DISTRIBUTION OF $\alpha n + \beta$ MODULO 1 OVER INTEGERS FREE FROM LARGE AND SMALL PRIMES

KAM HUNG YAU

Abstract. For any $\varepsilon > 0$, we obtain an asymptotic formula for the number of solutions $n \leq x$ to

$$\|\alpha n + \beta\| < x^{-\frac{1}{4} + \varepsilon}$$

where $n$ is $[y, z]$-smooth for infinitely many real number $x$. In addition, we also establish an asymptotic formula with an additional square-free condition on $n$. Moreover, if $\alpha$ is quadratic irrational then the asymptotic formulas holds for all sufficiently large $x$.

Our ingredients come from the Harman sieve which we adapt suitably to sieve for $[y, z]$-smooth numbers. The arithmetic information comes from estimates for exponential sums.

1. INTRODUCTION

A classical Theorem of Dirichlet [6, Theorem 185] states that if $\alpha$ is irrational then there exist infinitely many pairs of integers $(m, n)$ satisfying the inequality

$$|\alpha - \frac{m}{n}| < \frac{1}{n^2}.$$  \hfill (1.1)

Such problems fall in the area of Diophantine approximation. Equivalently by defining $\|x\| = \min_{v \in \mathbb{Z}} |x - v|$ to be the distance of the nearest integer of $x$, (1.1) implies

$$\|\alpha n\| < n^{-1}$$

for infinitely many positive integers $n$. A natural extension to this problem is to consider if there exist infinitely many solutions to

$$\|\alpha n + \beta\| < n^{-\kappa}. $$  \hfill (1.2)

Here $\beta \in \mathbb{R}$, $\kappa > 0$, $n \in N \subseteq \mathbb{N}$ and $N$ is some set of arithmetic interest.

Various results have been obtained when $N$ is the set of prime numbers. These include Harman [7, 9], Heath-Brown & Jia [12] and Jia [13]. Such results involve sieve method and bounds of exponential sums, while Vaughan [18] obtained his result by applying what is now known as the Vaughan Identity, see [19], together with bounds for exponential sums.

2010 Mathematics Subject Classification. 11K60, 11L07, 11N36.

Key words and phrases. Diophantine approximation, exponential sum, sieve method.
We remark that Harman’s sieve method first appeared when Harman studied this exact problem [7] to get the exponent $\kappa = 3/10$. For more details on the Harman sieve, see the monograph [10]. The best result to date is by Matomäki [14] with $\kappa < 1/3$ under the condition $\beta = 0$ and employs Harman’s sieve method where the arithmetic information comes from bounds for averages of Kloosterman sums.

Let $k > 1$ be a fixed positive integer and $\mathcal{N}$ the set of $k$-th power of primes. Baker & Harman [2] showed that we can take $\kappa < 3/20$ if $k = 2$ and $\kappa < (3 \cdot 2^{k-1})^{-1}$ if $k \geq 3$. In particular, when $k = 2$ it improves a result of Ghosh [5]. Later, Wong [20] provided an improvement of Baker & Harman [2] in the range $3 \leq k \leq 12$.

When $\mathcal{N}$ is the set of square-free numbers, the best result is due to Heath-Brown [11] with $\kappa < 2/3$ using an essentially elementary method. The result improves the previous work by Harman [8], Balog & Perelli [3], who both independently showed we can essentially take the exponent $\kappa < 1/2$.

In this paper, we consider the problem of establishing an asymptotic formula for the number of $[y, z]$-smooth $n \leq x$ solutions to (1.2) (numbers with prime factors in the interval $[y, z]$) with $\kappa = 1/4 - \varepsilon$, where $\varepsilon > 0$. We also consider a hybrid problem that interpolates between square-free and $[y, z]$-smooth integers.

We note that in the special case of smooth numbers we are able to obtain a non-trivial lower bound immediately. Indeed, for any fixed $\varepsilon > 0$, consider the set

$$\{n = ab \leq x : p|ab \implies p < x^{\varepsilon}, x^{\frac{1}{3}} \leq a < 2x^{\frac{1}{3}}\}.$$  

By applying Lemma 4.1 and our Type II information (Lemma 4.3 with (4.1)), we can show immediately for infinitely many $x$ chosen correctly there are at least $x^{2/3+\varepsilon-o(1)}$ integers up to $x$ which are $x^{\varepsilon}$-smooth and satisfies

$$\|\alpha n + \beta\| < x^{-\frac{1}{4}+\varepsilon}.$$  

The statement remains valid if we additionally impose these integers $n$ to be square-free.

2. Notation

For complex valued functions $f$ and $g$, we use the notation $f = O(g)$ and $f \ll g$ to mean there exist an absolute constant $C > 0$ such that $|f(x)| \leq C|g(x)|$ for all sufficiently large $x$.

