Metrical results on sums of squares and prime values of the integer parts of sequences

by

PAUL ROWE (Surrey)

1. Introduction. The notation $[x]$ will be used in what follows to represent the integer part of a number $x \in \mathbb{R}$. We reserve the letter p to denote a prime, and the variables r, s will always be numbers representable as the sum of two squares.

Harman has proved in [2] and [4, Theorem 3] respectively that:

1. *For almost all* $\alpha > 0$, *the number of integers of the form* $[10^n \alpha]$, $1 \leq n \leq N$, *which are prime numbers is*

$$\frac{\log N}{\log 10} + O((\log N)^{1/2}(\log \log N)^{(3/2)+\varepsilon})$$

for any $\varepsilon > 0$.

2. *For almost all* $(\alpha, \beta) \in \mathbb{R}^2$ *there are only finitely many* n *such that* $[10^n \alpha]$ *and* $[10^n \beta]$ *are simultaneously prime.*

These results naturally lead one to ask if one could prove a two-dimensional analogue of the first result for a denser set than the primes, say the set of numbers representable as the sum of two squares. Also, one is curious to know the structure of the set of measure zero for the second result. One would like to know its Hausdorff dimension and whether or not the set is dense. It clearly contains almost all points of the line $\beta = \alpha$ by the first result, but does it contain almost all points of any other lines? Here we shall give definitive answers to these questions as follows.

THEOREM 1. *For almost all* $(\alpha, \beta) \in \mathbb{R}^2$, *the number of solutions for* $n \leq N$ *to* $[10^n \alpha]$ *and* $[10^n \beta]$ *simultaneously equal to a sum of two squares is*

$$\frac{K^2}{\log 10} \log N + O((\log N)^{1/2}(\log(\log N + 2))^{(3/2)+\varepsilon})$$

for any $\varepsilon > 0$ as $N \to \infty$, where K is the constant

$$2^{-1/2} \prod_{p \equiv 3 \bmod 4} \left(1 - \frac{1}{p^2}\right)^{-1/2}.$$

REMARKS. The constant K enters naturally here via the well-known formula for the number of sums of two squares up to N:

$$(1) \qquad \sum_{\substack{r \leq N \\ r \text{ a sum of squares}}} 1 = \sum_{2 \leq n \leq N} \frac{K}{(\log n)^{1/2}} + O\left(\frac{N}{\log N}\right).$$

As in Harman's one-dimensional theorem for primes [2] we obtain an asymptotic formula with the expected main term.

THEOREM 2. *The set S of $(\alpha, \beta) \in [0,1)^2$ such that there are infinitely many n with $[10^n \alpha]$ and $[10^n \beta]$ simultaneously prime has Hausdorff dimension 2, is dense in $[0,1)^2$ and intersects any line in $\mathbb{R}^2$ except $\{(x,y) : x = y\}$ in a set of linear measure zero.*

The proofs of both theorems rely on our knowledge of the distribution of sums of two squares or primes in short intervals: we obtain an asymptotic formula if the intervals are not too short, and a correct order upper bound otherwise. A result related to Theorem 1 for square-free numbers is proved in the author's PhD thesis [7]. This gives, for every $k \geq 1$ and for almost all α, an asymptotic formula for the number of n with $[10^n \alpha^{a_1}], [10^n \alpha^{a_2}], \ldots,$ $[10^n \alpha^{a_k}]$ simultaneously square-free for $n \leq N$, where each $a_i \in \mathbb{N}$. The details are quite different for this result: one part is easier because the set of square-free numbers is better understood than the set of sums of two squares; one part is harder because we make use of the fact that the probability that a number n is a sum of two squares tends to zero as $n \to \infty$.

The first part of Theorem 2 can be generalised to higher dimensions and the last two parts could clearly be proved for all $(\alpha, \beta) \in \mathbb{R}^2$, but here $[0,1)^2$ is considered for simplicity. We note that another problem, that of finding the Hausdorff dimension of the set of α such that $[10^n \alpha]$ and $[10^n \alpha^2]$ are simultaneously prime infinitely often, seems to be much more difficult.

The reader should note that Harman proved his results for any sequence $[b^n \alpha]$ (indeed for even more general sequences) and there is no problem in replacing 10 in the above theorems by any real $b > 1$. We have chosen to state our results in the above form simply because they lend themselves to a natural interpretation in terms of the decimal expansions of pairs of reals chosen at random.

