Computer discovery of new mathematical facts and formulas

David H. Bailey
http://www.davidhbailey.com
Lawrence Berkeley National Laboratory (retired)
University of California, Davis, Department of Computer Science
In October 1992, I received this fax from the Simpsons TV show, requesting the 40,000th digit of π. I computed the first 40,000 digits, and faxed the result back (noting that the 40,000th digit is a 1).

In the episode airing May 6, 1993, Apu, the manager of a convenience store, was challenged by Marge’s attorney in a courtroom. He replied that he has a perfect memory. For example, he said, he can recite 40,000 digits of π, and the last digit is a 1.
The PSLQ integer relation algorithm

Let \((x_n)\) be a given vector of real numbers. An integer relation algorithm either finds integers \((a_n)\) such that

\[
a_1x_1 + a_2x_2 + \cdots + a_nx_n = 0
\]

(to within the “epsilon” of the arithmetic being used), or else finds bounds within which no relation can exist.

The “PSLQ” algorithm of mathematician-sculptor Helaman Ferguson (a Hamilton College graduate) is the most widely used integer relation algorithm.

Integer relation detection requires very high precision (at least \(n \times d\) digits, where \(d\) is the size in digits of the largest \(a_k\)), both in the input data and in the operation of the algorithm.

Helaman Ferguson’s “Syzygy” sculpture at Hamilton College
PSLQ, continued

- PSLQ constructs a sequence of integer-valued matrices $B_n$ that reduce the vector $y = x \cdot B_n$, until either the relation is found (as one of the columns of matrix $B_n$), or else precision is exhausted.

- A relation is detected when the size of smallest entry of the $y$ vector suddenly drops to roughly "epsilon" (i.e. $10^{-p}$, where $p$ is the number of digits of precision).

- The size of this drop can be viewed as a “confidence level” that the relation is not a numerical artifact: a drop of 20+ orders of magnitude almost always indicates a real relation.

Efficient variants of PSLQ:

- 2-level and 3-level PSLQ perform almost all iterations with only double precision, updating full-precision arrays as needed. They are hundreds of times faster than the original PSLQ.

- Multi-pair PSLQ dramatically reduces the number of iterations required. It was designed for parallel systems, but runs faster even on 1 CPU.
Decrease of $\log_{10}(\min |y_i|)$ in multipair PSLQ run
The binary algorithm for exponentiation

Problem: What is $3^{17}$ mod 10?

Algorithm A:

$3^{17} = 3 \times 3 \times 3 \times 3 \times 3 \times 3 \times 3 \times 3 \times 3 \times 3 \times 3 \times 3 \times 3 \times 3 \times 3 \times 3 = 129140163,$

so answer = 3.

Algorithm B (faster): $3^{17} = (((3^2)^2)^2) \times 3 = 129140163,$ so answer = 3.

Algorithm C (fastest):

$3^{17} = (((3^2 \mod 10)^2 \mod 10)^2 \mod 10)^2 \mod 10) \times 3 \mod 10 = 3.$

Note that in Algorithm C, we never have to deal with integers larger than $9 \times 9 = 81$. In general, if reducing mod $n$, we never have to deal with integers larger than $(n - 1)^2$. 
1996 result: How to compute arbitrary binary digits of log 2

Consider this well-known formula for log 2:

\[
\log 2 = \sum_{n=1}^{\infty} \frac{1}{n2^n} = 0.10110001011100100001011111101111010001110011110112
\]

Note that the binary digits of log 2 beginning after position \( d \) can be written as \( \{2^d \log 2\} \), where \( \{\cdot\} \) denotes fractional part. Thus we can write:

\[
\{2^d \log 2\} = \left\{ \sum_{n=1}^{d} \frac{2^{d-n}}{n} \right\} + \left\{ \sum_{n=d+1}^{\infty} \frac{2^{d-n}}{n} \right\}
\]

\[
= \left\{ \sum_{n=1}^{d} \frac{2^{d-n} \mod n}{n} \right\} + \left\{ \sum_{n=d+1}^{\infty} \frac{2^{d-n}}{n} \right\},
\]

where we have inserted \( \mod n \) since we are only interested in the fractional part when divided by \( n \). Now note that the numerator \( 2^{d-n} \mod n \) can be calculated very rapidly using the binary algorithm for exponentiation, using only standard arithmetic.

