TWO THEOREMS OF SÁRKÖZY

NEIL LYALL ALEX RICE

Abstract. In this note, we provide parallel expositions of two theorems of Sárközy, the qualitative versions of which state that any set of natural numbers of positive upper density necessarily contains two distinct elements which differ by a perfect square, as well as two elements which differ by one less than a prime number. We use simplified versions of Sárközy's original methods, and the proofs are self-contained with the exception of the Weyl Inequality, minor arc estimates of Vinogradov, and the Siegel-Walfisz Theorem on primes in arithmetic progressions.

1. Introduction

In a series of papers in the late 1970s, Sárközy [8],[9] confirmed conjectures of Lovász and Erdős, respectively, showing that any set $A \subseteq \mathbb{N}$ of *positive upper density*, i.e. satisfying

$$\limsup_{N\to\infty} \frac{|A \cap [1,N]|}{N} > 0,$$

necessarily contains two distinct elements which differ by a perfect square, as well as two elements which differ by one less than a prime number. Furstenberg [1] established the result for squares independently via ergodic theory, yielding no quantitative information, while Sárközy's approach was Fourier analytic, using the Hardy-Littlewood circle method to employ a density increment strategy inspired by Roth's proof of the analogous conjecture for three-term arithmetic progressions [6].

Here we use simplified versions of Sárközy original methods to establish the following quantitative estimates, using $A-A$ to denote the difference set $\{a-a' : a,a' \in A\}$, $[1,N]$ to denote $\{1,2,\dots,N\}$, and $\ll$ to denote "less than some constant times".

Theorem 1. *If $A \subseteq [1,N]$ and $n^2 \notin A-A$ for all $n \in \mathbb{N}$, then*

$$\frac{|A|}{N} \ll \frac{\log\log N}{\log N}. \tag{1}$$

Theorem 2. *If $A \subseteq [1,N]$ and $p-1 \notin A-A$ for all primes p, then,*

$$\frac{|A|}{N} \ll e^{-c(\log\log N)^{1/3}} \tag{2}$$

for some absolute constant $c>0$.

While these bounds are better than those obtained by Sárközy, they are not the best known. Pintz, Steiger, and Szemerédi [5] replaced (1) with

$$\frac{|A|}{N} \ll (\log N)^{-c\log\log\log\log N},$$

and the constant c has since been improved to $1/\log 3$. Ruzsa and Sanders [7] replaced (2) with

$$\frac{|A|}{N} \ll e^{-c(\log N)^{1/4}}$$

for some absolute constant $c>0$.

Technical Remark. We use the letters C and c to denote appropriately large or small absolute constants, which change from step to step. At the expense of the implied constants in (1) and (2), we are free to insist that the parameter N is sufficiently large, which we take as a perpetual hypothesis and abstain from including further.

2. Main Iteration Lemmas: Deducing Theorems 1 and 2

The principle behind a density increment strategy is that a set which lacks the desired arithmetic structure should spawn a new, significantly denser subset of a slightly smaller interval with an inherited lack of arithmetic structure. Iterating this procedure enough times for the density to reach 1 provides an upper bound on the density of the original set. With the following two lemmas, we make this idea precise and show how the particulars in each case yield the bounds claimed in Theorems 1 and 2.

Lemma 1. *Suppose $A \subseteq [1, N]$ with $|A| = \delta N$ and $\delta \geq N^{-1/20}$. If $n^2 \notin A - A$ for all $n \in \mathbb{N}$, then there exists $A' \subseteq [1, N']$ with*

$$N' \gg \delta^7 N, \quad |A'| \geq (\delta + c\delta^2)N', \quad \text{and} \quad n^2 \notin A' - A' \text{ for all } n \in \mathbb{N}.$$

2.1. **Proof of Theorem 1.** Suppose $A \subseteq [1, N]$ with $|A| = \delta N$ and $n^2 \notin A - A$ for all $n \in \mathbb{N}$.

