ON THE LEAST PRIME IN CERTAIN ARITHMETIC PROGRESSIONS

ANDREW GRANVILLE AND CARL POMERANCE

ABSTRACT

We find infinitely many pairs of coprime integers, a and q, such that the least prime congruent to a (modulo q) is unusually large. In so doing we also consider the question of approximating rationals by other rationals with smaller and coprime denominators.

1. *Introduction*

For any $x > x_0$ and for any positive valued function $g(x)$ define

$$R(x) = e^{\gamma} \log x \log_2 x \log_4 x/(\log_3 x)^2,$$
$$L(x) = \exp(\log x \log_3 x/\log_2 x),$$
$$E_g(x) = \exp(\log x/(\log_2 x)^{g(x)}).$$

Here $\log_k x$ is the k-fold iterated logarithm, γ is Euler's constant, and x_0 is chosen large enough so that $\log_4 x_0 > 1$.

The usual method used to find large gaps between successive prime numbers is to construct long sequences S of consecutive integers, each of which has a 'small' prime factor (so that they cannot be prime); then, the gap between the largest prime before S and the next, is at least as long as S.

Similarly if one wishes to find an arithmetic progression $a \pmod q$, with $\gcd(q, a) = 1$, in which the least prime is fairly large, then it suffices to ensure that each integer of the sequence $a, a+q, \ldots, a+kq$ has a 'small' prime factor. Let n be the product of those small primes (note that $\gcd(q, n) = 1$) and let the integer r be an inverse of $q \pmod n$. As $\gcd(ar+i, n) = \gcd(a+iq, n)$ we see that each of $a, a+q, \ldots, a+kq$ has a small prime factor if and only if each of $ar, ar+1, \ldots, ar+k$ does. Thus we are again considering long sequences of consecutive integers, each with a 'small' prime factor.

Jacobsthal's function $j(n)$ is defined to be the number of integers in the longest sequence of consecutive integers, each of which has a factor in common with n. Rankin **[10]** has shown that if n is the product of the first k primes then

$$j(n) \geqslant \{1+o(1)\} R(n) \tag{1}$$

as $k \to \infty$; and Maier and Pomerance [7] have recently improved this to

$$j(n) \geqslant \{c+o(1)\} R(n) \tag{2}$$

as $k \to \infty$ where c ($\approx 1.31246\ldots$) is the solution of $4/c - e^{-4/c} = 3$. As a consequence one knows that there are arbitrarily large pairs of successive prime numbers $q > p$ with difference as large as $\{c+o(1)\} R(p)$.

Received 21 March 1989.

1980 *Mathematics Subject Classification* (1985 *Revision*) 11N13.

The second author is partly supported by an NSF grant.

J. London Math. Soc. (2) 41 (1990) 193–200

On the other hand, in 1978 Iwaniec **[4]** proved Jacobsthal's original conjecture **[5]** which stated that, there exists a constant $\kappa > 0$ such that for all integers $n \geqslant 2$ we have

$$j(n) \leqslant \kappa \log^2 n, \tag{3}$$

which is analogous to Cramér's conjecture **[1]** that the largest gap between successive primes $q > p$ is $\ll \log^2 p$.

For any integer q with less than $\exp(\log y / \log_2 y)$ distinct prime factors, Pomerance **[8]** showed that

$$j(n_q) \geqslant \{1 + o(1)\} \frac{\phi(q)}{q} R(n_q), \tag{4}$$

where n_q is the product of the primes less than or equal to y that do not divide q. From this he deduced that for any q with less than $\exp(\log_2 q / \log_3 q)$ distinct prime factors, there exists an integer a, prime to q, for which

$$p(q, a) \geqslant \{1 + o(1)\}\, \phi(q)\, R(q), \tag{5}$$

where $p(q, a)$ is the least prime in the arithmetic progression $a \pmod q$.

By the method of [7] it is possible to improve (4) to

$$j(n_q) \geqslant \{c + o(1)\} \frac{\phi(q)}{q} R(n_q) \tag{6}$$

and deduce that (5) may be improved to

$$p(q, a) \geqslant \{c + o(1)\}\, \phi(q)\, R(q). \tag{7}$$

The main result of this paper is an improvement of (7) for infinitely many pairs q, a. However, rather than improving (4), we apply (4) to primes in arithmetic progressions in a different way.