Is there a similar formula and scheme for \( \pi \)?
The BBP formula for $\pi$

In 1996, a PSLQ program discovered this new formula for $\pi$:

$$\pi = \sum_{n=0}^{\infty} \frac{1}{16^k} \left( \frac{4}{8k+1} - \frac{2}{8k+4} - \frac{1}{8k+5} - \frac{1}{8k+6} \right).$$

This formula permits one to compute binary (or hexadecimal) digits of $\pi$ beginning at an arbitrary starting position.

This is the first known instance of a computer program discovering a new formula for $\pi$.

BBP-type formulas (also discovered using PSLQ) are now known for numerous other mathematical constants.

Some other BBP-type formulas found using PSLQ

\[
\pi^2 = \frac{1}{8} \sum_{k=0}^{\infty} \frac{1}{64^k} \left( \frac{144}{(6k+1)^2} - \frac{216}{(6k+2)^2} - \frac{72}{(6k+3)^2} - \frac{54}{(6k+4)^2} + \frac{9}{(6k+5)^2} \right)
\]

\[
\pi^2 = \frac{2}{27} \sum_{k=0}^{\infty} \frac{1}{729^k} \left( \frac{243}{(12k+1)^2} - \frac{405}{(12k+2)^2} - \frac{81}{(12k+4)^2} - \frac{27}{(27k+5)^2} - \frac{72}{(12k+6)^2} - \frac{9}{(12k+7)^2} - \frac{9}{(12k+8)^2} - \frac{5}{(12k+10)^2} + \frac{1}{(12k+11)^2} \right)
\]

\[
\zeta(3) = \frac{1}{1792} \sum_{k=0}^{\infty} \frac{1}{2^{12k}} \left( \frac{6144}{(24k+1)^3} - \frac{43008}{(24k+2)^3} + \frac{24576}{(24k+3)^3} + \frac{30720}{(24k+4)^3} - \frac{1536}{(24k+5)^3} + \frac{3072}{(24k+6)^3} + \frac{768}{(24k+7)^3} - \frac{3072}{(24k+9)^3} - \frac{2688}{(24k+10)^3} - \frac{192}{(24k+11)^3} - \frac{1536}{(24k+12)^3} - \frac{96}{(24k+13)^3} - \frac{672}{(24k+14)^3} - \frac{384}{(24k+15)^3} + \frac{24}{(24k+17)^3} + \frac{48}{(24k+18)^3} - \frac{12}{(24k+19)^3} + \frac{120}{(24k+20)^3} + \frac{48}{(24k+21)^3} - \frac{42}{(24k+22)^3} + \frac{3}{(24k+23)^3} \right)
\]

Normal numbers

Given integer $b \geq 2$, a real number $x$ is $b$-normal (or “normal base $b$”) if every $m$-long string of digits appears in the base-$b$ expansion of $x$ with limiting frequency $1/b^m$.

Using measure theory, it can be shown that almost all real numbers are $b$-normal for a given integer base $b$ (in fact, for all $b$ simultaneously).

These are widely believed to be $b$-normal, for all integer bases $b \geq 2$ and for all $m \geq 1$:

- $\pi = 3.14159265358979323846\ldots$
- $e = 2.7182818284590452354\ldots$
- $\sqrt{2} = 1.4142135623730950488\ldots$
- $\log(2) = 0.69314718055994530942\ldots$
- Every irrational algebraic number (this conjecture is due to Borel).

But there are no proofs of normality for any of these constants, not even for $b = 2$ and $m = 1$ (i.e., equal 0s and 1s in binary), much less for all $b$ and all $m$.

Until recently, normality proofs were known only for a few contrived examples such as Champernowne’s constant $= 0.12345678910111213141516\ldots$ (which is 10-normal).
BBP formulas and normality

Consider a general BBP-type constant (a formula that permits the BBP technique):

$$\alpha = \sum_{n=0}^{\infty} \frac{p(n)}{b^n q(n)},$$

where $p$ and $q$ are integer polynomials, $\text{deg } p < \text{deg } q$, and $q$ has no zeroes for nonnegative arguments.

