Integer partitions detect the primes

Jan-Willem van Ittersum
Korteweg-de Vries Institute for Mathematics
University of Amsterdam

November 5, 2025

The ancient dream: detecting the primes

- Mathematicians have long wondered whether the set of **prime numbers** can be described by a simple equation.
- In the 1970s, Yuri Matiyasevich proved that every "computably enumerable" set of integers can be captured by a **Diophantine equation** — that is, by a polynomial equation with integer solutions.
- Amazingly, this means that also the primes can be characterized in this way!
- A few years later, Jones, Sato, Wada, and Wiens actually wrote down such a polynomial — a **25th-degree monster in 26 variables** whose positive values are exactly the primes.
- It works, but it's... not exactly practical.

Is there an easier way to detect primes?

DIOPHANTINE REPRESENTATION OF THE SET OF PRIME NUMBERS

JAMES P. JONES, DAIHACHIRO SATO, HIDEO WADA AND DOUGLAS WIENS

1. Introduction. Martin Davis, Yuri Matijasevič, Hilary Putnam and Julia Robinson [4] [8] have proven that every recursively enumerable set is Diophantine, and hence that the set of prime numbers is Diophantine. From this, and work of Putnam [12], it follows that the set of prime numbers is representable by a polynomial formula. In this article such a prime representing polynomial will be exhibited in explicit form. We prove (in Section 2)

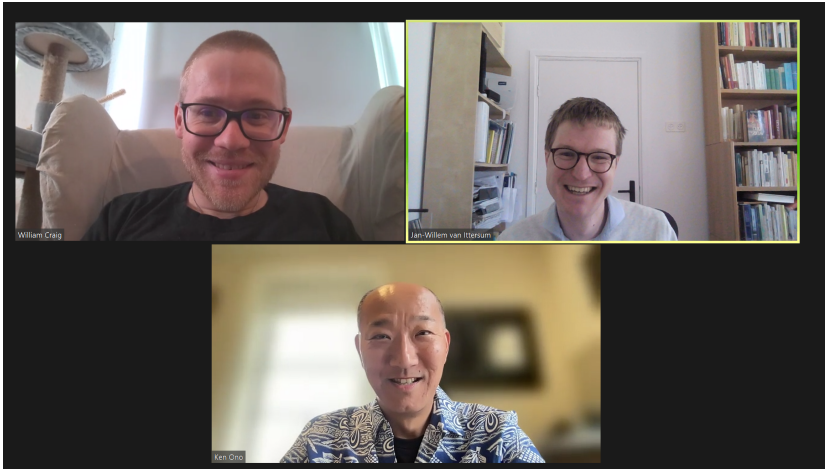
THEOREM 1. *The set of prime numbers is identical with the set of positive values taken on by the polynomial*

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

as the variables range over the nonnegative integers.

A 25-degree polynomial in 26 variables that detects primes...

Joint work with Will Craig and Ken Ono



Partitions: building blocks of additive number theory

Definition (Partition)

A *partition* of an integer n is a finite ordered sequence $(\lambda_1, \lambda_2, \ldots, \lambda_r)$ of integers $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_r \geq 1$ summing to n.

The conjugacy classes of the symmetric group S_n are labeled by partitions.

Example (Partitions of 4)

$(4), (3,1), (2,2), (2,1,1), (1,1,1,1)$

Write $p(n)$ for the number of partitions of n. Euler observed

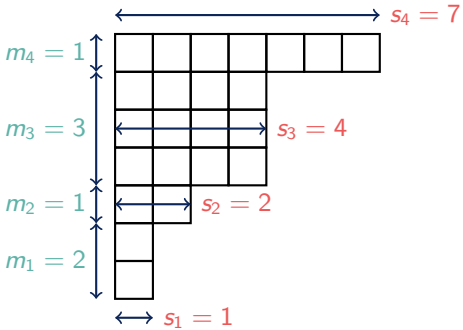
$$\sum_{n \geq 0} p(n)\, q^n = \frac{1}{\prod_{n \geq 1}(1 - q^n)}.$$

Partitions are determined by their part sizes and multiplicities

Note that a partition is uniquely determined by the different part sizes $s_1 < s_2 < \ldots < s_a$ and corresponding multiplicities $m_1, m_2, \ldots, m_a$.