Setting $A_0 = A$, $N_0 = N$, and $\delta_0 = \delta$, Lemma 1 yields, for each m, a set $A_m \subseteq [1, N_m]$ with $|A_m| = \delta_m N_m$ and $n^2 \notin A_m - A_m$ for all $n \in \mathbb{N}$ satisfying

$$N_m \geq c\delta^7 N_{m-1} \geq (c\delta^7)^m N \tag{3}$$

and

$$\delta_m \geq \delta_{m-1} + c\delta_{m-1}^2 \tag{4}$$

as long as

$$\delta_m \geq N_m^{-1/20}. \tag{5}$$

By (4), we see that the density δ_m will surpass 1, and hence (5) must fail, for $m = C\delta^{-1}$. In particular, by (3) we have

$$\delta \leq (c\delta^7)^{-C\delta^{-1}} N^{-1/20},$$

which can be rearranged to

$$N \leq (c\delta)^{-C\delta^{-1}}$$

and seen to imply

$$\delta \ll \frac{\log\log N}{\log N},$$

as required. □

We see from Lemma 1 that in the case of square differences, the new, denser set inherits the identical lack of structure, but things get slightly trickier in the other setting. Foreshadowing this issue, we let $\mathcal{P}$ denote the primes, and for $d \in \mathbb{N}$ we define $\Lambda_d = \{n \in \mathbb{N} : dn + 1 \in \mathcal{P}\}$.

Lemma 2. *Suppose $A \subseteq [1, N]$ with $|A| = \delta N$. If $(A - A) \cap \Lambda_d = \emptyset$ and*

$$d, \delta^{-1} \leq \log N, \tag{6}$$

then there exists $A' \subseteq [1, N']$ and $q \ll \delta^{-2}$ with

$$N' \gg \delta^5 N, \quad \frac{|A'|}{N'} \geq \delta + \frac{c\delta}{\log(C\delta^{-1})}, \quad \text{and} \quad (A' - A') \cap \Lambda_{qd} = \emptyset.$$

2.2. **Proof of Theorem 2.** Suppose $A \subseteq [1,N]$ with $|A| = \delta N$ and $p-1 \notin A-A$ for all $p \in \mathcal{P}$.

Setting $A_0 = A$, $N_0 = N$, $\delta_0 = \delta$, and $d_0 = 1$, Lemma 2 yields, for each m, a set $A_m \subseteq [1, N_m]$ with $|A_m| = \delta_m N_m$ and $(A_m - A_m) \cap \Lambda_{d_m} = \emptyset$ satisfying

$$N_m \geq c\delta^5 N_{m-1} \geq (c\delta^5)^m N, \tag{7}$$

$$\delta_m \geq \delta_{m-1} + \frac{c\delta_{m-1}}{\log(C\delta_{m-1}^{-1})}, \tag{8}$$

and

$$d_m \leq C\delta^{-2} d_{m-1} \leq (C\delta^{-2})^m \tag{9}$$

as long as

$$d_m, \delta_m^{-1} \leq \log N_m. \tag{10}$$

By (8), we see that the density δ_m will surpass 1 for $m = C(\log(C\delta^{-1}))^2$. Therefore, if

$$\delta \geq e^{-c(\log\log N)^{1/3}} \tag{11}$$

for an absolute constant $c > 0$, then (10) must fail for

$$m = C(\log\log N)^{2/3}. \tag{12}$$

However, we see that if c is sufficiently small, then (9), (11), and (12) imply

$$d_m \leq e^{3c(\log\log N)^{1/3} m} \leq e^{\log\log N/2} = \sqrt{\log N},$$

and similarly (7), (11), and (12) imply $N_m \geq N/\log N$. In particular (10) holds, yielding a contradiction, and the theorem follows. □

3. Preliminaries

3.1. **Fourier Analysis on $\mathbb{Z}$.** We embed our finite sets in $\mathbb{Z}$, on which we utilize the discrete Fourier transform. Specifically, for a function $F : \mathbb{Z} \to \mathbb{C}$, we define $\widehat{F} : \mathbb{T} \to \mathbb{C}$, where $\mathbb{T}$ denotes the circle parametrized by the interval $[0,1]$ with 0 and 1 identified, by

$$\widehat{F}(\alpha) = \sum_{n \in \mathbb{Z}} F(n) e^{-2\pi i n\alpha}.$$

Given $N \in \mathbb{N}$ and a set $A \subseteq [1,N]$ with $|A| = \delta N$, we examine the Fourier analytic behavior of A by considering the *balance function*, f_A, defined by

$$f_A = 1_A - \delta 1_{[1,N]}.$$

We analyze the behavior of $\widehat{f_A}$ using the Hardy-Littlewood circle method, decomposing the frequency space into two pieces: the points on the circle which are close to rationals with small denominator, and those which are not.