2. Theorem 1: Method of proof and preliminary lemmas. Without loss of generality we need only consider $(\alpha, \beta) \in [1,2)^2$. The method will

be to write

$$\mathcal{I}_n = \left(\bigcup_{10^n \leq r < 2 \cdot 10^n} \left[\frac{r}{10^n}, \frac{r+1}{10^n} \right) \right) \times \left(\bigcup_{10^n \leq s < 2 \cdot 10^n} \left[\frac{s}{10^n}, \frac{s+1}{10^n} \right) \right).$$

Then $(\alpha, \beta) \in \mathcal{I}_n$ precisely when $[10^n \alpha]$ and $[10^n \beta]$ are sums of two squares for $1 < \alpha, \beta < 2$ and therefore the number of n such that $[10^n \alpha]$ and $[10^n \beta]$ are simultaneously sums of two squares, for $n \leq N$, when $1 < \alpha < 2$, is $\sum_{n=1}^{N} \chi_n(\alpha, \beta)$ where we define

$$\chi_n(\alpha, \beta) = \chi_{\mathcal{I}_n}(\alpha, \beta) = \begin{cases} 1 & \text{if } (\alpha, \beta) \in \mathcal{I}_n, \\ 0 & \text{otherwise.} \end{cases}$$

We will consider, for any $U, V \leq N$, the expression

$$(2) \qquad \int_1^2 \int_1^2 \left(\sum_{n=U}^{V} \chi_n(\alpha, \beta) - \frac{K^2}{\log 10} \sum_{n=U}^{V} \frac{1}{n} \right)^2 d\alpha \, d\beta.$$

We expand out the integrand to give

$$= \int_1^2 \int_1^2 \left(\sum_{n=U}^{V} \chi_n(\alpha, \beta) \sum_{m=U}^{V} \chi_m(\alpha, \beta) \right.$$

$$\left. + \left(\frac{K^2}{\log 10} \sum_{m=U}^{V} \frac{1}{m} \right)^2 - 2 \frac{K^2}{\log 10} \sum_{m=U}^{V} \frac{1}{m} \sum_{n=U}^{V} \chi_n(\alpha, \beta) \right) d\alpha \, d\beta,$$

which leads to three double integrals to estimate: $I_1 + I_2 - 2I_3$, say. Since the expression in (2) is positive, it suffices to give either asymptotic formulae or upper bounds for each of the terms I_j. We use these bounds with the following lemma (Lemma 1.5 in [3]) to obtain the asymptotic formula required for Theorem 1.

LEMMA 1. *Let Y be a measure space with measure μ such that $0 < \mu(Y) < \infty$. Let $F(n) : Y \to \mathbb{R}$ $(n = 1, 2, \ldots)$ be a sequence of non-negative μ-measurable functions, and let x_n, ϕ_n be sequences of real numbers such that*

$$0 \leq x_n \leq \phi_n \quad (n = 1, 2, \ldots).$$

Write

$$\Phi(N) = \sum_{n=1}^{N} \phi_n,$$

and suppose that $\Phi(N) \to \infty$ as $N \to \infty$. Suppose that for arbitrary integers U, V $(1 \leq U < V)$ we have

$$\int_Y \left(\sum_{U \leq n < V} (F(n) - x_n) \right)^2 \leq C \sum_{U \leq n < V} \phi_n$$

for an absolute constant C. Then for any $\varepsilon > 0$, and for almost all points in Y, we have as $N \to \infty$,

$$\sum_{n=1}^{N} F(n) = \sum_{n=1}^{N} x_n + O(\Phi^{1/2}(N)(\log(\Phi(N)+2))^{(3/2)+\varepsilon} + \max_{1 \le n \le N} x_n).$$

Now I_2 is simply

$$\left(\frac{K^2}{\log 10} \sum_{m=U}^{V} \frac{1}{m} \right)^2,$$

while

$$I_3 = \left(\frac{K^2}{\log 10} \sum_{m=U}^{V} \frac{1}{m} \right) \left(\sum_{n=U}^{V} \lambda(\mathcal{I}_n) \right).$$

Here λ denotes planar Lebesgue measure, and

$$\sum_{n=U}^{V} \lambda(\mathcal{I}_n) = \sum_{n=U}^{V} \left(\sum_{10^n \le r < 2 \cdot 10^n} \frac{1}{10^n} \right) \left(\sum_{10^n \le s < 2 \cdot 10^n} \frac{1}{10^n} \right)$$

$$= \frac{1}{\log 10} \sum_{n=U}^{V} \left(\frac{K^2}{n} + O(n^{-3/2}) \right),$$

from (1). It thus remains to consider I_1 which is where the heart of the problem lies (the so-called "overlap estimates" studied for a range of problems in [3]). In the double sum over n and m, when $n < m/10$ (we could in fact replace $1/10$ with a similar fraction), we can use an analogue of Huxley's theorem for primes in short intervals and this will yield the main term. For the part of the proof which gives us a bound when $m/10 < n < m$ we rely on the following lemma:

