PAIR CORRELATION OF RATIONALS WITH PRIME DENOMINATORS

MAOSHENG XIONG AND ALEXANDRU ZAHARESCU⋆

ABSTRACT. We prove that the pair correlation of the sequence of rational numbers in the unit interval with prime denominators is Poissonian. The result is also true for rational numbers with denominators having exactly two distinct prime factors.

1. INTRODUCTION

After the appearance of the classical result of Dirichlet on rational approximation of real numbers, there has been interest in approximating irrationals by rationals satisfying various constraints. One such problem that has attracted a lot of attention is concerned with finding numbers $\tau > 0$ such that there are infinitely many primes p satisfying the diophantine inequality

$$\|\alpha p\| < p^{-\tau+\varepsilon}, \tag{1}$$

where $\|t\|$ denotes the distance from a real number t to the nearest integer. Vinogradov initially obtained $\tau = \frac{1}{5}$, and his result has been improved through the years by various authors such as Vaughan ([13]), Balog ([2]), Harman ([4],[6]), Jia ([8]-[11]) and Heath-Brown and Jia ([7]). The current record $\tau = \frac{1}{3}$ has been recently announced by Mikawa (see [11] for a detailed account of the history and recent progress).

2000 *Mathematics Subject Classification.* 11K06, 11J71.
Key words and phrases. pair correlation, prime denominators, Poisson distribution.
⋆The second author was supported by NSF grant number DMS-0456615, and by CNCSIS grant GR106/2007, code 1116, of the Romanian Ministry of Education and Research.

Since the inequality (1) is equivalent to the existence of a rational number $\frac{a}{p}$ with prime denominator p such that $|\alpha - \frac{a}{p}| < p^{-1-\tau+\varepsilon}$, all the above results on τ can be reinterpreted as quantitative statements on the gaps between rational numbers with prime denominators. Motivated by this connection, in this paper we would like to investigate some aspects of the statistical behavior of these gaps. More precisely, for each large integer Q, let

$$\mathscr{M}_Q = \left\{ \frac{a}{p} : 1 \leq a < p \leq Q, p \text{ is a prime} \right\}$$

be the set of rational numbers with prime denominators bounded by Q in the unit interval. As a first step in trying to understand the gaps between the elements of $\mathscr{M}_Q$, in the present paper we study the pair correlation of the sequence $\mathscr{M}_Q$ as Q goes to infinity. We remark that in the case of all fractions, that is, in case one drops the requirement on the denominators to be prime, the existence and explicit computation of the limiting pair correlation function were established in [3]. In that case one has a strong repulsion between elements of the sequence, stronger even than the repulsion between the zeros of the Riemann zeta function [12]. There is no such repulsion when the fractions are required to have prime denominators, in fact in this case the pair correlation becomes Poissonian, i.e., the limiting pair correlation function exists and is constant, equal to 1.

Theorem 1. *The limiting pair correlation function of the sequence* $(\mathscr{M}_Q)_{Q\in\mathbb{N}}$ *as* $Q \to \infty$ *exists and is constant equal to* 1.

There has also been interest in diophantine approximation problems with numbers having exactly two prime factors (see for example [5], pp. 23-37 and [1], Theorem 1). With this in mind, we also consider the pair correlation of the sequence

$$\mathscr{N}_Q = \left\{ \frac{a}{q} : 1 \leq a \leq q \leq Q, \gcd(a,q) = 1, q = p_1 p_2, p_1 \text{ and } p_2 \text{ distinct primes} \right\}$$

as Q tends to infinity. We prove that, as in the case above, the pair correlation becomes Poissonian as $Q \to \infty$.

