PROCEEDINGS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 104, Number 1, September 1988

GAPS BETWEEN PRIME NUMBERS

ADOLF HILDEBRAND AND HELMUT MAIER

(Communicated by Larry J. Goldstein)

ABSTRACT. Let $d_n = p_{n+1} - p_n$ denote the nth gap in the sequence of primes. We show that for every fixed integer k and sufficiently large T the set of limit points of the sequence $\{(d_n/\log n, \dots, d_{n+k-1}/\log n)\}$ in the cube $[0,T]^k$ has Lebesgue measure $\geq c(k)T^k$, where $c(k)$ is a positive constant depending only on k. This generalizes a result of Ricci and answers a question of Erdős, who had asked to prove that the sequence $\{d_n/\log n\}$ has a finite limit point greater than 1.

1. Introduction. Let p_n denote the nth prime, and let $d_n = p_{n+1} - p_n$ be the nth gap between consecutive primes. The prime number theorem implies that $p_n \sim n \log n$, as $n \to \infty$. Hence the average size of a gap d_n is $\log n$. However, there exist gaps that are much larger than $\log n$. In fact, as was shown by Westzynthius [**9**], we have

$$\limsup_{n\to\infty} \frac{d_n}{\log n} = \infty. \tag{1}$$

Erdős [**2**] proved that there exist infinitely many pairs of consecutive "large" gaps (d_n, d_{n+1}). In [**5**] the second author extended this result to an arbitrary number of consecutive gaps, showing that for any $k \geq 1$

$$\limsup_{n\to\infty} \frac{\min(d_n, \dots, d_{n+k-1})}{\log n} = \infty. \tag{2}$$

Our knowledge about small gaps is less satisfactory. As a counterpart to (1), one might expect that

$$\liminf_{n\to\infty} \frac{d_n}{\log n} = 0. \tag{3}$$

This would follow, if the twin prime conjecture were true. The prime number theorem trivially implies that the "lim inf" in (3) is at most 1. Erdős [**1**] was the first to obtain

$$\liminf_{n\to\infty} \frac{d_n}{\log n} \leq c \tag{4}$$

for some constant c strictly less than 1. A number of authors subsequently reduced the value of c in (4), the current record being $c = 0.248\dots$ [**6**]. A proof of (3), however, seems to be still out of reach.

Let S denote the set of limit points of the sequence $\{d_n/\log n\}$. A natural conjecture is that S consists of all nonnegative real numbers and the point ∞. By

Received by the editors July 23, 1987 and, in revised form, August 26, 1987.
1980 *Mathematics Subject Classification* (1985 *Revision*). Primary 11N05.
Work supported by NSF grants.

(1), ∞ is indeed a limit point of S, and Erdős' result (4) implies that S contains a real number strictly less than 1. Erdős and Ricci [**8**] proved that S has in fact positive Lebesgue measure. Erdős recently asked whether S contains a real number greater than 1. The purpose of this note is to prove the following theorem, which settles Erdős' question in the affirmative. Roughly speaking, the theorem asserts that a positive proportion of all real numbers belong to S, and that an analogous result holds for the limit points (in $\mathbf{R}^k$) of the sequence

$$\left(\frac{d_n}{\log n}, \dots, \frac{d_{n+k-1}}{\log n}\right) \qquad (n = 1, 2, 3, \dots). \tag{5}$$

THEOREM. *Let k be a positive integer, and let $S^{(k)}$ be the set of limit points in $\mathbf{R}^k$ of the sequence* (5). *Then we have, for any sufficiently large number T,*

$$\lambda(S^{(k)} \cap [0,T]^k) \geq c(k)T^k,$$

where $\lambda(\cdots)$ denotes the Lebesgue measure in $\mathbf{R}^k$ and $c(k)$ is a positive constant depending only on k.

The theorem immediately implies that the sequence $\{d_n/\log n\}$ has arbitrarily large finite limit points, thus answering the above-mentioned question of Erdős. In fact, noting that the set of points $(x_1, \dots, x_k)$ in $[0,T]^k$ satisfying $\min_{1\leq i\leq k} x_i < \varepsilon T$ has Lebesgue measure $< k\varepsilon T^k$, we obtain the following

COROLLARY. *Let k be a positive integer and let $\varepsilon = \varepsilon(k) = c(k)/k$, where $c(k)$ is the constant in the theorem. Then, for every sufficiently large T, the sequence* (5) *has a limit point in $[\varepsilon T, T]^k$.*

In particular, we obtain (2) as a consequence of the theorem.

