Reprinted from JOURNAL OF NUMBER THEORY
Vol. 2, No. 4, November 1970
Printed in Belgium

On the Distribution of the Convergents of Almost All Real Numbers

P. ERDÖS

Australian National University, Canberra, A.C.T., Australia

Communicated April 20, 1969; with revisions April 10, 1970

To the memory of Harold Davenport

Let $n_1 < n_2 < \cdots$ be an infinite sequence of integers. The necessary and sufficient condition that for almost all α the inequality $| \alpha - a_i/n_i | < \epsilon/n_i^2$ with $(a_i, n_i) = 1$ should have infinitely many solutions is that $\sum_{i=1}^{\infty} \varphi(n_i)/n_i^2 = \infty$.

The techniques used in the proof can perhaps be applied to prove an old conjecture of Duffin and Schaeffer.

Let $0 < \alpha < 1$,

$$\alpha = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cdots}}$$

be the development of α into a continued fraction (the a's are positive integers). $p_i^{(\alpha)}/q_i^{(\alpha)}$ $(1 \leqslant i < \infty)$ is the sequence of convergents belonging to α. Let $n_1 < n_2 < \cdots$ be any infinite sequence of integers. We are going to investigate the necessary and sufficient condition that for almost all α infinitely many of the $a_i^{(\alpha)}$ should occur amongst the n's. S. Hartman and P. Szüsz [1] proved that for almost all α every arithmetic progression contains infinitely many $q_i^{(\alpha)}$, and P. Szüsz [2] obtained an asymptotic formula for the number of $i \leqslant n$ for which $q_i^{(\alpha)}$ lies (for almost all α) in a given arithmetic progression.

We are going to prove the following

THEOREM I. *The necessary and sufficient condition that for almost all α infinitely many of the $q_i^{(\alpha)}$ are in the sequence $n_1 < n_2 < \cdots$ is that ($\phi(n)$ is Euler's ϕ function)*

$$\sum_{i=1}^{\infty} \phi(n_i)/n_i^2 = \infty. \tag{1}$$

The proof of the necessity is trivial and was of course well known. If p/q is a convergent of α then it is well known that $(p, q) = 1$ and

$$\left| \alpha - \frac{p}{q} \right| < \frac{1}{q^2} .$$

Thus the measure of the set in α for which a/n_i, $1 \leqslant a < n_i$, $(a, n_i) = 1$, is for some a a convergent of α is less than $2\phi(n_i)/n_i^2$. Hence by a simple and well-known argument (Borel-Cantelli Lemma) if the series (1) converges then for almost all α there are only finitely many $q_i^{(\alpha)}$ amongst the n_i.

The real difficulty is the proof of the sufficiency. It is well known that if

$$\left| \alpha - \frac{p}{q} \right| < \frac{1}{2q^2}$$

then p/q is a convergent of α. Thus to complete the proof of Theorem I it will suffice to prove the following.

THEOREM II. *Let $\epsilon > 0$ and assume that the series* (1) *diverges. Then for almost all α the inequality*

$$\left| \alpha - \frac{a}{n_i} \right| < \frac{\epsilon}{n_i^2}, \qquad (a, n_i) = 1. \tag{2}$$

has infinitely many solutions.

The proof of Theorem II will be long and difficult, and before we start it I want to make some remarks.

The well-known conjecture of Duffin and Schaeffer [3] contains our Theorems I and II as special cases. Their conjecture states that if $n_1 < n_2 < \cdots$ is a sequence of integers and $\delta_i > 0$ then the necessary and sufficient that for almost all α

$$\left| \alpha - \frac{a}{n_i} \right| < \frac{\delta_i}{n_i}, \qquad (a, n_i) = 1$$

should have infinitely many solutions is that

$$\sum_{i=1}^{\infty} \frac{\delta_i \phi(n_i)}{n_i}$$

diverges. Theorem II follows by putting $\delta_i = \epsilon/n_i$. It is very likely that our technique will also prove the above conjecture, but the details would be very much more complicated and so we do not investigate this question at present.

In Theorems I and II one could ask for the number of solutions in $i \leqslant N$. Put

$$A(N) = \sum_{i=1}^{N} \phi(n_i)/n_i^2.$$

Perhaps the following result holds: For almost all α the number of solutions of (2) for $1 \leqslant i \leqslant N$ equals $(1 + o(1))\, 2\epsilon A(N)$. Using a recent unpublished result of P. Szüsz one could make an analogous conjecture of Theorem I. The proof of these conjectures would in any case be probably very laborious and we do not consider them here.

Now we prove Theorem II. Theorem II will follow easily from the following

LEMMA 1. *Let (a_j, b_j) $a_j < b_j$, $j = 1,\ldots, T$ be a sequence of disjoint intervals in* (0, 1). *Denote the union of these intervals by S. Put*

$$\sum_{j=1}^{T} (b_j - a_j) = A.$$

Then there is an $\eta = \eta(A)$ so that if $n_1 < n_2 < \cdots < n_k$ is a sequence of sufficiently large integers satisfying

$$\sum_{i=1}^{k} \frac{\phi(n_i)}{n_i^2} = \eta_1 \leqslant \eta, \tag{3}$$

then the measure of the set in α where $\alpha \in S$ and for which

$$\left| \alpha - \frac{t}{n_i} \right| < \frac{\epsilon}{n_i^2}, \qquad (t, n_i) = 1 \tag{4}$$

is solvable for some i in $1 \leqslant i \leqslant k$ is greater than $\epsilon\eta_1 A$.