THEOREM 1. *Let $f(x)$ be a positive valued function that tends to 0 as $x \to \infty$. There are at least $x^2 / E_f(x)$ arithmetic progressions $a \pmod q$, with $(a, q) = 1$ and $1 \leqslant q \leqslant x$, for which*

$$p(q, a) \geqslant \{2 + o(1)\}\, qR(q).$$

Let $P(q) = \max_{(a,q)=1} p(q, a)$. Linnik **[6]** has shown that $P(q) \ll q^{c'}$ for some $c' > 0$, and Pomerance **[8]** that $P(q) \geqslant \{e^{\gamma} + o(1)\}\, \phi(q) \log q$ for all positive integers q. (Note that $P(q) \geqslant \{1 + o(1)\}\, \phi(q) \log q$ follows immediately from the prime number theorem.) We conjecture that $P(q) \gg \phi(q) \log^2 q$ for all q.

The weaker conjecture that $P(q)/\phi(q) \log q \to \infty$ is still unsolved, though, by (5), it can be seen to hold for almost all q. (This conjecture seems to be most difficult to prove when q is the product of the first k primes.)

Prachar **[9]** and Schinzel **[11]** have shown that there is some absolute constant $\alpha > 0$ such that for any fixed non-zero integer a, there are infinitely many integers q with $(q, a) = 1$ and

$$p'(q, a) \geqslant \{\alpha + o(1)\}\, qR(q),$$

where $p'(q, a)$ denotes the least prime $p \equiv a \pmod q$ with $p > a$. We now sketch a proof that $\alpha \geqslant c$. Let $\varepsilon > 0$ be fixed and arbitrarily small. By the method in [7], one can show that for all sufficiently large x, there are integers s_p for each prime $p \leqslant x$ such that

(i) $s_p = 0$ for each prime p that divides a,

(ii) $1 \leqslant s_p \leqslant p-1$ for each prime $p \leqslant x$, that does not divide a,
(iii) for each n in the range $1 \leqslant n \leqslant (c-\varepsilon) R(e^x)$, there is some prime $p \leqslant x$ for which $n \equiv s_p \pmod p$.

Choose q to be any integer in the range $\prod_{p \leqslant x} p < q < 2\prod_{p \leqslant x} p$, for which $q \equiv 1 \pmod p$ if p divides a, and $qs_p \equiv -a \pmod p$ if $p \leqslant x$ and p does not divide a.

As we may take x sufficiently large so that it lies in the range $q+a > x > |a|$, one can easily see that $\gcd(q,a) = 1$ and that for every n in the range

$$1 \leqslant n \leqslant (c-\varepsilon) R(e^x),$$

the number $qn+a$ is composite. Therefore

$$p'(q,a) \geqslant \{c-\varepsilon+o(1)\}\, qR(q).$$

We do not know how to further improve this result.

There are two main tools in the proof of Theorem 1. We start with the following technical improvement of (4).

PROPOSITION 1. *Fix $\frac{1}{5} > \delta > 0$ and let n be the product of the primes less than or equal to y. For any sufficiently large y (greater than y_δ) and any positive integer q with less than $\exp(\log y/\log_2 y)$ distinct prime factors, there exist at least $n/E_\delta(n)$ disjoint subintervals of $[1,n]$, each of length $(1-4\delta)(\phi(q)/q)R(n)$, that contain only integers which have some prime factor that does not divide q and is at most y.*

REMARK. Taking $q = 1$ in Proposition 1 we can see that if $f(x)$ is a positive valued function that tends to 0 as $x \to \infty$ then there are at least $x/E_f(x)$ disjoint subintervals of $[1,x]$ of length greater than or equal to $\{1+o(1)\}\, R(x)$ that contain only composite numbers. If we were to suppose that these subintervals are 'evenly' spread across the interval $[1,x]$ then we should expect that such a subinterval would occur in $[1, E_f(x)]$. Thus, for any z, there would be a pair of consecutive prime numbers less than or equal to z with difference $\gg \log z \log_2 z$.