LEMMA 2. *The number of solutions in r_1 and r_2 of $r_2 = 10^{m-n} r_1 + b$ where $r_2 \in [10^m, 2 \cdot 10^m]$ for $1 \le b < 10^{m-n}$ is*

$$\ll \frac{10^m}{n}.$$

Proof. Every sum of two squares can be written in the form $2^k v^2 w$, where $p \mid v \Rightarrow p \equiv 3 \bmod 4$, and $p \mid w \Rightarrow p \equiv 1 \bmod 4$. Write $r_j = 2^{k_j} v_j^2 w_j$. Clearly the contribution from those r_j with $2^{k_j} v_j > n$ can be neglected. It thus suffices to consider

$$2^{k_2} v_2^2 w_2 = 10^{m-n} 2^{k_1} v_1^2 w_1 + b$$

with fixed b, k_j, v_j, w_j where $2^{k_j} v_j \le n$. Let $g = (10^{m-n} 2^{k_1} v_1^2, 2^{k_2} v_2^2)$ (of course 10^{m-n} could be replaced with 2^{m-n} in this expression since $(5, v_2) = 1$). Write $A = 2^{k_2} v_2^2 / g$, $B = 10^{m-n} 2^{k_1} v_1^2 / g$, $C = b/g$. Then we

must bound the number of solutions (w_1, w_2) to

$$(3) \qquad\qquad Aw_2 = Bw_1 + C.$$

Now $(A, B) = 1$, and, for any fixed g, C only takes on $\ll 10^{m-n}/g$ values. The solutions to (3) have the form

$$w_1 = A\ell + J, \qquad w_2 = B\ell + K, \qquad \text{with } AK - BJ = C.$$

If $p \equiv 3 \bmod 4$ with $p \mid C$ and $p \mid AB$, then $p \mid w_1 w_2$, so there are no solutions to (3). Otherwise we can apply a standard sieve upper bound (Theorem 2.3 in [1] with $g = 2$, $\delta = 1/2$) to show that for each fixed triple (A, B, C) the number of solutions in w_1, w_2 is

$$\ll \frac{ABC}{\phi(ABC)} \frac{10^n g}{2^{k_1 + k_2} v_1^2 v_2^2 n}.$$

Here $\phi(x)$ is Euler's function, and we have used $2^{k_1 + k_2} v_1^2 v_2^2 \le n^4$ to note that $\log(10^n / 2^{k_1 + k_2} v_1^2 v_2^2) \gg n$.

Now $\phi(xy) \ge \phi(x)\phi(y)$ for any x, y. Also, for any X,

$$\sum_{C \le X} \frac{C}{\phi(C)} \ll X.$$

Using $\phi(2) = 1$ with the above we may sum over k_j, v_j and b to obtain an upper bound for the number of solutions of the form

$$\ll \sum_{k_1, k_2} \sum_{v_1, v_2} \frac{v_1 v_2}{\phi(v_1)\phi(v_2)} \frac{10^n g^2}{n 2^{k_1 + k_2} v_1^2 v_2^2} \sum_{C \le 10^{m-n}/g} \frac{C}{\phi(C)}$$

$$\ll \frac{10^m}{n} \sum_{k_1, k_2} 2^{-\max(k_1, k_2)} \sum_{v_1, v_2} \frac{(v_1, v_2)^2}{v_1 v_2 \phi(v_1)\phi(v_2)} \ll \frac{10^m}{n}$$

as required. Here we have noted that

$$\sum_{v_1, v_2} \frac{(v_1, v_2)^2}{v_1 v_2 \phi(v_1)\phi(v_2)} \le \sum_d d^2 \sum_{v_j} \frac{1}{d^2 v_1 v_2 \phi(dv_1)\phi(dv_2)}$$

$$\ll \sum_d d^2 \sum_{v_1, v_2} \frac{1}{(d^2 v_1 v_2)^{7/4}} \ll 1,$$

where we have used the simple bound $\phi(x) \gg x^{3/4}$.

3. Proof of Theorem 1. We are left to estimate

$$I_1 = \int\limits_1^2 \int\limits_1^2 \sum_{n=U}^{V} \sum_{m=U}^{V} \chi_n(\alpha, \beta) \chi_m(\alpha, \beta) \, d\alpha \, d\beta,$$

which is equal to

$$\sum_{n=U}^{V} \sum_{m=U}^{V} \lambda(\mathcal{I}_n \cap \mathcal{I}_m)$$

where $\mathcal{I}_n \cap \mathcal{I}_m$ is made up of a union of sets of the form

$$\left(\left[\frac{r_1}{10^n}, \frac{r_1+1}{10^n}\right) \cap \left[\frac{r_2}{10^m}, \frac{r_2+1}{10^m}\right)\right) \times \left(\left[\frac{s_1}{10^n}, \frac{s_1+1}{10^n}\right) \cap \left[\frac{s_2}{10^m}, \frac{s_2+1}{10^m}\right)\right).$$

We thus need to consider intersections of intervals of the form

$$\left[\frac{r_1}{10^n}, \frac{r_1+1}{10^n}\right) \cap \left[\frac{r_2}{10^m}, \frac{r_2+1}{10^m}\right)$$

and

$$\left[\frac{s_1}{10^n}, \frac{s_1+1}{10^n}\right) \cap \left[\frac{s_2}{10^m}, \frac{s_2+1}{10^m}\right).$$