For the proof of the theorem we shall use a method that was introduced by the second author in [**5**] to prove (2). The key idea is to construct a matrix, whose rows are intervals of consecutive integers, and which contains exceptionally few primes. The gaps between consecutive primes in the rows of this matrix are therefore larger than normal. One can in fact prescribe the ratio between the average size of a gap in the matrix and that of a "normal" gap by an appropriate choice of parameters. By letting this ratio be of order T, one obtains a large number of gaps d_n, for which the ratio $d_n/\log n$ is not greater than T, but also not substantially smaller than T. Using a sieve result, one can moreover show that these ratios actually fill out a positive proportion of the interval $[0,T]$. In this way, one obtains the assertion of the theorem for the case $k = 1$. A similar, though technically more complicated, argument yields the general case.

2. Lemmas. The proof of the theorem follows closely the argument of [**5**]. In this section we state four lemmas, all of which have their counterparts in [**5**]. We shall give a detailed proof only for the last lemma; the first three lemmas are obtained by minor modifications of the proofs in [**5**].

Given a constant $C > 0$, we call an integer $q > 1$ a good modulus, if $L(s,\chi) \neq 0$ for all nonprincipal characters $\chi \bmod q$ and all $s = \sigma + it$ satisfying

$$\sigma > 1 - \frac{C}{\log(q(|t|+1))}.$$

The following result can be derived from a large sieve type estimate of Gallagher [**3**] (cf. [**5**, Lemma 2]).

LEMMA 1. *Let q be a good modulus. Then we have, uniformly for $x \geq q^D$ and $(a,q)=1$,*

$$\pi(2x,q,a) - \pi(x,q,a) \gg \frac{x}{\varphi(q)\log x}.$$

Here D is a constant > 1 that depends only on the constant C implicit in the definition of a good modulus.

We shall apply this result with moduli q of the form

$$P(z) = \prod_{p \leq z} p.$$

From Page's theorem on exceptional characters one can deduce (cf. [**5**, Lemma 1]):

LEMMA 2. *If C is a sufficiently small constant, then there exist arbitrarily large values z, for which $q = P(z)$ is a good modulus in terms of C.*

The next lemma gives an upper bound for the number of prime g-tuples in arithmetic progressions. It can be deduced from any sufficiently general sieve upper bound, for example [**4**, Theorem 2.3]. The lemma generalizes Lemma 3 of [**5**], which corresponds to the case $g = 2$.

LEMMA 3. *Let g be a fixed positive integer. Let $z \geq 2$ and $s_1, \ldots, s_g$ be integers satisfying*

$$0 < |s_i - s_j| \leq z^2 \qquad (1 \leq i < j \leq g).$$

Then, for any $R \geq 2$, we have

$$\#\{1 \leq r \leq R\colon rP(z) + s_i \text{ prime for } i = 1, \ldots, g\} \ll_g \frac{R}{(V(z)\log R)^g},$$

where

$$V(z) = \prod_{p \leq z} \left(1 - \frac{1}{p}\right)$$

and the implied constant depends only on g.

Our final lemma is one of the key ingredients in our argument. It guarantees the existence of intervals in which the number of integers n satisfying $(n, P(z)) = 1$ is by a prescribed factor δ smaller than the expected number. A similar result was proved in [**5**, Lemma 6].

LEMMA 4. *Let $K \geq 2$ and $0 < \delta < 1$ be fixed constants. Then, for all sufficiently large numbers z, there exists a number y, $1 \leq y \leq 2P(z)$, such that the estimate*

$$\#\{y_1 < n \leq y_2\colon (n, P(z)) = 1\} \asymp \delta V(z)(y_2 - y_1) \tag{6}$$

holds for all y_1, y_2 satisfying

$$y \leq y_1 < y_2 \leq y + Kz, \quad y_2 - y_1 \geq \frac{z}{\log z}. \tag{7}$$

Here the constants implied in the symbol "$\asymp$" are absolute.