Theorem 2. *The limiting pair correlation function of the sequence* $(\mathscr{N}_Q)_{Q\in\mathbb{N}}$ *as* $Q \to \infty$ *exists and is constant equal to* 1*.*

2. Preliminaries

First let us review the definition of pair correlation. Let $\mathcal{F}$ be a finite set of cardinality N in $[0,1]$. The pair correlation measure $\mathcal{R}_{\mathcal{F}}(\mathbf{I})$ of an interval $\mathbf{I} \subset \mathbb{R}$ is defined as

$$\frac{1}{N}\#\left\{(x,y) \in \mathcal{F}^2 : x \neq y, x - y \in \frac{1}{N}\mathbf{I} + \mathbb{Z}\right\}.$$

Suppose that $(\mathcal{F}_n)_n$ is an increasing sequence of finite subsets of $[0,1]$ and that

$$\mathcal{R}(\mathbf{I}) = \lim_n \mathcal{R}_{\mathcal{F}_n}(\mathbf{I})$$

exists for every interval $\mathbf{I} \subset \mathbb{R}$. Then $\mathcal{R}$ is called the limiting pair correlation measure of $(\mathcal{F}_n)_n$. If the measure $\mathcal{R}$ is absolutely continuous with respect to the Lebesgue measure, say

$$\mathcal{R}(\mathbf{I}) = \int_{\mathbf{I}} g(x)\mathrm{d}x\,,$$

then g is called the limiting pair correlation function of $(\mathcal{F}_n)_n$. We denote

$$\mathcal{R}_F(\lambda) = 2^{-1}\mathcal{R}_{\mathcal{F}}([-\lambda,\lambda])\,.$$

Next, we collect several simple results which will be used later. Throughout the paper, p, q, p', q' stand for prime numbers.

Lemma 1. *For any integer* $n \geq 0$ *one has*

$$\sum_{p\leq Q} p^n = \frac{Q^{n+1}}{(n+1)\log Q}\left(1 + O\left((\log Q)^{-1}\right)\right),$$

as $Q \to \infty$*.*

Proof. This follows by partial summation from the prime number theorem. ∎

Lemma 2. *For any integer $n \geq 0$ one has*

$$\sum_{pq \leq Q} p^n q^n = \frac{2Q^{n+1} \log\log Q}{(n+1)\log Q} \left(1 + O\left((\log\log Q)^{-1}\right)\right),$$

as $Q \to \infty$.

Proof. This can be proved by combining Dirichlet's hyperbola method, Lemma 1 and Mertens' estimates.

The following result can be proved by simple Riemann integration.

Lemma 3. *Let $H : \mathbb{R} \to \mathbb{R}$ be a continuously differentiable function with* supp $H \subset (a, b)$ *for some real numbers a, b. Then for any $L > 0$ one has*

$$\sum_{l \in \mathbb{Z}} H\left(\frac{l}{L}\right) = L \int_{\mathbb{R}} H(x)\mathrm{d}x + O\left(\|DH\|_\infty \left(b - a + \frac{2}{L}\right)\right),$$

where

$$\|DH\|_\infty = \sup_{x \in \mathbb{R}} |H'(x)|.$$

3. Proof of Theorem 1

Our objective is to estimate, for any positive real number $\wedge$, the quantity

$$M_Q(\wedge) := \#\left\{(x, y) \in \mathscr{M}_Q{}^2 : x \neq y, x - y \in \frac{(0, \wedge)}{N} + \mathbb{Z}\right\},$$

as $Q \to \infty$. Let $M(Q) = \#(\mathscr{M}_Q)$. By Lemma 1 one has

$$\begin{aligned} M(Q) &= \sum_{p \leq Q}(p-1) = \sum_{p \leq Q} p - \sum_{p \leq Q} 1 \\ &= \frac{Q^2}{2\log Q}\left(1 + O\left((\log Q)^{-1}\right)\right). \end{aligned}$$

In the process of proving Theorem 1, we present a more general result.