We assume to start with that $n < m$. Since we only need an upper bound, it is sufficient to assume that the length of the intersection of the two intervals,

$$\left[\frac{r_1}{10^n}, \frac{r_1+1}{10^n}\right) \quad \text{and} \quad \left[\frac{r_2}{10^m}, \frac{r_2+1}{10^m}\right),$$

is the maximum possible whenever they intersect, i.e. $1/10^m$. This may appear rather wasteful, especially when n is close to m, but these terms only give a small contribution to our sums. The assumption we make is quite accurate when n is much less than m, which is the crucial case. Thus we obtain the conditions

$$\frac{r_1}{10^n} < \frac{r_2+1}{10^m} \quad \text{and} \quad \frac{r_2}{10^m} < \frac{r_1+1}{10^n}.$$

These imply that

(4) $$r_1 \cdot 10^{m-n} - 1 \le r_2 < r_1 \cdot 10^{m-n} + 10^{m-n}.$$

Likewise we do this for s replacing r. We therefore must bound

(5) $$\frac{1}{10^{2m}} \sum_{*} 1$$

where the sum $(*)$ is over r_1, r_2, s_1, s_2 satisfying (4) and its analogue for s_j.

We first consider the case when $m/10 \le n < m$. Now we have $r_2 = r_1 \cdot 10^{m-n} + b$ where $-1 \le b < 10^{m-n}$. Trivially the number of solutions with $b = -1$ or 0 is $\ll 10^n n^{-1/2}$ (the number of choices for r_1). For the remaining values of b we apply Lemma 2 to obtain a bound $\ll 10^m/n$. Since

we obtain the same bound for the equation in s_1, s_2 we have

$$\sum_{m=U}^{V} \sum_{m/10 \le n < m} \frac{1}{10^{2m}} \sum_{*} 1 \ll \sum_{m=U}^{V} \sum_{m/10 \le n < m} \frac{1}{10^{2m}} \left(\frac{10^m}{n} + \frac{10^n}{n^{1/2}} \right)^2$$

$$\ll \sum_{m=U}^{V} \frac{1}{m}$$

after an elementary calculation.

We now consider the case when $n < m/10$. We use the analogue for sums of two squares of Huxley's theorem for primes in short intervals [5]. This states that if $x^{(7/12)+\varepsilon} \le y < x$ for some $\varepsilon > 0$, then the number of sums of two squares in the interval $(x, x + y)$ is

$$\frac{Ky}{(\log x)^{1/2}} \left(1 + O\left(\frac{1}{(\log x)^{1/2}} \right) \right).$$

We take $x = 10^m$ and $y = 10^{m-n}$ so the condition $n < m/10$ implies that $x^{9/10} \le y < x$. Using this for both the inequalities in r_j and s_j we thus obtain an asymptotic formula for the expression (5):

$$\frac{1}{10^{2m}} \left(\left(\frac{K^2}{\log 10} \right)^2 \frac{10^{2m}}{mn} + O\left(\frac{10^{2m}}{mn^{3/2}} \right) \right).$$

Since $\sum n^{-3/2}$ converges, combining our two cases we obtain

$$I_1 = \sum_{m=U}^{V} \sum_{n=U}^{V} \frac{1}{mn} + O\left(\sum_{m=U}^{V} \frac{1}{m} \right).$$

We thus obtain

$$I_1 + I_2 - 2I_3 < C \sum_{m=U}^{V} \frac{1}{m}$$

for some constant C. Theorem 1 then follows from Lemma 1.

4. Theorem 2: Preliminaries. First recall that the diameter $d(A)$ of a set A in $\mathbb{R}^2$ is defined as

$$d(A) = \max_{x,y \in A} |x - y|.$$

We shall use the following definition of Hausdorff dimension.

DEFINITION. The set $S \subseteq \mathbb{R}^2$ has *Hausdorff dimension* $\delta > 0$ if both the following conditions hold:

(A) For any $\varepsilon > 0$, there exists a covering (ξ_i) of the set S with $d(\xi_i) < \varepsilon$ such that

$$\sum_{i=1}^{\infty} d(\xi_i)^{\gamma} < 1 \quad \text{for all } \gamma > \delta.$$

(B) There exists $\varepsilon > 0$ such that for all coverings (ξ_i) of S such that $d(\xi_i) < \varepsilon$ we have

$$\sum_{i=1}^{\infty} d(\xi_i)^{\gamma} \geq 1 \quad \text{for all } \gamma < \delta.$$

5. Proof of Theorem 2. For the first part of the theorem, to show that the set S has Hausdorff dimension 2 it clearly suffices to verify (B) above with $\delta = 2$.

