SQUAREFREE SMOOTH NUMBERS AND EUCLIDEAN PRIME GENERATORS

ANDREW R. BOOKER AND CARL POMERANCE

Abstract. We show that for each prime $p > 7$, every residue mod p can be represented by a squarefree number with largest prime factor at most p. We give two applications to recursive prime generators akin to the one Euclid used to prove the infinitude of primes.

1. Introduction

In [14], Mullin considered the sequence $\{p_k\}_{k=1}^{\infty}$ defined so that, for every $k \geq 0$, p_{k+1} is the smallest prime factor of $1 + p_1 \cdots p_k$. From the argument employed by Euclid to prove the infinitude of prime numbers, it follows that the p_k are pairwise distinct, and Mullin's sequence can thus be viewed as an explicit, constructive form of the proof. A natural question, which Mullin posed, is whether *every* prime eventually occurs in the sequence. Despite clear heuristic and empirical evidence that the answer must be yes, it appears to be very difficult to prove anything substantial to that end.[1]

With this setting in mind, in [5, Section 1.1.3] and [4], we (independently) described two variations of Euclid's argument that allow greater flexibility and lead to sequences that provably contain every prime number. We recall these constructions in detail in Section 5. The main focus of this article is the following question, which arises naturally as an ingredient in both constructions, but is possibly of independent interest:

For primes p, are all residue classes mod p represented by the positive integers that are both squarefree and p-smooth?

(Recall that an integer n is called *y-smooth* if every prime divisor of n is $\leq y$.) Since there are $2^{\pi(p)}$ squarefree, p-smooth, positive integers and only p residue classes mod p, one heuristically expects the answer to be yes, at least for large p. (However, note that $y = p$ is best possible, since the zero residue class mod p is not attained by a y-smooth number for any $y < p$.) We will show that, with two exceptions, this is indeed the case:

Theorem 1. *Let p be a prime different from 5 and 7, and $a \in \mathbb{Z}$. Then there is a squarefree, p-smooth, positive integer n such that $n \equiv a \pmod{p}$.*

The proof of Theorem 1 consists of three largely independent steps that we carry out in Sections 2–4. Our key tools include a numerically explicit form of the Pólya–Vinogradov inequality, see Frolenkov and Soundararajan [7], and a combinatorial result of Lev [12] on h-fold sums of dense sets. In Section 5 we apply Theorem 1 to variants of Euclid's argument, and thus show how one can generate all of the primes out of nothing. In the final section we mention a few related unsolved problems.

Date: November 4, 2016.

1991 *Mathematics Subject Classification.* Primary 11A41, Secondary 11A15, 11B25, 11L40.

[1] At least one of the authors thinks that Mullin's question is likely undecidable.

Related results on prime generators of Euclid type may be found in [3], [15], [19], [20], and the references of those papers.

2. Large p: character sums

For a prime p and a positive integer $d \mid p-1$, let

$$H_{d,p} = \{h \in (\mathbb{Z}/p\mathbb{Z})^* : \ h^{(p-1)/d} \equiv 1 \pmod{p}\}$$

denote the subgroup of $(\mathbb{Z}/p\mathbb{Z})^*$ of index d.

Proposition 2. *Let $p > 3 \times 10^8$ be a prime and suppose $d \mid p-1$ with $d < \log p + 1$. For each nonzero residue $m \pmod{p}$ there is some squarefree number $j < p$ with $j \in mH_{d,p}$.*

Proof. We may assume that $d > 1$, since otherwise we can take $j = 1$. Let χ be a character mod p of order d. Since χ^i for $i = 1, \ldots, d$ runs over all of the characters mod p of order dividing d, we have that

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

is the number of squarefree numbers $j < p$ with $\chi(j) = \chi(m)$, and so is the number of squarefree numbers smaller than p in the coset $mH_{d,p}$. The principle character (the term when $i = d$) contributes