Lemma 4. *For any function $H \in C_0^1(\mathbb{R})$ (continuously differentiable with compact support), define*

$$h(y) = \sum_{n\in\mathbb{Z}} H(M(Q)(y+n))$$

and

$$M_{Q,H} = \sum_{x,y\in\mathscr{M}_Q} h(x-y).$$

Then

$$M_{Q,H} \quad = \quad \frac{Q^2}{2\log Q}\int_{\mathbb{R}} H(x)\,\mathrm{d}x + O_H\left(\frac{Q^2}{(\log Q)^2}\right).$$

Note that assuming Lemma 4 and using the fact that the error term is $\ll Q^2/(\log Q)^2$, one has

$$\lim_{Q\to\infty}\frac{M_{Q,H}}{M(Q)} = \int_{\mathbb{R}} H(x)\,\mathrm{d}x.$$

Letting the smooth function H approach the characteristic function of the interval $(0,\wedge)$, by a standard approximation argument, we see that the pair correlation function of the sequence $\mathscr{M}_Q$ as $Q\to\infty$ is constant equal to 1. Thus in order to complete the proof of Theorem 1, it remains to prove Lemma 4.

Proof of Lemma 4. Let $e(y) = \exp(2\pi i y)$ for any $y\in\mathbb{R}$. First notice that for any integer r one has

$$\sum_{x\in\mathscr{M}_Q} e(rx) = \sum_{p\le Q}\sum_{a=1}^{p-1} e(ra/p) = \sum_{p\le Q, p|r} p - \pi(Q),$$

where $\pi(Q)$ is the number of primes in the interval $[1,Q]$. Next, the coefficients in the Fourier series

$$h(y) = \sum_{m\in\mathbb{Z}} c_m e(my)$$

of h are given by

$$\begin{aligned} c_m &= \int_0^1 h(y)e(-my)\mathrm{d}y = \sum_{n\in\mathbb{Z}} \int_0^1 H(M(Q)(y+n))e(-my)\mathrm{d}y \\ &= \sum_{n\in\mathbb{Z}} \int_n^{n+1} e(-mu)H(M(Q)u)\mathrm{d}u \\ &= \int_{\mathbb{R}} e(-mu)H(M(Q)u)\mathrm{d}u = \frac{1}{M(Q)}\widehat{H}\left(\frac{m}{M(Q)}\right), \end{aligned}$$

where $\widehat{H}$ is the Fourier transform of H defined by

$$\widehat{H}(x) = \int_{\mathbb{R}} H(y)e(-xy)\mathrm{d}y, \quad x \in \mathbb{R}.$$

One has

$$\begin{aligned} M_{Q,H} &= \sum_{x,y\in\mathscr{M}_Q} \sum_{m\in\mathbb{Z}} c_m e(m(x-y)) = \sum_m c_m \left| \sum_{x\in\mathscr{M}_Q} e(mx) \right|^2 \\ &= \sum_m c_m \left(\sum_{p,q\le Q, p|m, q|m} pq + \pi(Q)^2 - 2\pi(Q) \sum_{p\le Q, p|m} p \right) \\ &= A + \pi(Q)^2 \cdot B - 2\pi(Q) \cdot C, \end{aligned}$$

where

$$\begin{aligned} A &= \sum_{m\in\mathbb{Z}} c_m \sum_{p,q\le Q, p|m, q|m} pq = \sum_{p,q\le Q} pq \sum_{m\in\mathbb{Z}} C_{[p,q]m}, \\ B &= \sum_{m\in\mathbb{Z}} C_m, \quad C = \sum_{m\in\mathbb{Z}} c_m \sum_{p\le Q, p|m} p = \sum_{p\le Q} p \sum_{m\in\mathbb{Z}} C_{pm}\,. \end{aligned}$$

Consider for each $d > 0$ the function

$$H_d(x) = \frac{1}{d} H\left(\frac{M(Q)x}{d}\right), \quad x \in \mathbb{R}.$$

Using the Fourier transform and an appropriate change of variable we obtain that, for any $m \in \mathbb{Z}$,

$$\begin{aligned} \widehat{H}_d(m) &= \int_{\mathbb{R}} H_d(t) e(-mt)\,\mathrm{d}t = \int_{\mathbb{R}} \frac{1}{d} H\left(\frac{M(Q)t}{d}\right) e(-mt)\,\mathrm{d}t \\ &= \int_{\mathbb{R}} H(M(Q)t') e(-mdt')\,\mathrm{d}t' = c_{dm}. \end{aligned}$$