We will show that there exists $\varepsilon > 0$ such that for any $\delta > 0$ and for any collection of sets $(\xi_i) \subseteq [0,1)^2$ satisfying both

$$d(\xi_i) < \varepsilon \quad \text{for all } i, \quad \text{and} \quad \sum_{i=1}^{\infty} d(\xi_i)^{2-\delta} < 1,$$

the collection (ξ_i) does not cover the set S. We will do this by constructing a sequence of nested compact sets $\mathcal{I}_I \supseteq \mathcal{I}_{I+1} \supseteq \mathcal{I}_{I+2} \supseteq \cdots$ such that each $\mathcal{I}_i$ will consist of M_i boxes of the form

$$(6) \qquad \left[\frac{p_1}{10^{3^i}}, \frac{p_1 + \frac{3}{4}}{10^{3^i}}\right] \times \left[\frac{p_2}{10^{3^i}}, \frac{p_2 + \frac{3}{4}}{10^{3^i}}\right]$$

and so

$$\bigcap_{i=I}^{\infty} \mathcal{I}_i \subset S.$$

We need to show at each stage of our inductive construction that there are a large number of boxes, say $M_i > K_i$, which do not intersect any of the ξ_j of a certain size, say those with $\varepsilon_i \leq d(\xi_j) < \varepsilon_{i-1}$. The condition

$$\sum_{i=1}^{\infty} d(\xi_i)^{2-\delta} < 1$$

is vital for controlling the number of boxes from our construction which do not intersect any of the ξ_j. The reader will recognise this is a well-known technique used to establish the Hausdorff dimension of certain sets in metric number theory (see pages 271–275 of [3] for example).

We define K_i, ε_i for $i = 1, 2, \ldots$ now (the reader will see the reason for the choice as the proof proceeds):

$$K_i = \frac{10^{2 \cdot 3^i}}{4^{i^2+1}(\log 10)^{2i}} \quad \text{and} \quad \varepsilon_i = 10^{-i^2/\delta}$$

and choose I large enough so that both

$$(7) \qquad K_{i+1} < \frac{10^{2\cdot 3^{i+1}}}{2\cdot 4^{i^2+1}3^{2i+2}(\log 10)^{2i+2}} - \frac{4\cdot 10^{2\cdot 3^{i+1}}}{10^{i^2}3^{2i}(\log 10)^2} - 10^{2(i+1)^2/\delta}$$

and

$$(8) \qquad \left(\frac{3}{4}\cdot\frac{10^{2\cdot 3^i}}{3^{i+1}\log 10}(1+o(1))\right)^2 \geq \frac{1}{2}\cdot\frac{10^{4\cdot 3^i}}{3^{2i+2}(\log 10)^2}$$

hold for all $i \geq I$. Here the $o(1)$ in (8) means the $o(1)$ terms which arise below in the proof.

For the case $i = I$ we let $\mathcal{I}_I$ be the union of all boxes of the form (6) for $i = I$ and $10^{3^i} < p_1, p_2 < 2\cdot 10^{3^I}$. By the prime number theorem [6, p. 226], we have

$$M_I = (\pi(2\cdot 10^{3^I}) - \pi(10^{3^I}))^2 \sim \left(\frac{2\cdot 10^{3^I}}{\log 2 + 3^I \log 10} - \frac{10^{3^I}}{3^I \log 10}\right)^2 > K_I$$

where $\pi(x)$ represents the number of primes not greater than x. Also, if we let $\varepsilon = \varepsilon_I$ then no box in $\mathcal{I}_I$ intersects any ξ_j with $\varepsilon_I \leq d(\xi_j) < \varepsilon_{I-1}$ since $d(\xi_j) < \varepsilon_I$ for all j. This establishes the case $i = I$. For the inductive step we suppose we have $\mathcal{I}_i$. Then by Huxley's theorem [5] the number of primes in the interval

$$\left[10^{3^{i+1}}, 10^{3^{i+1}} + \frac{3}{4}\cdot 10^{3^{i+1}2/3}\right]$$

is

$$\frac{\frac{3}{4}\cdot 10^{2\cdot 3^i}}{\log 10^{3^{i+1}}}(1+o(1))$$

and therefore we can find

$$\geq \frac{1}{2}\cdot\frac{10^{4\cdot 3^i}}{(\log 10)^2 3^{2i+2}}$$

boxes of the form

$$\left[\frac{p_1}{10^{3^{i+1}}}, \frac{p_1 + \frac{3}{4}}{10^{3^{i+1}}}\right] \times \left[\frac{p_2}{10^{3^{i+1}}}, \frac{p_2 + \frac{3}{4}}{10^{3^{i+1}}}\right]$$

contained entirely within each box of $\mathcal{I}_i$ by (8). Thus the number of boxes in total is

$$> \frac{10^{2\cdot 3^i}}{4^{i^2+1}(\log 10)^{2i}}\cdot\frac{1}{2}\cdot\frac{10^{4\cdot 3^i}}{(\log 10)^2 3^{2i+2}} = \frac{10^{2\cdot 3^{i+1}}}{2\cdot 4^{i^2+1}3^{2i+2}(\log 10)^{2i+2}}$$

by the inductive hypothesis. The number of these boxes which intersect a set ξ_j of diameter d is

$$< \left(\frac{d\cdot 10^{3^{i+1}}}{3^i \log 10} + 1\right)^2 < \left(\frac{2d\cdot 10^{3^{i+1}}}{3^i \log 10}\right)^2 + 1$$