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

to the sum. This expression is $\sim \frac{6}{\pi^2}p/d$ as $p \to \infty$. One can get an explicit lower bound valid for all p via the Schnirelmann density of the squarefree numbers, see [17]. Thus,

$$\frac{1}{d}\sum_{j<p}\mu^2(j) \geq \frac{53}{88d}(p-1). \tag{2}$$

Our task is then to show that the other terms in (1) are small in comparison. By recognizing a squarefree number by an inclusion-exclusion over square divisors, we have for $1 \leq i \leq d$,

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

We may discard the term $v = 1$ since it is 0. For $v > \frac{1}{2}p^{1/4} + 1$, we may use the trivial estimate

$$\sum_{v>\frac{1}{2}p^{1/4}+1}\left|\sum_{j<p/v^2}\chi^i(j)\right| \leq \sum_{v>\frac{1}{2}p^{1/4}+1}\ \sum_{j<p/v^2} 1 < p\sum_{v>\frac{1}{2}p^{1/4}+1}\frac{1}{v^2} < 2p^{3/4}.$$

For $p \leq \frac{1}{2}v^{1/4} + 1$ we use an explicit form of the Pólya–Vinogradov inequality, see [7], where we may divide the estimate for even characters by 2 since our character sum is over an initial interval. This gives

$$\sum_{\substack{v>1\\ v\leq\frac{1}{2}p^{1/4}+1}}\left|\sum_{j<p/v^2}\chi^i(j)\right| \leq \sum_{\substack{v>1\\ v\leq\frac{1}{2}p^{1/4}+1}}\left(\frac{1}{2\pi}p^{1/2}\log p + p^{1/2}\right) \leq \frac{1}{4\pi}p^{3/4}\log p + \frac{1}{2}p^{3/4}.$$

Thus,

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

Hence

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

and we would like this expression to be smaller than the one in (2). That is, we would like the inequality

$$\frac{53(p-1)}{88d} > \left(1-\frac{1}{d}\right)p^{3/4}\left(\frac{1}{4\pi}\log p + \frac{5}{2}\right),$$

to hold true, or equivalently,

$$\frac{53(p-1)}{88p^{3/4}} > (d-1)\left(\frac{1}{4\pi}\log p + \frac{5}{2}\right).$$

Using $d < \log p + 1$, we see that this inequality holds for all $p > 3 \times 10^8$. □

Remark 3. Instead of [7] for our estimate of the character sum, we might have used [16] or we might have used the "smoothed" version in [13]. The former would require raising the lower limit 3×10^8 slightly, while the latter would likely lead to a reduction in the lower limit, but at the expense of a more complicated proof.

For a prime $p > 3 \times 10^8$ and an integer $d \mid p-1$ with $d < \log p + 1$, let $\mathcal{C}_{d,p}$ denote a set of squarefree coset representatives for $H_{d,p}$ smaller than p as guaranteed to exist by Proposition 2. Also, let $\mathcal{S}_{d,p}$ denote the set of primes that divide some member of $\mathcal{C}_{d,p}$ and let $\mathcal{S}_p$ be the union of all of the sets $\mathcal{S}_{d,p}$ for $d \mid p-1$, $d < \log p + 1$.

Let $\omega(n)$ denote the number of distinct prime divisors of n. It is known that $\omega(n) \le (1+o(1))\log n/\log\log n$ as $n \to \infty$. We have the weaker, but explicit inequality: $\omega(n) < \log n$ for $n > 6$. To see this, note that it is true for $\omega(n) \le 2$, since it holds for $n = 7$, and for $n \ge 8$ we have $\log n > 2$. If $\omega(n) = k \ge 3$, then $n \ge 6 \cdot 5^{k-2}$, so that $k \le (\log n + \log(25/6))/\log 5$, which is smaller than $\log n$ for $n \ge 11$. But $k \ge 3$ implies $n \ge 30$.

As a corollary, we conclude that under the hypotheses of Proposition 2, we have each $\#\mathcal{S}_{d,p} < d\log p$ and $\#\mathcal{S}_p < \frac{1}{2}(\log p + 1)^3$.

3. Proof of Theorem 1 for large p

Assume the prime p exceeds 3×10^8. Let $\mathcal{S}_p$ denote the set of primes identified at the end of the last section, let $N = p - 1$, and let

$$K = \pi(N) - \#\mathcal{S}_p > \pi(N) - \frac{1}{2}(\log p + 1)^3$$

denote the number of remaining primes smaller than p.