Denote $m \sim M$ as $M \leq m < 2M$, $e(x) = \exp(2\pi ix)$ and $\mu$ to be the Möbius function. Given a positive integer $n \geq 2$, we denote $P^-(n)$ and
$P^+(n)$ to be the smallest and largest prime factor of $n$ respectively. We always refer $p$ and their subscripts to be a prime.

3. Main Results

For any positive real numbers $y \leq z \leq x$, we denote

$$S(x; y, z) = \{a \in [1, x] \cap \mathbb{N} : p|a \implies p \in [y, z]\}$$

as the set of all integers in $[1, x]$ such that all of whose prime factors are contained in $[y, z]$. Moreover, we denote the cardinality of $S(x; y, z)$ by $\Psi(x; y, z)$.

For any positive integer $A > 0$, we write

$$I(A) = \{\alpha \in \mathbb{R}\setminus\mathbb{Q} : \alpha = [a_0; a_1, \ldots], |a_j| \leq A, j \geq 0\}.$$

We note that $\bigcup_{A \in \mathbb{N}} I(A)$ contains the set of all quadratic irrationals. We can now state our first result.

**Theorem 3.1.** There exists an increasing sequence of positive integers $(x_k)_{k \in \mathbb{N}}$ such that if $2 \leq y < x_k^{1/2}$, $y < z \leq x_k$, $\varepsilon > 0$ and $\delta = x_k^{-1/4+\varepsilon}$, then we have

$$\sum_{n \in S(x_k; y, z)} \|\alpha n + \beta\| < \delta^3 = 2\delta^2 \frac{x_k^{3/4+\varepsilon} + o(1))}.$$

Moreover, (3.1) holds for the sequence $(x_k)_{k \in \mathbb{N}} = (k)_{k \in \mathbb{N}}$ uniformly for all $\alpha \in I(A)$ and any fixed positive integer $A$.

For any positive real numbers $y \leq z \leq x$, we denote

$$S^*(x; y, z) = \{a \in [1, x] \cap \mathbb{N} : \mu^2(a) = 1, p|a \implies p \in [y, z]\}$$

as the set of all square-free integers in $[1, x]$ such that all of whose prime factors are contained in $[y, z]$. We also denote the cardinality of $S^*(x; y, z)$ by $\Psi^*(x; y, z)$. The next theorem is essentially Theorem 3.1, where we also impose on the integers we are counting to be square-free.

**Theorem 3.2.** There exists an increasing sequence of positive integers $(x_k)_{k \in \mathbb{N}}$ such that if $2 \leq y < x_k^{1/2}$, $y < z \leq x_k$, $\varepsilon > 0$ and $\delta = x_k^{-1/4+\varepsilon}$, then we have

$$\sum_{n \in S^*(x_k; y, z)} \|\alpha n + \beta\| < \delta^3 = 2\delta^2 \frac{x_k^{3/4+\varepsilon} + o(1))}.$$

Moreover, (3.2) holds for the sequence $(x_k)_{k \in \mathbb{N}} = (k)_{k \in \mathbb{N}}$ uniformly for all $\alpha \in I(A)$ and any fixed positive integer $A$. 
If we fix $0 < u_2 < u_1$ with $2 < u_1$, $u_2 < \lfloor u_1 \rfloor$ and set $y = x_k^{1/u_1}$ and $z = x_k^{1/u_2}$ then both the quantity $\Psi(x_k; y, z)$, $\Psi^*(x_k; y, z)$ in the main term of (3.1) and (3.2) respectively are bounded below by the number of integers $n = p_1 \ldots p_j$ that are products of $j = \lfloor u_1 \rfloor$ distinct primes with $p_i \in [y, z]$. This gives the lower bound

$$x_k^{1-o(1)} < \Psi(x_k; y, z), \Psi^*(x_k; y, z).$$

It follows that our Theorem 3.1 and 3.2 are non-trivial in this region.

We note that by [4, Theorem 1] of Friedlander, we can obtain an asymptotic formula for $\Psi(x_k; y, z)$ in certain regions. We also mention that Saias has studied extensively this quantity in a series of three papers [15, 16, 17]. In particular $\Psi(x_k; y, z)$ in our notation is $\theta(x_k, z, y)$ or $\Theta(x_k, z, y)$ in the notation of Friedlander [4] and Saias [15, 16, 17] respectively.

We remark that we assume $y < x_k^{1/2}$ in Theorem 3.1 and 3.2 since if $y \geq x_k^{1/2}$ and $z \geq x_k$ then both $S(x_k; y, z)$ and $S^*(x_k; y, z)$ essentially count primes in the interval $[y, x_k]$ and the result of Harman [10, Theorem 3.2] covers this case with $\delta = x_k^{-1/4+\epsilon}$.