since the number of primes in an interval of length $10^{3^{i+1}}d$ is

$$< \frac{3 \cdot 10^{3^{i+1}}d}{\log(10^{3^{i+1}}d)}$$

by the Brun–Titchmarsh inequality [1, Theorem 3.7], and we may need to count an extra prime outside the interval which nevertheless gives boxes that intersect the set ξ_j. Since we are assuming that $\sum_{i=1}^{\infty} d(\xi_i)^{2-\delta} < 1$, we have

$$\sum_{\substack{\xi_j \\ \varepsilon_{i+1} \le d < \varepsilon_i}} d^2 < \sum_{d < \varepsilon_i} d^2 < \varepsilon_i^{\delta} \sum_{d < \varepsilon_i} d^{2-\delta} < \varepsilon_i^{\delta} = 10^{-i^2},$$

and also

$$\sum_{\substack{\xi_j \\ d > \varepsilon_{i+1}}} 1 < \varepsilon_{i+1}^{\delta-2} \sum_{j=1}^{\infty} d(\xi_j)^{2-\delta} < \varepsilon_{i+1}^{\delta-2} < 10^{2(i+1)^2/\delta}.$$

Hence the number of boxes which do not intersect a set ξ_j with $\varepsilon_{i+1} \le d(\xi_j) < \varepsilon_i$ is

$$> \frac{10^{2 \cdot 3^{i+1}}}{2 \cdot 4^{i^2+1}3^{2i+2}(\log 10)^{2i+2}} - \frac{4 \cdot 10^{2 \cdot 3^{i+1}}}{3^{2i}10^{i^2}(\log 10)^2} - 10^{2(i+1)^2/\delta} > K_{i+1}$$

by (7). This proves the first part of the theorem.

For the second part, to show that S is dense, we need to show that any open set, B, of $[0,1)^2$ contains a point of S.

Let $(a, b) \times (c, d) \subseteq B$ where (a, b) and (c, d) are open intervals. We choose N_1 and N_2 such that the first N_1 digits of a and b correspond but the (N_1+1)st digits differ. Similarly N_2 is chosen for (c, d). We will construct two real numbers α and β such that $[10^n\alpha]$ and $[10^n\beta]$ are simultaneously prime infinitely often and $(\alpha, \beta) \in (a, b) \times (c, d)$.

We define the first N_1 digits of α to be the same as those of a and the first N_2 digits of β the same as those of c. The $(N_1 + 1)$st digit of α and the $(N_2 + 1)$st digit of β can then be chosen so that $(\alpha, \beta) \in (a, b) \times (c, d)$. Now we let $N > \max(N_1 + 1, N_2 + 1)$ and define the rest of the first N digits of α and β arbitrarily. We can define the next $n - N$ digits of α in such a way that $[10^n\alpha]$ is prime for this n as long as n is sufficiently large. This is because we need to find a prime in the interval $[x10^n, x10^n + 10^{n-N}]$, where x is a real number between 0 and 1, and by Huxley's theorem [5] there are

$$O\left(\frac{10^{n-N}}{\log 10^n}\right)$$

primes in this interval if

$$10^{n-N} \ge x^{(7/12)+\varepsilon}10^{(7n/12)+\varepsilon}$$

for any $\varepsilon > 0$. This condition holds if we choose $n > 12N/5$. We can define the corresponding $n - N$ digits of β similarly so that $[10^n \beta]$ is prime. Let this particular value of n be n_1. Now we repeat this process, defining a further $n_2 - n_1$ digits of α and of β so that $[10^n \alpha]$ and $[10^n \beta]$ are both prime for $n = n_2$. We can do this if $n_2 > 12n_1/5$. This process can be repeated indefinitely and so this proves the existence of α and β such that $[10^n \alpha]$ and $[10^n \beta]$ are simultaneously prime infinitely often with $(\alpha, \beta) \in B$. This proves the second part of the theorem.

Almost all points on the line $\{(\alpha, \alpha) : \alpha \in \mathbb{R}\}$ are in S by [2]. For the last part of the theorem, we will prove that this is the only line in $\mathbb{R}^2$ with the above property; all others intersect S in a set of measure zero, i.e. we will prove $(\alpha, A\alpha + B) \cap S$ has measure zero unless $A = 1$ and $B = 0$.

Without loss of generality assume that $A \geq 0$. We require $[10^n \alpha]$ and $[10^n (A\alpha + B)]$ simultaneously prime infinitely often and so we define the following two sequences of sets:

$$A_i = \left[\frac{p_{i1}}{10^i}, \frac{p_{i1} + 1}{10^i}\right) \cup \cdots \cup \left[\frac{p_{ik}}{10^i}, \frac{p_{ik} + 1}{10^i}\right), \quad i \geq 1,$$

where $p_{i1}, \ldots, p_{ik}$ are the primes between 10^i and $2 \cdot 10^i$, and

$$B_i = \left[\frac{q_{i1}}{10^i A} - \frac{B}{A}, \frac{q_{i1} + 1}{10^i A} - \frac{B}{A}\right) \cup \cdots \cup \left[\frac{q_{ik}}{10^i A} - \frac{B}{A}, \frac{q_{ik} + 1}{10^i A} - \frac{B}{A}\right), \quad i \geq 1,$$

where $q_{i1}, \ldots, q_{ik}$ are the primes between $10^i(A + B)$ and $10^i(2A + B)$. We thus need $\beta = A\alpha + B$ and $\alpha \in A_i \cap B_i$ for infinitely many i. To find the size of the intersection $A_i \cap B_i$ we consider the intersection of an interval from A_i with one from B_i. The intervals are of the form

$$\left[\frac{p_1}{10^i}, \frac{p_1 + 1}{10^i}\right) \quad \text{and} \quad \left[\frac{p_2}{10^i A} - \frac{B}{A}, \frac{p_2 + 1}{10^i A} - \frac{B}{A}\right).$$

