Acta Math. Hungar., **120 (1–2)** (2008), 79–102.
DOI: 10.1007/s10474-007-7107-1
First published online January 22, 2008

DIFFERENCE SETS AND SHIFTED PRIMES

J. LUCIER

Département de Mathématiques et de Statistiques, Université de Montréal,
CP 6128, succ. Centre-Ville, Montréal, Québec H3C 3J7, Canada
e-mail: lucier@crm.umontreal.ca

(Received June 8, 2007; accepted August 14, 2007)

Abstract. We show that if A is a subset of $\{1,\dots,n\}$ which has no pair of elements whose difference is equal to $p-1$ with p a prime number, then the size of A is $O\big(n(\log\log n)^{-c(\log\log\log\log\log n)}\big)$ for some absolute $c>0$.

1. Introduction

For a set of integers A we denote by $A-A$ the set of all differences $a-a'$ with a and a' in A, and if A is a finite set we denote its cardinality by $|A|$. Sárközy [12] proved, by the Hardy–Littlewood method, that if A is a subset of $\{1,\dots,n\}$ such that $A-A$ does not contain a perfect square, then

$$|A| \ll n(\log\log n)^{2/3}(\log n)^{-1/3}.$$

This estimate was improved by Pintz, Steiger and Szemerédi [10] to

$$|A| \ll n(\log n)^{-(1/12)\log\log\log\log n}.$$

This improvement was obtained using the Hardy–Littlewood method together with a combinatorial result concerning sums of rationals. Balog, Pelikán,

Key words and phrases: Hardy–Littlewood method, difference sets, shifted primes.
2000 Mathematics Subject Classification: 11P55, 11P32, 11B75.

Pintz and Szemerédi [1], elucidating the method in [10], proved for any fixed integer $k \geqq 2$, that if A is a subset of $\{1, \dots, n\}$ such that $A - A$ does not contain a perfect k-th power, then

$$|A| \ll_k n(\log n)^{-(1/4)\log\log\log\log n}.$$

In the works cited above the following basic property is used: if s is a perfect k-th power then so is $q^k s$ for every positive integer q. This multiplicative property is used in the following fashion. Suppose that B is a set of integers and $A = \{c + q^k b : b \in B\}$ for some integers c and $q \geqq 1$. If $A - A$ does not contain a perfect k-th power, then the same is true for $B - B$. This deduction is the basis of an iteration argument that plays a fundamental rôle in [1], [10], and [12].

Sárközy [13] also considered the set $\mathcal{S} = \{p - 1 : p \text{ a prime}\}$ of shifted primes, and showed that if A is a subset of $\{1, \dots, n\}$ such that $A - A$ does not contain an integer from $\mathcal{S}$ then

$$|A| \ll n\frac{(\log\log\log n)^3(\log\log\log\log n)}{(\log\log n)^2}.$$

The argument Sárközy used in [12] cannot be applied directly to the set $\mathcal{S}$ of shifted primes since it does not have a multiplicative property analogous to the one possessed by the set of perfect k-th powers. Sárközy got around this difficulty by not only considering the set $\mathcal{S}$ of shifted primes, but also the sets defined for each positive integer d by

$$\mathcal{S}_d = \left\{\frac{p-1}{d} : p \text{ a prime}, \ \ p \equiv 1 \pmod{d}\right\}.$$

In [13] Sárközy uses an iteration argument based on the following observation. Suppose B is a set of integers and $A = \{c + qb : b \in B\}$ for some integers c and $q \geqq 1$. If $A - A$ does not intersect S_d for some positive integer d, then $B - B$ does not intersect $\mathcal{S}_{dq}$.

In this article we show that the combinatorial argument presented in [1] and [10] can be carried out to improve Sárközy's result on the set $\mathcal{S}$ of shifted primes. We shall prove the following.

THEOREM. *Let n be a positive integer and A a subset of $\{1, \dots, n\}$. If there does not exist a pair of integers $a, a' \in A$ such that $a - a' = p - 1$ for some prime p, then*

$$|A| \ll n\left(\frac{(\log\log\log n)^3(\log\log\log\log n)}{(\log\log n)}\right)^{\log\log\log\log\log n}.$$

The set of perfect squares and the set $\mathcal{S}$ of shifted primes are examples of *intersective* sets. To define this class of sets we introduce some notation. Given a set of positive integers H we define $D(H,n)$, for any positive integer n, to be the maximal size of a subset A of $\{1,\dots,n\}$ such that $A-A$ does not intersect H. A set of positive integers H is called *intersective* if $D(H,n) = o(n)$.

Kamae and Mendès France [6] supplied a general criterion for determining if a set of positive integers is intersective. From their criterion they deduced the following.

(I) For any fixed integer a the set $\{p+a : p \text{ a prime}, p > -a\}$ is intersective if and only if $a = \pm 1$.

(II) Let h be a nonconstant polynomial with integer coefficients and whose leading coefficient is positive. The set $\{h(m) : m \geqq 1,\, h(m) \geqq 1\}$ is intersective if and only if for each positive integer d the modular equation $h(x) \equiv 0 \pmod d$ has a solution.

Let h be a polynomial as in (II) with degree $k \geqq 2$ and such that $h(x) \equiv 0 \pmod d$ has a solution for every positive integer d. The author [8] has shown that if A is a subset of $\{1,\dots,n\}$ such that $A-A$ does not intersect $\{h(m) : m \geqq 1,\, h(m) \geqq 1\}$, then $|A| \ll n(\log\log n)^{\mu/(k-1)}(\log n)^{-(k-1)}$, where $\mu = 3$ if $k=2$ and $\mu = 2$ if $k \geqq 3$. It is possible to improve this result with the method presented in this paper.

2. Preliminary lemmata

In this paper we use the following notations. For a real number x we write $e(x)$ for $e^{2\pi i x}$, and $[x]$ is used to denote the greatest integer less than or equal to x. The greatest common divisor of the integers u and v is given by (u,v). Euler's totient function is denoted, as usual, by ϕ. For any positive integer i we write $\log_i$ to denote the i-th iterated logarithm, that is, $\log_1 n = \log n$ and $\log_i n = \log(\log_{i-1} n)$ for every integer $i \geqq 2$.

A fundamental rôle is played by the following relations. For integers n and r, with n positive,

$$\sum_{t=0}^{n-1} e(rt/n) = \begin{cases} n & \text{if } n \mid r \\ 0 & \text{if } n \nmid r \end{cases}, \qquad \int_0^1 e(r\alpha)\,d\alpha = \begin{cases} 1 & \text{if } r = 0 \\ 0 & \text{if } r \neq 0. \end{cases}$$

Given a subset A of $\{1,\dots,n\}$ its generating function is given by

$$F(\alpha) = \sum_{a\in A} e(\alpha a), \quad \alpha \in \mathbb{R}.$$

Using the relations above we find that

$$\sum_{t=1}^{n} \left| F(t/n) \right|^2 = n|A|, \qquad \int_0^1 \left| F(\alpha) \right|^2 d\alpha = |A|.$$

Of course, these are particular cases of Parseval's identity.

Sárközy's method in [12] and [13] is based on Roth's work [11] on three-term arithmetic progressions in dense sets. Following this method Sárközy uses a functional inequality to derive his results concerning the set of perfect squares and the set $\mathcal{S}$ of shifted primes. Our approach here uses, like Gowers [3] and Green [4], a density increment argument. The next lemma tells us that if the generating function of a finite set A satisfies a certain size constraint, then it must be concentrated along an arithmetic progression. We use this result in Lemma 10 to obtain a density increment that we iterate in the final section of the paper to prove the theorem.