Example (Stanley coordinates)

For $\lambda = (7, 4, 4, 4, 2, 1, 1) = (7^1, 4^3, 2^1, 1^2)$, we have $\mathbf{s} = (1, 2, 4, 7)$ and $\mathbf{m} = (2, 1, 3, 1)$.



MacMahon partition function counts sums of products of multiplicities

Definition (MacMahon partition function, 1920)

For $a \geq 1$, define the *MacMahon partition function* by

$$M_a(n) = \sum_{\substack{n=m_1 s_1 + \cdots + m_a s_a \\ 0 < s_1 < s_2 < \cdots < s_a}} m_1 m_2 \cdots m_a$$

Consider

$$\Psi(n) := (n^2 - 3n + 2)M_1(n) - 8M_2(n).$$

Example (The value $\Psi(3)$)

$n = 3, a = 1$: $\lambda = (3)$ or $\lambda = (1^3)$, so $M_1(3) = 1 + 3 = 4$.
$n = 3, a = 2$: $\lambda = (2, 1)$, so $M_2(3) = 1 \cdot 1 = 1$.
Hence,

$$\Psi(3) = 2M_1(3) - 8M_2(3) = 0.$$

MacMahon partition function (continued)

Reminder

$$M_a(n) = \sum_{\substack{n=m_1s_1+\cdots+m_as_a \\ 0<s_1<s_2<\cdots<s_a}} m_1m_2\cdots m_a, \qquad \Psi(n) := (n^2-3n+2)M_1(n) - 8M_2(n).$$

Example (The value $\Psi(4)$)

$n=4, a=1$: $\lambda = (4), (2^2)$ or $\lambda = (1^4)$, so $M_1(4) = 1+2+4 = 7$.
$n=4, a=2$: $\lambda = (3,1)$ or $(2,1^2)$, so $M_2(4) = 1\cdot 1 + 1\cdot 2 = 3$.
Hence,

$$\Psi(4) = 6M_1(4) - 8M_2(4) = 18.$$

n	2	3	4	5	6	7	8	9	10	11	12	13	14	15	...
$\Psi(n)$	0	0	18	0	120	0	270	192	504	0	1680	0	1296	1536	...

The MacMahon partition function detects primes

Reminder

$$M_a(n) = \sum_{\substack{n=m_1s_1+\cdots+m_as_a \\ 0<s_1<s_2<\cdots<s_a}} m_1m_2\cdots m_a, \qquad \Psi(n) := (n^2-3n+2)M_1(n) - 8M_2(n).$$

Theorem (Craig–vI–Ono, '24)

For $n \geq 2$ we have

- $\Psi(n) \geq 0$
- $\Psi(n) = 0$ *if and only if n is prime.*

The MacMa

$$(n^2 - 3n + 2)M_1(n) - 8M_2(n)$$

$$(3n^3 - 13n^2 + 18n - 8)M_1(n) + (12n^2 - 120n + 212)M_2(n) - 960M_3(n)$$

$$(25n^4 - 171n^3 + 423n^2 - 447n + 170)M_1(n) + (300n^3 - 3554n^2 + 12900n - 14990)M_2(n) +$$
$$(2400n^2 - 60480n + 214080)M_3(n) - 725760M_4(n)$$

$$(126n^5 - 1303n^4 + 5073n^3 - 9323n^2 + 8097n - 2670)M_1(n) +$$
$$(3024n^4 - 48900n^3 + 288014n^2 - 737100n + 695490)M_2(n) +$$
$$(60480n^3 - 1510080n^2 + 10644480n - 23496480)M_3(n) +$$
$$(725760n^2 - 36288000n + 218453760)M_4(n) - 580608000M_5(n)$$