Employing Poisson's summation formula we find that

$$\sum_{m\in\mathbb{Z}} c_{[p,q]m} = \sum_{m\in\mathbb{Z}} \widehat{H}_{[p,q]}(m) = \sum_{m\in\mathbb{Z}} H_{[p,q]}(m) = \sum_{m\in\mathbb{Z}} \frac{1}{[p,q]} H\left(\frac{mM(Q)}{[p,q]}\right),$$

and similarly

$$\sum_{m\in\mathbb{Z}} c_{pm} = \sum_{m\in\mathbb{Z}} \frac{1}{p} H\left(\frac{mM(Q)}{p}\right), \quad \sum_{m\in\mathbb{Z}} c_m = \sum_{m\in\mathbb{Z}} H\left(mM(Q)\right).$$

We may suppose that supp $H \subset (0, \wedge)$ for some $\wedge > 0$. Since $p, q \leq Q$ and $M(Q) \asymp \frac{Q^2}{\log Q}$, if $m \neq 0$, then $|mM(Q)| \gg Q$ and

$$\left|\frac{mM(Q)}{p}\right| \gg \frac{Q}{\log Q},$$

for sufficiently large Q one has

$$H\left(\frac{mM(Q)}{p}\right) = H\left(mM(Q)\right) = 0,$$

thus $\sum_{m\in\mathbb{Z}} c_{pm} = \sum_{m\in\mathbb{Z}} c_m = 0$ and $B = C = 0$. Therefore

$$M_{Q,H} = A = \sum_{p,q\leq Q} \frac{pq}{[p,q]} \sum_{m\in\mathbb{Z}} H\left(\frac{mM(Q)}{[p,q]}\right).$$

In the above expression, for primes $p, q \leq Q$ to have non-trivial contribution, one must have $[p, q] > M(Q)/\wedge \gg Q^2/\log Q$, hence $p \neq q$ and

$$M_{Q,H} = \sum_{\substack{p,q\leq Q,\\ p\neq q,\\ pq>M(Q)/\wedge}} \sum_{m\in\mathbb{Z}} H\left(\frac{mM(Q)}{pq}\right).$$

For such p and q, by applying Lemma 3 we see that

$$\sum_{m\in\mathbb{Z}} H\left(\frac{mM(Q)}{pq}\right) = \frac{pq}{M(Q)} \int_{\mathbb{R}} H(x)\mathrm{d}x + O_H(1).$$

Therefore

$$M_{Q,H} = \frac{\int_{\mathbb{R}} H(x)\mathrm{d}x}{M(Q)} \sum_{\substack{p,q\leq Q,\\ p\neq q,\\ pq>M(Q)/\wedge}} pq + O\left(\sum_{p,q\leq Q} 1\right).$$

By the prime number theorem,

$$\sum_{\substack{p,q\leq Q,\\ p\neq q,\\ pq\leq M(Q)/\wedge}} pq \leq \frac{M(Q)}{\wedge} \sum_{p,q\leq Q} 1 \ll_H M(Q)\frac{Q^2}{(\log Q)^2},$$

hence

$$M_{Q,H} = \frac{\int_{\mathbb{R}} H(x)\mathrm{d}x}{M(Q)} \sum_{\substack{p,q\leq Q,\\ p\neq q,}} pq + O_H\left(\frac{Q^2}{(\log Q)^2}\right).$$

Using Lemma 1,

$$\begin{aligned} \sum_{\substack{p,q\leq Q,\\ p\neq q,}} pq &= \left(\sum_{p\leq Q} p\right)^2 - \sum_{p\leq Q} p^2 = \left(\frac{Q^2}{2\log Q}\left(1 + O\left((\log Q)^{-1}\right)\right)\right)^2 - O\left(\frac{Q^3}{\log Q}\right) \\ &= \frac{Q^4}{4(\log Q)^2}\left(1 + O\left((\log Q)^{-1}\right)\right). \end{aligned}$$