PROOF. Let K, δ and z be given as in the lemma and set $K_1 = 2K^{1/\delta}$. We shall prove the assertion of the lemma with $y = N + Kz$, where N is the least positive

solution to the system of congruences

$$N \equiv -1 \bmod p \qquad (p \le K_1),$$
$$N \equiv 0 \bmod p \qquad (K_1 < p \le z).$$

Since $1 \le N \le P(z)$, we have $1 \le y \le P(z) + Kz \le 2P(z)$, as required, provided z is sufficiently large, as we may assume.

The definition of N implies that for any integer n, $(n, P(z)) = 1$ holds if and only if $m = n - N$ satisfies the conditions

$$(*) \qquad \begin{cases} m \not\equiv 1 \bmod p & (p \le K_1), \\ m \not\equiv 0 \bmod p & (K_1 < p \le z). \end{cases}$$

The left-hand side of (6) is therefore equal to the number of integers m, $y_1 - N < m \le y_2 - N$, that satisfy $(*)$. Hence, putting $x_i = y_i - N$, we see that (6) is equivalent to

$$(6)' \qquad \#\{x_1 < m \le x_2 : (*)\} \asymp \delta V(z)(x_2 - x_1),$$

while the conditions (7) can be rewritten as

$$(7)' \qquad Kz \le x_1 < x_2 \le 2Kz, \qquad x_2 - x_1 \ge \frac{z}{\log z}.$$

To prove $(6)'$, we note that if $z > 2K$, as we may assume, then a positive integer $m \le 2Kz$ satisfying $(*)$ is either composed entirely of prime factors $\le K_1$, or of the form

$$(**) \qquad m = dp, \qquad p > z, \ dp \not\equiv 1 \bmod p' \ (p' \le K_1).$$

The first alternative holds for at most $(\log(Kz))^A$ integers $m \le Kz$, with a suitable $A = A(K_1)$. Hence, if z is sufficiently large, those integers contribute a negligible amount to the left-hand side of $(6)'$. The contribution of the remaining integers m (i.e., those satisfying $(**)$) is equal to

$$\#\{x_1 < m \le x_2 : (**)\} = \sum_{d \le x_2/z} S(d),$$

where

$$S(d) = \#\{\max(x_1/d, z) < p \le x_2/d : dp \not\equiv 1 \mod p' (p' \le K_1)\}.$$

For $d \le x_1/z$ $(\le 2K \le K_1)$, a straightforward application of the Eratosthenes sieve and the prime number theorem for arithmetic progressions shows that

$$S(d) \asymp \prod_{\substack{p' \le K_1 \\ p' \nmid d}} \left(1 - \frac{1}{p'}\right) \frac{x_2 - x_1}{d \log z} = V(K_1) \frac{x_2 - x_1}{\varphi(d) \log z} \asymp \frac{V(z)(x_2 - x_1)}{\varphi(d) \log K_1},$$

provided z is sufficiently large and x_1 and x_2 satisfy $(7)'$. Moreover, the upper bound implicit in this estimate remains valid for $x_1/z < d \le x_2/z$. Since, by $(7)'$, x_1/z and x_2/z are both of order K, we obtain

$$\sum_{d \le x_2/z} S(d) \asymp \left(\sum_{d \le x_1/z} \frac{1}{\varphi(d)} \right) \frac{V(z)(x_2 - x_1)}{\log K_1} \asymp \frac{\log K}{\log K_1} V(z)(x_2 - x_1) \asymp \delta V(z)(x_2 - x_1),$$

using the well-known estimate

$$\sum_{d\le u}\frac{1}{\varphi(d)} \asymp \log u \qquad (u \ge 2)$$

and the definition of K_1. This proves (6)′ and hence the lemma.

3. Proof of the theorem. We fix a positive integer k and a real number T, which we may assume to be sufficiently large. In what follows, the constants implied in the symbol "$\ll$" are allowed to depend on k, but not on T.