Let us assume that Lemma 1 has already been proved. We then easily deduce Theorem II. If Theorem II would be false there clearly would exist a set U of positive measure so that for all α in U (2) has only a finite number of solutions (it is easy to see that U is measurable). Hence by a simple argument there is an index i_0 and a $U_1 \subset U$ of positive measure so that for the α in U_1, (2) has no solutions with $i > i_0$. By the Lebesgue density theorem there is a sequence of disjoint intervals (a_j, b_j), $1 \leqslant j \leqslant T$, with

$$\sum_{j=1}^{T} (b_j - a_j) = A > \tfrac{1}{2} m(U_1).$$

($m(U_1)$ denotes the measure of U_1), so that

$$m[(a_j, b_j) \cap U_1] > \left(1 - \frac{\epsilon\eta}{2}\right)(b_j - a_j), \tag{5}$$

which implies that U_1 intersects the union of the T intervals (a_j, b_j), $j = 1,..., T$ in a set of measure greater than $A[1 - (\epsilon\eta/2)]$.

Now since (1) diverges there are arbitrarily large values of j_1 and j_2 so that $i_0 < j_1 < j_2$ and

$$\frac{\eta}{2} < \sum_{j_1}^{j_2} \phi(n_i)/n_i^2 < \eta. \tag{6}$$

Hence by Lemma 1 the measure of the set in α, $a_j < \alpha < b_j$, $1 \leqslant j \leqslant T$, for which

$$\left| \alpha - \frac{t}{n_i} \right| < \frac{\epsilon}{n_i^2}, \qquad (t, n_i) = 1$$

is solvable for $j_1 \leqslant i \leqslant j_2$ is greater than $\epsilon\eta A/2$. This contradicts (5) and thus Lemma 1 implies Theorem 2.

To complete our proof we now have to prove Lemma 1.

Denote by M the measure of the set in α, $a_j < \alpha < b_j$, $1 \leqslant j \leqslant T$, for which (4) is sovable for some i, $1 \leqslant i \leqslant k$. $m(n_i)$ denotes the measure of the set for which (4) is solvable for n_i, and $m(n_i, n_j)$ denotes the measure of the set for which (4) is solvable for both n_i and n_j. We have, by a simple sieve process,

$$M \geqslant \sum_{i=1}^{k} m(n_i) - \sum_{1 \leqslant i < j \leqslant k} m(n_i, n_j). \tag{6'}$$

First we estimate $m(n_i)$ from below. Denote by $\phi(n_i ; a_j, b_j)$ the number of integers t satisfying

$$a_j n_i < t \leqslant b_j n_i, \qquad (t, n_i) = 1.$$

By a simple sieve process we find, for sufficiently large n_i ($v(n)$ denotes the number of distinct prime factors of n),

$$\begin{aligned} \phi(n_i, a_j, b_j) &= \sum_{d \mid n_i} \mu(d) \left(\left[\frac{n_i b_j}{d} \right] - \left[\frac{n_i a_j}{d} \right] \right) \\ &> \sum_{d \mid n_i} \left(\mu(d) \frac{n_i(b_j - a_j)}{d} \right) - 2^{v(n_i)} = \phi(n_i)(b_j - a_j) - 2^{v(n_i)} \\ &= (1 + o(1))\, \phi(n_i)(b_j - a_j). \end{aligned} \tag{7}$$

From (7) and $A = \sum_{j=1}^{T} (b_j - a_j)$ we easily obtain (the length of the intervals (4) is $2\epsilon/n_i^2$)

$$m(n_i) \geqslant (1 + o(1)) \sum_{j=1}^{T} \frac{2\epsilon}{n_i^2} \phi(n_i, a_j, b_j)$$
$$= (1 + o(1))\, 2\epsilon A \frac{\phi(n_i)}{n_i^2}. \tag{8}$$

Thus from (3), (6), and (8)

$$M \geqslant (1 + o(1))\, 2\epsilon\eta_1 A - \sum_{1 \leqslant i \leqslant j \leqslant k} m(n_i, n_j). \tag{9}$$

Hence by (9), (3), and (6) to complete the proof of Lemma 1 we only have to show that if $\eta = \eta(A)$ is sufficiently small then (3) implies

$$\sum_{1 \leqslant i \leqslant j \leqslant k} m(n_i, n_j) < \eta_1 \epsilon A/2. \tag{10}$$

The proof of (10) will be long and difficult. First, for purposes of orientation, we remark that if n_j is large compared to n_i we have

$$m(n_i, n_j) = (1 + o(1))\, 4\epsilon^2 A \frac{\phi(n_i)\, \phi(n_j)}{n_i^2 n_j^2}. \tag{11}$$

The proof of (11) is easy. Consider a fixed interval

$$\left(\frac{t}{n_i} - \frac{\epsilon}{n_i^2}, \frac{t}{n_i} + \frac{\epsilon}{n_i^2}\right), \qquad (t, n_i) = 1. \tag{12}$$