Lastly, taking into account that

$$M(Q) = \frac{Q^2}{2\log Q}\left(1 + O\left((\log Q)^{-1}\right)\right),$$

one concludes that

$$M_{Q,H} = \frac{Q^2}{2\log Q}\int_{\mathbb{R}} H(x)\mathrm{d}x + O_H\left(\frac{Q^2}{(\log Q)^2}\right).$$

This completes the proof of Lemma 4. ∎

4. Proof of Theorem 2

Our objective is to estimate, for any positive real number $\wedge$, the quantity

$$N_Q(\wedge) := \#\left\{(x,y) \in \mathscr{N}_Q^{\,2} : x \neq y, x - y \in \frac{(0,\wedge)}{N} + \mathbb{Z}\right\},$$

as $Q \to \infty$. Letting $N(Q) = \#(\mathscr{N}_Q)$, one has

$$2N(Q) \quad = \quad \sum_{pq \leq Q, p\neq q} (p-1)(q-1) = \sum_{pq\leq Q} pq + O\left(\sum_{pq\leq Q} p + \sum_{p\leq\sqrt{Q}} p^2\right).$$

Applying Lemma 1 and Lemma 2 one sees that the big-O term above is

$$\ll \sum_{p\leq Q} p\frac{Q}{p} + \frac{Q^{3/2}}{\log Q} \ll \frac{Q^2}{\log Q},$$

and the main term is $\sum_{pq\leq Q} pq = \frac{Q^2 \log\log Q}{\log Q}\left(1 + O((\log\log Q)^{-1})\right)$. Therefore

$$N(Q) = \frac{Q^2\log\log Q}{2\log Q}\left(1 + O((\log\log Q)^{-1})\right).$$

Again, in the process of establishing Theorem 2, we present a more general result.

Lemma 5. *For any function $H \in C_0^1(\mathbb{R})$, define*

$$h(y) = \sum_{n\in\mathbb{Z}} H(N(Q)(y+n))$$

and

$$N_{Q,H} = \sum_{x,y\in\mathscr{N}_Q} h(x-y).$$

Then

$$N_{Q,H} \;=\; \frac{Q^2 \log\log Q}{2\log Q}\int_{\mathbb{R}} H(x)\,\mathrm{d}x + O_H\left(\frac{Q^2}{\log Q}\right).$$

If we assume Lemma 5 and use the fact that the error term is $\ll Q^2/\log Q$, we obtain

$$\lim_{Q\to\infty}\frac{N_{Q,H}}{N(Q)} = \int_{\mathbb{R}} H(x)\,\mathrm{d}x.$$

Then, letting the smooth function H approach the characteristic function of the interval $(0,\wedge)$, an approximation argument will show that the pair correlation function of the sets $\mathscr{N}_Q$ as $Q\to\infty$ is constant equal to 1. It remains to prove Lemma 5.

Proof of Lemma 5. Notice that for any integer r one has

$$\sum_{x\in\mathscr{N}_Q} e(rx) = \sum_{\substack{pq\le Q,\\ p<q}}\ \sum_{\substack{a=1\\ \gcd(a,pq)=1}}^{pq} e(ra/(pq)) = \sum_{\substack{pq\le Q,\\ p<q}}\sum_{a=1}^{pq} e(ra/(pq))\sum_{d|a,d|pq}\mu(d),$$

where μ is the Möbius function. Since p,q are distinct primes, the sum can be rewritten as

$$\begin{aligned}
\sum_{x\in\mathscr{N}_Q} e(rx) \;&=\; \sum_{\substack{pq\le Q,\\ p<q}}\left(\sum_{a=1}^{pq} e\left(ra/(pq)\right) - \sum_{a=1}^{p} e\left(ra/p\right) - \sum_{a=1}^{q} e\left(ra/q\right) + 1\right)\\
&=\; \sum_{\substack{pq\le Q,p<q\\ pq|r}} pq - \sum_{\substack{pq\le Q,p<q\\ p|r}} p - \sum_{\substack{pq\le Q,p<q\\ q|r}} q + \sum_{pq\le Q,p<q} 1\\
&=\; \sum_{\substack{pq\le Q,p<q\\ pq|r}} pq - \sum_{\substack{pq\le Q,p\ne q\\ p|r}} p + \sum_{pq\le Q,p<q} 1\,.
\end{aligned}$$

