(n).$$

Recíprocamente, sea $y = \psi_a(n)$ y tomemos $x = \chi_a(n)$ para satisfacer la ecuación de Pell en (1). Para que se verifique la ecuación de Pell en (2), tomamos $u = \chi_a(m)$, y $v = \psi_a(m)$, con $m = 2n\psi_a(n)$. Por el Lema 3.10 y la expresión (3.7), se sigue que $y^2 \mid v$, luego existe cierto r satisfaciendo (4). Ahora, como a viene dado y ya hemos establecido el valor de u, podemos tomar $b = a + u^2(u^2 - a)$, satisfaciendo (5). La ecuación de Pell en (3) se tiene tomando $s = \chi_b(n)$ y $t = \psi_b(n)$. Además, como $b > a$, $s = \chi_b(n) > \chi_a(n) = x$. Por el Lema 3.7 y usando que (5) implica que $b \equiv a \mod u$, se tiene que $\chi_b(n) = s \equiv x = \chi_a(n) \mod u$, luego existe c que satisfaga (6). Por último, gracias al Lema 3.16, $y = \psi_a(n) \geq n$, luego también se da (8). Este lema también nos dice que $t = \psi_b(n) \geq n$, y por el Corolario 3.1, $t \equiv n \mod b - 1$, que podemos escribir, usando que $b \equiv 1 \mod 4y$, como $t \equiv n \mod 4y$. Esto último nos asegura la existencia de cierto d verificando (7). ∎

Observación 3.2. La desigualdad en (8) es también diofántica. En efecto, esta relación de orden puede reescribirse fácilmente como una igualdad lineal, incluyendo una variable adicional de manera adecuada. Con la notación del segundo capítulo,

$$n \leq y \iff (\exists z)(n + z = y).$$

Notemos que podemos reducir el número de ecuaciones y variables eliminando b, s, t y v por sustitución en las ecuaciones del sistema (1) − (8) de la Proposición 3.2. De esta forma, obtenemos:

Corolario 3.2. Sean $a \geq 2$, $n \geq 1$ e $y \in \mathbb{N}$. Se tiene que $y = \psi_a(n)$ si y solo si existen enteros no negativos c, d, r, u, x tales que:

(1) $x^2 = (a^2 - 1)y^2 + 1$,
(2) $u^2 = 16(a^2 - 1)r^2y^4 + 1$,
(3) $(x + cu)^2 = ((a + u^2(u^2 - a))^2 - 1)(n + 4dy)^2 + 1$,
(4) $n \leq y$.

3.2 Teorema de Wilson

Ya señalamos en el primer capítulo que el Teorema de Wilson es una herramienta útil para establecer fórmulas relativas a los números primos. De hecho, resulta ser la clave que nos permitirá construir el conjunto de ecuaciones que nos llevará al polinomio enumerador de primos: será el pegamento en la equivalencia que presentaremos en la representación diofántica del conjunto de los números primos.

Teorema. *Sea $p \geq 2$. Entonces, p es primo si y solo si*

$$(p-1)! \equiv -1 \mod p.$$

Demostración. Sea $p \geq 2$ un número primo. Para $p = 2$ el resultado es obvio, pues $1 \equiv -1 \mod 2$. Supongamos entonces que p es un primo impar, $p > 2$. Como p es primo, $\mathbb{Z}/\mathbb{Z}p$ es un cuerpo, luego todo elemento no nulo es una unidad. Como todos los elementos son menores que p, no son divisibles por él, luego $\{1, 2, \ldots, p-1\}$ son unidades módulo p. Ahora bien, en $\mathbb{Z}/\mathbb{Z}p$ las únicas unidades que son inversas de sí mismas son 1 y $p-1$, pues son las raíces del polinomio $x^2 - 1$ en $(\mathbb{Z}/\mathbb{Z}p)[x]$. Por tanto, si desarrollamos la expresión del factorial

$$(p-1)! = 1 \cdot 2 \cdots (p-2) \cdot (p-1),$$

todos los factores del producto salvo 1 y $p-1$ tienen una inversa distinta de ellos mismos y que es otro elemento del producto, luego se cancelan entre sí y obtenemos el resultado módulo p

$$(p-1)! \equiv 1 \cdot (p-1) \equiv -1 \mod p.$$

Recíprocamente, sea $p \geq 2$ tal que $(p-1)! \equiv -1 \mod p$. Supongamos que p es compuesto, es decir, que existe $d \mid p$ con $1 < d < p$. Notemos que $d \mid (p-1)!$, pues como $d < p$, será uno de los factores en el producto. Además, de la hipótesis se sigue que $d \mid (p-1)! + 1$, por lo que $d \mid 1$. Pero $d > 1$, luego p es primo. ∎

El Teorema de Wilson se encuentra enunciado de diversas formas equivalentes en la literatura. El enunciado anterior es equivalente a este otro, que es precisamente el que utilizaremos en esta memoria:

Teorema 3.3 (Teorema de Wilson). *Para $k \geq 1$, se tiene que $k+1$ es primo si y solo si*

$$k + 1 \mid k! + 1.$$

Basta considerar $p = k + 1$ en el enunciado anterior del teorema, sumar uno a ambos lados y escribir la congruencia en términos de igualdad, obteniéndose de manera inmediata la relación de divisibilidad deseada.

3.3 Representación diofántica del factorial

Para la descripción diofántica del conjunto de los números primos necesitamos hacer uso del Teorema de Wilson, que caracteriza a los primos en función de su factorial. Por tanto, necesitaremos también una definición diofántica de esta función. Presentamos una serie de resultados previos que nos conducirán a esta definición del factorial.

El siguiente lema nos será de utilidad para forzar que una variable sea exponencialmente mayor que otra. En lo que sigue, $\square$ denotará un cuadrado perfecto.

Lema 3.22. Para $e \geq 2$, si

$$e^3(e+2)(n+1)^2 + 1 = \square, \tag{3.16}$$

entonces se tiene que $e - 1 + e^{e-2} \leq n$. Además, para todo e y $t > 0$, existe n satisfaciendo (3.16) tal que $t \mid n+1$.

Demostración. Sea $a = e+1$. Sustituyendo en (3.16), obtenemos una ecuación de Pell de la forma

$$(a^2-1)[(a-1)(n+1)]^2 + 1 = \square.$$

Si n es una solución de (3.16), entonces $(a-1)(n+1)$ será la segunda componente de un par que es solución de la ecuación de Pell, es decir, $\psi_a(j) = (a-1)(n+1)$ para algún $j > 0$. Por el Corolario 3.1, $(a-1)(n+1) = \psi_a(j) \equiv j \mod a-1 \Rightarrow a-1 \mid j$. Además, como $j \neq 0$, esto implica que $a-1 \leq j$, luego $\psi_a(a-1) \leq \psi_a(j)$. Usando el Lema 3.15,

$$(a-2)(a-1) + (a-1)^{a-2} < (2a-1)^{a-2} \leq \psi_a(a-1) \leq \psi_a(j) = (a-1)(n+1).$$

Quedándonos con los extremos de esta cadena de desigualdades, dividiendo entre $(a-1)$ a ambos lados y recuperando el valor de e, se obtiene la desigualdad deseada:

$$(a-2) + (a-1)^{a-3} < n+1 \Rightarrow e - 1 + e^{e-2} < n+1 \Rightarrow e - 1 + e^{e-2} \leq n.$$

La segunda afirmación del enunciado se deduce directamente del hecho de que existen infinitos z satisfaciendo la ecuación

$$(a^2-1)(a-1)^2t^2(z+1)^2 + 1 = \square.$$

Así, tomando $n+1 = t(z+1)$ existirán infinitos n tal que $t \mid n+1$ que verifiquen

$$(a^2-1)(a-1)^2(n+1)^2 + 1 = \square.$$

El Lema 3.25 será esencial en la demostración de la representación diofántica del factorial. En la prueba de este lema, necesitaremos dos desigualdades básicas:

Lema 3.23 (Desigualdad de Bernoulli). Para cualquier $x \in \mathbb{R}$ tal que $x \geq -1$, se tiene que $(1+x)^n \geq 1+nx$, donde n es un entero positivo.