It follows from (7) that for sufficiently large n_j the number of integers $(t', n_j) = 1$ satisfying

$$\frac{t}{n_i} - \frac{\epsilon}{n_i^2} < \frac{t'}{n_j} < \frac{t}{n_i} + \frac{\epsilon}{n_j^2}$$

is $(1 + o(1))\, \phi(n_j)\, 2\epsilon/n_i^2$. Since the number of the intervals (12) which are in $S$is by (7) $(1 + o(1))\, A\phi(n_i)$, and the length of the intervals of $m(n_j)$ is $2\epsilon/n_j^2$ we immediately obtain (11). (To clarify this sketch we remark that to get the exact formula for $m(n_i, n_j)$ one need not only count the t'/n_j which lie in an interval of $m(n_i)$, and on the other hand if t'/n_j lies in $m(n_i)$ sometimes not all of this interval of $m(n_j)$ is counted in $m(n_i, n_j)$. But it is clear that the error made by using the present simple counting process is negligible.)

If (11) would be true for all n_i and n_j then (10) would easily follow from (11), (9), and (3), if $\eta = \eta(A)$ is sufficiently small. The difficulty with our proof is that (11) is certainly not always true; thus to prove (10) we have to use very much more complicated arguments. If $m(n_i, n_j)$ is not 0 and if $n_i < n_j$ there must exist integers

$$1 \leqslant t_i < n_i, \qquad 1 \leqslant t_j < n_j, \qquad (t_i, n_i) = (t_j, n_j) = 1$$

satisfying

$$\left| \frac{t_i}{n_i} - \frac{t_j}{n_j} \right| < \frac{\epsilon}{n_i^2} + \frac{\epsilon}{n_j^2} < \frac{2\epsilon}{n_i^2}$$

or

$$| t_i n_j - t_j n_i | < 2\epsilon \frac{n_j}{n_i}.$$

Now denote by $f_\epsilon(n_i, n_j)$ the number of solutions of

$$\begin{gathered} 1 \leqslant t_i < n_i, \qquad 1 \leqslant t_j < n_j, \qquad (t_i, n_i) = (t_j, n_j) = 1 \\ | t_i n_j - t_j n_i | < 2\epsilon \frac{n_j}{n_i}. \end{gathered} \tag{13}$$

Observe that the overlap of the intervals

$$\left(\frac{t_i}{n_i} - \frac{\epsilon}{n_i^2}, \frac{t_i}{n_i} + \frac{\epsilon}{n_i^2} \right), \qquad \left(\frac{t_j}{n_j} - \frac{\epsilon}{n_j^2}, \frac{t_j}{n_j} + \frac{\epsilon}{n_j^2} \right)$$

is at most $2\epsilon/n_j^2$. Hence clearly

$$m(n_i, n_j) < \frac{2\epsilon}{n_j^2} f_\epsilon(n_i, n_j). \tag{14}$$

By the same method which we used to prove (11) we can show that if n_j/n_i is very large then

$$f_\epsilon(n_i, n_j) = (1 + o(1)) \frac{4\epsilon\phi(n_i)\, \phi(n_j)}{n_i^2}, \tag{15}$$

but again (15) is not always true. Thus we have to use much more complicated methods.

Let n be any integer. Define $g(n)$ as the smallest integer for which

$$\sum{}' \frac{1}{p} < 1, \tag{16}$$

where in $\sum'$ the summation is over all primes p with $p \mid n$, $p > g(n)$. Put

$$\sum_{1 \leqslant i < j \leqslant k} m(n_i, n_j) = \Sigma_1 + \Sigma_2, \tag{17}$$

where in Σ_1 the summation is extended over the i and j for which

$$\epsilon n_j / n_i > d 4^t t^t, \qquad \text{where } d = (n_i, n_j),\ t = \max(g(n_i), g(n_j)), \tag{18}$$

and in Σ_2 over the i and j which do not satisfy (18).

First we estimate Σ_1. We are going to prove that

$$\Sigma_1 < \frac{\eta_1 \epsilon A}{4}. \tag{19}$$

To prove (19) we first show that if n_i and n_j satisfy (18) then (the c's are suitable positive absolute constants)

$$f_\epsilon(n_i, n_j) < c_1 \epsilon \frac{\phi(n_i)\,\phi(n_j)}{n_i^2}. \tag{20}$$

Assume that (20) has already been proved. Then (19) easily follows. From (20) and (14) we have

$$m(n_i, n_j) < 4 c_1 \epsilon^2 \frac{\phi(n_i)\,\phi(n_j)}{n_i^2 n_j^2}.$$

Thus from (3)

$$\Sigma_1 < 4 c_1 \epsilon^2 \left(\sum_{i=1}^{k} \frac{\phi(n_i)}{n_i} \right)^2 < 4 c_1 \epsilon^2 \eta_1^2 < \frac{\eta_1 \epsilon A}{4}$$

if $\eta_1 = \eta_1(A)$ is sufficiently small.