Definition 1. Given $N \in \mathbb{N}$ and $\eta > 0$, we define, for each $q \in \mathbb{N}$ and $a \in [1,q]$,

$$\mathbf{M}_{a/q} = \mathbf{M}_{a/q}(N,\eta) = \left\{\alpha \in \mathbb{T} : |\alpha - \frac{a}{q}| < \frac{1}{\eta^2 N}\right\} \quad \text{and} \quad \mathbf{M}_q = \bigcup_{(a,q)=1} \mathbf{M}_{a/q}.$$

We then define $\mathfrak{M}$, the *major arcs*, by

$$\mathfrak{M} = \bigcup_{q=1}^{\eta^{-2}} \mathbf{M}_q,$$

and $\mathfrak{m}$, the *minor arcs*, by

$$\mathfrak{m} = \mathbb{T} \setminus \mathfrak{M}.$$

3.2. **Counting Primes in Arithmetic Progressions.** As indicated by the definition of Λ_d, we need some information about the distribution of primes in certain congruence classes. Classical estimates of this type come from the famous Siegel-Walfisz Theorem, which can be found for example in Corollary 11.19 of [4].

Lemma 3 (Siegel-Walfisz Theorem)**.** *If $q \leq (\log x)^D$ and $(a,q) = 1$, then*

$$\psi(x,a,q) := \sum_{\substack{p \in \mathcal{P}_x \\ p \equiv a \bmod q}} \log p = x/\phi(q) + O(xe^{-c\sqrt{\log x}})$$

for some constant $c = c(D) > 0$, where $\mathcal{P}_x = \mathcal{P} \cap [1,x]$.

4. Density Increment Strategy: Deducing Lemmas 1 and 2

Morally, a lack of square differences or Λ_d differences in a set A represents highly nonrandom behavior, which should be detectable in the Fourier analytic behavior of A. Specifically, we follow the approach of Lyall and Magyar [2] to locate a single small denominator q such that $\widehat{f_A}$ has large L^2 concentration around rationals with denominator q, then we use that information to show that A has increased density on a long arithmetic progression of step size q. From this correlation, we can quickly establish Lemmas 1 and 2.

Lemma 4 (L^2 concentration for squares)**.** *Suppose $A \subseteq [1,N]$ with $|A| = \delta N$, and let $\eta = c_0\delta$ for a sufficiently small constant $c_0 > 0$. If $n^2 \notin A - A$ for all $n \in \mathbb{N}$, $\delta \geq N^{-1/20}$, and $|A \cap (N/9, 8N/9)| \geq 3\delta N/4$, then there exists $q \leq \eta^{-2}$ such that*

$$\int_{\mathbf{M}_q} |\widehat{f_A}(\alpha)|^2 \mathrm{d}\alpha \gg \delta^3 N.$$

Lemma 5 (L^2 concentration for Λ_d)**.** *Suppose $A \subseteq [1,N]$ with $|A| = \delta N$, and let $\eta = c_0\delta$. If $d, \delta^{-1} \leq \log N$, $(A-A) \cap \Lambda_d = \emptyset$, and $|A \cap (N/9, 8N/9)| \geq 3\delta N/4$, then there exists $q \leq \eta^{-2}$ such that*

$$\int_{\mathbf{M}_q} |\widehat{f_A}(\alpha)|^2 \mathrm{d}\alpha \gg \frac{\delta^2 N}{\log(\eta^{-1})}.$$

Lemma 6 (Density Increment)**.** *Suppose $A \subseteq [1,N]$ with $|A| = \delta N$, and let $\eta = c_0\delta$. If $\sigma \leq 1/4\pi$ and*

$$\int_{\mathbf{M}_q} |\widehat{f_A}(\alpha)|^2 \mathrm{d}\alpha \geq \sigma\delta^2 N,$$

then there exists an arithmetic progression

$$P = \{x + \ell q : 1 \leq \ell \leq L\}$$

with $L = \lfloor \eta^2 \sigma N/q \rfloor$ and $|A \cap P|/L \geq \delta + \sigma\delta/32$.

4.1. **Proof of Lemma 1.** Suppose $A \subseteq [1,N]$, $|A| = \delta N$, $\delta \geq N^{-1/20}$, and $n^2 \notin A - A$ for all $n \in \mathbb{N}$.