Demostración. Procedamos por inducción en n. Para $n = 1$ se da la igualdad. Supongamos que se verifica la desigualdad para n, y veamos que se tiene para $n+1$:

$$(1+x)^{n+1} = (1+x)^n(1+x) \geq (1+nx)(1+x) = 1+(n+1)x+nx^2 \geq 1+(n+1)x.$$

∎

Lema 3.24. Si $0 \leq \alpha \leq 1/2$, entonces $(1-\alpha)^{-1} \leq 1+2\alpha$.

Demostración. Para $\alpha = 0$ y $\alpha = 1/2$ se tiene la igualdad. Si $0 < \alpha < 1/2$, se verifica la desigualdad estricta: la serie $\sum_{k=0}^{\infty}(2\alpha)^k$ es convergente, luego

$$(1-\alpha)^{-1} = \frac{1}{1-\alpha} < \frac{1}{1-2\alpha} = \sum_{k=0}^{\infty}(2\alpha)^k = 1+2\alpha+\ldots$$

∎

Lema 3.25. Para cualquier entero positivo k, si $(2k)^k \leq n$ y $n^k < p$, entonces

$$k! < \frac{(n+1)^k p^k}{\operatorname{rem}((p+1)^n, p^{k+1})} < k!+1.$$

Demostración. Estudiemos en primer lugar el resto que aparece en el denominador. Queremos ver el resto de la división euclídea de $(p+1)^n$ entre p^{k+1}. Vamos a expresarlo como es usual mediante el algoritmo de división como $(p+1)^n = p^{k+1}\cdot q + r$. Para ello, nos ayudaremos del Binomio de Newton, obteniendo así:

$$(p+1)^n = \sum_{i=0}^{n}\binom{n}{i}p^i = \sum_{i=0}^{k}\binom{n}{i}p^i + \sum_{i=k+1}^{n}\binom{n}{i}p^i = \sum_{i=0}^{k}\binom{n}{i}p^i + p^{k+1}\cdot\sum_{i=k+1}^{n}\binom{n}{i}p^{i-k-1}.$$

Veamos que $0 < \operatorname{rem}((p+1)^n, p^{k+1}) = \sum_{i=0}^{k}\binom{n}{i}p^i < p^{k+1}$, es decir, el algoritmo de división termina y el resto que buscamos tiene esa expresión, que no es cero:

$$\sum_{i=0}^{k}\binom{n}{i}p^i \leq \sum_{i=0}^{k} n^i p^i = \frac{(np)^{k+1}-1}{np-1}.$$

Podemos acotar el numerador como sigue

$$(np)^{k+1}-1 = n^k p^k np-1 \overset{\text{PH}}{\leq} (p-1)p^k np-1 = pp^k np-pp^k n-1 < pp^k pn-pp^k = (np-1)pp^k.$$

Luego

$$0 < \binom{n}{0}p^0 < \sum_{i=0}^{k}\binom{n}{i}p^i < \frac{(np)^{k+1}-1}{np-1} < pp^k = p^{k+1}.$$

Veamos ahora que se verifica la cota inferior del enunciado:

$$\begin{aligned} k!\left(\sum_{i=0}^{k}\binom{n}{i}p^i\right) &\leq k!\left(k\binom{n}{k-1}p^{k-1}+\binom{n}{k}p^k\right) \leq k!\left(k\frac{n^{k-1}}{(k-1)!}p^{k-1}+\frac{n^k}{k!}p^k\right) \\ &= k!\left(\frac{k^2n^{k-1}}{k!}p^{k-1}+\frac{n^k}{k!}p^k\right) = k^2n^{k-1}p^{k-1}+n^kp^k < kn^kp^{k-1}+n^kp^k \\ &< kpp^{k-1}+n^kp^k = (k+n^k)p^k \leq (1+n)^kp^k. \end{aligned}$$

Comprobemos por último la cota superior:

$$\begin{aligned} \frac{(n+1)^kp^k}{\sum_{i=0}^{k}\binom{n}{i}p^i} &= \frac{(n+1)^k}{\sum_{i=0}^{k}\binom{n}{i}p^{i-k}} < \frac{(n+1)^k}{\binom{n}{k}} < \frac{k!}{\frac{(n+1-k)^k}{(n+1)^k}} = \frac{k!}{\left(1-\frac{k}{n+1}\right)^k} < \frac{k!}{\left(1-\frac{k}{n}\right)^k} \\ &= k!\left(\left(1-\frac{k}{n}\right)^k\right)^{-1} \overset{3.23}{\leq} k!\left(1-\frac{k^2}{n}\right)^{-1} \overset{3.24}{\leq} k!\left(1+\frac{2k^2}{n}\right) \\ &\overset{\text{PH}}{\leq} k!\left(1+\frac{2k^2}{(2k)^k}\right) \leq k!\left(1+\frac{1}{k!}\right) = k!+1. \end{aligned}$$

Proposición 3.3 (Representación diofántica del factorial). Sean $f, k \in \mathbb{Z}_+$. Se tiene que $f = k!$ si y solo si existen enteros no negativos h, j, n, p, q, w, z tales que:

(1) $q = wz + h + j$,
(2) $z = f(h + j) + h$,
(3) $(2k)^3(2k + 2)(n + 1)^2 + 1 = \square$,
(4) $p = (n + 1)^k$,
(5) $q = (p + 1)^n$,
(6) $z = p^{k+1}$.

Demostración. Consideremos dados f, k enteros positivos y supongamos que existen enteros no negativos h, j, n, p, q, w, z que verifican las ecuaciones (1) – (6). Notemos en primer lugar que estamos en las condiciones del Lema 3.25, pues $k > 0$, $n^k < p = (n + 1)^k$ y al verificarse la ecuación (3), se tiene que $(2k)^k \leq n$ por el Lema 3.22. Gracias a las ecuaciones (2), (4) y (6), se tiene que $0 < h + j \leq z$. Además, esta

última desigualdad es estricta: si se diera la igualdad $h + j = z$, la ecuación (1) implicaría que $z = p^{k+1} \mid q = (p+1)^n$, lo cual contradice el Lema 3.25 (pues $\text{rem}(q, z) \neq 0$). Luego $0 < h + j < z$, es decir, lo que nos dice la ecuación (1) es que $\text{rem}(q, z) = h + j$. Entonces, por (2) y el lema anterior,

$$f \leq \frac{z}{h+j} \leq f + 1, \quad \text{y} \quad k! < \frac{pp^k}{\text{rem}(q,z)} = \frac{z}{h+j} < k! + 1,$$

y esto solo puede darse si $f = k!$, pues k y f son enteros positivos.

Recíprocamente, supongamos $f = k!$, con $k \geq 1$. Por el Lema 3.22, podemos tomar n tal que $(2k)^k \leq n$, verificando la ecuación (3). Sean $p = (n+1)^k$, $q = (p+1)^n$ y $z = p^{k+1}$, satisfaciendo (4), (5) y (6). Por último, para obtener (1) y (2) basta considerar $w = (q - \text{rem}(q, z))/z$, $h = z - f \cdot \text{rem}(q, z)$ y $j = \text{rem}(q, z) - h$.

3.4 Representación diofántica del conjunto de los números primos

Finalmente, tenemos todo el material que necesitamos para dar una definición diofántica del conjunto de los números primos que nos permite construir el polinomio de grado 25 en 26 variables del Teorema 3.1, enunciado al inicio de este capítulo.