In 2001, Richard Crandall and I proved that $\alpha$ is $b$-normal iff the sequence $x_0 = 0$, and

$$x_n = \left\{ b x_{n-1} + \frac{p(n)}{q(n)} \right\},$$

where $\{ \cdot \}$ again denote fractional part, is equidistributed in the unit interval. Here “equidistributed” means that the sequence visits each subinterval $(c, d)$ with limiting frequency $d - c$.

Two specific examples

Consider the sequence $x_0 = 0$ and

$$x_n = \left\{ 2x_{n-1} + \frac{1}{n} \right\}.$$ 

Then $\log 2$ is 2-normal iff this sequence is equidistributed in the unit interval.

Similarly, consider the sequence $x_0 = 0$ and

$$x_n = \left\{ 16x_{n-1} + \frac{120n^2 - 89n + 16}{512n^4 - 1024n^3 + 712n^2 - 206n + 21} \right\}.$$ 

Then $\pi$ is 16-normal (and hence 2-normal) iff this sequence is equidistributed in the unit interval.
A class of provably normal constants

Crandall and I also proved that an infinite class of constants are 2-normal, e.g.

\[
\alpha_{2,3} = \sum_{n=1}^{\infty} \frac{1}{3^n 2^{3^n}}
\]

\[
= 0.041883680831502985071252898624571682426096\ldots_{10}
\]

\[
= 0.0ab8e38f684bda12f684bf35ba781948b0fcd6e9e0\ldots_{16}
\]

This constant was proven 2-normal by Stoneham in 1971, but we have extended this to the case where \((2, 3)\) are any pair \((p, q)\) of relatively prime integers \(\geq 2\). We also extended this result to an uncountable class: For any real \(r\) in \([0, 1)\), the constant

\[
\alpha_{2,3}(r) = \sum_{n=1}^{\infty} \frac{1}{3^n 2^{3^n} + r_n}
\]

is 2-normal, where \(r_n\) is the \(n\)-th bit in the binary expansion of \(r\) in \([0, 1)\). These constants are all distinct, so the class is uncountable.

“Random walk” on the first 100 billion base-4 digits of $\pi$

http://gigapan.com/gigapans/106803
“Random walk” on pseudorandom data
“Random walk” on $\alpha_{2,3}$: Base 2 (normal) vs base 6 (nonnormal)
High-precision numerical integration using the tanh-sinh scheme

Given \( f(x) \) defined on \((-1, 1)\), define \( g(t) = \tanh(\pi/2 \sinh t) \). Setting \( x = g(t) \) yields

\[
\int_{-1}^{1} f(x) \, dx = \int_{-\infty}^{\infty} f(g(t))g'(t) \, dt \approx h \sum_{j=-N}^{N} w_j f(x_j),
\]

where \( x_j = g(hj) \) and \( w_j = g'(hj) \).

Features:

- Reducing \( h \) by half typically doubles the number of correct digits.
- Works well even for functions with blow-up singularities at endpoints.
- The cost of computing abscissas and weights increases only linearly with the number \( N \) of subdivisions, which is much faster than with most other methods.


Ising integrals from mathematical physics

We applied our methods to study three classes of integrals: $C_n$ are connected to quantum field theory, $D_n$ arise in the Ising theory of mathematical physics, while the $E_n$ integrands are derived from $D_n$:

\[
C_n := \frac{4}{n!} \int_0^\infty \cdots \int_0^\infty \frac{1}{\left(\sum_{j=1}^n (u_j + 1/u_j)\right)^2} \frac{du_1}{u_1} \cdots \frac{du_n}{u_n}
\]

\[
D_n := \frac{4}{n!} \int_0^\infty \cdots \int_0^\infty \prod_{i<j} \left(\frac{u_i - u_j}{u_i + u_j}\right)^2 \frac{du_1}{u_1} \cdots \frac{du_n}{u_n}
\]

\[
E_n = 2 \int_0^1 \cdots \int_0^1 \left(\prod_{1 \leq j < k \leq n} \frac{u_k - u_j}{u_k + u_j}\right)^2 dt_2 dt_3 \cdots dt_n
\]

where in the last line $u_k = t_1 t_2 \cdots t_k$.