LEMMA 1. *Let n be a positive integer and A a subset of $\{1,\dots,n\}$ with size δn. For any real α let $F(\alpha)$ denote the generating function of A. Let q be a positive integer and U a positive real number such that $2\pi qU \leqq n$. Let E denote the subset of $[0,1]$ defined by*

$$E = \left\{ \alpha \in [0,1] : \left| \alpha - \frac{a}{q} \right| \leqq \frac{U}{n} \text{ for some } 0 \leqq a \leqq q \right\}.$$

If θ is a positive number such that

$$\sum_{\substack{t=1 \\ t/n \in E}}^{n-1} \left| F(t/n) \right|^2 \geqq \theta |A|^2, \tag{1}$$

then there exists an arithmetic progression P in $\{1,\dots,n\}$ with difference q such that

$$|P| \geqq \frac{n}{32\pi qU} \quad \text{and} \quad |A \cap P| \geqq |P| \delta \left(1 + 8^{-1} \theta \right).$$

PROOF. This closely resembles Lemma 20 in [8] and can be proved in the same manner. □

We now state a combinatorial result presented by Balog, Pelikán, Pintz and Szemerédi in [1], the proof of which uses only elementary techniques. It is this result, that we use in Lemma 9, that allows us to improve Sárközy's result on the set $\mathcal{S}$ of shifted primes.

LEMMA 2 ([1], Lemma CR). *Let K and L be positive integers, and let τ be the maximal value of the divisor function up to KL. Let $\mathcal{K}$ be a nonempty* *subset of rationals such that if* $a/k \in \mathcal{K}$ *is in lowest terms then* $1 \leqq a \leqq k \leqq K$. *Suppose that for each* $a/k \in \mathcal{K}$ *there corresponds a subset of rationals* $\mathcal{L}_{a/k}$ *such that if* $b/l \in \mathcal{L}_{a/k}$ *is in lowest terms then* $1 \leqq b \leqq l \leqq L$. *Suppose further that* B *and* H *are positive integers such that*

$$|\mathcal{L}_{a/k}| \geqq H \quad \text{for all} \quad a/k \in \mathcal{K}$$

and

$$\left|\left\{b : \frac{b}{l} \in \bigcup \mathcal{L}_{a/k}\right\}\right| \leqq B \quad \text{for all} \quad l \leqq L.$$

Then the size of the set

$$\mathcal{Q} = \left\{\frac{a}{k} + \frac{b}{l} : \frac{a}{k} \in \mathcal{K},\ \frac{b}{l} \in \mathcal{L}_{a/k}\right\}$$

satisfies

$$|\mathcal{Q}| \geqq |\mathcal{K}| H \left(\frac{H}{LB\tau^8(1 + \log K)}\right).$$

3. Exponential sums over primes

Let d and n denote positive integers. As in [13], our application of the Hardy–Littlewood method employs exponential sums over numbers from the set $\mathcal{S}_d$ defined in the introduction. For any real number α set

$$S_{n,d}(\alpha) = \sum_{\substack{s \in \mathcal{S}_d \\ s \leqq n}} \log(ds + 1) e(\alpha s).$$

In this section we present some estimates related to $S_{n,d}(\alpha)$. Throughout this section we assume d and n satisfy

$$d \leqq \log n.$$

Lemma 3. *For* n *sufficiently large,*

$$S_{d,n}(0) \gg \frac{dn}{\phi(d)}.$$

PROOF. By the definition of $\mathcal{S}_d$ we find that

$$S_{d,n}(0)=\sum_{\substack{p\leqq dn+1\\ p\equiv 1 \bmod d}}\log p.$$

Since $d\leqq \log n$ the Siegel–Walfisz theorem says that this sum is asymptotic to $(dn+1)/\phi(q)$, from which the result follows. □

The next two lemmas provide estimates of $S(\alpha)$ derived by A. Sárközy.

LEMMA 4. *Let a and b be integers such that $(a,b)=1$ and $1\leqq b\leqq \log n$. There exists a positive real number c such that if α is a real number that satisfies*

$$\left|\alpha-\frac{a}{b}\right|\leqq\frac{\exp\left(c(\log n)^{1/2}\right)}{n},$$

and n is sufficiently large, then

$$\left|S_{d,n}(\alpha)\right|<\frac{dn}{\phi(d)\phi(b)},$$

furthermore, if $\alpha\neq a/b$ then

$$\left|S_{d,n}(\alpha)\right|<\frac{d}{\phi(d)\phi(b)}\left|\alpha-\frac{a}{b}\right|^{-1}.$$

PROOF. This is a restatement of Lemma 5 from [13]. □

Let R denote a real number that satisfies

$$3\leqq R\leqq \log n. \tag{2}$$

For integers a and b such that $(a,b)=1$ and $0\leqq a\leqq b\leqq R$ set

$$\mathfrak{M}(b,a)=\left\{\alpha\in[0,1]:\ \left|\alpha-\frac{a}{b}\right|\leqq\frac{R}{n\log\log R}\right\}. \tag{3}$$

Let $\mathfrak{m}$ denote the set of real numbers α for which there do not exist integers a and b such that $(a,b)=1$, $1\leqq b<R$, and $\alpha\in\mathfrak{M}(b,a)$.

LEMMA 5. *For $\alpha\in\mathfrak{m}$ and large n,*

$$S_{d,n}(\alpha)\ll\frac{dn}{\phi(d)}\cdot\frac{\log\log R}{R}. \tag{4}$$

PROOF. This is a restatement of Lemma 9 from [13]. □

LEMMA 6. *Let a and b be integers such that $0 \leqq a \leqq b \leqq R$ and $(a,b)=1$. Then for n sufficiently large*

$$\sum_{t/n\in\mathfrak{M}(b,a)} \left|S_{d,n}(t/n)\right| \ll \frac{dn}{\phi(d)\phi(b)}\log R.$$

PROOF. Suppose that $t/n \in \mathfrak{M}(b,a)$. Then

$$\left|\frac{t}{n}-\frac{a}{b}\right| \leqq \frac{R}{n\log\log R} \leqq \frac{\log n}{n},$$

and since $b \leqq R \leqq \log n$ we can, for large enough n, apply Lemma 4 with α replaced by t/n.