Teorema 3.4 (Representación diofántica de los números primos). *Para cualquier $k \geq 1$, se tiene que $k + 1$ es primo si y solo si existen enteros no negativos $a, b, c, d, e, f, g, h, i, j, l, m, n, o, p, q, r, s, t, u, v, w, x, y, z$ tales que*

(1) $q = wz + h + j$,
(2) $z = (gk + g + k)(h + j) + h$,
(3) $(2k)^3(2k+2)(n+1)^2 + 1 = f^2$,
(4) $e = p + q + z + 2n$,
(5) $e^3(e+2)(a+1)^2 + 1 = o^2$,
(6) $x^2 = (a^2 - 1)y^2 + 1$,
(7) $u^2 = 16(a^2 - 1)r^2y^4 + 1$,
(8) $(x + cu)^2 = ((a + u^2(u^2 - a))^2 - 1)(n + 4dy)^2 + 1$,
(9) $m^2 = (a^2 - 1)l^2 + 1$,
(10) $l = k + i(a - 1)$,
(11) $n + l + v = y$,
(12) $m = p + l(a - n - 1) + b(2a(n+1) - (n+1)^2 - 1)$,
(13) $x = q + y(a - p - 1) + s(2a(p+1) - (p+1)^2 - 1)$,
(14) $pm = z + pl(a - p) + t(2ap - p^2 - 1)$.

Demostración. Sea $k \geq 1$. Supongamos en primer lugar que existen $a, b, \ldots, z$ que verifican el sistema (1) − (14). Veamos que $k + 1$ es primo.
El Lema 3.22 aplicado a la ecuación (3) implica que $2k - 1 + (2k)^{2k-2} \leq n$, y como

$2k \geq 2$, acotando se obtiene que

$$2 \leq n, \text{ y} \tag{3.17}$$

$$k < n. \tag{3.18}$$

Sustituyendo en la ecuación (5) la variable e por su expresión en (4) y aplicando de nuevo el Lema 3.22, se tiene que

$$p + q + z + 2n - 1 + (p + q + z + 2n)^{p+q+z+2n-2} \leq a, \text{ y} \tag{3.19}$$

$$n < a. \tag{3.20}$$

Las ecuaciones (6), (7), (8) y (11) forman, precisamente, la caracterización diofántica de las soluciones de Pell dada en el Corolario 3.2, luego

$$(x, y) = (\chi_a(n), \psi_a(n)).$$

Como (9) es una ecuación de Pell en m y l, se dará $(m, l) = (\chi_a(k'), \psi_a(k'))$ para cierto k'. De la ecuación (11) se deduce que $l = \psi_a(k') < y = \psi_a(n)$, luego debe tenerse $k' < n$. Pero entonces, las desigualdades obtenidas en (3.18) y (3.20) implican que $k' < a - 1$ y $k < a - 1$. Y de la ecuación (10) y el Corolario 3.1 se sigue que $k' \equiv k$ mód $a - 1$, luego $k = k'$ y, por tanto,

$$(m, l) = (\chi_a(k), \psi_a(k)).$$

Ahora, de las expresiones (3.17), (3.18) y (3.19), se deduce que

$$p < a, \quad (n + 1)^k < a, \quad \text{y} \quad a < 2a(n + 1) - (n + 1)^2 - 1. \tag{3.21}$$

Combinando el Lema 3.14 y la ecuación (12), como $(m, l) = (\chi_a(k), \psi_a(k))$, se deduce que $p \equiv (n + 1)^k \mod 2a(n + 1) - (n + 1)^2 - 1$. Se sigue de las desigualdades (3.21) que, necesariamente,

$$p = (n + 1)^k. \tag{3.22}$$

Sigamos un procedimiento análogo para obtener dos condiciones más. De (3.17) y (3.19),

$$q < a, \quad (p + 1)^n < a, \quad \text{y} \quad a < 2a(p + 1) - (p + 1)^2 - 1. \tag{3.23}$$

De nuevo, por el Lema 3.14 combinado con la ecuación (13) donde $(x, y) = (\chi_a(n), \psi_a(n))$, se deduce que $q \equiv (n + 1)^n \mod 2a(p + 1) - (p + 1)^2 - 1$. Teniendo en cuenta las desigualdades (3.23),

$$q = (p + 1)^n. \tag{3.24}$$

Como $p \neq 0$ por la igualdad (3.22), de (3.17), (3.18) y (3.19) obtenemos

$$z < a, \quad p^{k+1} < a, \quad \text{y} \quad a < 2ap - p^2 - 1.$$

Y de nuevo, usando el Lema 3.14, la ecuación (14) y las desigualdades anteriores, debe tenerse

$$z = p^{k+1}. \tag{3.25}$$

Por último, las ecuaciones (1), (2), (3) junto con las condiciones (3.22), (3.24) y (3.25) obtenidas forman, precisamente, una definición diofántica del factorial (Proposición 3.3, tomando $f = gk+g+k$). Por tanto, $gk+g+k = k!$. Con una simple reescritura y sumando uno a ambos lados de la igualdad, $k!+1 = (g+1)(k+1)$, es decir, $k+1 \mid k!+1$ y el Teorema de Wilson 3.3 concluye la prueba en este sentido.

Recíprocamente, supongamos que $k+1$ es primo, con $k \geq 1$. Gracias al Teorema de Wilson, podemos obtener g despejando de $k! = gk+g+k$, conocido el primo $k+1$. La caracterización diofántica del factorial (Proposición 3.3) permite tomar f, h, j, n, p, q y w satisfaciendo las ecuaciones (1), (2), (3) bajo las condiciones

$$p = (n+1)^k, \qquad q = (p+1)^n, \qquad y \qquad z = p^{k+1}. \tag{3.26}$$

Tomemos $e = p + q + z + 2n$, satisfaciendo (4). Por el Lema 3.22 podemos encontrar $a \geq 2$ y o verificando la ecuación (5), resolviendo la ecuación de Pell. Consideramos $y = \psi_a(n)$, de modo que la caracterización del Corolario 3.2 nos proporciona c, d, r, u y x satisfaciendo (6), (7) y (8). Para satisfacer la ecuación de Pell en (9), tomamos $(m, l) = (\chi_a(k), \psi_a(k))$. Por el Corolario 3.1 y teniendo en cuenta que, como $a \geq 2$, se tiene que $k \leq \psi_a(k)$, existe cierto i que satisfaga (10). Y como, para $a \geq 2$, se tiene que $\psi_a(n) \geq \psi_a(n-1) + n$, de (3) se deduce que $k < n$, luego debe verificarse que $n + l \leq y$. Así, existe v que satisface la ecuación (11). Por último, como vimos en la demostración de la otra implicación del teorema, las ecuaciones (4) y (5) implican que

$$(n+1)^k < a, \qquad (p+1)^n < a, \qquad y \qquad p^k < a. \tag{3.27}$$

Luego por el Lema 3.14 y las condiciones obtenidas en (3.26), existen b, s y t satisfaciendo las ecuaciones restantes (12), (13) y (14). ∎

Una vez demostrado el carácter diofántico del conjunto de los números primos, y dado explícitamente un sistema de ecuaciones diofánticas que lo definen, para obtener el polinomio enumerador de primos del Teorema 3.1 basta sustituir k por $k+1$ en las ecuaciones del teorema anterior, sumar los cuadrados de las ecuaciones para obtener

el polinomio M en las condiciones de la Proposición 3.1 y aplicar el método de Putnam, esto es, tomar $P = (k+2)\{1 - M\}$, tal y como comentamos en los primeros párrafos de este capítulo.