We also need the following.

THEOREM 2. *For given positive integers m and n with n squarefree, we define $T_m(n)$ to be the set of integers t, $0 \leqslant t \leqslant n-1$, for which there do not exist integers p and q, with $0 \leqslant p \leqslant q \leqslant m$ and $(p,q) = (q,n) = 1$, for which*

$$\left|\frac{t}{n}-\frac{p}{q}\right| < \frac{1}{n}. \tag{8}$$

If $m \geqslant 2\sqrt{n}$ then

$$(n/m)^2 \ll \#T_m(n) \ll (n/m)^2 \log^4 n. \tag{9}$$

We believe that it should be possible to replace $\log^4 n$ by $\log^2 n$ in the upper bound in (7); it would be interesting to know the correct order of $\#T_m(n)$.

2. *The proof of Theorem* 1

We shall assume that both Proposition 1 and Theorem 2 are true. Fix $\varepsilon > 0$ and choose n as large as possible so that it is the product of the primes up to some value of y, yet with $n \leqslant x^2/E_{2\varepsilon}(x)$. Let $r = [(1-12\varepsilon) R(n)]$ (which is larger than $(2-25\varepsilon) R(x)+1$ for all sufficiently large x, by the prime number theorem).

By taking $q = 1$, $\delta = 3\varepsilon$ in Proposition 1 and $m = x$ in Theorem 2 we see that there are at least $C\,(= n/E_{3\varepsilon}(n) + O(x^2\log^4 x/E_{2\varepsilon}(x)^2))$ different values of t, less than n, with no two values less than $r+1$ apart, such that each of

$$t, t+1, \ldots, t+r$$

has a prime factor in common with n, and for which there exist positive integers p and q, with $p \leqslant q \leqslant x$ and $(p,q) = (q,n) = 1$, such that (8) holds. Then, by multiplying (8) through by qn, we note that there exists an integer a, $|a| < q$, such that $qt - pn = a$ and $(a,q) = 1$ (as $(q,pn) = 1$).

Now, if any one of

$$a, a+q, \ldots, a+rq \tag{10}$$

is prime then it must be a prime less than or equal to y, as $(a+jq, n) = (t+j, n) > 1$, and so there are at most $(r+1)\pi(y)$ such values of a, for any fixed q. Thus at most

$$(r+1)\,\pi(y)\,x = O(x\log^2 x)$$

such pairs a and q arise in this way.

Also observe that any two arithmetic progressions $a \pmod q$ corresponding to different values of t must themselves be different. (Else, if $qt - pn = a$ and $qt' - p'n = a'$ where $a \equiv a' \pmod q$, then $p = p'$, as $p \equiv p' \pmod q$ and $1 \leqslant p,\ p' \leqslant q$, and so $t' = t-1$, t or $t+1$, as $a' = a-q$, a or $a+q$, which contradicts $|t-t'| > r$.)

Discarding values of t that give rise to a prime value in (10), we are left with $C + O(x\log^2 x)$ (which is greater than $x^2/E_\varepsilon(x)$ for all sufficiently large x) distinct arithmetic progressions $a \pmod q$, with $(a,q) = 1$ and $q \leqslant x$, for which

$$p(q,a) > (2-25\varepsilon)\,R(x)\,q \geqslant (2-25\varepsilon)\,qR(q).$$

The result follows by letting $\varepsilon \to 0$.

3. *On the number of 'well-sieved' intervals*

Proof of Proposition 1. The proof is based on that of [**8**, Theorem 3]. Let

$$U = (1-4\delta)\frac{\phi(q)}{q}R(e^y), \qquad A = L(y)^{1-\delta}, \qquad B = y/\log^{2\delta} y.$$

The idea is to assign, in at least $n/E_\delta(n)$ different ways, arithmetic progressions $a_p \pmod p$ for each prime p dividing m (where $m = \prod_{p\leqslant y,\, p\nmid q} p$), in such a way that every integer in $[1, U]$ belongs to at least one of these progressions. Let t be the least positive integer for which $t \equiv -a_p \pmod p$ for each prime p dividing m; then every integer in the interval $[t+1, t+U]$ has a prime factor in common with m. (Note that if $j \equiv a_p \pmod p$ then p divides $t+j$.)