It is easy to see that the proof of Theorem 3.1 can be adjusted to prove Theorem 3.2, so we will only give full details for the proof of Theorem 3.1.

4. Preparations

For any $\delta > 0$, we define

$$\chi(r) = \begin{cases} 1 & \text{if } \|r\| < \delta, \\ 0 & \text{otherwise}. \end{cases}$$

We recall a Lemma from [11, Chapter 2], which provides us with a finite Fourier approximation to $\chi$. This converts the problem of detecting solutions to (1.2) to a problem about estimates for exponential sums.

**Lemma 4.1.** For any positive integer $L$, there exist complex sequences $(c_\ell^-)[\ell \leq L]$, $(c_\ell^+)[\ell \leq L]$ such that

$$2\delta - \frac{1}{L+1} + \sum_{0<|\ell| \leq L} c_\ell^- e(\ell \theta) \leq \chi(\theta) \leq 2\delta + \frac{1}{L+1} + \sum_{0<|\ell| \leq L} c_\ell^+ e(\ell \theta)$$

where

$$|c_\ell^\pm| \leq \min \left\{ 2\delta + \frac{1}{L+1}, \frac{3}{2|\ell|} \right\}.$$

Let $\alpha, x \in \mathbb{R}$ such that $|\alpha - a/q| < 1/q^2$ where $1 \leq a \leq q \ll x^{O(1)}$ and $(a, q) = 1$. Let the complex sequences $(a_m)_{m \in \mathbb{N}}$, $(b_n)_{n \in \mathbb{N}}$, $(c_\ell)_{\ell \in \mathbb{N}}$ satisfy
\[ |a_m| \leq \tau(m), \ |b_n| \leq \tau(n) \text{ and} \]

\[ |c_\ell| \leq \min \left\{ 2\delta + \frac{1}{\lfloor x \rfloor + 1}, \frac{3}{2\ell} \right\} \]

whenever \( m, n, \ell \in \mathbb{N} \). We will also use the classical divisor bound \( \tau(r) \ll x^{o(1)} \) for \( r \leq x \).

Lastly for \( \varepsilon > 0 \), we denote \( \delta = x^{-\frac{1}{2} + \varepsilon} \).

The next two Lemmas provide our Type I and II information respectively. They can be obtained by following the method of [10, Section 2.3].

**Lemma 4.2.** We have

\[ \sum_{\ell \leq x} c_\ell \sum_{\substack{mn \leq x \\atop m \sim M}} a_m e(\alpha \ell mn) \ll (M + xq^{-1} + \delta q)x^{o(1)}. \]

**Lemma 4.3.** We have

\[ (4.1) \quad \sum_{\ell \leq x} c_\ell \sum_{\substack{mn \leq x \\atop m \sim M}} a_m b_n e(\alpha \ell mn) \ll x^{1+o(1)} \left( \frac{\delta}{M} + \frac{1}{x} + \frac{q\delta}{q} \right)^{\frac{1}{2}} \]

for \( M \ll x^{1/2} \) and

\[ (4.2) \quad \sum_{\ell \leq x} c_\ell \sum_{\substack{mn \leq x \\atop m \sim M}} a_m b_n e(\alpha \ell mn) \ll x^{1+o(1)} \left( \frac{\delta}{M} + \frac{1}{x} + \frac{q\delta}{q} \right)^{\frac{1}{2}} \]

for \( M \gg x^{1/2} \).

4.1. **Estimates for Type I & II sum.** Denote

\[ \mathcal{B} = \{n \in \mathbb{N} : 2 \leq n \leq x\} \quad \text{and} \quad \mathcal{A} = \{n \in \mathcal{B} : \|\alpha n + \beta\| < \delta \}. \]

We will first state our Type I estimate.

**Lemma 4.4** (Type I estimate). For \( q = x^{2/3} \) we have

\[ \sum_{\substack{mn \in \mathcal{A} \\atop m \sim M}} a_m = 2\delta \sum_{\substack{mn \in \mathcal{B} \\atop m \sim M}} a_m + O(x^{\frac{3}{4} + \frac{3}{8}}) \]

whenever \( M \ll x^{3/4} \).

