Polynomial arithmetic and applications in number theory

David Harvey

September 26, 2008

What is my research about?

"Computational number theory."

Most of the time,

- I use **big computers**
- to multiply and divide **big polynomials**
- over coefficient rings that applied mathematicians usually **don't care about**
- to study problems of interest to number theorists.

Topics in polynomial arithmetic

- Algorithms for arithmetic in $(\mathbf{Z}/n\mathbf{Z})[x]$, where say $n < 2^{64}$ and $1 \leq \text{degree} \leq 10^9$
- Large-integer arithmetic (GMP library)
- Variants of Kronecker substitution
- Variants of FFT methods (Schönhage–Nussbaumer convolution, number-theoretic transforms, floating-point FFTs)
- Smooth performance with respect to degree
- Cache-friendly algorithms

Two applications

- Counting points on hyperelliptic curves over finite fields (half my Ph.D. thesis)
- Verification of the Kummer–Vandiver conjecture (joint work with Joe Buhler, Center for Communications Research, San Diego)

Hyperelliptic curves over finite fields

A hyperelliptic curve C of genus g over $\mathrm{GF}(p^n)$ is an equation

$$y^2 = f(x)$$

where $f \in \mathrm{GF}(p^n)[x]$ is monic, squarefree, degree $2g + 1$.

Basic problem: given f, compute the *zeta function* of C.

(Equivalently: count the number of solutions (x, y) to $y^2 = f(x)$ in $\mathrm{GF}(p^{nk})$ for $k = 1, \ldots, g$.)

Cryptographic applications

From the zeta function one can deduce the number of points $N \approx p^{ng}$ on the Jacobian of C over $\mathrm{GF}(p^n)$.

If we get lucky, N is prime, and then the curve can be used to construct a secure public-key cryptosystem (e.g. Diffie–Hellman key exchange in the Jacobian, where the discrete logarithm problem is presumably hard).

A common benchmark for 'cryptographic size' is $N = 2^{160}$.

Various algorithms

Many counting algorithms, too numerous to list. Here are a few:

- Naive counting: exponential in $\log p$, g, n
- Schoof–Pila (1992): polynomial in $\log p$, n, exponential in g
- Kedlaya (2001): soft-linear in p, polynomial in n, g
- My thesis (2008): soft-linear in $\sqrt{p}$, polynomial in n, g

All (but the first) depend heavily on asymptotically fast polynomial arithmetic.

A big example

Record genus 3 example computed with $\sqrt{p}$ algorithm:

$$y^2 = x^7 + 29723259490794204x^6 + 13669080989682802x^5 + 31024378462415735x^4 + 12535160111191415x^3 + 23344313901215683x^2 + 3192716983602209x + 16088167167540442$$

over $\mathrm{GF}(p)$ where $p = 2^{55} - 55 = 36028797018963913$.

Order of Jacobian is

$$N = 46768052141791550072336765593390889080855547973784 \approx 2^{165}.$$

Took 48.3 hours, used 90 GB RAM.

Unfortunately N is not prime :-(

The Kummer–Vandiver conjecture

Major open problem in algebraic number theory, proposed in 1849 by Kummer (and later by Vandiver). The claim is:

> *For all primes p, the class number of the maximal real subfield of the p-th cyclotomic field is not divisible by p.*

Lightning course in algebraic number theory

Remember unique factorisation in $\mathbf{Q}$?

$$30 = 2 \times 3 \times 5 = 5 \times 2 \times 3 = (-3) \times (-2) \times 5.$$

Unique factorisation is *broken* in some algebraic number rings:

$$6 = 2 \times 3 = (1 + \sqrt{-5}) \times (1 - \sqrt{-5}).$$

The elements 2, 3, $1 \pm \sqrt{-5}$ are all irreducible (but not 'prime') integers in $\mathbf{Q}(\sqrt{-5})$.

The class number

The *class number* $h(R)$ measures how badly unique factorisation is broken in R:

- $h(\mathbf{Q}) = 1$: not broken at all
- $h(\mathbf{Q}(\sqrt{-5})) = 2$: somewhat broken
- $h(\mathbf{Q}(\sqrt{-163})) = 1$: not broken at all
- $h(\mathbf{Q}(\sqrt[4]{-74})) = 100$: more badly broken

(Technically, $h(R) =$ number of equivalence classes of ideals in R under $I \sim J$ if $\alpha I = \beta J$ for some $\alpha, \beta \in R$.)

Cyclotomic fields

The p-th cyclotomic field is $\mathbf{Q}(\zeta_p)$ where $\zeta_p = e^{2\pi i/p}$.

The maximal real subfield of the p-th cyclotomic field is $\mathbf{Q}(\zeta_p)^+ = \mathbf{Q}(\zeta_p) \cap \mathbf{R} = \mathbf{Q}(\zeta_p + \zeta_p^{-1})$, about half the size of $\mathbf{Q}(\zeta_p)$ (algebraically speaking).

Kummer–Vandiver says that p never divides $h(\mathbf{Q}(\zeta_p)^+)$.

Testing Kummer–Vandiver

Extremely difficult to compute $h(\mathbf{Q}(\zeta_p)^+)$. Value is not definitively known beyond $p = 67$ (Schoof, 2003).

Can't compute $h(\mathbf{Q}(\zeta_p)^+)$ directly, but can indirectly test divisibility by p.

For each p, it comes down to multiplying together two certain polynomials of degree $\approx p/2$ with coefficients in $\mathbf{Z}/p\mathbf{Z}$, and checking which coefficients of the product are zero.

The computation

Previous record: all 788,060 primes less than 16,000,000 (Buhler et al, 2001).

Current project: aiming for all 7,603,553 primes less than $2^{27} = 134{,}217{,}728$.

Running on two supercomputers at the Texas Advanced Computing Center (TACC) at the University of Texas:

- Lonestar: 1460 $\times$ 4-core Intel Xeon $=$ 5,840 cores
- Ranger: 3936 $\times$ 16-core AMD Opteron $=$ 62,976 cores (world's 4th largest computer as of June 2008)

Total CPU time: $\approx$ 220,000 hours (half of it already burned up). We've done up to about $p = 88$ million so far.

Is Kummer–Vandiver true?

No counterexamples found so far, but many number theorists expect that Kummer–Vandiver is *false*!

Why??

Are number theorists always so peverse???

Heuristics

A probabilistic argument implies that the number of counterexamples up to X is about $\frac{1}{2} \log \log X$.

Up to 16,000,000, expect only about 1.40 counterexamples.

Up to 134,217,728, expect only about 1.46 counterexamples.

If you believe the heuristics, we have about a 4% chance of success this time around.

If you believe the heuristics *and* Moore's Law, perhaps one counterexample per 200 years.