$$(300n^8 - 1542n^7 - 33049n^6 + 377959n^5 - 1651959n^4 + 3726801n^3 - 4575760n^2 + 2903750n - 746500)M_1(n) +$$
$$(12000n^7 - 91008n^6 - 2799900n^5 + 50637162n^4 - 351366300n^3 + 1239098170n^2 - 2210467000n + 1585493500)M_2(n) +$$
$$(432000n^6 - 3548160n^5 - 236343840n^4 + 5133219840n^3 - 42370071840n^2 + 161101416000n - 236150560800)M_3(n) +$$
$$(12096000n^5 - 72817920n^4 - 17599680000n^3 + 396192142080n^2 - 3123876672000n + 8555162112000)M_4(n) +$$
$$(193536000n^4 - 1056513024000n^2 + 21310248960000n - 112944125952000)M_5(n) +$$
$$(-46495088640000n + 604436152320000)M_6(n) - 11158821273600000M_7(n)$$

Reminder

$M_a(n) =$ $8M_2(n)$.

Theorem (C

For $n \geq 2$ we

- $\Psi(n) \geq$
- $\Psi(n) =$

Generating series of MacMahon partition functions yields divisor sums

Let $\mathcal{U}_a(q) := \sum_{n\geq 0} M_a(n)\, q^n$ and $\sigma_{k-1}(n) = \sum_{d|n} d^{k-1}$. Observe

- $\mathcal{U}_1(q) = \displaystyle\sum_{n\geq 0} M_1(n)\, q^n = \sum_{n\geq 0} \sum_{\substack{n=m_1 s_1 \\ 0<s_1}} m_1\, q^n = \sum_{n\geq 1} \sigma_1(n)\, q^n.$
- $\mathcal{U}_1(q)^2 = \displaystyle\sum_{\substack{n\geq 0 \\ n=m_1 s_1+m_2 s_2 \\ 0<s_1,\, 0<s_2}} m_1 m_2\, q^n = \Bigg(\underbrace{\sum_{\substack{n\geq 0 \\ n=m_1 s_1+m_2 s_2 \\ 0<s_1<s_2}} + \sum_{\substack{n\geq 0 \\ n=m_1 s_1+m_2 s_2 \\ 0<s_2<s_1}}}_{2\mathcal{U}_2(q)} + \sum_{\substack{n\geq 0 \\ n=(m_1+m_2)s_1 \\ 0<s_1=s_2}} \Bigg) m_1 m_2\, q^n.$

Hence,

$$\mathcal{U}_1(q)^2 - 2\mathcal{U}_2(q) = \sum_{\substack{n\geq 0 \\ n=ms_1}} \left(\sum_{m_1+m_2=m} m_1 m_2 \right) q^n = \sum_{\substack{n\geq 0 \\ n=ms_1}} \frac{m^3-m}{6} q^n = \sum_{n\geq 1} \frac{\sigma_3(n)-\sigma_1(n)}{6} q^n.$$

Observation

Both $\mathcal{U}_1$ and $\mathcal{U}_2$ can be expressed in terms of divisor sums.

Theorem (Hoffman–Ihara, '17)

$$\sum_{a\geq 0}\mathcal{U}_a(q)\,x^a = \exp\left(\sum_{k\geq 1}\frac{(-1)^{k+1}}{k}x^k\sum_{n\geq 1}\sum_{d\mid n}\binom{d+k-1}{d-k}q^n\right).$$

Proof idea.

$\mathcal{U}_a$ is an element of a quasi-shuffle algebra.