**Proof.** By Lemma 4.1 with \( L = \lfloor x \rfloor \), we get

\[ \sum_{\substack{mn \in \mathcal{A} \\atop m \sim M}} a_m = \sum_{\substack{mn \in \mathcal{B} \\atop m \sim M}} a_m \chi(\alpha mn + \beta) = 2\delta \sum_{\substack{mn \in \mathcal{B} \\atop m \sim M}} a_m + O(E_1 + E_2) \]

where

\[ E_1 = \frac{1}{L + 1} \sum_{\substack{mn \leq x \\atop m \sim M}} a_m \ll x^{o(1)}. \]
and
\[ E_2 = \sum_{0 < |\ell| \leq L} |c_{\ell}^\pm| \left| \sum_{mn \leq x \atop m \sim M} a_m e(\ell(\alpha mn + \beta)) \right|. \]

Clearly there exists \( \xi_\ell \in \mathbb{C} \) with \( |\xi_\ell| = 1 \) such that
\[ E_2 = \sum_{0 < |\ell| \leq L} \xi_\ell c_{\ell}^\pm \sum_{mn \leq x \atop m \sim M} a_m e(\alpha \ell mn). \]

Applying Lemma 4.2 immediately gives the bound
\[ E_2 \ll (M + x q^{-1} + \delta q) x^{o(1)} \ll x^{3/4 + \frac{\delta}{2}} \]
whenever \( M \ll x^{3/4} \).

\[ \Box \]

Next, we state our Type II estimate.

**Lemma 4.5** (Type II estimate). For \( q = x^{2/3} \) we have
\[ \sum_{mn \in A \atop x^\gamma \leq m \leq x^\gamma + \tau} a_m b_n = 2\delta \sum_{mn \in B \atop x^\gamma \leq m \leq x^\gamma + \tau} a_m b_n + O(x^{3/4 + \frac{\delta}{2} + o(1)}) \]
uniformly for \( 1/4 \leq \gamma, \gamma + \tau \leq 3/4 \).

**Proof.** We follow the method of Lemma 4.4. Partition the summation over \( m \) into dyadic intervals and apply Lemma 4.3 \( \Box \)

4.2. **Sieve estimates.** For positive real numbers \( w, u, v \), we denote
\[ P(w) = \prod_{p < w} p \quad \text{and} \quad P(u, v) = \prod_{u < p \leq v} p. \]

We also set
\[ Y = x^{3/4 + \frac{\delta}{2} + o(1)}. \]

For any subset of positive integers \( A \subseteq [2, x] \) and any positive integer \( s \), we denote
\[ A_s = \{ n : ns \in A \}. \]

For positive real numbers \( y \leq z \leq x \), we denote
\[ S(A; y, z) = \# \{ a \in A : p | a \Rightarrow p \in [y, z] \} \]
and in the special case \( z = x \), we denote
\[ S(A; y) = S(A; y). \]

We state a variant of the Buchstab identity for \([y, z]\)-smooth numbers which is based on taking out the largest prime factor of the integers we are counting.
Lemma 4.6. For any $2 \leq y \leq z$ and any subset of positive integers $A \subseteq [2, x]$, we have
\[
S(A; y, z) = \sum_{y \leq p \leq z} S(A_p; y, p) + S(C; x^{\frac{1}{2}}) + O(x^{\frac{1}{2}})
\]
where $C = \{a \in A : y \leq a \leq z\}$.

Proof. Take $a \in \{a \in A : p | a \implies p \in [y, z]\}$. If $a$ is a prime then $a$ is counted in $\{a \in A : y \leq a \leq z, a$ is prime$\}$. Otherwise $a$ has at least two prime factors and we can write $a = P^+(a)n$ where $n > 1$ and the prime factors of $n$ lie in the interval $[y, P^+(a)]$, the result follows immediately. □

Next we state three Lemmas which give sieve estimates for different regions. In particular, the proofs will rely on ingredients coming from the Harman sieve, see [9, Lemma 2].

Our first sieve estimate is essentially based on an application of our Type II estimate (Lemma 4.5).

Lemma 4.7. For $x^{\frac{1}{4}} \leq y < z \leq x^{\frac{1}{2}}$, we have
\[
\sum_{y \leq p \leq z} S(A_p; y, p) = 2\delta \sum_{y \leq p \leq z} S(B_p; y, p) + O(Y).
\]

Proof. We have
\[
\sum_{y \leq p \leq z} S(A_p; y, p) = \sum_{y \leq p \leq z} \sum_{np \in A \atop (n,P(y)P(p,x)=1} 1 = \sum_{mn \in A \atop y \leq m \leq z} a(m, n).
\]

Here
\[
a(m, n) = \mathbb{1}_P(m) \sum_{(n,P(y)P(m,x)=1} 1 = \mathbb{1}_P(m) \sum_{d|n \atop d|P(y)P(m,x]} \mu(d)
\]
where $\mathbb{1}_P(\cdot)$ is the characteristic function for the primes. We may assume that $d$ is square-free otherwise it does not contribute to the sum $a(m, n)$.