Let u and v be integers such that

$$\frac{u}{n} < \frac{a}{b} < \frac{v}{n}, \quad v-u=2.$$

Applying Lemma 4 we obtain

$$\sum_{\substack{t/n\in\mathfrak{M}(b,a)\\ u/n\leqq t/n\leqq v/n}} \left|S_{d,n}(t/n)\right| \ll \frac{dn}{\phi(d)\phi(b)}.$$

For $t/n \in \mathfrak{M}(b,a)$ with $t/n < u/n$, Lemma 4 implies

$$\left|S_{d,n}(t/n)\right| \ll \frac{d}{\phi(d)\phi(b)}\left|\frac{t}{n}-\frac{a}{b}\right|^{-1} \ll \frac{d}{\phi(d)\phi(b)}\left|\frac{t}{n}-\frac{u}{n}\right|^{-1}.$$

Therefore

$$\sum_{\substack{t/n\in\mathfrak{M}(b,a)\\ t/n<u/n}} \left|S_{d,n}(t/n)\right| \ll \frac{dn}{\phi(d)\phi(b)} \sum_{\substack{t/n\in\mathfrak{M}(b,a)\\ t/n<u/n}} \frac{1}{|t-u|}$$

$$\ll \frac{dn}{\phi(d)\phi(b)} \sum_{1\leqq m\leqq R/\log\log R} \frac{1}{m} \ll \frac{dn}{\phi(d)\phi(b)}\log R.$$

Similarly

$$\sum_{\substack{t/n\in\mathfrak{M}(b,a)\\ v/n<t/n}} \left|S_{d,n}(t/n)\right| \ll \frac{dn}{\phi(d)\phi(b)}\log R. \quad \square$$

A multiplicative arithmetic function f is called strongly multiplicative if $f(p^k)=f(p)$ for every prime p and positive integer k. The next lemma contains a standard deduction on the average order over arithmetic progressions for certain strongly mutliplicative arithmetic functions.

LEMMA 7. *Let x be a real number such that $x \geqq 1$, and let d and r be positive integers. If f is a strongly multiplicative arithmetic function such that $f(m) \geqq 1$ for every positive integer m and $f(p) = 1 + O(p^{-1})$, then*

$$\sum_{\substack{m \leqq x \\ m \equiv r \bmod d}} f(m) \ll f\big((r,d)\big)\, \frac{x}{d}.$$

PROOF. Let g be the arithmetic function defined by

$$g(m) = \sum_{k|m} \mu\left(\frac{m}{k}\right) f(k),$$

where μ is the Möbius function. Using the fact that f is strongly multiplicative we deduce that

$$g(m) = \mu(m)^2 \prod_{p|m} \big(f(p) - 1\big).$$

Since $f(m) \geqq 1$ for every positive integer m it follows that g is a non-negative valued arithmetic function. By the Möbius inversion formula $f(m) = \sum_{k|m} g(k)$, therefore

$$\sum_{\substack{m \leqq x \\ m \equiv r \bmod d}} f(m) = \sum_{\substack{m \leqq x \\ m \equiv r \bmod d}} \sum_{k|m} g(k) = \sum_{k \leqq x} g(k) \sum_{\substack{m \leqq x \\ m \equiv r \bmod d \\ m \equiv 0 \bmod k}} 1.$$

The last sum above is zero if $(k,d) \nmid r$ and at most $x(d,k)/(dk)$ if $(k,d) \mid r$. This implies, since g is a non-negative valued function, that

$$\sum_{\substack{m \leqq x \\ m \equiv r \bmod d}} f(m) \leqq \frac{x}{d} \sum_{\substack{k \leqq x \\ (k,d)|r}} \frac{g(k)(k,d)}{k} = \frac{x}{d} \sum_{s|(r,d)} s \sum_{\substack{k \leqq x \\ (k,d)=s}} \frac{g(k)}{k}$$

$$= \frac{x}{d} \sum_{s|(r,d)} \sum_{\substack{l \leqq x/s \\ (l,d/s)=1}} \frac{g(sl)}{l}.$$

For positive integers u and v it can be verified that $g(uv) \leqq g(u)g(v)$, thus

$$\sum_{\substack{m \leqq x \\ m \equiv r \bmod d}} f(m) \leqq \frac{x}{d} \sum_{s|(r,d)} g(s) \sum_{l \leqq x} \frac{g(l)}{l}$$

$$\leqq f\big((r,d)\big)\,\frac{x}{d}\prod_{p\leqq x}\left(1+\frac{g(p)}{p}\right)=f\big((r,d)\big)\,\frac{x}{d}\prod_{p\leqq x}\left(1+\frac{f(p)-1}{p}\right).$$

Since $f(p)\geqq 1$ and $f(p)=1+O(p^{-1})$ the previous product is bounded from above by the absolutely convergent infinite product $\prod_p\big(1+p^{-1}\big(f(p)-1\big)\big)$. Therefore

$$\sum_{\substack{m\leqq x\\ m\equiv r \bmod d}} f(m)\ll f\big((r,d)\big)\,\frac{x}{d}. \quad \square$$

The next lemma is analogous to Proposition 11 of Green [4].

LEMMA 8.

$$\sum_{t=0}^{n-1}\big|\,S_{d,n}(t/n)\big|^4\ll\left(\frac{dn}{\phi(d)}\right)^4.$$

PROOF. By Gallagher's inequality [9, Lemma 1.2] we have

$$\sum_{t=0}^{n-1}\big|\,S_{d,n}(t/n)\big|^4\leqq n\int_0^1\big|\,S_{d,n}(\alpha)\big|^4\,d\alpha+2\int_0^1\big|\,S_{d,n}(\alpha)^3S'_{d,n}(\alpha)\big|\,d\alpha,$$

where $S'_{d,n}(\alpha)$ is the derivative of $S_{d,n}(\alpha)$ with respect to α. By Hölder's inequality

$$\int_0^1\big|\,S_{d,n}(\alpha)^3S'_{d,n}(\alpha)\big|\,d\alpha\leqq\left(\int_0^1\big|\,S_{d,n}(\alpha)\big|^4\,d\alpha\right)^{3/4}\left(\int_0^1\big|\,S'_{d,n}(\alpha)\big|^4\,d\alpha\right)^{1/4}.$$

Let $r_d(m)$ denote the number of pairs (p_1,p_2) where p_1 and p_2 are primes such that $p_1,p_2\equiv 1 \pmod d$ and

$$\frac{p_1-1}{d}+\frac{p_2-1}{d}=m.$$

By Parseval's identity,

$$\int_0^1\big|\,S_{d,n}(\alpha)\big|^4\,d\alpha\leqq(\log n)^4\sum_{m\leqq n}r_d(m)^2$$

and

$$\int_0^1\big|\,S'_{d,n}(\alpha)\big|^4\,d\alpha\leqq 2\pi(n\log n)^4\sum_{m\leqq n}r_d(m)^2.$$

From the above we deduce that

$$\sum_{t=0}^{n-1} \left| S_{d,n}(t/n) \right|^4 \ll n(\log n)^4 \sum_{m \leqq n} r_d(m)^2. \tag{5}$$

For each positive integer m we have

$$r_d(m) \leqq \left| \{ p : \ 1 < p \leqq dm + 2, \ p \equiv 1 \bmod d, \ dm + 2 - p \text{ is a prime} \} \right|.$$

To bound $r_d(m)$ we apply the combinatorial sieve to estimate the size of the set above. In particular, Corollary 2.4.1 of [5] implies

$$r_d(m) \ll \prod_{p \mid d(dm+2)} \left(1 - \frac{1}{p} \right)^{-1} \frac{dm+1}{\phi(d) \log^2 \big((dm+1)/d \big)}.$$

Note that

$$\prod_{p \mid d(dm+2)} \left(1 - \frac{1}{p} \right)^{-1} \leqq \frac{d}{\phi(d)} \left(\frac{dm+2}{\phi(dm+2)} \right),$$

therefore

$$r_d(m) \ll \frac{d^2 m}{\phi(d)^2 (\log m)^2} \left(\frac{dm+2}{\phi(dm+2)} \right).$$