Observación 3.3. Sabemos que todo número compuesto puede escribirse como una multiplicación de otros dos enteros de forma no trivial, como vimos en el segundo capítulo al dar el conjunto diofántico que forman los números compuestos. No se contemplaba la posibilidad de un resultado de este tipo para los números primos hasta la resolución del décimo problema de Hilbert. La caracterización diofántica del conjunto de los números primos nos permite dar una cota en el número de operaciones mediante las que se comprueba que, en efecto, cierto número es primo. En particular, del Teorema 3.4 se deduce que, si p es un número primo, entonces existe una prueba de su primalidad, consistente en 87 sumas y multiplicaciones.

En el próximo capítulo nos enfrentaremos a los problemas que surgen al tratar en la práctica con el polinomio enumerador de primos que acabamos de construir.

4 Cálculos efectivos

Si tratamos de evaluar el polinomio del Teorema 3.1 asociando números aleatorios a las variables, la posibilidad de obtener un valor primo es extremadamente remota. Aun así, podemos hacerlo para tener una idea de los valores negativos que representa este polinomio. Realizando estas pruebas aleatorias, hemos observado que muchos valores se repiten con elecciones distintas para las variables, como el -2188, y que los valores que representa son muy diversos: desde valores relativamente pequeños como -2058 y -444699, hasta valores gigantescos de una cantidad ingente de dígitos.

Nuestro objetivo es obtener valores primos. Para ello, hemos de establecer primero el primo que queremos que nos devuelva, y así poder seguir la guía que nos ofrece la demostración del Teorema 3.4 para calcular qué valores pueden tomar las variables fijado el parámetro k, que es el que determina el primo $k + 1$. Por este motivo, este tipo de polinomios carece de utilidad en la práctica para obtener números primos, además de que resulta computacionalmente imposible la resolución total del sistema: se requiere la resolución de ecuaciones de Pell, cuyas soluciones forman una sucesión creciente de valores que aumentan exponencialmente.

Veamos cómo podemos atacar las ecuaciones de Pell, dando un algoritmo para su resolución. Posteriormente, haremos uso de él para tratar de encontrar una solución del sistema que define al polinomio enumerador de primos 3.1 para el primo par 2.

4.1 Resolución de la ecuación de Pell

Hemos visto que las soluciones de la ecuación de Pell (3.1) vienen dadas por las sucesiones de Lucas (3.2), así como que pueden ser generadas a partir de su solución fundamental mediante la relación (3.4). Por ello, para resolver ecuaciones de Pell, lo primero que necesitamos es obtener su solución fundamental. Existen métodos diversos para este propósito, y existen libros completos dedicados a ello (véase [JW09]);

en este capítulo, obtendremos esta primera solución no trivial mediante fracciones continuas. Todos los resultados de esta sección, que presentamos sin demostraciones, vienen tratados en profundidad en [NZM91], [JW09] y [Rob06].

Definición 4.1 (Fracción continua simple). *Sea $\{a_i\}_{i\in\mathbb{N}}$ una sucesión (finita o no) de enteros, todos positivos salvo, quizá, a_0. Se denomina fracción continua simple a cualquier expresión de la forma*

$$a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\ddots}}}$$

que denotaremos de forma abreviada por $\langle a_0, a_1, a_2, \dots \rangle$.

Consideremos $\{a_i\}_{i\in\mathbb{N}}$ una sucesión de enteros, todos positivos salvo, quizá a_0. Definimos las sucesiones $\{h_n\}$ y $\{k_n\}$ de manera recursiva como sigue:

$$\begin{aligned} h_{-2} &= 0, \quad h_{-1} = 1, \quad h_i = a_i h_{i-1} + h_{i-2}, \quad \text{para } i \geq 0. \\ k_{-2} &= 1, \quad k_{-1} = 0, \quad k_i = a_i k_{i-1} + k_{i-2}, \quad \text{para } i \geq 0. \end{aligned} \tag{4.1}$$

Nótese la gran similitud entre estas y las sucesiones de Lucas (3.2). No es difícil obtener de estas expresiones la siguientes igualdades

$$\frac{h_i}{k_i} - \frac{h_{i-1}}{k_{i-1}} = \frac{(-1)^{i-1}}{k_i k_{i-1}}, \qquad \text{y} \qquad \frac{h_i}{k_i} - \frac{h_{i-2}}{k_{i-2}} = a_i \frac{(-1)^i}{k_i k_{i-2}}, \tag{4.2}$$

para $i > 1$. De aquí, se deduce que los h_i son coprimos con los k_i, pues cualquier divisor común de ambos también lo será de $(-1)^{i-1}$. También, que los $\frac{h_i}{k_i}$ forman una sucesión creciente en los i pares y decreciente en los impares. Es más, si $r_n := \frac{h_n}{k_n}$,

$$r_0 < r_2 < r_4 < r_6 < \cdots < r_7 < r_5 < r_3 < r_1. \tag{4.3}$$

Definición 4.2 (Convergentes de una fracción continua). *Sea $\langle a_0, a_1, \dots \rangle$ una fracción continua simple. Para cada $n \geq 0$, denotemos r_n al resultado de truncar por a_n la fracción continua, esto es, $r_n := \langle a_0, a_1, \dots, a_n \rangle$. Diremos que r_n es el $n-$ésimo convergente de la fracción continua.*

Lema 4.1. Para todo $x \in \mathbb{R}_+$, se tiene que

$$\langle a_0, a_1, \dots, a_{n-1}, x \rangle = \frac{x h_{n-1} + h_{n-2}}{x k_{n-1} + k_{n-2}}.$$

La fórmula se deduce fácilmente por inducción en n. Gracias a este lema, podemos probar que, en efecto, la notación que hemos usado es consistente. Sin más que tomar $x = a_n$ en el lema anterior, se obtienen las expresiones (4.1). Así,

Teorema 4.1. *Para todo $n \geq 0$, se tiene que $r_n = \frac{h_n}{k_n}$.*

Si a es un número racional, mediante el algoritmo de Euclides puede obtenerse su fracción continua finita correspondiente. Recíprocamente, dada una fracción continua finita, podemos obtener el valor que representa.

Teorema 4.2. *Toda fracción continua finita es un número racional. Recíprocamente, todo racional puede expresarse como una fracción continua finita, de exactamente dos formas.*

Ejemplo 4.1. $\frac{51}{22} = \langle 2, 3, 7\rangle = \langle 2, 3, 6, 1\rangle$.

Centrémonos en el caso infinito. ¿Qué es exactamente una fracción continua infinita? Resulta que estas son convergentes, y su valor viene dado por $\lim_{n\to\infty} r_n$, que siempre existe (consecuencia de (4.2) y (4.3)). Además, disponemos de una caracterización como en el caso finito, en este caso única:

Teorema 4.3. *Una fracción continua infinita simple converge a un valor irracional. Dos fracciones continuas distintas convergen a irracionales distintos.*

Un lema fundamental en la demostración del teorema anterior es el siguiente:

Lema 4.2. Sea $\theta = \langle a_0, a_1, \dots\rangle$ una fracción continua simple. Entonces, $a_0 = \lfloor\theta\rfloor$. Además, si $\theta_1 = \langle a_1, a_2, \dots\rangle$, entonces $\theta = a_0 + 1/\theta_1$.

Fijemos nuestra atención en un caso particular de fracciones continuas infinitas.

Definición 4.3 (Fracción continua periódica). *Una fracción continua infinita simple se dice periódica si existen $k \geq 0$ y $l > 0$ tales que $a_n = a_{n+l}$, para todo $n \geq k$. Las denotaremos por*

$$\langle a_0, \dots, a_{n-1}, \overline{a_n, \dots, a_{n+l}}\rangle := \langle a_0, \dots, a_n, \dots, a_{n+l}, a_n, \dots, a_{n+l}, \dots\rangle.$$

El conjunto $\{a_n, \dots, a_{n+l}\}$ se denomina período, y el menor l para el que se verifica la definición determina la longitud del período. Si el período comienza en a_0, se dice que la fracción continua es periódica pura.