If $|A \cap (N/9, 8N/9)| < 3\delta N/4$, then $\max\{|A \cap [1, N/9]|, |A \cap [8N/9, N]|\} > \delta N/8$. In other words, A has density at least $9\delta/8$ on one of these intervals.

Otherwise, Lemmas 4 and 6 apply, so in either case, letting $\eta = c_0\delta$, there exists $q \leq \eta^{-2}$ and an arithmetic progression

$$P = \{x + \ell q : 1 \leq \ell \leq L\}$$

with $qL \gg \delta^3 N$ and $|A \cap P|/L \geq \delta + c\delta^2$. Partitioning P into subprogressions of step size q^2, the pigeonhole principle yields a progression

$$P' = \{y + \ell q^2 : 1 \leq \ell \leq N'\} \subseteq P$$

with $N' \geq L/2q$ and $|A \cap P'|/N' \geq \delta + c\delta^2$. This allows us to define a set $A' \subseteq [1,N']$ by

$$A' = \{\ell \in [1,N'] : y + \ell q^2 \in A\},$$

which clearly satisfies $|A'| \geq (\delta + c\delta^2)N'$ and $N' \gg \delta^3 N/q^2 \gg \delta^7 N$. Moreover, one can easily check that due to our choice of a perfect square step size, A' inherits the lack of square differences from A. □

Lemma 2 follows from Lemmas 5 and 6 in a virtually identical fashion, except without the need to pass to a subprogression of square step size.

5. Proof of Lemmas 4, 5, and 6

5.1. **Proof of Lemma 4.** Suppose $A \subseteq [1,N]$ with $|A| = \delta N$, and let $\eta = c_0\delta$ and $M = \lfloor \sqrt{N}/3 \rfloor$. If $n^2 \notin A - A$ for all $n \in \mathbb{N}$, we see that

$$\begin{aligned}\sum_{\substack{n\in\mathbb{Z}\\1\le m\le M}} f_A(n)f_A(n+m^2) &= \sum_{\substack{n\in\mathbb{Z}\\1\le m\le M}} 1_A(n)1_A(n+m^2) - \delta \sum_{\substack{n\in\mathbb{Z}\\1\le m\le M}} 1_A(n)1_{[1,N]}(n+m^2)\\ &\quad - \delta \sum_{\substack{n\in\mathbb{Z}\\1\le m\le M}} 1_{[1,N]}(n-m^2)1_A(n) + \delta^2 \sum_{\substack{n\in\mathbb{Z}\\1\le m\le M}} 1_{[1,N]}(n)1_{[1,N]}(n+m^2)\\ &\le \Big(\delta^2 N - \delta(|A\cap[1,8N/9)| + |A\cap(N/9,N]|)\Big)M.\end{aligned}$$

Therefore, if $|A\cap(N/9,8N/9)| \ge 3\delta N/4$, we have

$$\sum_{\substack{n\in\mathbb{Z}\\1\le m\le M}} f_A(n)f_A(n+m^2) \le -\delta^2 NM/2. \tag{13}$$

One can easily check using the orthogonality relation

$$\int_0^1 e^{2\pi i n\alpha}\mathrm{d}\alpha = \begin{cases}1 & \text{if } n=0\\ 0 & \text{else}\end{cases}$$

that

$$\sum_{\substack{n\in\mathbb{Z}\\1\le m\le M}} f_A(n)f_A(n+m^2) = \int_0^1 |\widehat{f_A}(\alpha)|^2 S_M(\alpha)\mathrm{d}\alpha, \tag{14}$$

where

$$S_x(\alpha) = \sum_{m=1}^{x} e^{2\pi i m^2\alpha}.$$

From (13) and (14), we have

$$\int_0^1 |\widehat{f_A}(\alpha)|^2 |S_M(\alpha)|\mathrm{d}\alpha \ge \delta^2 NM/2. \tag{15}$$