Thus to prove (19) we only have to show (20). Denote by $H(u)$ the number of solutions in t_i, t_j of

$$t_i n_j - t_j n_i = u \qquad 1 \leqslant t_i \leqslant n_i,\ 1 \leqslant t_j \leqslant n_j,\ (t_i, n_i) = (t_j, n_j) = 1. \tag{21}$$

Then it is clear that

$$H(u) = 0 \quad \text{if} \quad d \nmid u, \quad \text{and} \quad H(u) \leqslant d \quad \text{if} \quad d \mid u. \tag{22}$$

We shall need stronger results than (22). We may assume that $d \mid u$, and we write

$$u = d\, d_u u_1 \tag{23}$$

where $(u_1, d) = 1$ and where d_u is composed of prime factors of d. This representation is clearly unique.

The notation $p^\alpha \| n$ will mean that $p^\alpha \nmid n$ but $p^{\alpha+1} \nmid n$, and we put

$$d_u(t) = \prod p^\alpha, \qquad p^\alpha \| d_u, \qquad p \leqslant t. \tag{24}$$

As usual, $\pi(t)$ will denote the number of primes not exceeding t. Finally we write

$$n_{ij} = \frac{n_i n_j}{d^2}. \tag{25}$$

We have

$$f_\epsilon(n_i, n_j) = \sum_{|u| < \epsilon n_j/n_i} H(u) = \sum_u{}' + \sum_u{}'', \tag{26}$$

where in $\sum'$, the summands u satisfy $|u| < 2\epsilon(n_j/n_i)$ and $d_u(t) \leqslant t^{\pi(t)}$, while in $\sum''$ they satisfy $|u| < 2\epsilon(n_j/n_i)$ and $d_u(t) > t^{\pi(t)}$.

LEMMA 2. $H(u) = 0$ *unless* $(u_1, n_{ij}) = 1$.

Proof. Suppose that $(u_1, n_{ij}) > 1$, and suppose $p \mid u_1, p \mid n_{ij}$. Since $(u_1, d) = 1$ we have $p \nmid d$; hence either $p \mid n_i$ or $p \mid n_j$. Assume that $p \mid n_i, p \nmid n_j$. If $H(u)$ were positive, there would be a solution of (21), whence $p \mid t_i n_j$. Since $p \nmid n_j$ we have $p \mid t_i$, which contradicts $(n_i, t_i) = 1$. This proves the lemma.

We now estimate the sum $\sum''$ in (26). By (22) and by Lemma 2 we may restrict ourselves to summands u with $d \mid u$ and with $u \neq 0$. Since $d_u(t) > t^{\pi(t)}$, $d_u(t)$ must have a prime factor $p \leqslant t$ with $p^\alpha \mid d_u(t)$ and $p^\alpha > t$. If α is even, $d_u(t)$ is divisible by a square greater than t. If α is odd, $d_u(t)$ is divisible by the square $p^{\alpha-1} > t^{1-(1/\alpha)} \geqslant t^{2/3}$.
Thus

$$\sum_u{}'' H(u) \leqslant d \sum_u{}'' 1 \leqslant 2d \sum_{\substack{x=1 \\ x \text{ div. by} \\ \text{a square} > t^{2/3}}}^{2\epsilon n_j/n_i d} 1 \leqslant 2d \sum_{k=t^{1/3}}^{\infty} \sum_{\substack{x=1 \\ k^2 | x}}^{2\epsilon n_j/n_i d} 1$$

$$\leqslant 4\epsilon \frac{n_i}{n_j} \sum_{k=t^{1/3}}^{\infty} \frac{1}{k^2} < 8\epsilon \frac{n_j}{n_i t^{1/3}}. \tag{27}$$

By (16), (18), and the theorem of Mertens we have

$$\frac{\phi(n_i)\,\phi(n_j)}{n_i n_j} > c_4 \prod_{p \leqslant t} \left(1 - \frac{1}{p}\right)^2 > c_5 (\log t)^{-2}. \tag{28}$$

The inequalities (27) and (28) imply that, for sufficiently large c_1,

$$\sum{}'' < c_1 \epsilon \phi(n_i)\ \phi(n_j)/2n_i^2. \tag{29}$$

Thus by (26), to complete the proof of (20) we only have to show that

$$\sum{}' < c_1 \epsilon \phi(n_i)\ \phi(n_j)/2n_i^2. \tag{30}$$

To do this we write

$$\sum{}' = \sum_s{}^* \sum_{\substack{|u|<2\epsilon n_j/n_i \\ d_u(t)=s}} H(u), \tag{31}$$

where the summand s in $\sum^*$ runs through all the divisors s of d which do not exceed $t^{\pi(t)}$ and all whose prime factors are not greater than t.

We now need a better estimate for $H(u)$ than (22). Let t_i', t_j' be the unique solution (if it exists) of

$$\begin{gathered} t_i' \frac{n_j}{d} - t_j' \frac{n_i}{d} = \frac{u}{d}, \\ 0 < t_i' < n_i/d, \quad 0 < t_j' < n_j/d, \quad (t_i', n_i/d) = (t_j', n_j/d) = 1. \end{gathered} \tag{32}$$

We obtain all the solutions of (21) by considering all the integers of the form

$$t_i' + X\frac{n_i}{d}, \qquad t_j' + X\frac{n_j}{d},$$

with the integer X satisfying

$$\left(t_i' + X\frac{n_i}{d}, d\right) = \left(t_j' + X\frac{n_j}{d}, d\right) = 1, \qquad 0 \leqslant X < d. \tag{33}$$

Then $H(u)$ is not greater than the number of solutions of (33). In fact, either $H(u) = 0$ or $H(u)$ equals the number of solutions of (33).