Limiting value of $C_n$: What is this number?

Key observation: The $C_n$ integrals can be converted to one-dimensional integrals involving the modified Bessel function $K_0(t)$:

$$C_n = \frac{2^n}{n!} \int_0^\infty t K_0^n(t) \, dt$$

High-precision numerical values, computed using this formula and tanh-sinh quadrature, approach a limit. For example:

$$C_{1024} = 0.6304735033743867961220401927108789043545870787\ldots$$

What is this number? We copied the first 50 digits into the online Inverse Symbolic Calculator (ISC) at [https://isc.carma.newcastle.edu.au](https://isc.carma.newcastle.edu.au). The result was:

$$\lim_{n \to \infty} C_n = 2e^{-2\gamma}.$$  

where $\gamma$ denotes Euler’s constant. This is now proven.
Other Ising integral evaluations found using PSLQ

\[
D_3 = 8 + 4\pi^2/3 - 27 L_{-3}(2)
\]
\[
D_4 = 4\pi^2/9 - 1/6 - 7\zeta(3)/2
\]
\[
E_2 = 6 - 8 \log 2
\]
\[
E_3 = 10 - 2\pi^2 - 8 \log 2 + 32 \log^2 2
\]
\[
E_4 = 22 - 82\zeta(3) - 24 \log 2 + 176 \log^2 2 - 256(\log^3 2)/3 + 16\pi^2 \log 2 - 22\pi^2/3
\]
\[
E_5 = 42 - 1984 \text{Li}_4(1/2) + 189\pi^4/10 - 74\zeta(3) - 1272\zeta(3) \log 2 + 40\pi^2 \log^2 2 - 62\pi^2/3 + 40(\pi^2 \log 2)/3 + 88 \log^4 2 + 464 \log^2 2 - 40 \log 2
\]

where \(\zeta\) is the Riemann zeta function and \(\text{Li}_n(x)\) is the polylog function. \(E_5\) remained a “numerical conjecture” for several years, but was proven in March 2014 by Erik Panzer.
Using PSLQ to find the minimal polynomial of an algebraic number

One simple but important application of PSLQ is to recognize a computed numerical value as the root of an integer polynomial of degree $m$.

Example: The following constant is suspected to be an algebraic number:

$$\alpha = 1.232688913061443445331472869611255647068988824547930576057634684778\ldots$$

What is its minimal polynomial?

Method: Compute the vector $(1, \alpha, \alpha^2, \cdots, \alpha^m)$ for $m = 30$, then input this vector to a PSLQ program.

Answer (using 250-digit arithmetic):

$$0 = 697 - 1440\alpha - 20520\alpha^2 - 98280\alpha^3 - 102060\alpha^4 - 1458\alpha^5 + 80\alpha^6 - 43920\alpha^7 + 538380\alpha^8 - 336420\alpha^9 + 1215\alpha^{10} - 80\alpha^{12} - 56160\alpha^{13} - 135540\alpha^{14} - 540\alpha^{15} + 40\alpha^{18} - 7380\alpha^{19} + 135\alpha^{20} - 10\alpha^{24} - 18\alpha^{25} + \alpha^{30}$$
The Poisson potential function

In 2012, Richard Crandall, while investigating techniques to sharpen images, noted that each pixel was given by a form of the 2-D Poisson potential function:

\[
\phi_2(x, y) = \frac{1}{\pi^2} \sum_{m,n \text{ odd}} \frac{\cos(m\pi x) \cos(n\pi y)}{m^2 + n^2}
\]

In a 2013 study, we numerically discovered, and then proved the intriguing fact that for rational numbers \(x\) and \(y\),

\[
\phi_2(x, y) = \frac{1}{\pi} \log \alpha
\]

where \(\alpha\) is algebraic, i.e., the root of a some integer polynomial of degree \(m\).

By computing high-precision numerical values of \(\phi_2(x, y)\) for various specific rational \(x\) and \(y\), and applying a multipair PSLQ program, we were able to produce the explicit minimal polynomials for \(\alpha\) in numerous specific cases.