For $\varepsilon > 0$, define an ε-neighborhood of $S^{(k)}$ as

$$S_\varepsilon^{(k)} = \left\{(t_1,\dots,t_k)\in \mathbf{R}^k\colon \max_{i=1}^{k}|t_i-\bar t_i|\le\varepsilon \text{ for some } (\bar t_1,\dots,\bar t_k)\in S^{(k)}\right\}.$$

We shall show that for every integer $N \ge 1$ the bound

$$\lambda(S_{1/N}^{(k)}\cap[0,T]^k) \ge c(k)T^k \tag{8}$$

holds with a positive constant $c(k)$ independent of N and T. Since the sets $S_{1/N}^{(k)}$ $(N = 1,2,3,\dots)$ form a decreasing sequence of sets, whose intersection is the closure of $S^{(k)}$ and hence $S^{(k)}$ itself (since $S^{(k)}$, as a set of limit points, is a closed set), we have

$$\lambda(S^{(k)}\cap[0,T]^k) = \lim_{N\to\infty}\lambda(S_{1/N}^{(k)}\cap[0,T]^k).$$

Thus, the asserted bound follows from the bound (8), and it remains to prove the latter one.

We fix an integer $N \ge 1$ and divide the cube $[0,T]^k$ into N^k boxes

$$B(\mathbf{n}) = \left[\frac{n_1-1}{N}T,\frac{n_1}{N}T\right]\times\cdots\times\left[\frac{n_k-1}{N}T,\frac{n_k}{N}T\right] \tag{9}$$
$$(\mathbf{n}=(n_1,\dots,n_k),\ 1\le n_i\le N).$$

We call a box good if it contains a point of $S^{(k)}$. It is clear from the definition of $S_\varepsilon^{(k)}$ that every good box is contained in the set $S_{1/N}^{(k)}\cap[0,T]^k$. Since the boxes $B(\mathbf{n})$ are disjoint apart from a set of volume (i.e., Lebesgue measure in $\mathbf{R}^k$) zero, and each of these boxes has volume $(T/N)^k$, we have

$$\lambda(S_{1/N}^{(k)}\cap[0,T]^k) \ge \#\{\text{good boxes}\}\cdot\left(\frac{T}{N}\right)^k.$$

Thus, to obtain (8), it suffices to show

$$\#\{\text{boxes } B(\mathbf{n}) \text{ containing a point of } S^{(k)}\} \ge c(k)N^k. \tag{10}$$

We now construct the matrix mentioned in the introduction. We fix a number z, for which $q = P(z)$ is a good modulus in the sense of Lemma 2. The lemma guarantees the existence of arbitrarily large numbers z with this property. We then apply Lemma 4 with $K = T$ and $\delta = c/T$, where c is a constant depending on k that will be specified presently. The hypotheses of the lemma are satisfied, provided $T > \max(2,c)$ and z is sufficiently large in terms of T, as we may assume.

We therefore obtain a positive number $y \le 2P(z)$, such that (6) holds, whenever (7) is satisfied. We now define an integer matrix $A = (a_{rs})$ by

$$a_{rs} = rP(z) + s \qquad (R < r \le 2R, y < s \le y + Tz),$$

where

$$R = P(z)^{D-1},$$

D being the constant of Lemma 1, applied with $q = P(z)$.

We first estimate from below the number of primes in A. The columns in the matrix A are the arithmetic progressions

$$\{x < n \le 2x : n \equiv s \bmod q\}, \qquad y < s \le y + Tz,$$

where

$$q = P(z), \qquad x = P(z)^D = RP(z).$$

Only columns with $(s, q) = (s, P(z)) = 1$ can contain primes. By Lemma 4, the number of such columns is

$$\asymp \delta V(z) Tz = cV(z)z.$$

Moreover, by Lemma 1 and our assumption that $q = P(z)$ is a good modulus, the number of primes in each of these columns is

$$\gg \frac{x}{\varphi(q)\log x} = \frac{x}{P(z)V(z)\log P(z)^D} \asymp \frac{R}{V(z)z},$$

since

$$\log P(z)^D = D\sum_{p \le z} \log p \asymp z.$$

Hence the entire matrix A contains

$$\gg cV(z)z \cdot \frac{R}{V(z)z} = cR$$

primes, where the implied constant is absolute. If we now choose the constant c sufficiently large, then we have

$$\#\{\text{primes in } A\} \ge 3kR. \tag{11}$$

By (11), a row in A contains on average at least $3k$ primes. Call a row "good" if it contains at least $2k$ primes, and "bad" otherwise. Since the matrix A has $[2R] - [R] \le R + 1$ rows, at most $(R+1)(2k-1)$ of the primes in A can be located in bad rows. In view of (11), we therefore have

$$\#\{\text{primes in good rows of } A\} \ge 3kR - (2k-1)(R+1) \ge kR, \tag{12}$$

provided $R \ge 2k$, as we may assume.