It follows from traditional Weyl sum estimates that if $\delta \ge N^{-1/20}$, then

$$|S_M(\alpha)| \ll q^{-1/2}M \quad \text{if} \quad \alpha \in \mathbf{M}_q \subseteq \mathfrak{M} \tag{16}$$

and

$$|S_M(\alpha)| \le C\eta M \le \delta M/4 \quad \text{for all} \quad \alpha \in \mathfrak{m}, \tag{17}$$

provided we choose $c_0 \le 1/4C$. We will discuss these estimates in more detail in Appendix A. By (17) and Plancherel's Identity, we have

$$\int_{\mathfrak{m}} |\widehat{f_A}(\alpha)|^2 |S_M(\alpha)|\mathrm{d}\alpha \le \delta^2 NM/4,$$

which by (15) yields

$$\int_{\mathfrak{M}} |\widehat{f_A}(\alpha)|^2 |S_M(\alpha)|\mathrm{d}\alpha \ge \delta^2 NM/4. \tag{18}$$

Finally, by (16) and (18) we have

$$\delta^2 N \ll \Big(\sum_{q=1}^{\eta^{-2}} q^{-1/2}\Big) \max_{q\le\eta^{-2}} \int_{\mathbf{M}_q} |\widehat{f_A}(\alpha)|^2\mathrm{d}\alpha \ll \eta^{-1} \max_{q\le\eta^{-2}} \int_{\mathbf{M}_q} |\widehat{f_A}(\alpha)|^2\mathrm{d}\alpha,$$

and the lemma follows. □

5.2. **Proof of Lemma 5.** Suppose $A \subseteq [1,N]$ with $|A| = \delta N$, and let $\eta = c_0\delta$ and $M = \lfloor N/9 \rfloor$. If $(A-A) \cap \Lambda_d = \emptyset$, then we define a function ν on $\mathbb{Z}$ by

$$\nu(m) = \frac{\phi(d)}{d}\log(dm+1)1_{\Lambda_d}(m)$$

and we see just as in the proof of Lemma 4 that if $|A \cap (N/9, 8N/9)| \geq 3\delta N/4$, then

$$\sum_{\substack{n\in\mathbb{Z}\\ 1\leq m\leq M}} f_A(n)f_A(n+m)\nu(m) \leq -\frac{\delta^2 N}{2}\sum_{m=1}^{M}\nu(m). \tag{19}$$

If $d \leq \log N$, then it follows from Lemma 3 that

$$\sum_{m=1}^{M}\nu(m) = \phi(d)\psi(dM+1,1,d)/d \geq M/2. \tag{20}$$

Again by orthogonality, we have

$$\sum_{\substack{n\in\mathbb{Z}\\ 1\leq m\leq M}} f_A(n)f_A(n+m)\nu(m) = \int_0^1 |\widehat{f_A}(\alpha)|^2 W_M(\alpha)\mathrm{d}\alpha, \tag{21}$$

where

$$W_x(\alpha) = \sum_{m=1}^{x}\nu(m)e^{2\pi i m\alpha}.$$

From (19), (20), and (21), we have

$$\int_0^1 |\widehat{f_A}(\alpha)|^2 |W_M(\alpha)|\mathrm{d}\alpha \geq \delta^2 NM/4. \tag{22}$$

It follows from Lemma 3 and work of Vinogradov that if $\delta \geq 1/\log N$, then

$$|W_M(\alpha)| \ll M/\phi(q) \quad \text{if} \quad \alpha \in \mathbf{M}_q \subseteq \mathfrak{M} \tag{23}$$

and

$$|W_M(\alpha)| \leq C\eta M \leq \delta M/8 \quad \text{for all} \quad \alpha \in \mathfrak{m}, \tag{24}$$

provided we choose $c_0 \leq 1/8C$. We will discuss these estimates in more detail in Appendix B. Just as before, we can apply (24) and Plancherel's Identity to conclude

$$\int_{\mathfrak{M}} |\widehat{f_A}(\alpha)|^2 |W_M(\alpha)|\mathrm{d}\alpha \geq \delta^2 NM/8. \tag{25}$$