This implies

$$\sum_{m \leqq n} r_d(m)^2 \ll \frac{d^4 n^2}{\phi(d)^4 (\log n)^4} \sum_{\substack{u \leqq dn+2 \\ u \equiv 2 \bmod d}} \left(\frac{u}{\phi(u)} \right)^2.$$

Let $f(u) = \big(u/\phi(u) \big)^2$. It can be verified that f is a strongly multiplicative arithmetic function such that $f(u) \geqq 1$ for every positive integer u and $f(p) = 1 + O(p^{-1})$. Thus, we can apply Lemma 7 to obtain

$$\sum_{\substack{u \leqq dn+2 \\ u \equiv 2 \bmod d}} \left(\frac{u}{\phi(u)} \right)^2 \ll n.$$

Therefore

$$\sum_{m \leqq n} r_d(m)^2 \ll \frac{d^2 n^3}{\phi(d)^2 (\log n)^4},$$

and thus, on account of (5), the result follows. $\square$

4. A density increment

Throughout this section n denotes a positive integer and A a subset of $\{1,\dots,n\}$. For any real α set

$$F(\alpha)=\sum_{a\in A}e(\alpha a),\qquad F_1(\alpha)=\sum_{\substack{a\in A\\ a\leqq n/2}}e(\alpha a).$$

Denote by C_1 a fixed positive constant. This constant will be used throughout the rest of the paper. We will need C_1 to be sufficiently large, but it should be noted that the size of C_1 will never be determined by n or A. Let δ denote the density of A, that is, $|A|=\delta n$. The following parameters are defined in terms of C_1 and δ:

$$R(\delta)=(C_1\delta^{-1})^{(\log\log C_1\delta^{-1})^{7/8}},\tag{6}$$

$$\theta(\delta)=(C_1\delta^{-1})^{-4(\log\log\log C_1\delta^{-1})^{-1}}.\tag{7}$$

$$Q_1=(C_1\delta^{-1})^{(\log\log C_1\delta^{-1})^{1/8}},\tag{8}$$

$$\Lambda=\left[\frac{3}{4}\log\log\log C_1\delta^{-1}\right],\tag{9}$$

With $R=R(\delta)$ let $\mathfrak{M}(q,a)$ be defined as in (3), and for any positive integer $q\leqq R$ set

$$\mathfrak{M}(q)=\bigcup_{\substack{a=0\\ (a,q)=1}}^{q}\mathfrak{M}(q,a).$$

LEMMA 9. *Let d be a positive integer such that $d\leqq\log n$. Suppose that $A-A$ does not intersect $\mathcal{S}_d$ and that*

$$C_1\delta^{-1}\leqq e^{(\log\log n)^{1/2}},\tag{10}$$

provided C_1 and n are sufficiently large there exists a positive integer $q\leqq R(\delta)$ such that

$$\sum_{\substack{t=1\\ t/n\in\mathfrak{M}(q)}}^{n-1}\left|F(t/n)\right|^2\geqq\theta(\delta)|A|^2.\tag{11}$$

PROOF. Here we adopt the method used in [1]. Given any positive integer λ we make the following definitions. For integers a and k, with $k \geqq 1$, define

$$\mathfrak{M}_\lambda(k,a) = \left\{ \alpha \in [0,1] : \left| \alpha - \frac{a}{k} \right| \leqq \frac{\lambda R}{n \log\log R} \right\},$$

and for real numbers $K, U \geqq 1$,

$$\mathcal{P}_\lambda(K,U) = \left\{ \frac{a}{k} : 1 \leqq a \leqq k \leqq K, (a,k) = 1, \max_{t/n \in \mathfrak{M}_\lambda(k,a)} \left| F_1(t/n) \right| \geqq |A|/U \right\}.$$

Furthermore, set

$$Q_\lambda = Q_1^{2^\lambda - 1} \tag{12}$$

and

$$\mu_\lambda = \max_{\substack{1 \leqq K \leqq Q_\lambda \\ 1 \leqq U}} \frac{\left| \mathcal{P}_\lambda(K,U) \right|}{U^2}.$$

Let K_λ and U_λ denote a pair for which μ_λ takes its maximum. As $K = U = 1$ is considered in the definition of μ_λ we have

$$1 \leqq \mu_\lambda \leqq \frac{K_\lambda^2}{U_\lambda^2}. \tag{13}$$

It follows that

$$1 \leqq U_\lambda \leqq K_\lambda \leqq Q_\lambda. \tag{14}$$

For each $\lambda \leqq \Lambda$ we want the intervals $\mathfrak{M}_\lambda(k,a)$ with $k \leqq Q_\lambda$ to be pairwise disjoint. It can be verified that this will happen if

$$\frac{2\lambda R}{n \log\log R} < \frac{1}{Q_\lambda^2} \qquad (\text{for } \lambda \leqq \Lambda). \tag{15}$$

To show this we estimate λ, R, and Q_λ for $\lambda \leqq \Lambda$. By (9) and (10) we deduce that

$$\lambda \leqq \frac{3}{4} \log\log\log\log n \qquad (\text{for } \lambda \leqq \Lambda).$$

By (9) we find that $2^\lambda \leqq (\log\log C_1\delta^{-1})^{3/4}$, and hence by (8) and (12) we find that

$$\log Q_\lambda \leqq 2^\lambda \log Q_1 \leqq (\log\log C_1\delta^{-1})^{7/8} \log C_1\delta^{-1}.$$

By (6) this implies $\log Q_\lambda \leqq \log R$, and so

$$Q_\lambda \leqq R. \tag{16}$$

By (6) and (10) we find, for n large enough, that

$$3 \leqq R \leqq \log n. \tag{17}$$

From the above estimates for λ, R, and Q_λ we deduce that (15) holds for sufficiently large n. Therefore, when $\lambda \leqq \Lambda$ we have

$$\mu_\lambda |A|^2 = \left|\mathcal{P}_\lambda(K_\lambda, U_\lambda)\right| \frac{|A|^2}{U_\lambda^2} \leqq \sum_{t=0}^{N-1} \left|F_1(t/n)\right|^2 \leqq n|A|.$$

So

$$\delta \leqq \mu_\lambda^{-1}. \tag{18}$$

Let us assume, to obtain a contradiction, that

$$\sum_{\substack{t=1 \\ t/n \in \mathfrak{M}(q)}}^{n-1} \left|F(t/n)\right|^2 < \theta(\delta)|A|^2 \quad \text{(for all } 1 \leqq q \leqq R). \tag{19}$$

By using Lemma 2 and (19) we will show, provided C_1 and n are sufficiently large, that

$$\mu_{\lambda+1} \geqq \theta(\delta)^{-1/2} \mu_\lambda \quad \text{(for } 1 \leqq \lambda \leqq \Lambda). \tag{20}$$

Assuming for now that (20) holds we show how a contradiction is obtained, thus proving that the assumption (19) is false. Since $\mu_1 \geqq 1$, it follows from (20) that $\mu_{\Lambda+1} \geqq \theta(\delta)^{-(1/2)\Lambda}$, and thus by (18) we have $\delta \leqq \theta(\delta)^{(1/2)\Lambda}$. We can take C_1 to be large enough so that (9) implies $\Lambda \geqq (1/4)\log_3 C_1\delta^{-1}$, then by (7) we find that $\delta \leqq C_1^{-1}\delta < \delta$, a contradiction. Therefore (19) cannot hold for all $1 \leqq q \leqq R$.