Next, we estimate the number of $(k+1)$-tuples

$$(a_{rs_1}, \dots, a_{rs_{k+1}}) = (p_n, \dots, p_{n+k}) \tag{13}$$

of *consecutive* primes in the rows of our matrix. A row with $m > k$ primes contains exactly $m - k$ such $(k+1)$-tuples. If the row is good, i.e., if $m \ge 2k$, then $m - k \ge m/2$, so that the number of $(k+1)$-tuples is at least half the number of primes in the row. Thus, using (12), we see that

$$\#\{\text{tuples (13) in } A\} \ge \tfrac{1}{2}\#\{\text{primes in good rows of } A\} \ge \tfrac{1}{2}kR. \tag{14}$$

With each of the $(k+1)$-tuples (13) we can associate a k-tuple of differences between consecutive primes

$$(a_{rs_2} - a_{rs_1}, \dots, a_{rs_{k+1}} - a_{rs_k}) = (s_2 - s_1, \dots, s_{k+1} - s_k)$$
$$= (p_{n+1} - p_n, \dots, p_{n+k} - p_{n+k-1}) = (d_n, \dots, d_{n+k-1})$$

as well as a "normalized" k-tuple of differences

$$\left(\frac{s_2 - s_1}{\log x}, \dots, \frac{s_{k+1} - s_k}{\log x}\right) = \left(\frac{d_n}{\log x}, \dots, \frac{d_{n+k-1}}{\log x}\right). \tag{15}$$

Since for every tuple (13)

$$y < s_1 < \cdots < s_{k+1} \leq y + Tz \tag{16}$$

and

$$\log x = D \log P(z) \geq z,$$

if, as we may assume, z is sufficiently large, each of the tuples (15) is contained in the cube $[0,T]^k$, and hence in one of the N^k boxes $B(\mathbf{n})$ defined in (9). We shall show

$$\#\{\text{boxes } B(\mathbf{n}) \text{ containing a } k\text{-tuple (15)}\} \gg N^k, \tag{17}$$

where the implied constant depends only on k.

Having proved (17), the proof of (10), and hence that of the theorem, can be easily completed. To this end, we repeat the above construction with a sequence of values z tending to infinity. By choosing a suitable subsequence and using (17), we obtain a *fixed* collection of $\gg N^k$ boxes $B(\mathbf{n})$, each of which contains a tuple (15) for all values z in this subsequence. Hence, each of those boxes contains a limit point of the tuples (15). Since the elements a_{rs} of our matrix A have order of magnitude x, we have, for any k-tuple (15) associated with a $(k+1)$-tuple (13),

$$\log x \sim \log a_{rs_1} = \log p_n \sim \log n.$$

Thus, every limit point of the tuples (15) is also a limit point of the tuples (5), hence belongs to $S^{(k)}$, and (10) follows.

To prove (17), we estimate from above the number of $(k+1)$-tuples (13), for which the associated tuple (15) falls into a *fixed* box $B(\mathbf{n})$, i.e., which satisfies

$$\frac{n_i - 1}{N} T \log x \leq s_{i+1} - s_i \leq \frac{n_i}{N} T \log x \qquad (i = 1, \dots, k). \tag{18}$$

We shall show that, for each of the boxes $B(\mathbf{n})$,

$$\#\{\text{tuples (13) in } A \text{ satisfying (18)}\} \ll RN^{-k}, \tag{19}$$

with the implied constant depending only on k. Since, by (14), the matrix A contains $\gg R$ $(k+1)$-tuples (13), we see that (19) implies (17).