For an integer m with $0 < m < p$, let $f(m)$ denote the number of unordered pairs of distinct primes q, r with $q, r < p$, $qr \equiv m \pmod{p}$ and $q, r \notin \mathcal{S}_p$. Set

$$\mathcal{A} = \{m \in (0,p) : f(m) > K/\sqrt{p}\} \cup \{1\}, \quad A = \#\mathcal{A}.$$

Lemma 4. *For $p > 3 \times 10^8$, we have*

$$A > \frac{N}{\log N} + 2.$$

Proof. Evidently,

$$\sum_{m=1}^{p-1} f(m) = \binom{K}{2} = \frac{1}{2}K(K-1). \tag{3}$$

Further, if two pairs q, r and q', r' are counted by $f(m)$, then either they have no prime in common or they are the same pair. Thus, for each m,

$$f(m) \le \frac{1}{2}K. \tag{4}$$

Since

$$\sum_{m \notin \mathcal{A}} f(m) \le (N - A)K/\sqrt{p},$$

we have by (3) that

$$\sum_{m \in \mathcal{A}} f(m) \ge \frac{1}{2}K(K-1) - (N-A)K/\sqrt{p}.$$

Thus, from (4),

$$A \ge \frac{1}{K/2} \sum_{m \in \mathcal{A}} f(m) \ge K - 1 - 2(N-A)/\sqrt{p} > K - 2\sqrt{p} - 1.$$

Since $K > \pi(N) - \frac{1}{2}(\log p + 1)^3$, by using inequality (3.1) in [18] and $p > 3 \times 10^8$, we have the inequality in the lemma. $\square$

For a positive integer k, let $\mathcal{A}^k$ denote the set of k-fold products of members of $\mathcal{A}$.

Lemma 5. *There are positive integers $d < \log p + 1$, $k < 2\log p + 3$ such that $\mathcal{A}^k$ contains a subgroup $H_{d,p}$ of $(\mathbb{Z}/p\mathbb{Z})^*$.*

Proof. Let g be a primitive root modulo p and let $\mathcal{A}'$ denote the set of discrete logarithms of members of $\mathcal{A}$ to the base g. That is, $j \in \mathcal{A}'$ with $0 \le j < N$ if and only if $g^j \pmod{p} \in \mathcal{A}$. We now apply a theorem of Lev [12, Theorem 2′] to the set $\mathcal{A}'$. With $\kappa := \lceil (N-1)/(A-2) \rceil$, this result implies that there are positive integers $d' \le \kappa$ and $k \le 2\kappa + 1$ such that $k\mathcal{A}'$ contains N consecutive multiples of d'. Here, $k\mathcal{A}'$ denotes the set of integers that can be written as the sum of k members of $\mathcal{A}'$. Thus, reducing mod N, the set $k\mathcal{A}'$ contains a subgroup of $\mathbb{Z}/N\mathbb{Z}$ of index $d := (d', N)$. Hence, $\mathcal{A}^k$ contains the subgroup $H_{d,p}$ of $(\mathbb{Z}/p\mathbb{Z})^*$. Further, from Lemma 4, we have $d \le d' \le \kappa < \log p + 1$, which completes the proof. $\square$

Lemma 6. *For the subgroup $H_{d,p}$ of $(\mathbb{Z}/p\mathbb{Z})^*$ produced in Lemma 5, each member of $H_{d,p}$ has a representation modulo p as a squarefree number involving primes smaller than p and not in $\mathcal{S}_p$ (and so not in $\mathcal{S}_{d,p}$).*