Write
\[
a(m, n) = \mathbb{1}_P(m) \sum_{d_1d_2|m \atop d_2|P(y)P^{-1}(d_1)>m} \mu(d_1d_2).
\]

We may assume $d_1 > 1$, the case $d_1 = 1$ follows immediately. We now get rid of the cross condition $P^{-1}(d_1) > m$ by appealing to the truncated Perron
Applying the above formula with \( \rho = \log(P^-(d) - \frac{1}{2}) \), \( \gamma = \log m \) and \( T = x^2 \delta^{-1} \), we have

\[
\sum_{\substack{mn \in A \\ y \leq m \leq z}} a(m, n) = M(A) + O(R).
\]

Here

\[
M(A) = \frac{1}{\pi} \int_{-T}^{T} \sum_{\substack{mn \in A \\ y \leq m \leq z}} \mathbb{1}_P(m) e^{i\gamma t} \left( \sum_{d_1, d_2 \mid n} \mu(d_1 d_2) \sin(\rho t) \right) \frac{dt}{t}
\]

and

\[
R = T^{-1} \sum_{\substack{mn \in A \\ y \leq m \leq z}} \mathbb{1}_P(m) \sum_{d_1, d_2 \mid n} \mu(d_1 d_2) \left| \frac{\log(P^-(d) - \frac{1}{2}) - \log m}{P^-(d) - \frac{1}{2} - m} \right|
\]

We will consider the remainder term first. By the mean value theorem, we get

\[
\frac{1}{\log(P^-(d) - \frac{1}{2}) - \log m} = \frac{\eta}{P^-(d) - \frac{1}{2} - m}
\]

where \( \eta \in [m, P^-(d) - \frac{1}{2}] \) or \( [P^-(d) - \frac{1}{2}, m] \). In any case, \( \eta \leq \max\{m, n - \frac{1}{2}\} \) so we bound

\[
\frac{\eta}{P^-(d) - \frac{1}{2} - m} \ll \max\{m, n\} \leq x.
\]

Therefore the remainder term \( R \ll T^{-1} x^{2 + o(1)} \ll \delta x^{o(1)} \).

It remains to estimate the main term. Note that the integral in the main term between \(-1/T\) and \(1/T\) can be trivially bounded by \( \ll T^{-1} x^{1 + o(1)} \ll \delta x^{-1} \). Applying our Type II estimate (Lemma 4.5) in the integral for the
region $\mathcal{R}(T) = (-T, -1/T) \cup (T, T)$, we get

$$M(A) = \frac{2\delta}{\pi} \int_{\mathcal{R}(T)} \sum_{m \in B \atop y \leq m \leq z} 1_{\mathbb{P}(m)} m^it \sum_{d_1d_2 | n \atop d_2 | \mathbb{P}(y)} \mu(d_1d_2) \sin\left(t \log \left( P^{-}(d_1) - \frac{1}{2} \right) \right) \frac{dt}{t}$$

$$+ O\left(Y \int_{\mathcal{R}(T)} \frac{dt}{t}\right) + O(\delta x^{-1})$$

$$= \frac{2\delta}{\pi} \int_{-T}^{T} \sum_{m \in B \atop y \leq m \leq z} 1_{\mathbb{P}(m)} m^it \sum_{d_1d_2 | n \atop d_2 | \mathbb{P}(y)} \mu(d_1d_2) \sin\left(t \log \left( P^{-}(d_1) - \frac{1}{2} \right) \right) \frac{dt}{t}$$

$$+ O(Y)$$

$$= 2\delta M(B) + O(Y)$$

where the last line follows from the truncated Perron formula once again. The result follows immediately. \(\square\)

The next sieve estimate is based on an idea which dates back to Vinogradov who applied it to estimate sum over primes. The idea is to systematically take out the largest prime factor until we get sums which we can estimate by our Type II estimate (Lemma 4.5).

**Lemma 4.8.** For $2 \leq y < z < x^{\frac{1}{4}}$, we have

$$\sum_{y \leq p \leq z} S(A_p; y, p) = 2\delta \sum_{y \leq p \leq z} S(B_p; y, p) + O(Y).$$

**Proof.** By the method of Lemma 4.6 we have

$$\sum_{y \leq p \leq z} S(A_p; y, p) = \sum_{y \leq p_1 \leq p \leq z} S(A_{pp_1}; y, p_1) + O(x^{\frac{1}{2}}).$$

We will first consider the sum on the right with $pp_1 > x^{\frac{1}{4}}$. Clearly we have

$$\sum_{y \leq p_1 \leq p \leq z \atop pp_1 > x^{1/4}} S(A_{pp_1}; y, p_1) = \sum_{y \leq p_1 \leq z \atop pp_1 > x^{1/4}} \sum_{n \in A \atop npp_1 \neq 1} a_m \sum_{d_1d_2 | n \atop d_1 | \mathbb{P}(n) \atop d_2 | \mathbb{P}(y)} \mu(d_1d_2)$$