Finally, by (23) and (25) we have

$$\delta^2 N \ll \Big(\sum_{q=1}^{\eta^{-2}}\frac{1}{\phi(q)}\Big)\max_{q\leq\eta^{-2}}\int_{\mathbf{M}_q}|\widehat{f_A}(\alpha)|^2\mathrm{d}\alpha \ll \log(\eta^{-1})\max_{q\leq\eta^{-2}}\int_{\mathbf{M}_q}|\widehat{f_A}(\alpha)|^2\mathrm{d}\alpha,$$

where the last inequality comes from the estimate

$$\sum_{q=1}^{X}\frac{1}{\phi(q)} \ll \log X,$$

and the lemma follows. □

5.3. **Proof of Lemma 6.** Suppose $A \subseteq [1,N]$ with $|A| = \delta N$, and let $\eta = c_0\delta$. Suppose further that

$$\int_{\mathbf{M}_q} |\widehat{f_A}(\alpha)|^2 \mathrm{d}\alpha \geq \sigma\delta^2 N, \tag{26}$$

and let $P = \{q, 2q, \ldots, Lq\}$ with $L = \lfloor \eta^2\sigma N/q \rfloor$. We will show that some translate of P satisfies the conclusion of Lemma 6. We note that for $\alpha \in [0,1]$,

$$|\widehat{1_P}(\alpha)| = \Big| \sum_{\ell=1}^{L} e^{-2\pi i \ell q \alpha} \Big| \geq L - \sum_{\ell=1}^{L} |1 - e^{-2\pi i \ell q \alpha}| \geq L - 2\pi L^2 \|q\alpha\|, \tag{27}$$

where $\|\cdot\|$ denotes the distance to the nearest integer. Further, if $\alpha \in \mathbf{M}_q$, then

$$\|q\alpha\| \leq q/\eta^2 N \leq \sigma/L \leq 1/4\pi L, \tag{28}$$

provided $\sigma \leq 1/4\pi$. Therefore, by (27) and (28) we have

$$|\widehat{1_P}(\alpha)| \geq L/2 \quad \text{for all} \quad \alpha \in \mathbf{M}_q. \tag{29}$$

By (26), (29), and Plancherel's Identity we see

$$\sigma\delta^2 N \leq \int_{\mathbf{M}_q} |\widehat{f_A}(\alpha)|^2 \mathrm{d}\alpha \leq \frac{4}{L^2} \int_0^1 |\widehat{f_A}(\alpha)|^2 |\widehat{1_P}(\alpha)|^2 \mathrm{d}\alpha = \frac{4}{L^2} \sum_{n\in\mathbb{Z}} |f_A * \widetilde{1_P}(n)|^2, \tag{30}$$

where $\widetilde{1_P}(n) = 1_P(-n)$ and

$$f_A * \widetilde{1_P}(n) = \sum_{m\in\mathbb{Z}} f_A(m) 1_P(m-n) = |A \cap (P+n)| - \delta |(P+n) \cap [1,N]|. \tag{31}$$

We now take advantage of the fact that f_A, and consequently $f_A * \widetilde{1_P}$, has mean value zero. In other words,

$$\sum_{n\in\mathbb{Z}} f_A * \widetilde{1_P}(n) = 0. \tag{32}$$

As with any real valued function, we can write

$$|f_A * \widetilde{1_P}| = 2(f_A * \widetilde{1_P})_+ - f_A * \widetilde{1_P}, \tag{33}$$

where $(f_A * \widetilde{1_P})_+ = \max\{f_A * \widetilde{1_P}, 0\}$.

For the purposes of proving Lemma 6, we can assume that $f_A * \widetilde{1_P}(n) \leq 2\delta L$ for all $n \in \mathbb{Z}$, as otherwise the progression $P + n$ would more than satisfy the conclusion. Combined with the trivial upper bound $f_A * \widetilde{1_P}(n) \geq -\delta L$, we can assume

$$|f_A * \widetilde{1_P}(n)| \leq 2\delta L \quad \text{for all} \quad n \in \mathbb{Z}. \tag{34}$$

By (30), (32), (33), and (34), we have

$$\sum_{n\in\mathbb{Z}} (f_A * \widetilde{1_P})_+(n) = \frac{1}{2} \sum_{n\in\mathbb{Z}} |f_A * \widetilde{1_P}| \geq \frac{1}{4\delta L} \sum_{n\in\mathbb{Z}} |f_A * \widetilde{1_P}|^2 \geq \frac{\sigma\delta NL}{16}. \tag{35}$$