Suppose now that

$$p \mid d,\ p \nmid d_u,\ p \nmid n_{ij}.$$

Then (33) implies that

$$X\frac{n_i}{d} \not\equiv -t_i' \pmod{p}, \qquad X\frac{n_j}{d} \not\equiv -t_j' \pmod{p}.$$

Since $p \nmid (n_i/d)$ and $p \nmid (n_j/d)$, each of these relations exclude a residue class mod p for the variable X. The determinant

$$t_i' \frac{n_j}{d} - t_j' \frac{n_i}{d} = \frac{u}{d} = u_1 d_u$$

by (32), and since $p \nmid u_1 d_u$, it is $\equiv 0 \pmod p$. Hence two distinct residue classes mod p are excluded for X. In general, if $p \mid d$ but if $p \mid d_u$ or $p \mid n_{ij}$, we can conclude that one residue class is excluded for X. Hence

$$H(u) \leqslant d \prod_{\substack{p \mid d \\ p \nmid d_u \\ p \nmid n_{ij}}} \left(1 - \frac{2}{p}\right) \prod_{\substack{p \mid d \\ p \mid d_u n_{ij}}} \left(1 - \frac{1}{p}\right). \tag{34}$$

Now if $p \mid n_{ij}$ and $p \mid d_u$, then (32) has no solution. We may therefore in the sum $\sum_s^*$ in (31) restrict ourselves to summands s which are not divisible by primes p with $p \mid n_{ij}$.

We have

$$\sum_{\substack{|u| < 2\epsilon n_j/n_i \\ d_u(t) = s}} H(u) \leqslant d \prod_{\substack{p \mid d \\ p \leqslant t \\ p \nmid n_{ij} s}} \left(1 - \frac{2}{p}\right) \prod_{\substack{p \mid d \\ p \leqslant t \\ p \mid n_{ij} s}} \left(1 - \frac{1}{p}\right) \sum_{\substack{|u| < 2\epsilon n_j/n_i \\ u \text{ of type (23)} \\ d_u(t) = s \\ H(u) > 0}} 1. \tag{35}$$

By Lemma 2, the condition $H(u) > 0$ implies that $(u_1, n_{ij}) = 1$. We have

$$u = d\, d_u u_1 = d\, d_u(t) \left(\frac{d_u}{d_u(t)} u_1\right) = ds \left(\frac{d_u}{d_u(t)} u_1\right) = dsy,$$

say. Now $(u_1, d) = (u_1, n_{ij}) = 1$, and $d_u/d_u(t)$ is not divisible by a prime $\leqslant t$. Hence y is not divisible by any prime p with $p \leqslant t$ and $p \mid dn_{ij}$. Hence the sum on the right hand side of (35) is bounded by

$$\sum_{\substack{1 \leqslant |y| \leqslant 2\epsilon n_j/n_i ds \\ y \text{ not div. by primes } p \\ \text{with } p \leqslant t \text{ and } p \mid dn_{ij}}} 1. \tag{36}$$

Thus by the sieve of Eratosthenes it is bounded by

$$4\left(\frac{\epsilon n_j}{n_i\, ds} \prod_{\substack{p \leqslant t \\ p \mid dn_{ij}}} \left(1 - \frac{1}{p}\right) + 2^{\pi(t)}\right).$$

By (18), by the formula of Mertens, and since $s \leqslant t^{\pi(t)}$, we have

$$\frac{n_j}{n_i\, ds} \prod_{p \leqslant t} \left(1 - \frac{1}{p}\right) - \frac{4^t t^t}{s} (\log t)^{-2} > 2^{\pi(t)},$$

and thus the sum on the right hand side of (36) is bounded by

$$c_6 \frac{\epsilon n_j}{n_i\, ds} \prod_{\substack{p \leqslant t \\ p \mid dn_{ij}}} \left(1 - \frac{1}{p}\right).$$

In view of the definition of t in (18), this is

$$< c_7 \frac{\epsilon n_j}{n_i\, ds} \prod_{p \mid dn_{ij}} \left(1 - \frac{1}{p}\right). \tag{37}$$

Using (35) and the definition of t again we obtain

$$\begin{aligned} \sum_{\substack{|u| < 2\epsilon n_j/n_i \\ d_u(t) = s}} H(u) &< \frac{c_8 \epsilon n_j}{n_i s} \prod_{p \mid dn_{ij}} \left(1 - \frac{1}{p}\right) \prod_{\substack{p \mid d \\ p \nmid n_{ij}s}} \left(1 - \frac{2}{p}\right) \prod_{\substack{p \mid d \\ p \mid n_{ij}s}} \left(1 - \frac{1}{p}\right) \\ &< \frac{c_9 \epsilon n_j}{n_i s} \prod_{p \mid d} \left(1 - \frac{1}{p}\right) \prod_{p \mid dn_{ij}} \left(1 - \frac{1}{p}\right) \prod_{\substack{p \mid d \\ p \nmid n_{ij}s}} \left(1 - \frac{1}{p}\right). \end{aligned}$$