Set $m = pp_1$ and note that $m < x^{\frac{1}{2}}$. It follows

$$\sum_{y \leq p_1 \leq p \leq z \atop pp_1 > x^{1/4}} S(A_{pp_1}; y, p_1) = \sum_{y \leq p_1 \leq p \leq z \atop pp_1 > x^{1/4}} a_m \sum_{m \in A \atop x^{1/4} < m < x^{1/2}} \sum_{d_1d_2 | n \atop d_2 | \mathbb{P}(n) \atop P^{-}(d_2) > P^{-}(m)} \mu(d_1d_2)$$

where $a_m$ is 1 if $m = pp_1$ with $p, p_1 \in [y, z]$ and 0 otherwise. Applying the method of Lemma 4.7, we get

$$\sum_{y \leq p_1 \leq p \leq z \atop pp_1 > x^{1/4}} S(A_{pp_1}; y, p_1) = 2\delta \sum_{y \leq p_1 \leq p \leq z \atop pp_1 > x^{1/4}} S(B_{pp_1}; y, p_1) + O(Y).$$
The only part left to consider is the sum

\[
\sum_{y \leq p \leq z \atop pp \leq x^{1/4}} S(A_{pp}; y, p_1).
\]

If the sum above is zero then we are done, otherwise we take out the next largest prime factor to get

\[
\sum_{y \leq p \leq z \atop pp \leq x^{1/4}} S(A_{pp}; y, p_1) = \sum_{y \leq p \leq z \atop pp \leq x^{1/4}} S(A_{pp, p_2}; y, p_2) + O(x^{1/2}).
\]

The sum on the right with \(pp_1p_2 > x^{1/4}\) can be dealt with again by the method of Lemma 4.7. By induction this can go on for at most \(O(\log x)\) steps. Since we have an asymptotic formula for every sum, the result follows. \(\square\)

For our next sieve estimate, we note that our Type II estimates are not sufficient in this region. Instead we bypass this complication by a role reversal that minimises our length of summation in exchange for sifting primes.

**Lemma 4.9.** For \(x^{3/4} \leq z \leq x\), we have

\[
\sum_{x^{3/4} < p \leq z} S(A_p; y, p) = 2\delta \sum_{x^{3/4} < p \leq z} S(B_p; y, p) + O(Y).
\]

**Proof.** Take \(x^{3/4} < p \leq z\) and an element \(np \in A\) such that if \(q | n\) then \(q \in [y, p]\). This gives

\[
(4.3) \quad \sum_{x^{3/4} < p \leq z} S(A_p; y, p) = \sum_{n < x^{1/4}} c_n S(A'_n; (x/n)^{1/2}) + O(x^{o(1)}).
\]

Here

\[
A'_n = \{\max\{y + \Delta(y), x^{3/4}\} < m \leq z : mn \in A\} = \{\max\{y + \Delta(y), x^{3/4}\} < m \leq \min\{z, x/n\} : \|\alpha m + \beta\| < \delta\}
\]

where \(c_n\) is 1 if all prime factors of \(n\) are at least \(y\) and 0 otherwise and \(\Delta(y)\) is -1/2 if \(y\) is an integer and zero otherwise. We also recall

\[
S(W; k) = \#\{w \in W : (w, P(k)) = 1\}.
\]

We note that the sum in (4.3) may be empty depending on the choice of \(y\). To exhibit (4.3), take \(m \in S(A'_n; (x/n)^{1/2})\) and suppose that \(m\) has two prime factors say \(q_1, q_2\) then we must have

\[
m \geq q_1q_2 \geq (x/n)^{1/2}(x/n)^{1/2} = x/n.
\]
This is a contradiction unless $mn = x$ and this occurs at most $O(x^{o(1)})$ times. Moreover, it is evident that
\[
\sum_{n < x^{1/4}} c_n S(A'_n; (x/n)^{1/2}) = \sum_{n < x^{1/4}} c_n (S(A'_n; x^{1/2}) + O(x^{1/2})) \\
= \sum_{n < x^{1/4}} c_n S(A'_n; x^{1/2}) + O(x^{3/4}).
\]

Let $n \sim N$ where $N \ll x^{1/4}$ and $\max\{y + \Delta(y), x^{3/4}\} n \ll M(n) = M \ll \min\{zn, x\}$ and consider the set
\[
A^{(M)} = \{m \sim M : \|am + \beta\| < \delta\}.
\]

Then we have
\[
A^{(M)}_n = \{m \sim M/n : \|am + \beta\| < \delta\}
\]
and by the Harman sieve \cite[Lemma 2]{Harman} using the Type I and II estimate (Lemma 4.4 and 4.5), we get
\[
\sum_{n \sim N} c_n S(A^{(M)}_n; x^{1/2}) = 2\delta \sum_{n \sim N} c_n S(B^{(M)}_n; x^{1/2}) + O(Y).
\]

Therefore by summing over $N, M$ we obtain
\[
\sum_{n < x^{1/4}} c_n S(A'_n; x^{1/2}) = 2\delta \sum_{n < x^{1/4}} c_n S(B'_n; x^{1/2}) + O(Y),
\]
which is what we needed to show. \hfill \Box

For subset $C \subseteq [2, x]$ and integers $2 \leq y < z \leq x$, we denote
\[
T(C; y, z) = \#\{c \in C : \mu^2(c) = 1, p|c \implies p \in [y, z]\}.
\]

The next three results are square-free analogue of Lemma 4.7, 4.8 and 4.9.