CASE 1. Suppose that

$$\frac{p_1}{10^i} \leq \frac{p_2}{10^i A} - \frac{B}{A}.$$

Then we also need

$$\frac{p_2}{10^i A} - \frac{B}{A} < \frac{p_1 + 1}{10^i}.$$

Suppose first that $A > 1$. The length of the intersection is thus $\leq 1/10^i A$. We need

$$p_1 \leq \frac{p_2}{A} - \frac{B \cdot 10^i}{A} < p_1 + 1,$$

or equivalently

$$0 < \frac{p_2}{A} - p_1 - \frac{B \cdot 10^i}{A} < 1.$$

We thus need to determine the number of solutions to

$$\left| \frac{p_2}{A} - p_1 - \frac{B \cdot 10^i}{A} \right| < 1,$$

where $10^i \leq p_1, p_2 < 2 \cdot 10^i$ are primes. To do this, we modify the proof of Lemma 8.6 in [3] to take into account the $B \cdot 10^i/A$ term: By Dirichlet's theorem there exist $a, q \in \mathbb{Z}$ with $1 \leq a \leq q \leq 10^{3i/4}$ and $(a, q) = 1$ such that

$$\left| \frac{1}{A} - \frac{a}{q} \right| < \frac{1}{q \cdot 10^{3i/4}}.$$

Let

$$\varrho = \frac{1}{A} - \frac{a}{q} > 0.$$

We split the range for p_2 into blocks $[H, H + Z]$ where $Z = q \cdot 10^{i/20}$. Then in this range we have

$$\frac{p_2}{A} - p_1 - \frac{B \cdot 10^i}{A} = \frac{p_2 a}{q} - p_1 + H\varrho - \frac{B \cdot 10^i}{A} + O(10^{-7i/10}).$$

If we let

$$C = \left[H\varrho - \frac{B \cdot 10^i}{A} + \frac{1}{2} \right]$$

then we need an upper bound for the number of solutions of

$$\left| \frac{p_2 a}{q} - p_1 + C \right| < 2.$$

This is equivalent to the number of solutions of

$$p_2 a = (p_1 - C)q + b \quad \text{where } |b| < 2q.$$

For each b the solutions have the form

$$p_2 = \bar{a}b + xq, \quad p_1 = C + xa + b\left(\frac{a\bar{a} - 1}{q} \right)$$

where $a\bar{a} = 1 \bmod q$ and $\bar{a}$ is the least positive residue. We can use Theorem 2.3 of [1] to obtain a bound for the number of solutions in x of this equation. This is

$$\ll \frac{Z}{q \log^2 10^i} \frac{aq}{\phi(aq)} \sum_{\substack{|b|<2q \\ (b,q)=1=(qC+b,a)}} \frac{qC + b}{\phi(qC + b)}$$

$$\ll \frac{Z}{q \log^2 10^i} \frac{aq}{\phi(aq)} \sum_{\substack{(n,aq)=1 \\ |n-Cq|<2q}} \frac{n}{\phi(n)} \ll \frac{Z}{i^2}$$

if C is not so large that $\log C$ is of order equal to a power of q, since by Lemma 7.4 of [3],

$$\sum_{\substack{(n,aq)=1 \\ |n-Cq|<2q}} \frac{n}{\phi(n)} = 4q\,\frac{\phi(aq)}{aq} \prod_{p \nmid aq} \left(1 + \frac{1}{p(p-1)}\right)$$

$$+ O(\tau(aq)\log 2(Cq + 2q))$$

where $\tau(q)$ represents the number of divisors of q. Now by summing over the blocks we obtain $O(10^i/i^2)$ as required. In the case when C is large relative to q, we make use of the averaging over C. Write

$$C_0 = \left[\varrho \cdot 10^i - \frac{B \cdot 10^i}{A} + \frac{1}{2}\right], \quad C_1 = C_0 - 2, \quad C_2 = \left[2\varrho \cdot 10^i - \frac{B \cdot 10^i}{A}\right] + 3.$$