Teorema 4.4 (Lagrange, 1770). *Un número es un irracional cuadrático si y solo si la fracción continua simple que lo representa es periódica.*

Es más, si dicho irracional cuadrático es reducido, es decir, si es mayor que 1 y su conjugado se encuentra en el intervalo $(0, 1)$, entonces se corresponde con una fracción continua periódica pura.

Ejemplo 4.2. Si $\theta = \langle\overline{2, 3}\rangle = 2 + \frac{1}{3+\frac{1}{\theta}}$, podemos calcular su valor, pues esto no es más que una ecuación cuadrática en θ. Quedándonos con la raíz positiva, $\theta = ((3 + \sqrt{15})/3)$. Ahora, podemos calcular también el valor de $\xi = \langle 4, 1, \overline{2, 3}\rangle = \langle 4, 1, \theta\rangle$

$$\xi = 4 + (1 + \theta^{-1})^{-1} = 4 + \frac{\theta}{\theta + 1} = \frac{29 + \sqrt{15}}{7}.$$

Obsérvese que, si (x, y) es una solución de $x^2 - dy^2 = 1$, entonces $\frac{x^2}{y^2} = d + \frac{1}{y^2}$, por lo que, para y suficientemente grande, $\frac{x}{y}$ es una buena aproximación de $\sqrt{d}$. El siguiente resultado presenta la forma de su fracción continua correspondiente.

Teorema 4.5. *Sea $d > 0$ un entero no cuadrado. La fracción continua de $\sqrt{d}$ es periódica. Más precisamente, si l denota la longitud del período,*

$$\sqrt{d} = \langle a_0, \overline{a_1, \ldots, a_{l-1}, 2a_0}\rangle, \text{ con } a_0 = \lfloor\sqrt{d}\rfloor.$$

Puede probarse que toda solución de Pell se corresponde con un convergente de $\sqrt{d}$:

Teorema 4.6. *Si (x, y) es una solución de la ecuación de Pell, entonces $\frac{x}{y}$ es un convergente de la fracción continua de $\sqrt{d}$.*

Por otro lado, no todos los convergentes de $\sqrt{d}$ se corresponden con una solución, pero sí infinitos de ellos, como se muestra en el siguiente teorema:

Teorema 4.7. *Sea l la longitud del periodo de la fracción continua de $\sqrt{d}$. Para todo $n \geq 0$, se tiene que*

$$h_{nl-1}^2 - dk_{nl-1}^2 = (-1)^{nl}.$$

Notemos que, cuando nl sea par, tendremos exactamente la ecuación de Pell. Así, obtenemos infinitas soluciones dadas por los convergentes (h_{nl-1}, k_{nl-1}).

Teorema 4.8. *Sea l la longitud del período de la fracción continua de $\sqrt{d}$. La solución fundamental (x_1, y_1) de $x^2 - dy^2 = 1$ es*

$$(x_1, y_1) = \begin{cases} (h_{l-1}, k_{l-1}) & \text{si } l \text{ es par,} \\ (h_{2l-1}, k_{2l-1}) & \text{si } l \text{ es impar.} \end{cases}$$

Para su demostración, consúltese [Rob06]. Es este resultado el que vamos a implementar en Sage para que calcule la solución fundamental de la ecuación de Pell que queramos resolver. Seguiremos los siguientes pasos:

1. Comprobamos que $d \neq \square$ para asegurarnos de estar en las condiciones del teorema.
2. Definimos el cuerpo cuadrático sobre el que trabajamos, $\mathbb{Q}(\sqrt{d})$.
3. Almacenamos la fracción continua de $\sqrt{d}$, obtenemos el período y calculamos su longitud.
4. Formulamos las instrucciones que nos proporciona el teorema. Si la longitud es par, la solución fundamental viene dada por el convergente $r_{l-1} = \frac{h_{l-1}}{k_{l-1}}$. Si es impar, por r_{2l-1}.

Sage dispone de todas las herramientas que necesitamos para ello. Así, obtenemos el siguiente programa de elaboración propia:

```
sage: def convergent_fundamentalsol(d):

        if d.is_square()== False:

           Qd.<sqrtd>=QuadraticField(d)
           cfrac=continued_fraction(sqrtd)
           per=cfrac.period()
           lperiod=len(per)

           if mod(lperiod,2)==0:
              return cfrac.numerator(lperiod-1), cfrac.denominator(
                     lperiod-1)
           else:
              return cfrac.numerator(2*lperiod-1), cfrac.denominator(
                     2*lperiod-1)
        else:
           print("Error: d es cuadrado")
```

Ejemplo 4.3. Consideremos la ecuación de Pell $x^2 - 3y^2 = 1$. En este caso, es fácil ver, sin necesidad de cálculo, que la primera solución no trivial es $(2, 1)$. En Sage:

```
sage: d=3
      Qd.<sqrtd>=QuadraticField(d)
      per=continued_fraction(sqrtd).period()
      lper=len(per)
      show("El periodo de la fraccion continua de ", sqrt(d),
           " es ", per, ". Su longitud es l= ", lper)
```

Así, obtenemos que la fracción continua es $\sqrt{3} = \langle 1, \overline{1, 2} \rangle$, en la forma del Teorema 4.5. La longitud de su período es $l = 2$. Como l es par, `convergent_fundamentalsol(d)` nos da la solución que se corresponde con el convergente de índice $l - 1$. En efecto,

```
sage: convergent_fundamentalsol(d)
(2,1)
```

4.2 Evaluación del polinomio enumerador de primos

Ahora que tenemos un método para resolver las ecuaciones de Pell, tratemos de resolver el sistema de ecuaciones que define el polinomio enumerador de primos del Teorema 3.1 para un valor concreto del parámetro k. Hemos de satisfacer el sistema:

(I) $q = wz + h + j$,
(II) $z = (gk + g + k)(h + j) + h$,
(III) $(2k)^3(2k + 2)(n + 1)^2 + 1 = f^2$,
(IV) $e = p + q + z + 2n$,
(V) $e^3(e + 2)(a + 1)^2 + 1 = o^2$,
(VI) $x^2 = (a^2 - 1)y^2 + 1$,
(VII) $u^2 = 16(a^2 - 1)r^2y^4 + 1$,
(VIII) $(x + cu)^2 = ((a + u^2(u^2 - a))^2 - 1)(n + 4dy)^2 + 1$,
(IX) $m^2 = (a^2 - 1)l^2 + 1$,
(X) $l = k + i(a - 1)$,
(XI) $n + l + v = y$,
(XII) $m = p+l(a-n-1)+b(2a(n+1)-(n+1)^2-1)$,
(XIII) $x = q+y(a-p-1)+s(2a(p+1)-(p+1)^2-1)$,
(XIV) $pm = z + pl(a - p) + t(2ap - p^2 - 1)$.

La evaluación de un sistema equivalente a este para el primo 2 también se realiza en [Gup03]. Aunque la idea que proponemos aquí para la resolución es esencialmente la misma, lo resolvemos de manera independiente mediante los resultados obtenidos a lo largo de esta memoria. Seguiremos la demostración que hemos dado del sistema de ecuaciones que define al primo $k + 1$, y resolveremos las ecuaciones de Pell mediante el programa formulado en la sección anterior.

El Teorema 3.4 establece que, si $k+1$ es primo, entonces existe solución de este sistema.