**Lemma 4.10.** For $x^{1/2} \leq y < z \leq x^{1/2}$, we have
\[
\sum_{y \leq p \leq z} T(A_p; y, p - 1) = 2\delta \sum_{y < p \leq z} T(B_p; y, p - 1) + O(Y).
\]

**Proof.** By writing
\[
\sum_{y \leq p \leq z} T(A_p; y, p - 1) = \sum_{y \leq p \leq z} \mu^2(np) \sum_{np \in A} \mathbb{1}_{(n, P(y), P(p-1, x)) = 1} \\
= \sum_{mn \in A} \mathbb{1}_P(m) \mu^2(n) \sum_{d_1d_2|n} \mu(d_1d_2) \\
\sum_{d_1|P(y)} \sum_{d_2 | P^{-1}(d_2) \geq m} \mu(d_1d_2).
\]

We see that the method of Lemma 4.7 gives the result. \hfill \Box
Lemma 4.11. For \(2 \leq y < z < x^{1/4}\), we have
\[
\sum_{y \leq p \leq z} T(A_p; y, p - 1) = 2\delta \sum_{y \leq p \leq z} T(B_p; y, p - 1) + O(Y).
\]

Proof. We follow as in Lemma 4.8 but we successively take out the largest distinct prime factors. \(\square\)

Lemma 4.12. For \(x^{3/4} \leq z \leq x\), we have
\[
\sum_{x^{3/4} < p \leq z} T(A_p; y, p - 1) = 2\delta \sum_{x^{3/4} < p \leq z} T(B_p; y, p - 1) + O(Y).
\]

Proof. This follows immediately from the proof of Lemma 4.9 by taking \(c_n\) there to be 1 if \(n\) is square-free and all prime factors of \(n\) is at least \(y\) or 0 otherwise. \(\square\)

5. Proof of Theorem 3.1

By Dirichlet’s theorem, we have \(|\alpha - a/q| < 1/q^2\) for infinitely many \(q\), let \((q_k)_{k \in \mathbb{N}}\) be an increasing sequence of such denominators. Therefore, we can choose an increasing sequence \((x_k)_{k \in \mathbb{N}}\) so that \(q_k = x_k^{2/3}\). Now we take \(x_k\) to be a sufficiently large element in the sequence \((x_k)_{k \in \mathbb{N}}\).

By Lemma 4.6, we assert
\[
S(A; y, z) = \sum_{y \leq p \leq z} S(A_p; y, p) + S(C; x_k^{1/2}) + O(x_k^{1/2}).
\]

By [18, Theorem 2] and partial summation, we have
\[
S(C; x_k^{1/2}) = 2\delta S(D; x_k^{1/2}) + O(Y)
\]
where \(D = \{n \in \mathbb{N} : y \leq n \leq z\}\) and \(Y = x_k^{3/4+\epsilon/2+o(1)}\). Therefore by substituting this into the above, we obtain
\[
S(A; y, z) = \sum_{y \leq p \leq z} S(A_p; y, p) + 2\delta S(D; x_k^{1/2}) + O(Y).
\]

Suppose \(y < x_k^{1/4}\). Then we write
\[
S(A; y, z) = \left( \sum_{y \leq p \leq z < x_k^{1/4}} + \sum_{y \leq p < x_k^{1/4}} + \sum_{x_k^{1/4} \leq p \leq z < x_k^{3/4}} + \sum_{x_k^{1/4} \leq p \leq x_k^{3/4}} \right) S(A_p; y, p) + 2\delta S(D; x_k^{1/2}) + O(Y).
\]

We note that some of the sums above may be empty depending on the choice of \(z\). Applying Lemma 4.7, 4.8, 4.9 then Lemma 4.6 to the above we get
\[
S(A; y, z) = 2\delta S(B; y, z) + O(Y).
\]
The case $x_k^{1/4} \leq y < x_k^{1/2}$ is similar to the case above and the first result follows.

