Euclidean prime generators

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

U. Georgia Number Theory Seminar, March 22, 2017

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 $I \le (\log E)^{O(\log\log\log E)}$, 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 $p \ge 7$ 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 $p < 3 \times 10^8$.

Fix some large prime p. Say a subgroup H of $(\mathbb{Z}/p\mathbb{Z})^*$ is “good” if each coset of H contains a squarefree number smaller than p. We prove that if H is large, i.e., has small index, then it is good.

The plan: Say H has index d and let χ be a character of $(\mathbb{Z}/p\mathbb{Z})^*$ of order d. Then $H = \ker \chi$. And $a \in mH$ if and only if $\chi(a) = \chi(m)$. Thus,

$$\frac{1}{d}\sum_{i=1}^{d} \chi^i(j)\bar{\chi}^i(m)$$

is the indicator function for mH. Hence

$$\frac{1}{d}\sum_{i=1}^{d}\sum_{j<p} \mu^2(j)\chi^i(j)\bar{\chi}^i(m)$$

is the number of squarefree integers $j < p$ with $j \in mH$.

The philosophy with character sums is that the principal character should give the main information, and the task at hand is to show that the other characters mostly cancel out.

In our case, the principal character in

$$\frac{1}{d}\sum_{i=1}^{d}\sum_{j<p}\mu^2(j)\chi^i(j)\bar{\chi}^i(m)$$

is χ^d, and the contribution to the sum when $i = d$ is

$$\frac{1}{d}\sum_{j<p}\mu^2(j).$$

We know that the asymptotic density of the squarefree integers is $6/\pi^2$, but to get a universal inequality we need the "Schnirelmann density". This is 53/88.

Note, the Schnirelmann density of a set of positive integers $\mathcal{A}$ is

$$\inf_{n\geq 1} \frac{1}{n} \sum_{a\in\mathcal{A}\cap[1,n]} 1.$$

This density is useful in additive number theory: if $\mathcal{A}$ has positive Schnirelmann density then there is some positive integer k such that every number is contained in $k\mathcal{A}$, the set of k-fold sums of elements of $\mathcal{A}$.

So, the contribution of the principal character is $\frac{1}{d}\sum_{j<p}\mu^2(j)$, which is at least $\frac{53}{88}(p-1)/d$.

So, what should be done with the non-principal characters in

$$\frac{1}{d}\sum_{i=1}^{d}\sum_{j<p}\mu^2(j)\chi^i(j)\bar{\chi}^i(m) \text{ ?}$$

We'd like to use a character sum estimate but μ^2 is in the way. So, first use inclusion-exclusion:

$$\sum_{j<p}\mu^2(j)\chi^i(j) = \sum_{v\geq 1}\mu(v)\chi^i(v^2)\sum_{j<p/v^2}\chi^i(j).$$

The $v = 1$ term is 0. For $v > p^{1/4}$, we use the trivial estimate p/v^2 in the inner sum to get a total contribution of $< p^{3/4}$. For $1 < v < p^{1/4}$ we use the Pólya–Vinogradov inequality in the inner sum, getting a total contribution that is $O(p^{3/4}\log p)$.

Being a bit more careful with these last steps, and using a numerically explicit version of the Pólya–Vinogradov inequality due to Frolenkov and Soundararajan, we have

$$\frac{1}{d}\sum_{i=1}^{d-1}\left|\sum_{j<p}\mu^2(j)\chi^i(j)\bar{\chi}^i(m)\right| \le \frac{d-1}{d}p^{3/4}\left(\frac{1}{4\pi}\log p+\frac{5}{2}\right).$$

We want this expression to be smaller than $\frac{53}{88}(p-1)/d$. For $d < \log p + 1$, this is true once $p > 3 \times 10^8$.

We have proved: *For $p > 3 \times 10^8$ and $d \mid p-1$ with $d < \log p + 1$, and for the subgroup H of $(\mathbb{Z}/p\mathbb{Z})^*$ of index d, every coset contains a squarefree number smaller than p.*

Our goal is to show that every residue mod p contains a p-smooth, squarefree number.