To obtain (19), we note that the number of $(k+1)$-tuples to be estimated is at most equal to the number of tuples $(a_{rs_1}, \dots, a_{rs_{k+1}})$ in our matrix, that consist entirely of primes (though not necessarily consecutive primes), and where $s_1, \dots, s_{k+1}$ are subject to (16) and (18). Such tuples of primes can only exist, if

$$(s_i, P(z)) = 1 \qquad (i = 1, \dots, k+1). \tag{20}$$

For *fixed* $s_1, \dots, s_{k+1}$ satisfying (16), (18) and (20), the number of such tuples is by Lemma 3

$$\#\{R < r \le 2R\colon a_{rs_i} = rP(z) + s_i \text{ prime for } i = 1, \dots, k+1\} \ll \frac{R}{(V(z)\log R)^{k+1}} \asymp \frac{R}{(V(z)z)^{k+1}}.$$

Moreover, the number of tuples $(s_1, \dots, s_{k+1})$ that satisfy (16), (18) and (20) can be estimated by Lemma 4. To this end, we note that the conditions (16) and (18) restrict s_1 to the interval $(y, y+Tz]$, and each of the numbers s_i, $2 \le i \le k+1$, to a subinterval of $(y, y+Tz]$ of length $(T\log x)/N \asymp Tz/N$ (which is $\ge z/\log z$ for sufficiently large z). Lemma 4 therefore gives

$$\#\{(s_1, \dots, s_{k+1})\colon (16), (18), (20)\} \ll (\delta TzV(z))\left(\delta\frac{T}{N}zV(z)\right)^k = \frac{(czV(z))^{k+1}}{N^k}.$$

Altogether, the number of $(k+1)$-tuples to be estimated in (19) is bounded by

$$\ll \frac{R}{(V(z)z)^{k+1}} \cdot \frac{(czV(z))^{k+1}}{N^k} \asymp \frac{R}{N^k},$$

as required.

The proof of the theorem is now complete.

4. Concluding remarks. Rankin [7] proved a stronger form of (1), namely

$$\limsup_{n\to\infty} \frac{d_n}{L_0(n)} > 0$$

with

$$L_0(n) = \frac{\log n \log_2 n \log_4 n}{(\log_3 n)^2},$$

where $\log_k n$ denotes the k times iterated logarithm. An analogous strengthening of (2) was proved in [5]. Thus, (1) and (2) remain valid, when the function $\log n$ is replaced by any function $L(n)$ satisfying

$$\log n \le L(n) = o(L_0(n)) \qquad (n \to \infty). \tag{21}$$

One might therefore ask if one can similarly replace $\log n$ by $L(n)$ in the definition of the set $S^{(k)}$ in the theorem. By modifying slightly the present proof and using some additional arguments from [5] and [7], one can indeed show that the result remains valid with any *slowly oscillating* function $L(n)$ satisfying (21) in place of $\log n$.

It is an open problem to find a *specific* real number that is a limit point of the sequence $\{d_n/\log n\}$. Our method is, like earlier methods, nonconstructive and yields only the *existence* of (sufficiently many) limit points. To show that a given real number is a limit point of $\{d_n/\log n\}$ would probably require completely new ideas.

REFERENCES

1. P. Erdős, *On the difference between consecutive primes*, Quart. J. Oxford **6** (1935), 124–128.
2. ——, *Problems and results on the difference of consecutive primes*, Publ. Math. Debrecen **1** (1949–1950), 33–37.

3. P. X. Gallagher, *A large sieve density estimate near* $\sigma = 1$, Invent. Math. **11** (1970), 329–339.
4. H. Halberstam and H.-E. Richert, *Sieve methods*, Academic Press, New York, 1974.
5. H. Maier, *Chains of large gaps between consecutive primes*, Adv. in Math. **39** (1981), 257–269.
6. ——, *Small differences between prime numbers*, Preprint.
7. R. A. Rankin, *The difference between consecutive prime numbers*, J. London Math. Soc. **13** (1938), 242–247.
8. G. Ricci, *Recherches sur l'allure de la suite* $\{(p_{n+1} - p_n)/\log p_n\}$, Colloque sur la Théorie des Nombres, Bruxelles, 1955, pp. 93–106.
9. E. Westzynthius, *Über die Verteilung der Zahlen, die zu den n ersten Primzahlen teilerfremd sind*, Comm. Phys. Math. Helsingfors **25** (1931), 1–37.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF ILLINOIS, URBANA, ILLINOIS 61801

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF GEORGIA, ATHENS, GEORGIA 30602