We now proceed to show that (20) holds. To this end, let us fix λ with $1 \leqq \lambda \leqq \Lambda$. For now we also fix a rational a/k in $\mathcal{P}_\lambda(U_\lambda, K_\lambda)$. We associate with a/k a fraction $u/n \in \mathfrak{M}_\lambda(k, a)$ such that $\left|F(u/n)\right| \geqq |A|/U_\lambda$. Such a u/n exists by the way a/k was chosen.

Since $A - A$ contains no integers from $\mathcal{S}_d$ we find that

$$\sum_{t=0}^{n-1} F_1(u/n + t/n) F(-t/n) S_{d,n}(t/n) = 0.$$

By the triangle inequality, Lemma 3, and the way u/n was chosen we find that

$$\frac{|A|^2}{U_\lambda} \cdot \left(\frac{dn}{\phi(d)}\right) \ll \sum_{t=1}^{n-1} \bigl|F_1(u/n+t/n)\bigr| \, \bigl|F(t/n)\bigr| \, \bigl|S_{d,n}(t/n)\bigr|. \tag{21}$$

Set

$$Y = (C_1\delta^{-1})^{3/2} Q_\lambda^2 \tag{22}$$

and let $\mathcal{N}$ denote the set of t/n such that $\bigl|F(t/n)\bigr| \leqq |A|/Y$. By two applications of the Cauchy–Schwarz inequality, Parseval's identity, and Lemma 8 we find that

$$\sum_{t/n\in\mathcal{N}} \bigl|F_1(u/n+t/n)\bigr| \, \bigl|F(t/n)\bigr| \, \bigl|S_{d,n}(t/n)\bigr|$$

$$\leqq \left(\sum_{t=0}^{n-1} \bigl|F_1(u/n+t/n)\bigr|^2\right)^{1/2} \left(\sum_{t/n\in\mathcal{N}} \bigl|F(t/n)\bigr|^4\right)^{1/4} \left(\sum_{t=0}^{n-1} \bigl|S_{d,n}(t/n)\bigr|^4\right)^{1/4}$$

$$\ll \frac{dn^{3/2}|A|^{1/2}}{\phi(d)} \left(\sum_{t/n\in\mathcal{N}} \bigl|F(t/n)\bigr|^4\right)^{1/4}.$$

Now

$$\left(\sum_{t/n\in\mathcal{N}} \bigl|F(t/n)\bigr|^4\right)^{1/4} \leqq \max_{t/n\in\mathcal{N}} \bigl|F(t/n)\bigr|^{1/2} \left(\sum_{t=0}^{n-1} \bigl|F(t/n)\bigr|^2\right)^{1/4}$$

$$\leqq \frac{|A|^{1/2}}{Y^{1/2}} \bigl(n|A|\bigr)^{1/4} = \frac{n^{1/4}|A|^{3/4}}{Y^{1/2}}.$$

Therefore

$$\sum_{t/n\in\mathcal{N}} \bigl|F_1(u/n+t/n)\bigr| \, \bigl|F(t/n)\bigr| \, \bigl|S_{d,n}(t/n)\bigr| \ll \frac{dn^{7/4}|A|^{5/4}}{\phi(d)Y^{1/2}}.$$

By (14) and (22) we find that

$$Y^{-1/2} = C_1^{-3/4}\delta^{3/4}Q_\lambda^{-1} \leqq C_1^{-3/4}|A|^{3/4}n^{-3/4}U_\lambda^{-1},$$

thus

$$\sum_{t/n\in\mathcal{N}} \bigl|F_1(u/n+t/n)\bigr|\,\bigl|F(t/n)\bigr|\,\bigl|S_{d,n}(t/n)\bigr| \ll C_1^{-3/4}\frac{|A|^2}{U_\lambda}\left(\frac{dn}{\phi(d)}\right). \tag{23}$$

Let $\mathcal{N}_1$ denote the set of t/n such that $\bigl|F_1(u/n+t/n)\bigr| \leqq |A|/Y$. By the same reasoning used in the deduction of (23) we find that

$$\sum_{t/n\in\mathcal{N}_1} \bigl|F_1(u/n+t/n)\bigr|\,\bigl|F(t/n)\bigr|\,\bigl|S_{d,n}(t/n)\bigr| \ll C_1^{-3/4}\frac{|A|^2}{U_\lambda}\left(\frac{dn}{\phi(d)}\right). \tag{24}$$

For $\lambda \leqq \Lambda$ we have $Q_{\lambda+1}/Q_\lambda < R$. Indeed, (9) and (12) imply

$$\frac{Q_{\lambda+1}}{Q_\lambda} \leqq Q_1^{2^\Lambda} \leqq (C_1\delta^{-1})^{(\log\log C_1\delta^{-1})^{3/4}} < R.$$

Let $\mathfrak{m}^*$ denote the union of the $\mathfrak{M}(q)$ with $Q_{\lambda+1}/Q_\lambda \leqq q \leqq R$. By the Cauchy–Schwarz inequality we find that

$$\sum_{t/n\in\mathfrak{m}^*} \bigl|F_1(u/n+t/n)\bigr|\,\bigl|F(t/n)\bigr|\,\bigl|S_{d,n}(t/n)\bigr| \leqq \bigl(n|A|\bigr) \sup_{t/n\in\mathfrak{m}_\lambda^*}\bigl|S_{d,n}(t/n)\bigr|. \tag{25}$$

We are now going to show that

$$\sup_{t/n\in\mathfrak{m}_\lambda^*}\bigl|S_{d,n}(t/n)\bigr| \ll C_1^{-1}U_\lambda^{-1}\delta\left(\frac{dn}{\phi(d)}\right). \tag{26}$$

Suppose that $t/n\in\mathfrak{m}^*$, then $t/n\in\mathfrak{M}(q,a)$ for some integers a and q such that $0\leqq a\leqq q$, $(a,q)=1$, and $Q_{\lambda+1}/Q_\lambda \leqq q \leqq R$. Since $q\leqq R\leqq \log n$, we deduce from Lemma 4 that

$$S_{d,n}(t/n) \ll \frac{dn}{\phi(d)\phi(q)}.$$

Using the well-known estimate

$$\phi(q) \gg \frac{q}{\log\log q}, \tag{27}$$

(see for example [7, Theorem 328]), we obtain

$$S_{d,n}(t/n) \ll \left(\frac{dn}{\phi(d)}\right)\frac{\log\log q}{q}. \tag{28}$$

The lower bound on q implies

$$\frac{\log\log q}{q} \ll \frac{\log\log Q_{\lambda+1}/Q_\lambda}{Q_{\lambda+1}/Q_\lambda}. \tag{29}$$

By (12) we have $Q_{\lambda+1}/Q_\lambda = Q_\lambda Q_1 = Q_1^{2^\lambda}$, thus

$$\frac{\log\log Q_{\lambda+1}/Q_\lambda}{Q_{\lambda+1}/Q_\lambda} = \frac{\log\log Q_1^{2^\lambda}}{Q_\lambda Q_1} = \frac{\lambda(\log 2) + \log\log Q_1}{Q_\lambda Q_1}.$$

Using (8) and (9) we find that $\lambda \ll \log\log Q_1$, by this and (14) we obtain

$$\frac{\log\log Q_{\lambda+1}/Q_\lambda}{Q_{\lambda+1}/Q_\lambda} \ll \frac{\log\log Q_1}{U_\lambda Q_1}.$$