Proof. Suppose that $m \in \mathcal{A}^k$, so that

$$m = m_1 m_2 \dots m_k \equiv (q_1 r_1)(q_2 r_2) \dots (q_k r_k) \pmod{p}, \tag{5}$$

where each $m_i \in \mathcal{A}$ and $m_i \equiv q_i r_i \pmod{p}$. This last product over primes is p-smooth, but is not necessarily squarefree. However, each $m_i \in \mathcal{A}$ has many representations as $q_i r_i$, in fact at least $K/\sqrt{p}$ representations, with each representation involving two new primes. So, if k is small enough, there will be a representation of each m_i so that the product of primes in (5) is indeed squarefree. Now $k < 2\log p + 3$, so having at least $2(k-1)+1 < 4\log p + 5$ representations for each member of $\mathcal{A}$ is sufficient. The number of representations exceeds $K/\sqrt{p}$ and by the same calculation that gave us the last step in Lemma 4, we have this expression exceeding $\sqrt{p}/\log p$. This easily exceeds $4\log p + 5$ for $p > 3 \times 10^8$. □

It is now immediate that every residue class mod p contains a squarefree number with prime factors at most p. Indeed this is true for $0 \pmod{p}$ — take p as the representative. For a nonzero class $j \pmod{p}$, find that member m of $\mathcal{C}_{d,p}$ with $j \in mH_{d,p}$, and write $j = mh$ with $h \in H_{d,p}$. We have seen that each member of $H_{d,p}$ has a squarefree representative using primes smaller than p and not in $\mathcal{S}_{d,p}$. Since m is squarefree and uses only primes in $\mathcal{S}_{d,p}$ it follows that mh is also squarefree using only primes smaller than p. This completes the proof of Theorem 1 for primes $p > 3 \times 10^8$.

4. Verification of Theorem 1 for small p

It remains only to verify the theorem for $p < 3 \times 10^8$. For $p > 10^4$ we use the following simple strategy: Compute a primitive root $g \pmod{p}$, and find pairwise coprime, squarefree, p-smooth numbers m_i such that $m_i \equiv g^{2^i} \pmod{p}$ for each nonnegative integer $i \le \log_2(p-2)$. If this is possible then, given any nonzero residue $n \pmod{p}$, we have $n \equiv g^k \pmod{p}$ for some integer $k \in [0, p-2]$. Expressing k in binary, viz. $k = \sum_{0 \le i \le \log_2(p-2)} b_i 2^i$ for $b_i \in \{0,1\}$, we have $n \equiv \prod_{0 \le i \le \log_2(p-2)} m_i^{b_i} \pmod{p}$. Since the m_i are pairwise coprime, the residue class of n is thus represented by a squarefree p-smooth number, as desired.

It is convenient to choose m_i of the form $q_i r_i$ for primes q_i, r_i. To find these efficiently, for each $i = 0, 1, 2, \ldots$ we search through small primes q, compute the smallest positive $r \equiv q^{-1} g^{2^i} \pmod{p}$, test whether r is prime, and ensure that qr is coprime to m_j for $j < i$. The only essential ingredient needed to carry this out is a fast primality test; we used a strong Fermat test to base 2 coupled with the classification [6] of small strong pseudoprimes, which would allow us, in principle, to handle any $p < 2^{64}$. Heuristically, one can expect this method to succeed using $O(\log^3 p)$ arithmetic operations on numbers of size p, and we found it to be very fast in practice; it takes just minutes to verify the theorem for all $p \in (10^4, 3 \times 10^8)$ on a modern multicore processor.

For $p < 10^4$ we fall back on a brute-force algorithm: For each integer $a \in [1, p-1]$, consider each of the numbers $a, a+p, a+2p, \ldots$ until encountering one that divides $\prod_{\substack{q \text{ prime} \\ q<p}} q$. This takes only seconds to check for all $p < 10^4$ other than 5 and 7. (For $p \in \{5, 7\}$ one can see directly that $4 + p\mathbb{Z}$ is not represented, but all other residue classes are.)

5. Generating all of the primes from nothing

We give two applications of Theorem 1 to Euclidean prime generators. The first was described without proof in [5, Section 1.1.3]; we supply the short proof here.