Para buscar una posible solución para el primo 2, debemos tomar

$$k = 1.$$

Reconstruyamos los pasos de la demostración para encontrar un valor que pueda tomar cada una de las variables, de manera consistente con el resto:

(1.) Esta caracterización de los números primos se basa, en primer lugar, en la caracterización diofántica del factorial dada en la Proposición 3.3, estableciendo $k! = gk + g + k$ por el Teorema de Wilson. Dado $k = 1$, podemos despejar el valor de g. Así,

$$g = 0.$$

De la caracterización del factorial salen las ecuaciones (I)-(III) y tres condiciones adicionales que deben satisfacerse:

$$p = (n+1)^k, \quad q = (p+1)^n, \quad z = p^{k+1}. \tag{4.4}$$

Es más, en su demostración vimos que para que se verifiquen las ecuaciones (I) y (II), hemos de tomar

$$w = \frac{q - \mathrm{rem}(q,z)}{z}, \quad h = z - \mathrm{rem}(q,z), \quad j = \mathrm{rem}(q,z) - h. \tag{4.5}$$

La ecuación (III), sustituyendo el valor de k, resulta $32(n+1)^2 + 1 = f^2$, que no es más que una ecuación de Pell en las variables $(f, n+1)$. Tomaremos como posibles valores los de su solución fundamental, que podemos calcular mediante el programa que proporcionamos en la sección anterior, obteniendo así

$$(f, n+1) = (17, 3) \Rightarrow (f, n) = (17, 2).$$

Podemos sustituir en las condiciones (4.4) para obtener los valores de p, q y z

$$p = 3, \quad q = 4^2 = 16, \quad z = 3^2 = 9.$$

Del mismo modo, sustituimos ahora en (4.5) para obtener w, h y j

$$w = \frac{16 - 7}{9} = 1, \quad h = 9 - 7 = 2, \quad j = 7 - 2 = 5.$$

(2.) Las ecuaciones (IV) y (V) están basadas en el lema de carácter exponencial 3.22. Conocidos n, p, q y z, podemos despejar el valor de e de la ecuación (IV). Así,

$$e = 32.$$

Una vez sustituido el valor de e en la ecuación (V), tenemos una ecuación de Pell en las variables $(o, a+1)$. De nuevo, obtenemos su solución fundamental en Sage y tomamos

$$o = 834035301564579468329946270481226888212608613465610836377 7,$$
$$a = 790169035809889616168555687974994918632638071340929091 2.$$

(3.) Las ecuaciones (VI), (VII), (VIII) y (XI) resultan de la descripción diofántica de las soluciones de la ecuación de Pell (Corolario 3.2). Además, esta caracterización de las soluciones, junto con la ecuación (X), determina los valores que toman las variables de (IX). Dado n, se establece $(x, y) = (\chi_a(n), \psi_a(n))$. Aplicando esto a nuestra elección, como $n = 2$, vemos que en este caso no podemos considerar su solución fundamental, que ya vimos en el capítulo anterior que es $(\chi_a(1), \psi_a(1)) = (a, 1)$. Necesitamos calcular la siguiente solución, y podemos hacerlo, por ejemplo, mediante la expresión de las soluciones como sucesiones de Lucas:

$$\chi_a(0) = 1, \quad \chi_a(1) = a, \quad \chi_a(2) = 2a\chi_a(1) - \chi_a(0) = 2a \cdot a - 1 = 2a^2 - 1,$$
$$\psi_a(0) = 0, \quad \psi_a(1) = 1, \quad \psi_a(2) = 2a\psi_a(1) - \psi_a(0) = 2a \cdot 1 - 0 = 2a.$$

Por tanto, sustituyendo a por su valor, calculado previamente, consideramos

$$(x, y) = (2a^2 - 1, 2a).$$

En la demostración vimos que la ecuación de Pell en (IX) se satisface en el índice k de sus correspondientes sucesiones de Lucas. En nuestro caso, $k = 1$, luego lo que tenemos, esta vez sí, es la solución fundamental

$$(m, l) = (a, 1).$$

Conocidos los valores de a, k, l, n e y, podemos despejar de la ecuación (X) el valor de i y de la ecuación (XI) el valor de v. Así,

$$i = 0, \qquad v = 2a - 3.$$

Ahora, si sustituimos el valor de y en la ecuación (VII), nos queda la siguiente ecuación de Pell en las variables (u, r):

$$u^2 = \theta r^2 + 1, \text{ donde } \theta = 32a(a^2 - 1).$$

Si sustituimos el valor de a, θ tiene un valor gigantesco, y al proporcionárselo al programa para obtener la solución fundamental de esta ecuación, resulta

computacionalmente imposible; no somos capaces de obtener posibles valores para las variables u y r. Esto imposibilita también la obtención de valores para las variables x y c, pues no disponemos de un valor de u para poder tratar de resolver la ecuación (VII). En [Gup03] se analiza la complejidad de estos cuatro valores restantes, haciendo uso de los lemas sobre las propiedades de las soluciones de las ecuaciones de Pell.

(4.) Por último, obtengamos valores para b, s y t a partir de las ecuaciones (XII), (XIII) y (XIV). Despejando,

$$b = 0 \qquad s = 1, \qquad t = 0.$$

Tratando de resolver el sistema para el primo $k + 1 = 2$, hemos conseguido obtener posibles valores para 22 de las 26 variables. Es importante destacar que hemos intentado tomar los menores valores posibles de cada variable según las elecciones previas. Esto no asegura que cada variable tenga el menor valor posible para el primo 2, pues al no ser capaces de resolver totalmente el sistema, cabe la posibilidad de que este no tenga solución para estas elecciones.

Queda reflejado en estos cálculos lo complicado que resulta obtener valores primos mediante el polinomio enumerador de primos. Por un lado, dado k que viene determinado por el primo $k + 1$, para aplicar el Teorema de Wilson, hemos de calcular $k!$. Esto, para el primo 2 no resulta un problema, pero si consideramos primos grandes, puede llegar a serlo. Por otro lado, tenemos que lidiar con ecuaciones de Pell cuyas soluciones fundamentales son valores excesivamente grandes e incluso incalculables. Es más: el resto de posibles valores para las variables de las ecuaciones de Pell del sistema son aún mayores, pues son potencias de esta primera solución no trivial.

5 ¡Reducto!

En la sección introductoria a los polinomios enumeradores de primos afirmamos la existencia de polinomios con grado menor o con un número menor de variables que el que hemos construido en el tercer capítulo. Existen resultados universales para cualquier conjunto diofántico que nos aseguran la existencia de polinomios enumeradores de primos de menor grado o con un número menor de variables. No obstante, en general no es posible acotar ambos parámetros de manera simultánea; la disminución del grado resulta en un aumento de las variables, y viceversa. Este hecho es consecuencia de los resultados universales que presentamos a continuación.

Todo conjunto diofántico es recursivamente enumerable (r.e.), y vimos en el segundo capítulo que el recíproco también es cierto (Teorema DPRM), es decir, que todo conjunto r.e. D se puede representar como

$$x \in D \iff (\exists z_1, \dots, z_\nu)(P(x, z_1, \dots, z_\nu) = 0).$$

En [Dav73] se muestra que puede definirse $\{D_n\}_{n\geq 1}$, con ayuda de funciones de paridad, de modo que la sucesión $D_1, D_2, \dots$ contiene a todos los conjuntos r.e. Es más, la relación $x \in D_n$ es también r.e. (*Universality Theorem* en [Dav73]). En consecuencia, el Teorema DPRM implica que, para ciertos δ y ν, existe un polinomio U de grado δ en $z_1, \dots, z_\nu$ tal que, para cualesquiera x y n,

$$x \in D_n \iff (\exists z_1, \dots, z_\nu)(U(x, n, z_1, \dots, z_\nu) = 0). \tag{5.1}$$

Esta ecuación U define a todos los conjuntos diofánticos de manera simultánea. Así, fijado un n, que determina un conjunto diofántico concreto, se tiene que su ecuación diofántica correspondiente $P = 0$ es equivalente a otra ecuación $U = 0$ con el mismo parámetro (o parámetros, si hay más de uno), siendo el polinomio U de grado δ en $\nu + 1$ variables.