We can assume that $C_0 \geq q^2$. Then our bound is now

$$\ll \sum_{0 \leq m \leq 10^i/Z} \frac{Z}{qi^2} \frac{aq}{\phi(aq)} \sum_{\substack{|b|<2q \\ (b,q)=1=(qC+b,a) \\ C=[(10^i+mZ)\varrho - (B\cdot 10^i/A)+1/2]}} \frac{qC+b}{\phi(qC+b)}$$

$$\ll \frac{Z}{qi^2} \frac{aq}{\phi(aq)} \frac{1}{Z\varrho} \sum_{\substack{(n,aq)=1 \\ qC_1 \leq n \leq qC_2}} \frac{n}{\phi(n)} \ll \frac{10^i}{i^2}$$

since by Lemma 7.4 of [3],

$$\sum_{\substack{(n,aq)=1 \\ qC_1 \leq n \leq qC_2}} \frac{n}{\phi(n)} = \left(2\varrho \cdot 10^i - \frac{B \cdot 10^i}{A} + 3 - \varrho \cdot 10^i + \frac{B \cdot 10^i}{A} + 2\right) q\,\frac{\phi(aq)}{q}$$

$$+ O\left(\tau(aq)\log 2\left(2\varrho \cdot 10^i - \frac{B \cdot 10^i}{A} + 3\right)q\right).$$

For the case $A < 1$, the range of the intersection is $\leq 1/10^i$ and we need the number of solutions of $0 < p_2 - Ap_1 - B \cdot 10^i < A$. By the same method this gives the required result.

Suppose now that $A = 1$. In this case we need to find the number of solutions of

$$|p_2 - p_1 - B \cdot 10^i| < 1.$$

By Lemma 8.5 of [3] this is at most

$$\frac{K \cdot 10^i}{i_2} \sum_{|B \cdot 10^i - n|<2} \frac{n}{\phi(n)}$$

provided that $|B| \cdot 10^i > 3$. This is equal to

$$\frac{K \cdot 10^i}{i^2} \frac{630\zeta(3)}{\pi^4} + O(\log(|B| \cdot 10^i + 2))$$

by Lemma 2.5 of [3]. For any $B \neq 0$, we can choose i sufficiently large to ensure that $|B| \cdot 10^i \geq 3$ and so we obtain the required bound.

There are three other ways the intervals can intersect.

CASE 2. We have

$$\frac{p_2}{10^i A} - \frac{B}{A} \leq \frac{p_1}{10^i} < \frac{p_2 + 1}{10^i A} - \frac{B}{A}.$$

We write this as

$$0 \leq Ap_1 - p_2 + B \cdot 10^i < 1 \quad \text{if } A < 1$$

and as

$$0 \leq p_1 - \frac{p_2}{A} + \frac{B \cdot 10^i}{A} < \frac{1}{A} \quad \text{if } A > 1.$$

CASE 3. We have

$$\frac{p_2}{10^i A} - \frac{B}{A} \leq \frac{p_1}{10^i} \leq \frac{p_2 + 1}{10^i A} - \frac{B}{A} - \frac{1}{10^i},$$

which gives

$$0 \leq Ap_1 - p_2 + B \cdot 10^i \leq 1 - A$$

where we must have $A < 1$ in this case.

CASE 4. We have

$$\frac{p_1}{10^i} \leq \frac{p_2}{10^i A} - \frac{B}{A} \leq \frac{p_1 + 1}{10^i} - \frac{1}{10^i A},$$

which gives

$$0 \leq \frac{p_2}{A} - p_1 + \frac{B \cdot 10^i}{A} \leq 1 - \frac{1}{A}$$

where we must have $A > 1$ in this case.

It is clear that these cases give the same result as before so we deduce that

$$\lambda(A_i \cap B_i) = O\left(\frac{10^i}{i^2} \frac{1}{A \cdot 10^i}\right) = O(1/i^2)$$

and therefore since $\sum_{i=1}^{\infty} \lambda(A_i \cap B_i)$ converges, by the first Borel–Cantelli lemma [3, p. 8], we find that almost all α belong to only finitely many $A_i \cap B_i$. Hence the set of α that belong to infinitely many of the $A_i \cap B_i$ has measure zero. This completes the proof of the final part of the theorem.

References

[1] H. Halberstam and H.-E. Richert, *Sieve Methods*, London Math. Soc. Monogr. 4, Academic Press, London, 1974.
[2] G. Harman, *Metrical theorems on prime values of the integer parts of real sequences*, Proc. London Math. Soc. (3) 75 (1997), 481–496.
[3] —, *Metric Number Theory*, London Math. Soc. Monogr. (N.S.) 18, Oxford Univ. Press, 1998.
[4] —, *Metrical theorems on prime values of the integer parts of real sequences II*, J. London Math. Soc. (2) 64 (2001), 287–298.
[5] M. N. Huxley, *On the difference between consecutive primes*, Invent. Math. 15 (1972), 164–170.
[6] H. E. Rose, *A Course in Number Theory*, 2nd ed., Oxford Univ. Press, 1988, 1994.
[7] P. M. D. Rowe, *Contributions to metric number theory*, PhD thesis, Royal Holloway College, Univ. of London, 2002.

St. John's Seminary
Wonersh
Guildford
Surrey GU5 0QX
England, U.K.
E-mail: michael.rowe2@tesco.net