Corollary 7. *Starting from the emptyset, recursively define a sequence of primes, where if n is the product of the primes generated so far, take*

$$p = \min\{q \text{ prime} : q \nmid n,\ q \mid d+1 \text{ for some } d \mid n\}$$

as the next prime. This sequence begins with $2, 3, 7, 5$*, and then produces the primes in order.*

Proof. Let $p \geq 11$ be the least prime not yet produced after the 4th step and let n be the product of the primes smaller than p. By Theorem 1, there exists some $d \mid n$ with $d \equiv -1 \pmod{p}$. Thus, we generate p at the next step. $\square$

The second variant was described in [4]. With Theorem 1 in hand, we can give a shorter proof. (As will be clear from the proof, there is an obstruction preventing the terms from appearing in strict numerical order in this case, so the conclusion is weaker than that of Corollary 7.)

Corollary 8. *Starting from the empty set, recursively generate a sequence of primes where if n is the product of the primes generated so far, take as the next prime some prime factor of some $d + n/d$, where $d \mid n$. Such a sequence can be chosen to contain every prime.*

Proof. One such sequence begins with 2, 3, 5, 13, 7. Suppose $p > 7$ is a prime number and that the sequence constructed so far contains every prime smaller than p and perhaps some primes larger than p, but it does not contain p. Let n be the product of the primes generated so far. If $\left(\frac{-n}{p}\right) = 1$ then there exists $a \in (\mathbb{Z}/p\mathbb{Z})^*$ such that $a + n/a = 0 \pmod{p}$. By Theorem 1 there exists $d \mid n$ belonging to the class of a, so we can choose p as the next prime. On the other hand, since $p > 5$, if $\left(\frac{-n}{p}\right) = -1$, then [4, Lemma 3(i)] guarantees the existence of $a \in (\mathbb{Z}/p\mathbb{Z})^*$ such that $\left(\frac{a+n/a}{p}\right) = -1$. By Theorem 1 there exists $d \mid n$ belonging to the class of a, and by multiplicativity it follows that $d + n/d$ has a prime factor q satisfying $\left(\frac{q}{p}\right) = -1$. Choosing q as the next prime, we thus have $\left(\frac{-qn}{p}\right) = 1$, so by the above argument we can now take p. This completes the proof. $\square$

6. Comments and problems

The prime generator of Corollary 7 has its roots in a construction in the primality test of [1]. There, a finite initial set of primes is given with product I, and then one takes the product E of all primes $p \nmid I$ with $p \mid d+1$ for some $d \mid I$. The primality test can be used for numbers $n < E^2$ and runs in time $I^{O(1)}$. Further, it is shown that there are choices of I, E with $I = (\log E)^{O(\log\log\log E)}$, so the test runs in "almost" polynomial time. The same I, E construction (with I no longer required to be squarefree) is used in the finite fields primality test of Lenstra [11].

In Theorem 1 we insist that the squarefree p-smooth integers used be positive. If negatives are allowed, then the primes 5 and 7 are no longer exceptional cases. Further, if "d" is allowed to be negative in the context of the prime generator in Corollary 7, the primes are generated in order. (For this to be nontrivial, d should not be chosen as -1.)

Suppose we use the generator of Corollary 8 by always returning the least prime possible, and say this sequence of primes is $q_1, q_2, \ldots$. Does $\{q_k\}$ contain every prime? Is there way of choosing the sequence in Corollary 8 such that every prime is generated and the kth prime generated is asymptotically equal to the kth prime? Is it true that any sequence containing all primes as in Corollary 8 cannot contain the primes in order starting from some point? These questions might all be asked if we allow prime factors of $d \pm n/d$ in Corollary 8 instead of just $d + n/d$.

Presumably in Theorem 1, when p is large, residues $a \pmod{p}$ have many representations as squarefree p-smooth integers. Say we try to minimize the largest squarefree p-smooth used.

For $p > 7$, let $M(p)$ be the smallest number such that every residue mod p can be represented by a squarefree p-smooth number at most $M(p)$. Our proof shows that $M(p) \le p^{O(\log p)}$. We conjecture that $M(p) \le p^{O(1)}$.