La ecuación $U = 0$ se denomina ecuación diofántica universal, y el par (ν, δ), par universal. Existen pares fijos de variables–grado de modo que todo conjunto diofántico

puede ser expresado como en (5.1). Más información sobre este tema puede encontrarse en [Jon82], donde además proporciona pares universales (ν, δ) (véase el Teorema 4 del artículo citado). Entre otros, son pares universales $(38, 8)$ y $(11, 4.6 \times 10^{44})$. De este modo, fijado cualquier par universal (ν, δ), existen polinomios enumeradores de primos de grado $\delta + 1$ en $\nu + 1$ variables.

5.1 Reducción del número de variables

Centrémonos primero en la reducción de las variables. ¿Cuál es el mínimo número de variables necesario para construir un polinomio enumerador de primos? Esto se oculta bajo el concepto de dimensión de un conjunto diofántico:

Definición 5.1 (Dimensión de un conjunto diofántico). *Se define la dimensión de un conjunto diofántico D como el menor $n \in \mathbb{N}$ para el que existe un polinomio P tal que*

$$D = \{y \mid (\exists x_1, \dots, x_n)(P(x_1, \dots, x_n, y) = 0)\}.$$

En [Dav73] (pág. 263) encontramos un resultado que nos asegura la existencia de una cota superior en la dimensión de todo conjunto diofántico, esto es, en el número de variables en los que puede definirse una ecuación diofántica para ese conjunto:

Teorema 5.1. *Existe un entero n para el que todo conjunto conjunto diofántico tiene dimensión menor o igual que n.*

En 1975, Julia Robinson y Yuri Matijasevič demuestran que todo conjunto diofántico puede definirse con $n = 14$ variables (ref. [MR75]). Para la demostración de este resultado, se usan gran parte de las técnicas tratadas en el tercer capítulo. En particular, aportan una definición diofántica de las soluciones de la ecuación de Pell alternativa a la dada en la Proposición 3.2:

Proposición 5.1. Sean $A > 1$, $B > 1$ y $C > 0$. Entonces, $\psi_A(B) = C$ si y solo si existen $i, j, k, D, E, F, G, H, I$ verificando

(1) $DFI = \square$, $F \mid H - C$, $B \leq C$,
(2) $D = (A^2 - 1)C^2 + 1$,
(3) $E = 2(i + 1)D(k + 1)C^2$,
(4) $F = (A^2 - 1)E^2 + 1$,
(5) $G = A + F(F - A)$,
(6) $H = B + 2(j + 1)C$,
(7) $I = (G^2 - 1)H^2 + 1$.

El método de prueba es similar al que hemos seguido en este tipo de resultados; la demostración puede consultarse en la referencia citada. Recordemos que ya hicimos notar en la Observación 3.2 que las relaciones de orden usadas son diofánticas. Del mismo modo, la relación de divisibilidad que aparece en este sistema también lo es. En efecto, $x \mid y \iff (\exists z)(xz = y)$.

Esta nueva definición diofántica de las soluciones de la ecuación de Pell es la que utilizan Jones et al. para demostrar la existencia de un polinomio enumerador de primos en 12 variables. En el tercer capítulo empleamos lo que denominan *congruence method*. Para la construcción de este polinomio, aplican algunos de los lemas que recogen las propiedades de las soluciones de la ecuación de Pell, pero principalmente hacen uso de nuevos resultados que derivan una construcción diferente: utilizan la técnica desarrollada en [MR75], que recogen bajo el nombre de *ratio method*. Este método es, generalmente, más económico respecto al número de variables, aunque el grado aumenta sustancialmente. Enunciaremos aquí solo el resultado final: un sistema diofántico que define un polinomio enumerador de primos en 12 variables.

Teorema 5.2. *Para cualquier entero positivo k, se tiene que $k+1$ es primo si y solo si el sistema* (1) − (21) *tiene solución en los enteros no negativos:*

(1) $(2k+2)^3(2k+4)(n+1)^2+1 = \square$,
(2) $(2n+2)^3(2n+4)(x+1)^2+1 = \square$,
(3) $M = 16nx(w+2)+1$,
(4) $A = M(x+1)$,
(5) $B = n+1$,
(6) $C = m+B$,
(7) $DFI = \square, F \mid H-C$,
(8) $D = (A^2-1)C^2+1$,
(9) $E = 2(i+1)DC^2$,
(10) $F = (A^2-1)E^2+1$,
(11) $G = A+F(F-A)$,
(12) $H = B+2(j+1)C$,
(13) $I = (G^2-1)H^2+1$,
(14) $\left[\dfrac{R}{\left(\frac{C}{KL}-(w+1)x\right)\left(1-\frac{R}{C}\right)^2 L}-(S+1)\right]^2 < \frac{1}{4}$,
(15) $(M^2-1)K^2+1 = \square$,
(16) $(M^2x^2-1)L^2+1 = \square$,
(17) $(M^2n^2x^2-1)R^2+1 = \square$,
(18) $K = n-k+1+p(M-1)$,
(19) $L = k+1+l(Mx-1)$,
(20) $R = k+1+r(Mnx-1)$,
(21) $S = (z+1)(k+1)-2$.

La prueba se encuentra en [JSWW76]. Las variables $A, B, C, D, E, F, G, H, I, K, L, M, R$ y S pueden ser eliminadas del sistema por sustitución, resultando un sistema compuesto por 10 variables $i, j, l, m, n, p, r, w, x, z$, el parámetro k que determina el primo, 6 expresiones igualadas a un cuadrado, una condición de divisibilidad y una desigualdad. Estas condiciones pueden ser definidas mediante una sola variable gracias al *relation-combining theorem*, expuesto en [MR75]. Así, siguiendo el método de Putnam, podemos elevar al cuadrado todas las ecuaciones y unirlas en una única ecua-

ción $M = 0$, obteniendo un polinomio enumerador de primos $P = (k+2)(1-M)$ en 12 variables. El grado del polinomio se indica que es 148864, aunque mediante una serie de consideraciones, combinaciones y modificaciones puede quedar reducido a grado 13697.

Al año de la publicación de este resultado, Matijasevič escribió un artículo que fue posteriormente publicado en 1981. En él demuestra la existencia de un polinomio enumerador de primos que reduce en dos el número de variables con respecto al que produce este sistema, esto es, existe un polinomio enumerador de primos en 10 variables (véase [Mat81]).

No hemos encontrado polinomios enumeradores de primos con un número menor de variables, a excepción de un resultado sujeto a una conjetura. Jones afirmó lo siguiente: supuesto cierto el recíproco del Teorema de Wolstenholme, dicho teorema proporcionaría otra caracterización para los primos, alternativa a la del Teorema de Wilson, y sería posible la construcción de un polinomio enumerador de primos en 8 variables mediante una representación diofántica del coeficiente binomial del Teorema de Wolstenholme, $\binom{2p-1}{p-1}$. Esta idea se desarrolla en [VO15].

En los resultados de universalidad, el menor número de variables ν para el que se ha demostrado la existencia de una ecuación diofántica universal es $\nu = 9$ variables, dada en [Jon82], y es evidente que al menos dos variables son necesarias. En este rango el problema sigue abierto, tanto en el caso de los primos como el caso universal.