□

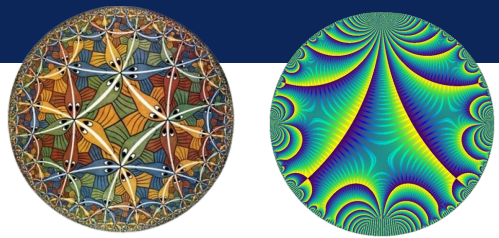
For $k \geq 2$, write

- $B_k := k$th Bernoulli number;
- $\sigma_{k-1}(n) := \sum_{d\mid n} d^{k-1}$;
- $G_k(q) := -\dfrac{B_k}{2k} + \displaystyle\sum_{n\geq 1}\sigma_{k-1}(n)q^n$;
- $\widetilde{M} := \mathbb{Q}[G_2, G_4, G_6, \ldots]$.

Corollary (Andrews–Rose, '11)

For all $a \geq 1$ we have $\mathcal{U}_a \in \widetilde{M}$.

Elements of $\widetilde{M}$ are *quasimodular forms*



Artistic impressions of a hyperbolic tiling (Escher) and a modular form (Lowry-Duda)

Definition (Modular form)

A q-series $f \in \mathbb{Q}[\![q]\!]$ is *modular* of weight k if

- f is holomorphic for $|q| < 1$;
- $f(q) = \tau^k f(\widetilde{q})$
  for $q = e^{2\pi i \tau}$, $\widetilde{q} = e^{-\frac{2\pi i}{\tau}}$, $\mathrm{Im}(\tau) > 0$.

Examples

For $k \geq 4$ even, the series G_k is modular.
Also $\Delta(q) := q \prod_{n=1}^{\infty}(1-q^n)^{24}$ is modular.

Note that $G_2(q) = \tau^2 G_2(\widetilde{q}) - \frac{1}{4\pi i}\tau$.
Elements of $\widetilde{M}$ are called *quasimodular forms*.

The operator $D := q\frac{d}{dq}$ acts on $\widetilde{M}$:

Theorem (Kaneko–Zagier, '95)

The algebra $\widetilde{M}$ is a graded differential algebra, freely generated by G_2, G_4 and G_6.

Example (Discriminant modular form)

$$\Delta = \frac{(240\,G_4)^3 - (504\,G_6)^2}{1728}$$

One can prove various identities using modular forms

Example

$$DG_2 = -2G_2^2 + \tfrac{5}{6}G_4, \quad \text{that is} \quad n\,\sigma_1(n) = \frac{5}{6}\sigma_3(n) + \frac{1}{6}\sigma_1(n) - 2\sum_{a+b=n}\sigma_1(a)\,\sigma_1(b).$$

Example

$$G_8 = 120\,G_4^2, \quad \text{that is} \quad \sigma_7(n) = \sigma_3(n) + 120\sum_{a+b=n}\sigma_3(a)\,\sigma_3(b).$$

Example (Lagrange's four-square theorem)

$$\sum_{a,b,c,d\in\mathbb{Z}} q^{a^2+b^2+c^2+d^2} = 8G_2(q) - 32G_2(q^4), \quad \text{that is} \quad r_4(n) = 8\sigma_1(n) - 32\sigma_1(n/4),$$

where $r_4(n)$ denotes the number of ways to write n as a sum of four squares.

Some quasimodular forms detect primes

Consider

$$f_{k,\ell}(q) := (D^\ell + 1)G_{k+1} - (D^k + 1)G_{\ell+1} = c_0 + \sum_{n\geq 1}\left(\sum_{d|n}(n^\ell+1)d^k - (n^k+1)d^\ell\right)q^n.$$

Note that for $d = 1$ one gets $(n^\ell + 1) \quad - (n^k + 1) \quad = \quad n^\ell - n^k$.
for $d = n$ one gets $(n^\ell + 1)n^k - (n^k + 1)n^\ell = -n^\ell + n^k$.

Hence,

$$f_{k,\ell}(q) = c_0 + \sum_{n\geq 1}\Bigg(\sum_{\substack{d|n\\ 0<d<n}}(n^\ell+1)d^k - (n^k+1)d^\ell\Bigg)q^n.$$

Lemma (Lelièvre, '04)

The coefficients of $f_{k,\ell}$ vanish at primes.