In each of our assignments we shall take the arithmetic progressions $0 \pmod p$ for $p \in (A, B]$ and so the values of t will be congruent modulo r (where $r = \prod_{A<p\leqslant B,\, p\nmid q} p$). As $U < r$ for all sufficiently large y, so all of the intervals are distinct and we have proved Proposition 1.

Define $\psi(x;z)$ to be the number of positive integers less than or equal to x, free of prime factors greater than z. In [2], de Bruijn showed that $\psi(x;z) = x/s^{s+o(s)}$ as s tends to infinity, uniformly in the range $x \geqslant z \geqslant \exp((\log x)^{\frac{3}{4}})$, where $s = \log x/\log z$. From this one can immediately deduce that

$$\psi(U;A) = U/(\log y)^{1/(1-\delta)+o(1)} \tag{11}$$

as $(1-5\delta)y\log_3 y/\log_2 y < U < y\log y$.

Now, after we remove all multiples of primes dividing r from $[1, U]$, we are left with the $\psi(U;A)$ integers composed only of primes less than or equal to A, the $\sum_{p>B}[U/p]$ integers divisible by primes greater than B, and the remaining integers that are divisible by some prime from $(A, B]$ which divides q. Now

$$\begin{aligned}\sum_{p>B}\left[\frac{U}{p}\right] &\leqslant U\sum_{B<p\leqslant U}\frac{1}{p} \leqslant U\log\left(\frac{\log U}{\log B}\right)\{1+o(1)\}\\ &\leqslant \{1+2\delta+o(1)\}\,U\frac{\log_2 y}{\log y},\end{aligned}$$

and the third set of integers above has cardinality at most

$$\omega(q)\frac{U}{A} = o\left(U\frac{\log_2 y}{\log y}\right),$$

where $\omega(q)$ is the number of distinct prime factors of q. By taking these estimates together with (11) we see that we have

$$R_1 \leqslant U\frac{\log_2 y}{\log y}\{1+2\delta+o(1)\}$$

integers left.

In our 'second sieving' we choose the arithmetic progression $a_p \pmod p$, for each successive prime $p \leqslant A$ which does not divide q, so that we cover as many of the remaining unsieved integers as possible. Then the number of integers left is

$$\begin{aligned}R_2 &\leqslant R_1\prod_{\substack{p\leqslant A\\ p\nmid q}}\left(1-\frac{1}{p}\right) \leqslant R_1\frac{q}{\phi(q)}\prod_{p\leqslant A}\left(1-\frac{1}{p}\right)\\ &\leqslant \{1+2\delta+o(1)\}(1-4\delta)\frac{\phi(q)}{q}\frac{e^{\gamma}y\log y\log_3 y}{(\log_2 y)^2}\frac{\log_2 y}{\log y}\frac{q}{\phi(q)}\frac{e^{-\gamma}}{(1-\delta)}\frac{\log_2 y}{\log y\log_3 y}\\ &\leqslant (1-\delta+o(1))\frac{y}{\log y}.\end{aligned}$$

Let P be the set of primes in $(B, y]$ that do not divide q, which has cardinality

$$\Pi = \frac{y}{\log y}\left(1-\frac{1}{\log^{2\delta} y}+O\left(\frac{1}{\log y}\right)\right).$$

Note that $R_2 \leqslant \Pi$, and then select $\Pi - R_2$ different integers in $[1, U]$ to take together with those integers remaining from our sievings, so that we now have a set N of Π distinct integers in $[1, U]$. Any bijection θ from P to N assigns an arithmetic progression for each of the remaining primes (that is $\theta(p) \pmod p$ for each prime p in P), so that every integer in $[1, U]$ belongs to at least one of our arithmetic progressions. There are $\Pi!$ such bijections. It is possible, however, that many such bijections may lead to the same set of arithmetic progressions and so we must take account of this. Each prime p in P is greater than B and so there are no more than $(U/B)+1$ elements of N in any given arithmetic progression $a_p \pmod p$. Therefore at most $\{(U/B)+1\}^{\Pi}$ bijections give rise to the same value of t. Therefore the number of *distinct* values of t that we get is at least

$$\Pi! \Big/ \left(\frac{U}{B}+1\right)^{\Pi} = \exp\{y(1-\log^{-2\delta} y + O(\log_2 y/\log y))\}$$
$$\geqslant n/\exp(\log n/(\log\log n)^{\delta})$$

for y sufficiently large, as $\log n = y + O(y/\log y)$.