Suppose the Fourier series expansion of h is $h(y) = \sum_{m\in\mathbb{Z}} c_m e(my)$. Then one has

$$\begin{aligned}
N_{Q,H} &= \sum_{x,y\in\mathcal{N}_Q} \sum_{m\in\mathbb{Z}} c_m e(m(x-y)) = \sum_m c_m \left| \sum_{x\in\mathcal{N}_Q} e(mx) \right|^2 \\
&= \sum_m c_m \left(\sum_{\substack{pq\le Q, p<q \\ pq|r}} pq - \sum_{\substack{pq\le Q, p\ne q \\ p|r}} p + \sum_{pq\le Q, p<q} 1 \right)^2 \\
&= I + E,
\end{aligned}$$

where

$$I = \sum_m c_m \left(\sum_{\substack{pq\le Q, p<q \\ pq|r}} pq \right)^2 = \sum_{\substack{pq, p'q'\le Q, \\ p<q, p'<q'}} pqp'q' \sum_{m\in\mathbb{Z}} c_{[pq,p'q']m},$$

and

$$E = \sum_m c_m \left(2 \sum_{\substack{pq\le Q, p<q \\ pq|r}} pq \left(- \sum_{\substack{pq\le Q, p\ne q \\ p|r}} p + \sum_{pq\le Q, p<q} 1 \right) + \left(- \sum_{\substack{pq\le Q, p\ne q \\ p|r}} p + \sum_{pq\le Q, p<q} 1 \right)^2 \right).$$

We will see below that $E \ll Q^2/\log Q$. Let us analyze I first. As in the proof of Lemma 4,

$$\sum_{m\in\mathbb{Z}} c_{[pq,p'q']m} = \frac{1}{[pq,p'q']} H\left(\frac{N(Q)m}{[pq,p'q']} \right),$$

hence

$$\begin{aligned}
I &= \sum_{\substack{pq, p'q'\le Q \\ p<q, p'<q'}} \gcd(pq, p'q') \sum_{m\in\mathbb{Z}} H\left(\frac{N(Q)m}{[pq,p'q']} \right) \\
&= \sum_{\substack{pq, p'q'\le Q \\ p<q, p'<q' \\ \gcd(pq,p'q')=1}} \sum_{m\in\mathbb{Z}} H\left(\frac{N(Q)m}{pqp'q'} \right) + \sum_{\substack{pq, p'q'\le Q \\ p<q, p'<q' \\ \gcd(pq,p'q')>1}} \gcd(pq, p'q') \sum_{m\in\mathbb{Z}} H\left(\frac{N(Q)m}{[pq,p'q']} \right).
\end{aligned}$$

Suppose supp $H \subset (0, \wedge)$, then

$$0 < \frac{N(Q)m}{[pq, p'q']} < \wedge \Longrightarrow 0 < m < \frac{\wedge[pq, p'q']}{N(Q)}.$$

It follows that the second term in the above expression of I is

$$\begin{aligned}
&\ll_H \sum_{\substack{pq, p'q' \le Q \\ p<q, p'<q' \\ \gcd(pq, p'q')>1}} \gcd(pq, p'q') \frac{\wedge[pq, p'q']}{N(Q)} \ll_H \frac{1}{N(Q)} \sum_{\substack{pq, p'q' \le Q \\ p<q, p'<q' \\ \gcd(pq, p'q')>1}} pqp'q' \\
&\ll \frac{1}{N(Q)} \sum_{pq, pq' \le Q} p^2 qq' = \frac{1}{N(Q)} \left(\sum_{p \le \sqrt{Q}} p^2 \left(\sum_{q \le Q/p} q \right)^2 + \sum_{\sqrt{Q} < p \le Q} p^2 \left(\sum_{q \le Q/p} q \right)^2 \right) \\
&\ll \frac{\log Q}{Q^2 \log\log Q} \left(\sum_{p \le \sqrt{Q}} p^2 \left(\frac{(Q/p)^2}{\log(Q/p)} \right)^2 + \sum_{\sqrt{Q} < p \le Q} p^2 \left((Q/p)^2 \right)^2 \right) \\
&\ll \frac{\log Q}{Q^2 \log\log Q} \left(\frac{Q^4}{(\log Q)^2} \sum_p p^{-2} + Q^4 \sum_{p > \sqrt{Q}} p^{-2} \right) \ll \frac{Q^2}{\log Q}.
\end{aligned}$$