Using (8) we find, by taking C_1 large enough, that

$$\log\left(\frac{\log\log Q_1}{Q_1}\right) \leqq -\log C_1\delta^{-1},$$

and thus

$$\frac{\log\log Q_1}{Q_1} \leqq C_1^{-1}\delta.$$

From (29) and the subsequent estimates we obtain

$$\frac{\log\log q}{q} \ll C_1^{-1}U_\lambda^{-1}\delta, \tag{30}$$

Since $t/n \in \mathfrak{m}^*$ is arbitrary (28) and (30) imply that (26) is true. By (25) and (26) we have

$$\sum_{t/n\in\mathfrak{m}^*} \left|F_1(u/n+t/n)\right| \left|F(t/n)\right| \left|S_{d,n}(t/n)\right| \ll C_1^{-1}\frac{|A|^2}{U_\lambda}\left(\frac{dn}{\phi(d)}\right). \tag{31}$$

The contribution to the sum in (21) coming from the terms with $t/n \in \mathfrak{m}$ can similarly be bounded. By the Cauchy–Schwarz inequality and Lemma 5 we find that

$$\sum_{t/n\in\mathfrak{m}} \left|F_1(u/n+t/n)\right| \left|F(t/n)\right| \left|S_{d,n}(t/n)\right| \leqq \big(n|A|\big) \sup_{t/n\in\mathfrak{m}} \left|S(t/n)\right|$$

$$\ll \left(n|A|\right)\left(\frac{dn}{\phi(d)}\right)\frac{\log\log R}{R}.$$

Since $R \geqq Q_{\lambda+1}/Q_\lambda$ the argument used in the previous paragraph implies

$$\sum_{t/n\in\mathfrak{m}} \left|F_1(u/n+t/n)\right| \left|F(t/n)\right| \left|S_{d,n}(t/n)\right| \ll C_1^{-1}\frac{|A|^2}{U_\lambda}\left(\frac{dn}{\phi(d)}\right). \tag{32}$$

Let $\mathfrak{N}(b,a)$ be the set of $t/n \in \mathfrak{M}(b,a)$ with $t/n \neq 0$ such that

$$\left|F(t/n)\right| \geqq \frac{|A|}{Y}, \qquad \left|F_1(u/n+t/n)\right| \geqq \frac{|A|}{Y}.$$

By (23), (24), (31), and (32) it follows for C_1 large enough that

$$\frac{d|A|^2 n}{\phi(d)U_\lambda} \ll \sum_{b\leqq Q_{\lambda+1}/Q_\lambda}\ \sum_{(a,b)=1}\ \max_{t/n\in\mathfrak{N}(b,a)}\left|F(t/n)\right|$$

$$\times \max_{t/n\in\mathfrak{N}(b,a)}\left|F_1(u/n+t/n)\right| \sum_{t/n\in\mathfrak{M}(b,a)}\left|S_{d,n}(t/n)\right|.$$

Since $d \leqq \log n$ we can apply Lemma 6 to the inner sum above to obtain

$$\frac{|A|^2}{U_\lambda \log R}$$

$$\ll \sum_{b\leqq Q_{\lambda+1}/Q_\lambda}\frac{1}{\phi(b)}\sum_{(a,b)=1}\ \max_{t/n\in\mathfrak{N}(b,a)}\left|F(t/n)\right| \max_{t/n\in\mathfrak{N}(b,a)}\left|F_1(u/n+t/n)\right|.$$

Let $\mathcal{L}(L,V,W)$ denote the set of reduced fractions $b/l \in [0,1]$ such that

$$\frac{L}{2} \leqq l \leqq L, \quad \frac{|A|}{V} \leqq \max_{t/n\in\mathfrak{M}(l,b)}\left|F(t/n)\right| \leqq 2\frac{|A|}{V},$$

$$\frac{|A|}{W} \leqq \max_{t/n\in\mathfrak{M}(l,b)}\left|F_1(u/n+t/n)\right| \leqq 2\frac{|A|}{W}.$$

For $b/l \in \mathcal{L}(L,V,W)$, we have

$$\frac{1}{\phi(l)}\max_{t/n\in\mathfrak{M}(l,b)}\left|F(t/n)\right| \max_{t/n\in\mathfrak{M}(l,b)}\left|F_1(u/n+t/n)\right| \ll \frac{(\log\log 3L)|A|^2}{LVW}$$

by (27). Therefore

$$\frac{|A|^2}{U_\lambda \log R} \ll \sum_L \sum_V \sum_W \left|\mathcal{L}(L,V,W)\right| \frac{(\log\log 3L)|A|^2}{LVW}.$$

where L runs through all the powers of 2 in the interval $[1, 2Q_{\lambda+1}/Q_\lambda]$, and V and W run through all the powers of 2 in the interval $[1, 2Y]$. There must exist a triple (L,V,W) of indices such that

$$\left|\mathcal{L}(L,V,W)\right| \gg \frac{LVW}{U_\lambda(\log\log 3L)(\log R)}.$$

We associate this triple with a/k.

The number of possible triples (L,V,W) is $\ll \log(Q_{\lambda+1}/Q_\lambda)(\log Y)^2$, which by (16) and (22) is $\ll (\log R)^3$. Therefore there exists a subset $\mathcal{K} \subset \mathcal{P}_\lambda$, satisfying

$$|\mathcal{K}| \gg \frac{\left|P_\lambda(K_\lambda, U_\lambda)\right|}{(\log R)^3}, \tag{33}$$

such that to each $a/k \in \mathcal{K}$ we associate the same triple, say (L,V,W).

Let $a/k \in \mathcal{K}$, then together with the associated fraction $u/n \in \mathfrak{M}_\lambda(k,a)$, we associate a set $\mathcal{L}_{a/k}$ of rationals b/l, $0 \leqq b \leqq l$, $(b,l)=1$, $L/2 \leqq l \leqq L$, such that

$$|\mathcal{L}_{a/k}| \gg \frac{LVW}{U_\lambda(\log\log 3L)(\log R)}, \tag{34}$$

$$\frac{|A|}{V} \leqq \max_{v/n \in \mathfrak{M}(l,b)} \left|F(v/n)\right| \leqq \frac{2|A|}{V}, \tag{35}$$

$$\frac{|A|}{W} \leqq \max_{w/n \in \mathfrak{M}(l,b)} \left|F_1(u/n + w/n)\right| \leqq \frac{2|A|}{W}. \tag{36}$$

Set

$$\mathcal{Q} = \left\{ \frac{a}{k} + \frac{b}{l} : \frac{a}{k} \in \mathcal{K},\ \frac{b}{l} \in \mathcal{L}_{a/k} \right\}.$$

Let us estimate the cardinality of $\mathcal{Q}$. Since $L \leqq Q_{\lambda+1}/Q_\lambda \leqq R$, assumption (19) and (35) imply

$$\left|\left\{ b : \frac{b}{l} \in \bigcup \mathcal{L}_{a/k} \right\}\right| \left(\frac{|A|}{V}\right)^2 \leqq \sum_{t/n \in \mathfrak{M}(l)} \left|F(t/n)\right|^2 \leqq \theta(\delta)|A|^2.$$

So that

$$\left|\left\{b:\ \frac{b}{l}\in\bigcup\mathcal{L}_{a/k}\right\}\right|\ll\theta(\delta)V^2.$$