4. *Approximation of rationals by rationals with coprime denominators*

Proof of Theorem 2. Fix r and consider the set of fractions a/b with

$$0 \leqslant a \leqslant b \leqslant r$$

and $(a,b)=1$. We order them

$$\frac{0}{1} = \frac{a_1}{b_1} < \frac{a_2}{b_2} < \ldots < \frac{a_k}{b_k} = \frac{1}{1},$$

so that, by the theory of Farey fractions (see Hardy and Wright [**3**, pp. 23–24]), we have

$$b_i + b_{i+1} \geqslant r, \tag{12}$$
$$\gcd(b_i, b_{i+1}) = 1, \tag{13}$$

and

$$\frac{a_{i+1}}{b_{i+1}} - \frac{a_i}{b_i} = \frac{1}{b_i b_{i+1}} \tag{14}$$

for each $i = 1, 2, \ldots, k-1$.

Clearly any rational of the form t/n, with $0 \leqslant t \leqslant n-1$, lies in such an interval $[a_i/b_i, a_{i+1}/b_{i+1}]$ for some i.

Lower bound. Let $r=m$ and consider any $b_i \leqslant n/4m$. There are at least $n/mb_i - 3$ integers t for which

$$\frac{t}{n} \in \left[\frac{a_i}{b_i}+\frac{1}{n}, \frac{a_{i+1}}{b_{i+1}}-\frac{1}{n}\right],$$

as $a_{i+1}/b_{i+1} - a_i/b_i \geqslant 1/mb_i (\geqslant 4/n)$; and

$$\min_{\substack{0 \leqslant p \leqslant q \leqslant m \\ (q,n)=1}} \left|\frac{t}{n}-\frac{p}{q}\right| \geqslant \min_{j=i \text{ or } i+1} \left|\frac{t}{n}-\frac{a_j}{b_j}\right| \geqslant \frac{1}{n}$$

for any such t. Therefore

$$\#T_m(n) \geqslant \sum_{b \leqslant n/4m} \sum_{(a,b)=1} \frac{n}{mb} - 3$$
$$\gg \frac{n}{m} \sum_{b \leqslant n/4m} \frac{\phi(b)}{b} \gg \left(\frac{n}{m}\right)^2.$$

Upper bound. We shall prove the following.

LEMMA. *There exists a constant $\kappa > 0$ such that if a, b, c, d and n are positive integers, with n squarefree and $bc-ad = 1$, then there exist integers u and v with $(u,v) = (v,n) = 1$, $a/b \leqslant u/v \leqslant c/d$ and $v \leqslant \kappa(b+d)\log^2 n$.*

Now let $r = m/(2\kappa \log^2 n)$ and suppose that $t/n \in [a_i/b_i, a_{i+1}/b_{i+1}]$, where $b_i b_{i+1} > n$. By (14) and the lemma, there exist integers p and q with $(p,q) = (q,n) = 1$, $p/q \in [a_i/b_i, a_{i+1}/b_{i+1}]$ and $q \leqslant m$, and so

$$\left|\frac{t}{n}-\frac{p}{q}\right| \leqslant \left|\frac{a_{i+1}}{b_{i+1}}-\frac{a_i}{b_i}\right| = \frac{1}{b_i b_{i+1}} < \frac{1}{n}.$$

Therefore

$$\# T_m(n) \leqslant \sum_{\substack{i=1 \\ b_i b_{i+1} \leqslant n}}^{k-1} \# \left\{\text{integers } t : \frac{t}{n} \in \left[\frac{a_i}{b_i}, \frac{a_{i+1}}{b_{i+1}}\right]\right\}. \tag{15}$$