5.2 Reducción del grado

Centrémonos ahora en la reducción del grado de los polinomios enumeradores de primos. ¿Cuál es el mínimo grado para el que existe un polinomio enumerador de primos de tal grado? Presentamos el concepto del grado de un conjunto diofántico:

Definición 5.2 (Grado de un conjunto diofántico). *Se define el grado del conjunto D como el menor $m \in \mathbb{N}$ para el que existe un polinomio P de grado m satisfaciendo la definición diofántica, esto es,*

$$D = \{y \mid (\exists x_1, \dots, x_n)(P(x_1, \dots, x_n, y) = 0)\}.$$

El siguiente resultado universal, basado en el método de sustitución de Skolem, proporciona una cota superior en el grado de un conjunto diofántico.

Teorema 5.3. *Todo conjunto diofántico tiene grado menor o igual que 4.*

Demostración. El grado del polinomio P que define el sistema diofántico D puede ser reducido incluyendo variables adicionales z_j como sigue:

$$\begin{aligned} z_j &= x_i x_k \\ z_j &= x_k^2 \\ z_j &= y x_i \\ z_j &= y^2. \end{aligned}$$

Realizando sucesivas sustituciones, el grado de P puede reducirse a 2. De este modo, la ecuación $P = 0$ es equivalente a un sistema de ecuaciones simultáneas de grado a lo sumo 2. Sumando los cuadrados, se obtiene una ecuación de grado a lo sumo 4.

Este resultado proporciona un método efectivo para reducir el grado de un polinomio enumerador de primos. Recordemos que los números primos forman un conjunto diofántico definido por la ecuación $M = 0$, siendo M la suma de los cuadrados de las ecuaciones del sistema del Teorema 3.4. Es a este polinomio M al que hemos de aplicar estas reducciones de cuadrados y de productos que se proponen en la demostración, quedando M reducido a grado 2. Así, la ecuación $M = 0$ es equivalente a un sistema de ecuaciones de grado 2, y sumando los cuadrados se obtiene una nueva ecuación $M = 0$ de grado 4. El polinomio enumerador de primos resulta de grado 5 como consecuencia del método de Putnam, esto es, de tomar $P = (k+2)\{1 - M\}$.

No hemos encontrado polinomio enumeradores de primos de grado menor. Gracias a los resultados de universalidad, hemos probado que el conjunto de los números primos tiene grado menor o igual que 4. Además, resulta que no existen ecuaciones diofánticas universales de grado $\delta = 2$. Sin embargo, las ecuaciones diofánticas de segundo grado son decidibles, y por tanto definen únicamente conjuntos recursivos (Siegel, [Sie72]). El conjunto de los números primos es, en efecto, recursivo, por lo que, en principio, puede que exista un polinomio enumerador de primos cuadrático. La existencia de una ecuación diofántica universal de grado $\delta = 3$ es un problema que sigue abierto.

Bibliografía

[AdRJ12] Juan Arias de Reyna and Toulisse Jeremy. The n-th prime asymptotically. *arXiv e-prints*, pages arXiv–1203, 2012.

[Buc46] R Creighton Buck. Prime-representing functions. *The American Mathematical Monthly*, 53(5):265–265, 1946.

[CC10] Chris K Caldwell and Yuanyou Furui Cheng. Determining mills' constant and a note on honaker's problem. *arXiv preprint arXiv:1010.4883*, 2010.

[Dav53] Martin Davis. Arithmetical problems and recursively enumerable predicates. *The journal of symbolic logic*, 18(1):33–41, 1953.

[Dav73] Martin Davis. Hilbert's tenth problem is unsolvable. *The American Mathematical Monthly*, 80(3):233–269, 1973.

[DPR61] Martin Davis, Hilary Putnam, and Julia Robinson. The decision problem for exponential diophantine equations. *Annals of Mathematics*, 74(3):425–436, 1961.

[DSW94] Martin Davis, Ron Sigal, and Elaine J Weyuker. *Computability, complexity, and languages: fundamentals of theoretical computer science*. Elsevier, 1994.

[FGG+19] Dylan Fridman, Juli Garbulsky, Bruno Glecer, James Grime, and Massi Tron Florentin. A prime-representing constant. *The American Mathematical Monthly*, 126(1):70–73, 2019.

[G31] Kurt Gödel. *On Formally Undecidable Propositions of Principia Mathematica and Related Systems*. Basic Books, New York, NY, USA, 1931.

[Gup03] Nachiketa Gupta. *Finding a Solution to the Diophantine Representation of the Primes*. PhD thesis, University of Pennsylvania, 2003.

[Hil02] David Hilbert. Mathematical problems (transl. mw newson). *Bull. Amer. Math. Soc*, 8:437–479, 1902.

[HW79] Godfrey Harold Hardy and Edward Maitland Wright. *An introduction to the theory of numbers*. Oxford university press, 1979.

[Jon75] James P. Jones. Formula for the nth prime number. *Canadian Mathematical Bulletin*, 18(3):433–434, 1975.

[Jon79] James P. Jones. Diophantine representation of mersenne and fermat primes. *Acta Arithmetica*, 35(3):209–221, 1979.

[Jon82] James P. Jones. Universal diophantine equation. *The Journal of Symbolic Logic*, 47(3):549–571, 1982.

[JSWW76] James P Jones, Daihachiro Sato, Hideo Wada, and Douglas Wiens. Diophantine representation of the set of prime numbers. *The American Mathematical Monthly*, 83(6):449–464, 1976.

[JW09] Michael J Jacobson and Hugh C Williams. *Solving the Pell equation*. Springer, 2009.

[Mat70] Yuri Matijasevič. Enumerable sets are diophantine. In *Soviet Math. Dokl.*, volume 11, pages 354–358, 1970.

[Mat71] Yuri Matiyasevič. Diophantine representation of the set of prime numbers. In *Dokl. Akad. Nauk SSSR*, volume 196(4), pages 770–773, 1971.

[Mat81] Yuri Matijasevič. Primes are nonnegative values of a polynomial in 10 variables. *Journal of Soviet Mathematics*, 15:33–44, 1981.

[Mil47] William H Mills. A prime-representing function. *Bull. Amer. Math. Soc*, 53(6):604, 1947.

[MR75] Yuri Matijasevič and Julia Robinson. Reduction of an arbitrary diophantine equation to one in 13 unknowns. *Acta Arithmetica*, 1(27):521–553, 1975.

[NZM91] Ivan Niven, Herbert S Zuckerman, and Hugh L Montgomery. *An introduction to the theory of numbers*. John Wiley & Sons, 1991.

[OEI] OEIS Foundation Inc. The On-Line Encyclopedia of Integer Sequences. Published electronically at `http://oeis.org`.

[Put60] Hilary Putnam. An unsolvable problem in number theory. *The Journal of Symbolic Logic*, 25(3):220–232, 1960.

[Rei43] Irving Reiner. Functions not formulas for primes. *The American Mathematical Monthly*, 50(10):619–621, 1943.

[Rib12] Paulo Ribenboim. *The new book of prime number records*. Springer Science & Business Media, 2012.

[Rob69] Julia Robinson. Unsolvable diophantine problems. *Proceedings of the American Mathematical Society*, 22(2):534–538, 1969.

[Rob06] Neville Robbins. *Beginning number theory*. Jones & Bartlett Learning, 2006.

[San08] Kamron Saniee. A simple expression for multivariate lagrange interpolation. *SIAM undergraduate research online*, 1(1):1–9, 2008.

[Sie72] C.L. Siegel. *Zur Theorie Der Quadratischen Formen, Von C.L. Siegel*. Akademie der Wissenschaften, Göttingen. Mathematisch-Physikalische Klasse. Nachrichten, Jahrg. 1972, Nr. 3. 1972.

[VO15] Luca Vallata and Eugenio G. Omodeo. A diophantine representation of wolstenholme's pseudoprimality. In *Italian Conference on Computational Logic*, 2015.

[Zyk] Bradley Zykoski. An approach to pell's equation via algebraic number theory.