Thus by (31),

$$\sum{}' < \frac{c_9 \epsilon n_j}{n_i} \prod_{p \mid d} \left(1 - \frac{1}{p}\right) \prod_{p \mid dn_{ij}} \left(1 - \frac{1}{p}\right) \sum_s{}^{**} \frac{1}{s} \prod_{\substack{p \mid d \\ p \nmid n_{ij}s}} \left(1 - \frac{1}{p}\right),$$

where the sum is over all the divisors s of d with $(s, n_{ij}) = 1$. We obtain

$$\sum{}' < \frac{c_9 \epsilon n_j}{n_i} \prod_{p \mid d} \left(1 - \frac{1}{p}\right)^2 \prod_{p \mid dn_{ij}} \left(1 - \frac{1}{p}\right) \sum_{\substack{s \mid d \\ (s, n_{ij}) = 1}} \frac{1}{s} \prod_{\substack{p \mid d \\ p \mid n_{ij}s}} \left(1 - \frac{1}{p}\right)^{-1}. \tag{38}$$

Now

$$\begin{aligned} &\sum_{\substack{s \mid d \\ (s, n_{ij}) = 1}} \frac{1}{s} \prod_{\substack{p \mid d \\ p \mid u_{ij}s}} \left(1 - \frac{1}{p}\right)^{-1} \\ &\quad = \Bigg(\prod_{\substack{p \mid d \\ p \mid n_{ij}}} \left(1 - \frac{1}{p}\right)^{-1} \Bigg) \sum_{\substack{s \mid d \\ (s, n_{ij}) = 1}} \frac{1}{s} \prod_{p \mid s} \left(1 - \frac{1}{p}\right)^{-1} \\ &\quad < \Bigg(\prod_{\substack{p \mid d \\ p \mid n_{ij}}} \left(1 - \frac{1}{p}\right)^{-1} \Bigg) \Bigg(\prod_{\substack{p \mid d \\ p \nmid n_{ij}}} \left(1 + \left(\frac{1}{p} + \frac{1}{p^2} + \cdots\right)\left(1 - \frac{1}{p}\right)^{-1}\right) \Bigg) \\ &\quad < c_{10} \prod_{p \mid d} \left(1 - \frac{1}{p}\right)^{-1}. \end{aligned}$$

Thus (38) yields

$$\sum' < \frac{c_{11}\epsilon n_j}{n_i} \prod_{p|d}\left(1-\frac{1}{p}\right) \prod_{p|dn_{ij}}\left(1-\frac{1}{p}\right) = \frac{c_{11}\epsilon n_j}{n_i}\,\frac{\phi(n_i)\,\phi(n_j)}{n_i n_j}$$
$$= c_{11}\epsilon\,\frac{\phi(n_i)\,\phi(n_j)}{n_i^2},$$

which proves (30) and hence (20) by virtue of (27), (26) and (29). Thus (19) follows.

Next we prove

$$\sum\nolimits_2 < n_i\epsilon A/4. \tag{39}$$

If (39) is proved then (19) and (17) implies (10) which completes the proof of Lemma 1 and hence also proves Theorems 1 and 2. The proof of (39) is not quite simple.

By the definition of $f_\epsilon(n_i, n_j)$ and by (22) we have $f_\epsilon(n_i, n_j) \leqslant 2\epsilon n_j/n_i$, and using (14) we obtain

$$m(n_i, n_j) < \frac{4\epsilon^2}{n_i n_j}. \tag{40}$$

By (40) we have

$$\sum\nolimits_2 < 4\epsilon^2 \sum{}' \frac{1}{n_i n_j},$$

where in $\sum'$ the summation is extended over the n_i and n_j which do not satisfy (18). Thus to complete the proof of (39) we only have to show that

$$\sum{}' \frac{1}{n_i n_j} < \frac{n_1 A}{16\epsilon}. \tag{41}$$

Let $t_0 = \eta_1^{-1}$. We have

$$\sum{}' \frac{1}{n_i n_j} = \sum\nolimits_1' \frac{1}{n_i n_j} + \sum\nolimits_2' \frac{1}{n_i n_j} \tag{42}$$

where in $\sum_1'$, $t \leqslant t_0$ and in $\sum_2'$, $t > t_0$ (t is defined by (18)). The estimation of $\sum_1'$ is trivial. If $t \leqslant t_0$ we have by the theorem of Mertens and by (16).

$$\frac{\phi(n_i)\,\phi(n_j)}{n_i n_j} > c_{12} \prod_{p<t_0}\left(1-\frac{1}{p}\right)^2 > \frac{c_{13}}{(\log t_0)^2}. \tag{43}$$

Thus by (3)

$$\sum\nolimits_1' \frac{1}{n_i n_j} < \Big(\sum_{i=1}^{k} \frac{\phi(n_i)}{n_i^2}\Big)^2 \frac{(\log t_0)^2}{c_{13}} < \frac{\eta_1 A}{32\epsilon}$$

for sufficiently small η_1.