For each $d \mid p-1$ with $d < \log p + 1$ and subgroup H of $(\mathbb{Z}/p\mathbb{Z})^*$ of index d, let $\mathcal{C}_{d,p}$ be a fixed set of squarefree coset representatives smaller than p, and let $\mathcal{S}_{d,p}$ be the set of primes that divide a member of $\mathcal{C}_{d,p}$. Further, let $\mathcal{S}_p$ be the union of the sets $\mathcal{S}_{d,p}$ for $d \mid p-1$, $d < \log p + 1$.

Then each $\#\mathcal{S}_{d,p} < d \log p$ and $\#\mathcal{S}_p < \frac{1}{2}(\log p + 1)^3$.

We use this as follows. Take pairs of distinct primes $q, r < p$ which are not in $\mathcal{S}_p$ and consider residues $m \equiv qr \pmod{p}$. Let $\mathcal{A}$ denote the set of such m that arise in at least $\sqrt{p}/\log p$ ways as a product qr. By an averaging argument we show that $\#\mathcal{A} > p/\log p + 2$.

We would like to use members of $\mathcal{A}$ to fill up all of $(\mathbb{Z}/p\mathbb{Z})^*$ by multiplying them together. Failing that, we would at least like to fill up a large subgroup.

We use an additive result of V. Lev: *If $\mathcal{A}' \subset \{0, 1, \ldots, N\}$, then there are positive integers*

$$d \le \kappa := \lceil N/(\#\mathcal{A}' - 2) \rceil, \quad k \le 2\kappa + 1$$

such that $k\mathcal{A}'$ contains N consecutive multiples of d.

We apply Lev's theorem not to our set $\mathcal{A}$ of size $> p/\log p + 2$, but to the set $\mathcal{A}'$ of discrete logarithms of members of $\mathcal{A}$ with respect to some primitive root g mod p. That is, for each $a \in \mathcal{A}$, we take $a' \in \{0, 1, \ldots, p-2\}$ where $g^{a'} \equiv a \pmod{p}$. So the set $k\mathcal{A}'$ corresponds to $\mathcal{A}^k$, the set of k-fold products of members of $\mathcal{A}$. And $k\mathcal{A}'$ having N $(= p-2)$ consecutive multiples of d corresponds to $\mathcal{A}^k$ having N consecutive powers of g^d.

Replacing d with $(d, p-1)$, we have that $d \le \kappa$, $k \le 2\kappa + 1$, and $\mathcal{A}^k$ contains the subgroup H of $(\mathbb{Z}/p\mathbb{Z})^*$ of index d. Finally note that $\kappa < \log p + 1$.

Recall that $\mathcal{A}$ consists of numbers qr where q, r are distinct primes smaller than p. Now k-fold products of numbers qr need not be squarefree, but each $m \equiv qr \pmod{p}$ has many representations as qr, so we can choose the representation to have the product squarefree.

Now we fill up the full group $(\mathbb{Z}/p\mathbb{Z})^*$ using our squarefree coset representatives, noting that none of the primes involved are used in the pairs qr.

This proves that for every prime $p > 3 \times 10^8$, each residue class mod p has a p-smooth, squarefree representative.

We would like to close the gap and show this holds for all p in $[11, 3 \times 10^8]$.

We do this by brute force for $p < 10^4$.

For $p > 10^4$, we proceed as follows. Let g be a primitive root mod p. Each nonzero residue is of the form g^h, where $h \in [1, p-1]$. Suppose that we have each $g^{2^i} \equiv q_i r_i \pmod{p}$ for each $2^i \le p-1$, where all of the primes q_i, r_i are distinct and $< p$. Then we're done, since each $h \le p-1$ has a binary representation.

To search for a pair q, r for a given i, we let q run over small primes not already used, until $r = q^{-1} g^{2^i} \mod p$ is a prime that's not already used. We are not guaranteed beforehand that a suitable pair q, r will be found for g^{2^i}, but heuristically it seems that it should work well, and in practice it did work well.

This completes our proof that every prime $p \ge 11$ has every residue class represented by a squarefree, p-smooth number.

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 $p>7$, we have found all of the primes below p 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