The first term in the expression of I is, by using Lemma 3,

$$I' = \sum_{\substack{pq, p'q' \le Q \\ p<q, p'<q' \\ \gcd(pq, p'q')=1 \\ pqp'q' > N(Q)/\wedge}} \sum_{m \in \mathbb{Z}} H\left(\frac{N(Q)m}{pqp'q'} \right) = \sum_{\substack{pq, p'q' \le Q \\ p<q, p'<q' \\ \gcd(pq, p'q')=1 \\ pqp'q' > N(Q)/\wedge}} \left(\frac{pqp'q'}{N(Q)} \int_{\mathbb{R}} H(x) \mathrm{d}x + O_H(1) \right).$$

By using Lemma 2, one has

$$\sum_{\substack{pq, p'q' \le Q \\ p<q, p'<q' \\ \gcd(pq, p'q')=1 \\ pqp'q' > N(Q)/\wedge}} 1 \le \left(\sum_{pq \le Q} 1 \right)^2 \ll \left(\frac{Q \log\log Q}{\log Q} \right)^2 \ll \frac{Q^2}{\log Q},$$

and

$$\sum_{\substack{pq,p'q'\le Q\\ p<q,p'<q'\\ \gcd(pq,p'q')=1\\ pqp'q'\le N(Q)/\wedge}} \frac{pqp'q'}{N(Q)} \ll_H \sum_{\substack{pq,p'q'\le Q\\ p<q,p'<q'}} 1 \ll \frac{Q^2}{\log Q}.$$

Hence

$$I' = \int_{\mathbb{R}} H(x)\mathrm{d}x \sum_{\substack{pq,p'q'\le Q\\ p<q,p'<q'\\ \gcd(pq,p'q')=1}} \frac{pqp'q'}{N(Q)} + O_H\left(\frac{Q^2}{\log Q}\right).$$

On the other hand, by the previous arguments one has

$$\sum_{\substack{pq,p'q'\le Q\\ p<q,p'<q'\\ \gcd(pq,p'q')>1}} \frac{pqp'q'}{N(Q)} \ll \frac{1}{N(Q)} \sum_{\substack{pq,pq'\le Q\\ p<q,p'<q'\\ \gcd(pq,p'q')>1}} pqp'q' \ll \frac{Q^2}{\log Q},$$

and

$$\sum_{\substack{pq,p'q'\le Q\\ p=q}} \frac{pqp'q'}{N(Q)} = \frac{1}{N(Q)} \sum_{p\le\sqrt{Q}} p^2 \sum_{p'q'\le Q} p'q' \ll \frac{Q^2}{\log Q}.$$

Therefore one concludes that

$$\begin{aligned} I &= \int_{\mathbb{R}} H(x)\mathrm{d}x \left(\sum_{\substack{pq,p'q'\le Q\\ p<q,p'<q'}} \frac{pqp'q'}{N(Q)} + O\left(\frac{Q^2}{\log Q}\right)\right) + O_H\left(\frac{Q^2}{\log Q}\right)\\ &= \int_{\mathbb{R}} H(x)\mathrm{d}x \left(\frac{1}{4}\sum_{pq,p'q'\le Q} \frac{pqp'q'}{N(Q)} + O\left(\frac{Q^2}{\log Q}\right)\right) + O_H\left(\frac{Q^2}{\log Q}\right)\\ &= \frac{Q^2\log\log Q}{2\log Q}\int_{\mathbb{R}} H(x)\mathrm{d}x + O_H\left(\frac{Q^2}{\log Q}\right). \end{aligned}$$