Thus to prove (41) and (39) we only have to prove that

$$\sum\nolimits_2' \frac{1}{n_i n_j} < \frac{\eta_1 A}{32\epsilon}. \tag{45}$$

The proof of (45) will be the main difficulty. Let u and v be integers satisfying

$$t_0 = \eta_1^{-1} < u \leqslant v \tag{46}$$

and put

$$A_{u,v} = \sum_{u,v} \frac{1}{n_i n_j}, \tag{47}$$

where in $\sum_{u,v}$, $g(n_i) = u$, $g(n_j) = v$ (see (16)). We then have

$$\sum\nolimits_2' \frac{1}{n_i n_j} = \sum A_{u,v}, \qquad t_0 < u \leqslant v. \tag{48}$$

We have to estimate $A_{u,v}$. We remind the reader that in $A_{u,v}$ n_i and n_j run through the integers of (3) for which $g(n_i) = u$, $g(n_j) = v$ ($g(n)$ is defined by (16)) and (18) is not satisfied.

If $g(n_i) = u$ we have as in (43),

$$\frac{\phi(n_i)}{n_i} > c_{14} \prod_{p<u} \Big(1 - \frac{1}{p}\Big) > \frac{c_{15}}{\log u}. \tag{49}$$

Thus by (3)

$$A_{u,v} < \frac{\log u}{c_{15}} \sum_{i=1}^{k} \frac{\phi(n_i)}{n_i} \sum\nolimits_v \frac{1}{n_j} < \eta_1 \frac{\log u}{c_{15}} \sum\nolimits_v \frac{1}{n_j}, \tag{50}$$

where in $\sum_v$, n_j runs through the integers not satisfying (18) for which $g(n_j) = v$ (since $g(n_j) \geqslant g(n_i)$ we have $g(n_j) = v = t$).

Since (18) is not satisfied we have for the n_j in $\sum_v$

$$\frac{n_i}{d(4v)^v} < n_j < n_i\, d(4v)^v, \qquad (n_i, n_j) = d. \tag{51}$$

First of all we can improve (51). By (22) $H(u) = 0$ unless $u \equiv 0 \pmod d$. Thus we can assume $n_j/n_i \geqslant d$ or $n_i/n_j \geqslant d$. In other words, instead of (51) we may assume that

$$\frac{n_i}{d(4v)^v} < n_j \leqslant \frac{n_i}{d} \qquad \text{or} \qquad n_i d \leqslant v < n_i\, d(4v)^v, \tag{51$'$}$$

and d runs through all divisors of n_i. Write

$$n_i = dn_i', \qquad n_j = dn_j', \qquad (n_i', n_j') = 1.$$

We have

$$\sum\nolimits_v \frac{1}{n_j} = \sum\nolimits_v' \frac{1}{n_j} + \sum\nolimits_v'' \frac{1}{n_j} \tag{52}$$

where in $\sum_v'$ the n_j run through all the n's not satisfying (18) with $g(n_j) = v$ and for which

$$\sum_{\substack{p \mid n_j' \\ p > v}} \frac{1}{p} < \frac{1}{2}. \tag{53}$$

By (53) and the definition of $g(n_j) = v$, we have for the n_j in $\sum_v'$

$$\sum_{\substack{p \mid d \\ p > v}} \frac{1}{p} > \frac{1}{4}. \tag{54}$$

Thus we have

$$\sum_v{}' \frac{1}{n_j} \leqslant \sum{}' \frac{1}{d} \sum \frac{1}{n_j'}, \tag{55}$$

where in $\sum'$, d runs through all divisors of n_i satisfying (54) and

$$n_i/d^2(4v)^v < n_j' \leqslant n_i/d^2 \qquad \text{or} \qquad n_i \leqslant n_j' < n_i(4v)^v. \tag{56}$$

Now by a simple calculation (l runs through all the integers of the two intervals (56))

$$\sum \frac{1}{n_j'} \leqslant \sum \frac{1}{l} < v^2. \tag{57}$$

Thus from (56) and (57),

$$\sum_v{}' \frac{1}{n_j} < v^2 \sum{}' \frac{1}{d}. \tag{58}$$

Now we have to estimate $\sum' 1/d$. Let $q_1 < \cdots < q_n$ be the prime factors of n_i, which are greater than $v = g(n_j)$. We have for the d in $\sum'$ (by (54))

$$\sum_{q_i \mid d} \frac{1}{q_i} > \frac{1}{4}.$$

It easily follows from the prime-number theorem (or a more elementary theorem) that for sufficiently large v the integer d has to be divisible by more than v of the q's.

Writing $d = d_1 d_2$ where d_2 is divisible by precisely v of the q's we obtain

$$\sum' \frac{1}{d} \leqslant \Big(\sum_{d_1 \mid n_i} \frac{1}{d_1} \Big) \Big(\sum_{d_2} \frac{1}{d_2} \Big) = \Big(\sum_{d \mid n_i} \frac{1}{d} \Big) \Big(\Big(\sum \frac{1}{q_i} \Big)^v \Big/ v! \Big). \tag{59}$$

Now by (16) (as in (43) and (49)).

$$\sum_{d \mid n_i} \frac{1}{d} < c_{16} \log u. \tag{60}$$

By $g(n_j) \geqslant g(n_i)$ we have $\sum 1/q_i < 1$. Thus from (59),

$$\sum' \frac{1}{d} < (c_{16} \log u) \frac{1}{v!}. \tag{61}$$

Inequalities (58) and (61) imply

$$\sum\nolimits_v \frac{1}{n_j} < c_{16} v^2 (\log u) \frac{1}{v!} < \frac{1}{2^v} \tag{62}$$

for sufficiently large v.