Lemma 2 then implies

$$|\mathcal{Q}|\gg|\mathcal{K}|\cdot\frac{L^2V^2W^2}{U_\lambda^2(\log\log 3L)^2(\log R)^2}\cdot\frac{\theta(\delta)^{-1}}{LV^2\tau^8(1+\log K_\lambda)}.$$

From (14) and (16) we obtain $\log K_\lambda \leqq \log R$, by this and (33) it follows that

$$|\mathcal{Q}|\gg W^2\left(\frac{\theta(\delta)^{-1}}{\tau^8(\log R)^6}\right)\frac{\left|\mathcal{P}_\lambda(K_\lambda,U_\lambda)\right|}{U_\lambda^2}. \tag{37}$$

Note that $\mathcal{Q}$ is a subset of $(0,2]$. Let $\mathcal{Q}_1=\mathcal{Q}\cap(0,1]$ and $\mathcal{Q}_2=\mathcal{Q}\cap(1,2]$. Let us assume without loss of generality that $|\mathcal{Q}_1|\geqq(1/2)|\mathcal{Q}|$. If this is not the case, then $|\mathcal{Q}_2|\geqq(1/2)|\mathcal{Q}|$, and we can replace $\mathcal{Q}_1$ in the argument below by the rational numbers in $\mathcal{Q}_2$ shifted to the left by 1. Since $|\mathcal{Q}_1|\geqq(1/2)|\mathcal{Q}|$ we see that (37) is still valid with $\mathcal{Q}$ replaced by $\mathcal{Q}_1$

Let $r/s=a/k+b/l$ be in $\mathcal{Q}_1$. For $u/n\in\mathfrak{M}_\lambda(k,a)$ and $w/n\in\mathfrak{M}(l,b)$ we have

$$\left|\frac{r}{s}-\left(\frac{u}{n}+\frac{w}{n}\right)\right|\leqq\left|\frac{u}{n}-\frac{a}{k}\right|+\left|\frac{w}{n}-\frac{b}{l}\right|\leqq\frac{(\lambda+1)R}{n\log\log R},$$

and therefore $u/n+w/n\in\mathfrak{M}_{\lambda+1}(s,r)$. Thus, by (36) we deduce that

$$\max_{t/n\in\mathfrak{M}_{\lambda+1}(s,r)}\left|F_1(t/n)\right|\geqq\frac{|A|}{W}\quad(\text{for } r/s\in\mathcal{Q}_1). \tag{38}$$

We now estimate the size of the denominator of r/s. Certainly $s\leqq kl\leqq K_\lambda L$. By (14) we have $K_\lambda\leqq Q_\lambda$ and L was chosen to satisfy $L\leqq Q_{\lambda+1}/Q_\lambda$. Therefore $s\leqq Q_{\lambda+1}$ whenever $r/s\in\mathcal{Q}_1$. By this and (38) we obtain

$$\mathcal{Q}_1\subset\mathcal{P}_{\lambda+1}(Q_{\lambda+1},W). \tag{39}$$

By (37), with $\mathcal{Q}$ replaced by $\mathcal{Q}_1$, and (39) we find that

$$\frac{\left|\mathcal{P}_{\lambda+1}(Q_{\lambda+1},W)\right|}{W^2}\gg\left(\frac{\theta(\delta)^{-1}}{\tau^8(\log R)^6}\right)\frac{\left|\mathcal{P}_\lambda(K_\lambda,U_\lambda)\right|}{U_\lambda^2}.$$

This implies

$$\mu_{\lambda+1}\gg\frac{\theta(\delta)^{-1}}{\tau^8(\log R)^6}\mu_\lambda. \tag{40}$$

We now estimate τ, the maximum of the divisor function up to $K_\lambda L \leqq Q_{\lambda+1}$. If $d(m)$ is the number of divisors of m then

$$\log d(m) \ll \frac{\log m}{\log\log m},$$

(see [7, Theorem 317]). Thus, by (12), we have

$$\log \tau \ll \frac{\log Q_{\lambda+1}}{\log\log Q_{\lambda+1}} \ll \frac{2^\lambda \log Q_1}{\log\log Q_1},$$

and since $\lambda \leqq \Lambda$ we deduce from (8) and (9) that

$$\log \tau \ll \frac{\log C_1\delta^{-1}}{(\log\log C_1\delta^{-1})^{1/4}}.$$

It follows from (7) that

$$\log \tau = o\big(\log \theta(\delta)^{-1}\big) \qquad (\text{for } C_1\delta^{-1} \to \infty). \tag{41}$$

We also find from (6) and (7) that

$$\log\log R = o\big(\log \theta(\delta)^{-1}\big) \qquad (\text{for } C_1\delta^{-1} \to \infty). \tag{42}$$

Since $\theta(\delta)^{-1}$ tends to infinity as $C_1\delta^{-1}$ tends to infinity, we deduce from (40), (41), and (42) that for C_1 sufficiently large

$$\mu_{\lambda+1} \geqq \theta(\delta)^{-1/2}\mu_\lambda.$$

Since $\lambda \leqq \Lambda$ was arbitrary, (20) is true and as shown earlier, the lemma can be deduced from this. □

We now derive a density increment argument that will be iterated in the next section to prove our theorem.

LEMMA 10. *Let d be a positive integer such that $d \leqq \log n$. Suppose that $A - A$ does not intersect $\mathcal{S}_d$ and that δ, the density of A, satisfies* (10). *Provided C_1 and n are sufficiently large there exist positive integers d' and n', and a subset A' of $\{1, \dots, n'\}$ of size $\delta' n'$, such that $A' - A'$ does not intersect $\mathcal{S}_{d'}$, and moreover*

$$d \leqq d' \leqq R(\delta)d, \quad R(\delta)^{-2}n \leqq n' \leqq n, \quad \delta' \geqq \delta\big(1 + 8^{-1}\theta(\delta)\big).$$

PROOF. By the hypotheses, Lemma 9 implies that there exists a positive integer $q \leqq R(\delta)$ such that (11) is true. With this q and $U = R(\delta)/\log\log R(\delta)$ let E be defined as in Lemma 1. Note that $\mathfrak{M}(q) \subset E$. The inequality (17) is still valid, thus $2\pi qU \leqq 2\pi R(\delta)^2 \leqq n$ for sufficiently large n. Therefore, we can apply Lemma 1 with $\theta = \theta(\delta)$ to deduce that there exists an arithmetic progression P with difference q such that

$$|P| \geqq \frac{n \log\log R(\delta)}{32\pi q R(\delta)} \tag{43}$$

and

$$|A \cap P| \geqq |P|\delta\bigl(1 + 8^{-1}\theta(\delta)\bigr). \tag{44}$$

Let $n' = |P|$. Then there exists an integer c and subset A' of $\{1, \dots, n'\}$ such that $A \cap P = \{c + qa' : a' \in A'\}$. Put $d' = dq$. Since $A - A$ does not intersect $\mathcal{S}_d$, we deduce that A' does not intersect $\mathcal{S}_{dq}$. Let the size of A' be $\delta' n'$. Then (44) implies

$$\delta' \geqq \delta\bigl(1 + 8^{-1}\theta(\delta)\bigr).$$

To finish we need to estimate n' and d'. Since $q \leqq R(\delta)$ we find by (43) and for C_1 large enough that $n' \geqq R(\delta)^{-2} n$, and clearly, $n' \leqq n$. Now, again by the fact that $q \leqq R(\delta)$, we obtain $q \leqq d' = dq \leqq R(\delta)q$. □