Samples of minimal polynomials found by PSLQ

Minimal polynomial corresponding to $x = y = 1/s$:

5. $1 + 52\alpha - 26\alpha^2 - 12\alpha^3 + \alpha^4$

6. $1 - 28\alpha + 6\alpha^2 - 28\alpha^3 + \alpha^4$

7. $-1 - 196\alpha + 1302\alpha^2 - 14756\alpha^3 + 15673\alpha^4 + 42168\alpha^5 - 111916\alpha^6 + 82264\alpha^7$
   $-35231\alpha^8 + 19852\alpha^9 - 2954\alpha^{10} - 308\alpha^{11} + 7\alpha^{12}$

8. $1 - 88\alpha + 92\alpha^2 - 872\alpha^3 + 1990\alpha^4 - 872\alpha^5 + 92\alpha^6 - 88\alpha^7 + \alpha^8$

9. $-1 - 534\alpha + 10923\alpha^2 - 342864\alpha^3 + 2304684\alpha^4 - 7820712\alpha^5 + 13729068\alpha^6$
   $-22321584\alpha^7 + 39775986\alpha^8 - 44431044\alpha^9 + 19899882\alpha^{10} + 3546576\alpha^{11}$
   $-8458020\alpha^{12} + 4009176\alpha^{13} - 273348\alpha^{14} + 121392\alpha^{15}$
   $-11385\alpha^{16} - 342\alpha^{17} + 3\alpha^{18}$

10. $1 - 216\alpha + 860\alpha^2 - 744\alpha^3 + 454\alpha^4 - 744\alpha^5 + 860\alpha^6 - 216\alpha^7 + \alpha^8$

These computations are very expensive:

The case $x = y = 1/32$ required 10,000-digit arithmetic and ran for 45 hours.

Other runs, using even higher precision, ultimately failed, evidently due to subtle program bugs. HELP!
Kimberley’s formula for the degree of the polynomial

Based on our preliminary results, Jason Kimberley of the University of Newcastle, Australia observed that the degree $m(s)$ of the minimal polynomial associated with the case $x = y = 1/s$ appears to be given by the following:

Set $m(2) = 1/2$. Otherwise for primes $p$ congruent to 1 mod 4, set $m(p) = \text{int}^2(p/2)$, where int denotes greatest integer, and for primes $p$ congruent to 3 mod 4, set $m(p) = \text{int}(p/2)(\text{int}(p/2) + 1)$. Then for any other positive integer $s$ whose prime factorization is $s = p_1^{e_1}p_2^{e_2} \cdots p_r^{e_r}$,

$$m(s) = 4^{r-1} \prod_{i=1}^{r} p_i^{2(e_i-1)} m(p_i).$$

Does Kimberley’s formula hold for larger $s$? Why?

What is the true mathematical connection between the pair of rationals $(x, y)$ and the algebraic number $\alpha$?
Three improvements to the Poisson polynomial computation program

   ▶ Speedup: 3X

2. A new 3-level multipair PSLQ program.
   ▶ Speedup: 4.2X

3. Parallel implementation on a 16-core system.
   ▶ Speedup: 12.2X

Overall speedup: 156X

Total size of code: 190,000 lines
A fast algorithm to compute the Poisson potential function $\phi_2(x, y)$

$$\phi_2(x, y) = \frac{1}{2\pi} \log \left| \frac{\theta_2(z, q)\theta_4(z, q)}{\theta_1(z, q)\theta_3(z, q)} \right|,$$

where $q = e^{-\pi} = 0.0432139\ldots$, and $z = \frac{\pi}{2}(y + ix)$. Compute the four theta functions using the following rapidly convergent formulas involving complex variables:

$$\theta_1(z, q) = 2 \sum_{k=1}^{\infty} (-1)^{k-1} q^{(2k-1)^2/4} \sin((2k - 1)z),$$

$$\theta_2(z, q) = 2 \sum_{k=1}^{\infty} q^{(2k-1)^2/4} \cos((2k - 1)z),$$

$$\theta_3(z, q) = 1 + 2 \sum_{k=1}^{\infty} q^{k^2} \cos(2kz),$$

$$\theta_4(z, q) = 1 + 2 \sum_{k=1}^{\infty} (-1)^k q^{k^2} \cos(2kz).$$
High-level computational algorithm

1. Given rationals $x = p/s$ and $y = q/s$, select a conjectured minimal polynomial degree $m(s)$ (using Kimberley’s formula) and other parameters for the run.