Now we estimate $\sum_v'' 1/n_j$. We prove the following

Lemma 3. *The number of integers $m < x$ for which*

$$\sum_{\substack{p \mid m \\ p > v}} \frac{1}{p} \geqslant \frac{1}{2} \tag{63}$$

is less than $x/v!$ for sufficiently large v.

We split the integers $m < x$ satisfying (63) into two classes. In the first class are the integers m which have at least $2v$ distinct prime factors in the interval (v, e^{v^2}). The number of integers of the first class is clearly less than

$$x \Big(\sum_{v < p < \exp v^2} \frac{1}{p} \Big)^{2v} \Big/ (2v)! < x \frac{(3 \log v)^{2v}}{(2v)!} < \frac{x}{2 \cdot v!} \tag{64}$$

for sufficiently large v.

Let $p_1 < \cdots < p_{2v}$ be the first $2v$ primes greater than v. It follows from the prime-number theorem (or a more elementary theorem) that for sufficiently large v

$$\sum_{i=1}^{2v} \frac{1}{p_i} < \frac{1}{4}. \tag{65}$$

Thus from (63) and (65) if m is in the second class we have

$$f(m) = \sum_{\substack{p \mid m \\ p > \exp v^2}} \frac{1}{p} > \frac{1}{4}. \tag{66}$$

We evidently have

$$\sum_{m=1}^{x} f(m) = \sum_{p > \exp v^2} \frac{1}{p} \left[\frac{x}{p}\right] < \sum_{p > \exp v^2} \frac{1}{p^2} < \frac{x}{\exp v^2}. \tag{67}$$

From (66) and (67) the number of integers of the second class is less than

$$\frac{4x}{\exp v^2} < \frac{x}{2v!}, \tag{68}$$

(64) and (68) complete the proof of Lemma 3.

By the same method, we could prove the following sharpening of Lemma 3: denote by $N(\alpha, v, x)$ the number of integers $m \leqslant x$ satisfying

$$\sum_{\substack{p \mid m \\ p > v}} \frac{1}{p} > \alpha.$$

Put $\log \beta = \alpha$. For every ϵ and α there is a $v_0 = v_0(\epsilon, \alpha)$ so that for every $v > v_0$

$$x/\exp v^{\beta(1+\epsilon)} < N(\alpha, v, x) < x/\exp v^{\beta(1-\epsilon)}.$$

From Lemma 3 we immediately obtain

$$\sum{}' \frac{1}{m} < \frac{2}{v!} \tag{69}$$

where in $\sum'$ the summation is extended over the integers $x < m \leqslant 2x$ which satisfy (63). From (57) and (69) and (60) we obtain by a simple calculation (in $\sum_v'' 1/n_j'$ the n_j run through the integers satisfying (63) and (56))

$$\sum_v{}'' \frac{1}{n_j} \leqslant \sum_{d \mid n_i} \frac{1}{d} \sum_v{}'' \frac{1}{n_j'} < \frac{2v^2}{v!} \sum_{d \mid n_i} \frac{1}{d} < 2c_{16} v^2 \frac{\log u}{v!} < \frac{1}{2^v} \tag{70}$$

for sufficiently large v. Thus from (52), (62), and (70),

$$\sum_{v} \frac{1}{n_j} < \frac{1}{2^{v-1}}. \tag{71}$$

Inequalities (70) and (71) imply that for sufficiently large v

$$A_{u,v} < \frac{\eta_1 \log u}{c_{15} 2^{v-1}} < \frac{\eta_1}{2^{v/2}}, \tag{72}$$

which with (48) finally imply that for sufficiently small $\eta (t_0 = \eta_1^{-1}$ and $0 < \eta_1 < \eta$, by (46) v is large if η is small)

$$\sum_{2}' \frac{1}{n_i n_j} < \eta_1 \sum_{\eta_1^{-1} < u \leqslant v} \frac{1}{2^{v/2}} < \eta_1{}^2 < \frac{\eta_1 A}{32\epsilon}.$$

This proves (45), hence (41), and hence (39). Thus Lemma 1 is proved, and therefore also Theorems I and II.

Acknowledgment

Finally I would like to thank W. Schmidt who improved the paper in several places—in particular, his proof of (20) is given here, which replaced my somewhat sketchy and not quite accurate earlier proof.

References

1. S. Hartman and P. Szüsz, On congruence classes of denominators of convergents, *Acta Arith.* **6** (1960), 179–184.
2. P. Szüsz, Über die metrische Theorie der diophantischen Approximation, II, *Acta Arith.* **8** (1963), 225–241.
3. Duffin and Schaeffer, Khintchene's problem in metric diophantine approximation, *Duke Math. J.* **8** (1941), 243–255.

PRINTED IN BELGIUM BY THE ST. CATHERINE PRESS, TEMPELHOF, 37, BRUGES, LTD.