5. Proof of Theorem

Let us assume, for a contradiction, that the theorem is false. Then for C_1 and n sufficiently large, there exists a subset A of $\{1, \dots, n\}$ of size δn, such that $A - A$ does not intersect $\mathcal{S}$ and

$$\delta \geqq C_1 \left(\frac{\log_2 n}{(\log_3 n)^2 (\log_4 n)} \right)^{-\log_5 n}. \tag{45}$$

Set

$$Z = \bigl[64\,\theta(\delta)^{-1} \log C_1 \delta^{-1}\bigr], \tag{46}$$

and put $d_0 = 1$, $n_0 = n$, $A_0 = A$, and $\delta_0 = \delta$. By using Lemma 10 repeatedly we can show that for each integer k, with $1 \leqq k \leqq Z$, there are integers d_k and n_k and a subset A_k of $\{1, \dots, n_k\}$ of size $\delta_k n_k$ such that $A_k - A_k$ does not intersect $\mathcal{S}_{d_k}$. Moreover, d_k, n_k, and δ_k satisfy

$$d_{k-1} \leqq d_k \leqq R(\delta_{k-1}) d_{k-1}, \qquad R(\delta_{k-1})^{-2} n_{k-1} \leqq n_k \leqq n_{k-1},$$
$$\delta_k \geqq \delta_{k-1}\bigl(1 + 8^{-1}\theta(\delta_{k-1})\bigr).$$

Since $d_0 = 1$ and $n_0 = n$, these estimates imply

$$d_k \leqq R(\delta)^k, \quad n_k \geqq R(\delta)^{-2k} n, \quad \delta_k \geqq \delta\big(1 + 8^{-1}\theta(\delta)\big)^k. \tag{47}$$

Let us show that we can actually perform this iteration Z times. Let $0 \leqq l \leqq Z - 1$, and suppose that we have performed this iteration l times. To show that Lemma 10 can be applied an $(l+1)$-st time we need to show that n_l is sufficiently large, $d_l \leqq \log n_l$, and that (10) is satisfied with δ replaced by δ_l.

We begin by estimating n_l. By (47) we obtain

$$\log n_l \geqq \log n - 2l \log R(\delta). \tag{48}$$

Since $l < Z$, (6) and (46) imply

$$l \log R(\delta) \leqq 64\, \theta(\delta)^{-1} (\log C_1 \delta^{-1})^2 (\log_2 C_1 \delta^{-1})^{7/8}.$$

By (45) we obtain

$$(\log C_1 \delta^{-1})^2 (\log_2 C_1 \delta^{-1})^{3/4} \leqq 2 (\log_3 n)^2 (\log_4 n)^{7/8} (\log_5 n)^2$$

for large enough n. By (7) and (45) we find, for n and C_1 sufficiently large, that

$$\log \theta(\delta)^{-1} = \frac{4 \log C_1 \delta^{-1}}{\log_3 C_1 \delta^{-1}} \leqq \log \left(\frac{\log_2 n}{(\log_3 n)^2 (\log_4 n)} \right).$$

(Here we used that $(\log x)(\log_3 x)^{-1}$ is eventually increasing.) Therefore

$$\theta(\delta)^{-1} \leqq \frac{\log_2 n}{(\log_3 n)^2 (\log_4 n)}.$$

From the above we deduce, for n and C_1 large enough, that

$$l \log R(\delta) \leqq \log_2 n. \tag{49}$$

Therefore, by (48),

$$\log n_l \geqq \log n - 2 \log_2 n = \log \left(\frac{n}{(\log n)^2} \right),$$

and so

$$n_l \geqq \frac{n}{(\log n)^2} \tag{50}$$

for $l < Z$. This shows that by taking n to be arbitrarily large, the same is true for n_l.

We now show that $d_l \leqq \log n_l$. By (47) we have $\log d_l \leqq l \log R(\delta)$, and thus by (49) we obtain $\log d_l \leqq (1/2) \log_2 n$. For large n this implies

$$d_l \leqq (\log n)^{1/2} \leqq \log \frac{n}{(\log n)^2} \leqq \log n_l$$

by (50).

We leave it to the reader to verify that (45) and (50) imply, for n and C_1 sufficiently large, that (10) is satisfied with δ and n replaced by δ_l and n_l respectively. Finally, since $A_l - A_l$ does not intersect $\mathcal{S}_{d_l}$ we can apply Lemma 10 to obtain the desired outcome.

Since (47) is true with $k = Z$ we find that

$$\log \delta_Z \geqq Z \log\left(1 + 8^{-1}\theta(\delta)\right) - \log C_1\delta^{-1}.$$

Since $8^{-1}\theta(\delta) < 1$, this implies

$$\log \delta_Z \geqq 16^{-1} Z\theta(\delta) - \log C_1\delta^{-1}. \tag{51}$$

(Here we used $\log(1 + x) \geqq x/2$ for $0 \leqq x \leqq 1$.) For C_1 large enough $Z \geqq 32\theta(\delta)^{-1} \log C_1\delta^{-1}$, thus

$$\log \delta_Z \geqq 2 \log C_1\delta^{-1} - \log C_1\delta^{-1} > 0.$$

This implies $\delta_Z > 1$, a contradiction, since by definition $\delta_Z \leqq 1$. This contradiction establishes the theorem.

Acknowledgement. The author was supported by a postdoctoral fellowship from the Centre de Recherches Mathématiques at Montréal.

References

[1] A. Balog, J. Pelikán, J. Pintz and E. Szemerédi, Difference sets without κ-powers, *Acta Math. Hungar.,* **65** (1994), 165–187.

[2] H. Furstenberg, Ergodic behaviour of diagonal measures and a theorem of Szemerédi on arithmetic progressions, *J. Analyse Math.,* **31** (1977), 204–256.

[3] W. T. Gowers, A new proof of Szemerédi's theorem, *Geom. Funct. Anal.,* **11** (2001), 465–588.

[4] B. Green, On arithmetic structures in dense sets of integers, *Duke Math. J.,* **114** (2002), 215–238.

[5] H. Halberstam and H.-E. Richert, *Sieve Methods,* Academic Press (London, 1974).

[6] T. Kamae and M. Mendès France, Van der Corput's difference theorem, *Isreal J. Math.,* **31** (1978), 335–342.

[7] G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers,* Oxford University Press (Oxford, 1960).
[8] J. Lucier, Intersective sets given by a polynomial, *Acta Arith.,* **123** (2006), 57–95.
[9] H. L. Montgomery, *Topics in Multiplicative Number Theory*, Lecture Notes in Math., 127, Springer-Verlag (Berlin, 1971).
[10] J. Pintz, W. L. Steiger and E. Szemerédi, On sets of natural numbers whose difference set contains no squares, *J. London Math. Soc.,* **37** (1988), 219–231.
[11] K. F. Roth, On certain sets of integers, *J. London Math. Soc.,* **28** (1953), 104–109.
[12] A. Sárközy, On difference sets of sequences on integers. I, *Acta Math. Acad. Sci. Hungar.,* **31** (1978), 125–149.
[13] A. Sárközy, On difference sets of sequences on integers. III, *Acta Math. Acad. Sci. Hungar.,* **31** (1978), 355–386.