Overview: why the expression Ψ detects primes

- The expression $\Psi(n) = (n^2 - 3n + 2)M_1(n) - 8M_2(n)$ are the coefficients of the quasimodular form
$$F = (D^2 - 3D + 2)G_2 - G_4.$$

- We have $(D + 1)F = f_{1,3}$, which is prime-detecting.
- Hence, for $n \geq 2$ we have $\Psi(n) = 0$ iff n is prime.
- For the other four expressions the same strategy works.

Question

Can't there be more prime-detecting expressions in the MacMahon functions?

Cusp forms cannot detect primes

Let $f = \sum_{n \geq 0} a_n q^n$ be *modular* of weight k. Assume $a_0 = 0$ (f is a cusp form).

Theorem (Deligne, '74)

For all p prime, $|a_p| \leq 2p^{\frac{k-1}{2}} |a_1|$.

The space of cusp forms admits a natural basis of so-called *Hecke eigenforms*.

Theorem (Eichler–Shimura, Igusa, Deligne, '71)

For each Hecke eigenform f and prime ℓ, there exists an irreducible continuous group homomorphism

$$\rho_f : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}_2(\overline{\mathbb{Q}}_\ell)$$

such that if p prime with $p \neq \ell$

$$\mathrm{Tr}(\rho_f(\mathrm{Frob}_p)) = a_p \quad \text{and} \quad \det(\rho_f(\mathrm{Frob}_p)) = p^{k-1}.$$

Theorem (Serre, Murthy, '83)

The sequence $\{a_p\}_{p \text{ prime}}$ has infinitely many sign changes.

Theorem (Serre, Ribet, Ono–Skinner, '98)

For all sufficiently large primes ℓ and $0 \leq a < \ell$, we have

$$a_p \equiv a \mod \ell$$

for infinitely many primes p.

There are no further prime-detecting combinations of MacMahon functions

Question

Can't there be more prime-detecting expressions in the MacMahon functions?

- Cusp forms cannot detect primes.
- Also, 'quasimodular cusp forms' cannot detect primes.
- In the five examples, the MacMahon functions were quasimodular not involving cusp forms.
- 'Generalically', linear combinations of MacMahon functions admit a cuspidal part and hence cannot detect primes.

$(n^2-3n+2)M_1(n)-8M_2(n)$
$(3n^3-13n^2+18n-8)M_1(n)+(12n^2-120n+212)M_2(n)-960M_3(n)$
$(25n^4-171n^3+423n^2-447n+170)M_1(n)+(300n^3-3554n^2+12900n-14990)M_2(n)+$ $(2400n^2-60480n+214080)M_3(n)-725760M_4(n)$
$(126n^5-1303n^4+5073n^3-9323n^2+8097n-2670)M_1(n)+$ $(3024n^4-48900n^3+288014n^2-737100n+695490)M_2(n)+$ $(60480n^3-1510080n^2+10644480n-23496480)M_3(n)+$ $(725760n^2-36288000n+218453760)M_4(n)-580608000M_5(n)$
$(300n^8-1542n^7-33049n^6+377959n^5-1651959n^4+3726801n^3-4575760n^2+2903750n-746500)M_1(n)+$ $(12000n^7-91008n^6-2799900n^5+50637162n^4-351366300n^3+1239098170n^2-2210467000n+1585493500)M_2(n)+$ $(432000n^6-3548160n^5-236343840n^4+5133219840n^3-42370071840n^2+161101416000n-236150560800)M_3(n)+$ $(12096000n^5-72817920n^4-17599680000n^3+396192142080n^2-3123876672000n+8555162112000)M_4(n)+$ $(193536000n^4-1056513024000n^2+21310248960000n-112944125952000)M_5(n)+$ $(-46495088640000n+604436152320000)M_6(n)-1115882127360000M_7(n)$