Next, we need to show that $E \ll Q^2/\log Q$. One of the terms in E is

$$\begin{aligned} E' &= \sum_{m} c_m \sum_{\substack{pq\le Q, p<q \\ pq|r}} pq \sum_{\substack{pq\le Q, p\ne q \\ p|r}} p = \sum_{\substack{pq, p'q'\le Q, \\ p<q, p'\ne q'}} pqp' \sum_{m\in\mathbb{Z}} c_{[pq,p']m} \\ &= \sum_{\substack{pq, p'q'\le Q, \\ p<q, p'\ne q'}} \frac{pqp'}{[pq,p']} \sum_{m\in\mathbb{Z}} H\left(\frac{N(Q)m}{[pq,p']}\right) \ll_H \sum_{\substack{pq, p'q'\le Q, \\ p<q, p'\ne q'}} \frac{pqp'}{[pq,p']} \cdot \frac{\wedge[pq,p']}{N(Q)} \\ &\ll_H \sum_{pq\le Q} \frac{pq}{N(Q)} \cdot \sum_{p'\le Q} p' \sum_{q'\le Q/p'} 1 \ll \sum_{p'\le Q} p' \frac{Q}{p'} \ll \frac{Q^2}{\log Q}. \end{aligned}$$

The others terms in E can be treated in a similar way. This completes the proof of Lemma 5 and also the proof of Theorem 2. ∎

References

[1] E. Alkan, G. Harman, A. Zaharescu, *Diophantine approximation with mild divisibility constraints*, J. Number Theory 118 (2006), 1–14.

[2] A. Balog, *A remark on the distribution of αp modulo one*, Analytic and elementary number theory (Marseille, 1983), 6–24, Publ. Math. Orsay, 86-1, Univ. Paris XI, Orsay, 1986.

[3] F.P. Boca, A. Zaharescu, *The correlations of Farey fractions*, J. London Math. Soc. (2) 72 (2005), no. 1, 25–39.

[4] G. Harman, *On the distribution of αp modulo one*, J. London Math. Soc. (2) 27 (1983), no. 1, 9–18.

[5] G. Harman, *Trigonometric sums over primes. II*, Glasg. Math. J. 24(1) (1983), 23–37.

[6] G. Harman, *On the distribution of αp modulo one. II*, Proc. London Math. Soc. (3) 72 (1996), no. 2, 241–260.

[7] D.R. Heath-Brown, C. Jia, *The distribution of αp modulo one*, Proc. London Math. Soc. (3) 84 (2002), no. 1, 79–104.

[8] C. Jia, *On the distribution of αp modulo one*, Lecture notes in contemporary mathematics, 1989, 39–48, Science Press, Beijing, 1990.

[9] C. Jia, *On the distribution of αp modulo one*, J. Number Theory 45 (1993), no. 3, 241–253.

[10] C. Jia, *On the distribution of αp modulo one. II*, Sci. China Ser. A 43 (2000), no. 7, 703–721.

[11] C. Jia, *On the distribution of αp modulo one*, Number theoretic methods (Iizuka, 2001), 151–157, Dev. Math., 8, Kluwer Acad. Publ., Dordrecht, 2002.

[12] H. Montgomery, *The pair correlation of zeros of the zeta function*, Proc. Sym. Pure Math. 24, AMS, 181–193, 1973.

[13] R.C. Vaughan, *On the distribution of αp modulo* 1, Mathematika 24 (1977), no. 2, 135–141.

Maosheng Xiong: Department of Mathematics, Eberly College of Science, Pennsylvania State University, McAllister Building, University Park, PA 16802 USA

E-mail address: xiong@math.psu.edu

Alexandru Zaharescu: Institute of Mathematics of the Romanian Academy, P.O. Box 1–764, 70700 Bucharest, Romania, and Department of Mathematics, University of Illinois at Urbana-Champaign, 273 Altgeld Hall, MC-382, 1409 W. Green Street, Urbana, IL 61801 USA

E-mail address: zaharesc@math.uiuc.edu