Euclidean prime generators

Andrew R. Booker, **Bristol University**
Bristol, UK
Carl Pomerance, **Dartmouth College** (emeritus)
Hanover, New Hampshire, USA
(U. Georgia, emeritus)

Integers Conference, Carrollton, Georgia, October 6, 2016

We all recall Euclid's proof that there are infinitely many primes:
Assume there are only finitely many, multiply them all together, add 1, and take a prime factor.

Starting from the empty product, that is, 1, we get

2, 3, 7, 43, ...

The next step is $2 \times 3 \times 7 \times 43 + 1 = 1807 = 13 \times 139$, so we have a choice of taking 13 as the next prime, 139, or both.

A. A. Mullin suggested in 1963 to look at the sequence of primes formed with Euclid's construction where we always take the least prime factor of the product plus 1.

There is a heuristic argument of **D. Shanks** that this sequence contains every prime.

Mullin also suggested to always take the largest prime factor of the product plus 1. It would seem obvious that this sequence omits infinitely many primes, but it is not trivial to prove this since the product plus 1 could conceivably be a power of the least prime not found so far.

However, **Booker** (2012) was able to prove this second sequence does omit infinitely many primes, and a simplified proof was given by **P. Pollack** and **E. Treviño** (2014).

Euclid's construction of new primes from old motivated a definition that appeared in the **APR** (Adleman, P, Rumely) primality test: Start with a bunch of small primes, maybe all of the primes to some point. Call these the *initial* primes and let I denote their product. Then consider the primes p of the form $a+1$ where $a \mid I$. Call these the *Euclidean* primes and let E denote their product.

The **APR** primality test runs in time $I^{O(1)}$ on numbers $n < E^2$. Thus, one wants I small and E large. It's shown that with a judicious choice of I one has $I \leq (\log E)^{O(\log\log\log E)}$, and so the test is *almost* polynomial time.

The same I, E construction is used in the **Lenstra** finite fields primality test.

Euclid's construction can be modified in several possible ways, the first motivated by initial and Euclidean primes:

1. If n is the product of the primes so far, choose as the next prime the least new prime dividing some $a+1$, with $a \mid n$. That is,

$$\min\{p : p \nmid n,\ p \mid a+1 \text{ for some } a \mid n\}.$$

2. If n is the product of the primes so far, choose as the next prime some prime factor of some $a+b$, where $ab = n$.

Booker (2016) gave a proof that there are valid choices in the second sequence so that every prime is generated. I claimed about 20 years ago, but never wrote up, that the first sequence contains every prime, essentially in order.

To illustrate the first sequence, we start with the empty product 1, and find the primes $2, 3$. Can we now get 5? Well none of $1+1$, $2+1$, $3+1$, $6+1$ is divisible by 5.

To illustrate the first sequence, we start with the empty product 1, and find the primes $2, 3$. Can we now get 5? Well none of $1+1$, $2+1$, $3+1$, $6+1$ is divisible by 5.

But we can get 7. And then we get 5 via $14+1$. And then we get 11 via $10+1$.

Continuing, the least new prime that can made from these: $7 \times 11 + 1$ has the prime factor 13. We can pick up 17 from $3 \times 11 + 1$. We can pick up 19 from $2 \times 5 \times 17 + 1$, etc.

So, it really does seem that this sequence picks up every prime in order, except that 5 and 7 are reversed. This assertion immediately follows from: *Every prime $p \geq 7$ has the residue class -1 represented by a squarefree number all of whose primes are smaller than p.*

We prove the following stronger result:

Every prime $p > 7$ has each residue class mod p *represented by a squarefree number all of whose prime factors are at most p.*

Not only does this assertion immediately prove that the first sequence contains every prime (and in order starting with 11), it also allows a short proof that the second sequence contains every prime.

We prove the assertion via a combinatorial result of **V. Lev** on sumsets and a numerically explicit **Pólya–Vinogradov** inequality. Some computation is required for $p < 3 \times 10^8$.

To close the talk, lets see the short proof that the second sequence contains every prime.

Recall: If n is the product of the primes found so far, then we choose a prime factor of some $a+b$ where $ab=n$.

Say we have found all of the primes below $p>7$ and have not found p yet. Let n be the product of the primes found so far. If $(-n/p)=1$, then there is a solution a to $a^2+n\equiv 0 \pmod{p}$, so that

$$a+n/a\equiv 0 \pmod{p}.$$

By our assertion, we can represent $a \pmod{p}$ as a squarefree product of the primes less than p, and then $a \mid n$, with $p \mid a+n/a$.

So assume that $(-n/p) = -1$. This case is trickier, but there's a short proof that there is a solution a to

$$\left(\frac{a + n/a}{p}\right) = -1. \tag{1}$$

Assuming so, represent a as a squarefree product of primes $< p$, so that $a \mid n$. Then choose q as any prime factor of $a + n/a$ with $(q/p) = -1$ (at least one such q must exist), and take it as the next prime in the sequence. The new product is qn and we have $(-qn/p) = 1$, so we can find p with one more step.

Here's why (1) is solvable. It's equivalent to $a^3 + an$ being a quadratic nonresidue mod p. The elliptic curve $y^2 = x^3 + nx$ has at most $p + 2p^{1/2}$ solutions mod p, and all but 1 of them occur in pairs $(x, \pm y)$, so there are values of x *not* corresponding to a point on the curve; let a be one of them.

Thank you