Now suppose $\alpha \in \mathcal{I}(A)$ then we can assume $\alpha \in (0, 1) \setminus \mathbb{Q}$. Indeed, we can write $\alpha = b + r$ where $b \in \mathbb{Z}$ and $r \in (0, 1) \setminus \mathbb{Q}$ so that $\|\alpha n + \beta\| = \|rn + \beta\|$. Therefore we have the continued fraction expansion

$$\alpha = [0; a_1, a_2, \ldots]$$

with $0 < a_i \leq A$ for some absolute constant $A$. The convergents $p_s/q_s$ to $\alpha$ satisfy the inequality $|\alpha - p_s/q_s| < 1/q_s^2$. Clearly $(q_s)_{s \geq 1}$ is a strictly increasing sequence. Therefore, take any sufficiently large $x$ there exist $s \in \mathbb{N}$ such that $q_{s-1} \leq x^{2/3} \leq q_s$. Since $q_s = a_s q_{s-1} + q_{s-2}$ for $s \geq 2$, we get $q_s \leq (A + 1)q_{s-1} \leq 2Ax^{2/3} \ll x^{2/3}$, since $A$ is fixed. Hence $x^{2/3} \ll q_s \ll x^{2/3}$ and the result follows by the argument above.

6. Proof of Theorem 3.2

We proceed as in the proof of Theorem 3.1 but instead we apply Lemma 4.10, 4.11 and 4.12.

Acknowledgement

The author is grateful to the anonymous referee for their excellent comments which resulted in a complete rewrite of this paper; B. Kerr and I. E. Shparlinski for helpful comments on a previous draft. Lastly the author thanks T. Evans and N. Poznanović for stimulating conversations on a related topic at the beach. The author was supported by an Australian Government Research Training Program (RTP) Scholarship.

References

[1] R. C. Baker, Diophantine Inequalities, Clarendon Press, Oxford, 1986.
[2] R. C. Baker and G. Harman, On the distribution of $\alpha p^k$ modulo one, Mathematica 38 (1991), 170-184.
[3] A. Balog and A. Perelli, Diophantine approximation by square-free numbers, Ann. Sc. Norm. Super. Pisa Cl. Sci. (4) 11(3) (1984), 353-359.
[4] J. Friedlander, Integers free from large and small primes, Proc. London Math. Soc. (2) 33 (1976), 565-576.
[5] A. Ghosh, The distribution of $\alpha p^2$ modulo 1, Proc. Lond. Math. Soc. 3 (1981), 252-269.
[6] G. H. Hardy, E. M. Wright, An Introduction to the Theory of Numbers, 6th ed., Oxford University Press, Oxford (2008). Revised by D. R. Heath-Brown and J. Silverman; With a foreword by Andrew Wiles.
[7] G. Harman, *On the distribution of $\alpha p$ modulo one*, J. Lond. Math. Soc. (2) 27 (1983), 9-18.

[8] G. Harman, *Diophantine approximation with square-free integers*, Math. Proc. Cambridge Philos. Soc. 95 (1984), no. 3, 381-388.

[9] G. Harman, *On the distribution of $\alpha p$ modulo one* II, Proc. Lond. Math. Soc. (3) 72 (1996), 241-260.

[10] G. Harman, *Prime-Detecting Sieves*, London Math. Soc. Monogr. Ser. 33 (2007).

[11] D. R. Heath-Brown *Diophantine approximation with square-free numbers*, Math. Z. 187 (1984), no. 3, 335-344.

[12] D. R. Heath-Brown and C. Jia, *The distribution of $\alpha p$ modulo one*, Proc. Lond. Math. Soc. (3) 84 (1) (2002), 79-104.

[13] C. Jia, *On the distribution of $\alpha p$ modulo one* (II), Sci. China Ser. A 43 (7) (2000), 703-721.

[14] K. Matomäki, *The distribution of $\alpha p$ modulo one*, Math. Proc. Cambridge Philos. Soc. 147 (2) (2009), 267-283.

[15] É. Saias, *Entiers sans grand ni petit facteur premier*. I., Acta Arith. 61 (1992), no. 4, 347-374.

[16] É. Saias, *Entiers sans grand ni petit facteur premier*. II., Acta Arith. 63 (1993), no. 4, 287-312.

[17] É. Saias, *Entiers sans grand ni petit facteur premier*. III., Acta Arith. 71 (1995), no. 4, 351-379.

[18] R. C. Vaughan, *On the distribution of $\alpha p$ modulo 1*, Mathematika 24 (1977), 135-141.

[19] R. C. Vaughan, *Sommes trigonométriques sur les nombres premiers*, C. R. Acad. Sci. Paris Sér. A-B 285 (16) (1977), A981-A983.

[20] K. C. Wong, *On the distribution of $\alpha p^k$ modulo 1*, Glasgow Math. J. 39 (1997), no. 2, 121-130.

DEPARTMENT OF PURE MATHEMATICS, UNIVERSITY OF NEW SOUTH WALES, SYDNEY, NSW 2052, AUSTRALIA

E-mail address: kamhung.yau@unsw.edu.au