2. Calculate $\phi_2(x, y)$ to $P_2$-digit precision using the formulas from two viewgraphs above. When done, calculate $\alpha = \exp(8\pi \phi_2(x, y))$ and generate the $(m + 1)$-long vector $X = (1, \alpha, \alpha^2, \cdots, \alpha^m)$, to $P_2$-digit precision.

3. Apply the three-level multipair PSLQ algorithm to $X$. For larger problems, employ a parallel version of the three-level multipair PSLQ code, using the OpenMP construct DO PARALLEL to perform certain time-intensive loops in parallel.

4. If no numerically significant relation is found, try again with a larger degree $m$ or higher precision $P_2$. If a relation is found, employ the polynomial factorization facilities in Mathematica and Maple to ensure that the polynomial is irreducible.
This polynomial has degree 192, with coefficients as large as $10^{85}$. This computation required 18,000-digit arithmetic and 34 CPU-hours.

The case $(1/37, 1/37)$ required 51,000-digit arithmetic and 90 CPU-days (5.6 days on a 16-core parallel system).

Kimberley's formula was upheld for $(1/s, 1/s)$, for all $s$ up to 52 (except for $s = 41, 43, 47, 49, 51$, which were too expensive), and also for $s = 60$ and $s = 64$. 
From our results, in the case \((1/s, 1/s)\) where \(s\) is even, the resulting polynomial is always palindromic \((a_k = a_{m-k})\). For instance, when \(s = 16\),

\[
p_{16}(\alpha) = 1 - 1376\alpha^1 - 12560\alpha^2 - 3550496\alpha^3 + 81241720\alpha^4 - 169589984\alpha^5 + 1334964944\alpha^6 - 24307725984\alpha^7 + 238934926108\alpha^8 - 1043027124704\alpha^9 + 2328675366384\alpha^{10} - 3219896325280\alpha^{11} + 4238551472456\alpha^{12} - 10247414430048\alpha^{13} + 28552105805904\alpha^{14} - 55832851687968\alpha^{15} + 70020268309062\alpha^{16} - 55832851687968\alpha^{17} + 28552105805904\alpha^{18} - 10247414430048\alpha^{19} + 4238551472456\alpha^{20} - 3219896325280\alpha^{21} + 2328675366384\alpha^{22} - 1043027124704\alpha^{23} + 238934926108\alpha^{24} - 24307725984\alpha^{25} + 1334964944\alpha^{26} - 169589984\alpha^{27} + 81241720\alpha^{28} - 3550496\alpha^{29} - 12560\alpha^{30} - 1376\alpha^{31} + \alpha^{32}
\]

Nitya Mani, an undergraduate student at Stanford University, observed that if \(\alpha\) is a root of a palindromic polynomial such as this, then \(\alpha + 1/\alpha\) is a root of a transformed polynomial of half the degree.
Proofs of Kimberley’s formula and the palindromic property

- On March 16, DHB presented our results at a seminar at the University of California, Berkeley.
- Following the presentation, Watson Ladd, a graduate student in mathematics, brought to our attention the fact that some of our conjectures should follow from results in the theory of elliptic curves, Gaussian integers and ideals.
- After some effort, Ladd produced proofs of Kimberley’s formula and the palindromic property, which proofs were then included in our paper and returned to the journal.
The computer as an active agent of discovery

- Thirty years ago, computations were derided in the mathematical research world: “Real mathematicians don't compute.”
- Nowadays, with Mathematica, Maple and Sage, almost any mathematical concept can be explored computationally.
- A new generation of computer-literate mathematicians is now using these tools, and many are writing their own programs.
- Hundreds of papers have been published with results discovered by computer.

Everyone from first-year students to senior faculty, from research mathematicians to computer scientists, can participate in world-class mathematical research.

Thanks! This talk is available at http://www.davidhbailey.com/dhbtalks/dhb-hamilton-math.pdf.