We mentioned in the Introduction that the condition "p-smooth" in Theorem 1 cannot be relaxed to "y-smooth" for any $y < p$, since otherwise the residue class $0 \pmod p$ will not be represented. However, we may ask for the smallest number $y = y(p)$ such that every nonzero residue class mod p can be represented by a y-smooth squarefree number. Via the Burgess inequality, it is likely that one can show that $y(p) \le p^{1/(4\sqrt{e})+o(1)}$ as $p \to \infty$. Assuming the Generalized Riemann Hypothesis for Kummerian fields (as Hooley [9] did in his GRH-conditional proof of Artin's conjecture), it is likely that one can prove that $y(p) = O((\log p)^2)$. We note that if one drops the "squarefree" condition then these statements follow from work of Harman [8] unconditionally and Ankeny [2] under GRH; see also the recent paper [10] for a strong, numerically explicit version of the latter.

Acknowledgments

We thank Enrique Treviño and Kannan Soundararajan for some helpful comments. The first-named author was partially supported by EPSRC Grant EP/K034383/1.

References

[1] L. M. Adleman, C. Pomerance, and R. S. Rumely, *On distinguishing prime numbers from composite numbers*, Ann. of Math. **117** (1983), 173–206.
[2] N. C. Ankeny, *The least quadratic non residue*, Ann. of Math. **55** (1952), 65–72.
[3] A. R. Booker, *On Mullin's second sequence of primes*, Integers **12** (2012), 1167–1177.
[4] A. R. Booker, *A variant of the Euclid-Mullin sequence containing every prime*, J. Integer Seq., Volume 19 (2016), Article 16.6.4.
[5] R. E. Crandall and C. Pomerance, *Prime numbers: A computational perspective*, second ed., Springer, New York, 2005.
[6] J. Feitsma, *Pseudoprimes*, http://www.janfeitsma.nl/math/psp2/index.
[7] D. A. Frolenkov and K. Soundararajan, *A generalization of the Pólya–Vinogradov inequality*, Ramanujan J. **31** (2013), 271–279.
[8] G. Harman, *Integers without large prime factors in short intervals and arithmetic progressions*, Acta Arith. **91** (1999), 279–289.
[9] C. Hooley, *On Artin's conjecture*, J. Reine Angew. Math. **225** (1967), 209–220.
[10] Y. Lamzouri, X. Li, K. Soundararajan, *Conditional bounds for the least quadratic non-residue and related problems*, Math. Comp. **84** (2015), 2391–2412.
[11] H. W. Lenstra, Jr., *Galois theory and primality testing*, in "Orders and their applications (Oberwolfach, 1984), Lecture Notes in Math. vol. 1142 (1985), Springer-Verlag, Berlin, pp. 169–189.
[12] V. F. Lev, *Optimal representations by sumsets and subset sums*, J. Number Theory **62** (1997), 127–143.
[13] M. Levin, C. Pomerance, and K. Soundararajan, *Fixed points for discrete logarithms*, ANTS IX Proceedings, Springer LNCS **6197** (2010), 6–15.
[14] A. A. Mullin, *Recursive function theory*, Bull. Amer. Math. Soc. **69** (1963), 737.
[15] P. Pollack and E. Treviño, *The primes that Euclid forgot*, Amer. Math. Monthly **121** (2014), 433–437.
[16] C. Pomerance, *Remarks on the Pólya–Vinogradov inequality*, Proc. Integers Conf. Oct. 2009, Integers **11A** (2011), Article 19, 11 pp.
[17] K. Rogers, *The Schnirelmann density of the squarefree integers*, Proc. Amer. Math. Soc. **15** (1964), 515–516.
[18] J. B. Rosser and L. Schoenfeld, *Approximate formulas for some functions of prime numbers*, Illinois J. Math. **6** (1962), 64–94.

[19] S. S. Wagstaff, Jr., *Computing Euclid's primes*, Bull. Inst. Combin. Appl. **8** (1993), 23–32.
[20] T. D. Wooley, *A superpowered Euclidean prime generator*, 2016, preprint.

School of Mathematics, Bristol University, University Walk, Bristol, BS8 1TW, UK
E-mail address: `andrew.booker@bristol.ac.uk`

Mathematics Department, Dartmouth College, Hanover, NH 03755, USA
E-mail address: `carl.pomerance@dartmouth.edu`