Now, whenever $b_i b_{i+1} \leqslant n$, we have

$$\# \left\{\text{integers } t : \frac{t}{n} \in \left[\frac{a_i}{b_i}, \frac{a_{i+1}}{b_{i+1}}\right]\right\} \leqslant \frac{n}{b_i b_{i+1}} + 2 \leqslant \frac{3n}{b_i b_{i+1}} \tag{16}$$

and $\min\{b_i, b_{i+1}\} \leqslant 2n/r$, as $\max\{b_i, b_{i+1}\} \geqslant r/2$ by (12). Therefore, by (15) and (16),

$$\begin{aligned} \# T_m(n) &\leqslant \sum_{b \leqslant 2n/r} \sum_{(a,b)=1} \frac{12n}{br} \\ &\leqslant \sum_{b \leqslant 2n/r} \frac{12n}{r} = \frac{24n^2}{r^2} \ll \frac{n^2}{m^2} \log^4 n. \end{aligned}$$

Proof of the lemma. Let $g = (d,n)$, $m = n/g$ and e be an inverse of $d \pmod m$. We know that there exists a positive integer $r < \kappa \log^2 m$ for which $(be+r, m) = 1$, by (3), and so

$$(b+rd, m) = (bed+rd, m) = (be+r, m) = 1$$

as $(d,m) = 1$. Furthermore, $(b+rd, g) = (b,g)$ which divides $(b,d) = 1$ (as g divides d), and so $(b+rd, n) = 1$.

Let $u = a+rc$, $v = b+rd$. Then $(v,n) = 1$, $a/b \leqslant u/v \leqslant c/d$, $v \leqslant \kappa(b+d)\log^2 n$, and finally $(u,v) = 1$ as $cv-du = 1$.

REMARK. We examined at most $\kappa \log^2 n$ possible values for r and s when finding $u = cr+as$ and $v = dr+bs$ such that $(u,v) = (v,n) = 1$. If, instead of as above, we let r and s *both* go through the positive integers $\ll \log n$ then perhaps we would find such values for u and v. If so, then $v \ll (b+d)\log n$ and, by taking $r \asymp m/\log n$ in the above, we would get $\# T_m(n) \ll (n \log n/m)^2$ in (9).

Acknowledgement. We should like to thank the referee for a number of helpful observations.

References

1. H. CRAMÉR, 'On the order of magnitude of the difference between consecutive prime numbers', *Acta Arith.* 2 (1936) 396–403.
2. N. G. DE BRUIJN, 'On the number of positive integers $\leqslant x$ and free of prime factors $> y$', *Indag. Math.* 12 (1951) 50–60.
3. G. H. HARDY and E. M. WRIGHT, *An introduction to the theory of numbers*, 5th edition (University Press, Oxford, 1979).
4. H. IWANIEC, 'On the problem of Jacobsthal', *Demonstratio Math.* 11 (1978) 225–231.
5. E. JACOBSTHAL, 'Über Sequenzen ganzer Zahlen, von denen keine zu *n* teilerfremd ist, I–III', *Norske Vid. Selsk. Forh.* (*Trondheim*) 33 (1960) 117–139.
6. U. V. LINNIK, 'On the least prime in an arithmetic progression. II. The Deuring–Heilbronn phenomenon', *Mat. Sb. N.S.* 15 (57) (1944) 347–368.
7. H. MAIER and C. POMERANCE, 'Unusually large gaps between consecutive primes', *Trans. Amer. Math. Soc.* to appear.
8. C. POMERANCE, 'A note on the least prime in an arithmetic progression', *J. Number Theory* 12 (1980) 218–223.
9. K. PRACHAR, 'Über die kleinste Primzahl einer arithmetischen Reihe', *J. Reine Angew. Math.* 206 (1961) 3–4.
10. R. A. RANKIN, 'The difference between consecutive prime numbers V', *Proc. Edinburgh Math. Soc.* 13 (2) (1962/63) 331–332.
11. A. SCHINZEL, 'Remark on the paper of K. Pracher "Über die kleinste Primzahl einer arithmetischen Reihe"', *J. Reine Angew. Math.* 210 (1962) 121–122.

Department of Mathematics
University of Toronto
Toronto
Ontario, M5S 1A1
Canada

Department of Mathematics
University of Georgia
Athens
Georgia 30602
USA