By (31), we see that $f_A * \widetilde{1_P}(n) = 0$ if $n \notin [-qL, N]$. Letting $E = \{n \in \mathbb{Z} : P + n \subseteq [1,N]\}$ and $F = [-qL, N] \setminus E$, we see that $|F| \leq 2qL \leq 2\eta^2\sigma N$. Therefore, by (34) and (35) we have

$$\sum_{n\in E} (f_A * \widetilde{1_P})_+(n) \geq \frac{\sigma\delta NL}{16} - 2\delta L|F| \geq \frac{\sigma\delta NL}{16} - 4\eta^2\sigma\delta NL > \frac{\sigma\delta NL}{32}, \tag{36}$$

provided $c_0 < 1/8$. Recalling that $|E| \leq N$ and $f_A * \widetilde{1_P}(n) = |A \cap (P+n)| - \delta L$ for all $n \in E$, we have that there exists $n \in \mathbb{Z}$ with

$$|A \cap (P+n)|/L \geq \delta + \sigma\delta/32,$$

as required. □

Appendix A. Exponential Sum Estimates over Squares: Proof of (16) and (17)

In this section, we return to the setting of the proof of Lemma 4, recalling all relevant notation and assumptions. We begin with a version of the usual major arc asymptotic for Weyl sums, which is completely standard in the context of Waring's Problem, for example.

Lemma 7. *If $\alpha = a/q + \beta$ with $q \leq M^{1/4}$ and $|\beta| \leq M^{-7/4}$ then*

$$S_M(\alpha) = q^{-1}G(a,q)\int_0^M e^{2\pi i x^2\beta}\mathrm{d}x + O(M^{1/2}),$$

where

$$G(a,q) = \sum_{r=0}^{q-1} e^{2\pi i r^2 a/q}.$$

Proof. First we see that for any $x \geq 0$ we have

$$S_x(a/q) = \sum_{m=1}^{x} e^{2\pi i m^2 a/q} = \sum_{r=0}^{q-1} e^{2\pi i r^2 a/q}\Big|\{1 \leq m \leq x \ : \ m \equiv r \bmod q\}\Big| = q^{-1}G(a,q)x + O(q).$$

Then, fixing α as in Lemma 7, we apply integration by parts twice to yield

$$\begin{aligned} S_M(\alpha) = \sum_{m=1}^{M} e^{2\pi i m^2 a/q}e^{2\pi i m^2\beta} &= S_M(a/q)e^{2\pi i M^2\beta} - \int_0^M S_x(a/q)(4\pi i x\beta)e^{2\pi i x^2\beta}\mathrm{d}x \\ &= q^{-1}G(a,q)\Big(Me^{2\pi i M^2\beta} - \int_0^M x(4\pi i x\beta)e^{2\pi i x^2\beta}\mathrm{d}x\Big) + O(q(1+M^2\beta)) \\ &= q^{-1}G(a,q)\int_0^M e^{2\pi i x^2\beta}\mathrm{d}x + O(M^{1/2}), \end{aligned}$$

as required. □

To establish the cancellation on the major arcs promised in (16), we need the following standard estimate on the Gauss sum $G(a,q)$.

Lemma 8. *If $(a,q) = 1$, then $|G(a,q)| \ll \sqrt{q}$.*

Proof. Using a change of variables ($r = s + h$) and the orthogonality relation

$$\sum_{s=0}^{q-1} e^{2\pi i s t/q} = \begin{cases} q & \text{if } q \mid t \\ 0 & \text{else} \end{cases}, \tag{37}$$

we see

$$|G(a,q)|^2 = \sum_{r,s=0}^{q-1} e^{2\pi i(r^2-s^2)a/q} = \sum_{s,h=0}^{q-1} e^{2\pi i(2sh+h^2)a/q} = \sum_{h=0}^{q-1} e^{2\pi i h^2 a/q}\begin{cases} q & \text{if } q \mid 2ha \\ 0 & \text{else} \end{cases}.$$

In particular, if $(a,q) = 1$, then $|G(a,q)|^2 \leq 2q$.

□

A.1. **Proof of (16).** Since $\delta \gg M^{-1/10}$, the hypotheses of Lemma 7 are satisfied when $\alpha \in \mathbf{M}_q \subseteq \mathfrak{M}$. Trivially bounding the integral in Lemma 7 and applying Lemma 8, the result follows. □

For the minor arcs, we make further use of Lemma 7 as well as the following well-known result, which roughly says that being close to rational with small denominator is the only obstruction to a great deal of cancellation in a Weyl sum.

Lemma 9 (Weyl Inequality)**.** *If $|\alpha - a/q| < q^{-2}$ with $(a,q) = 1$, then*

$$|S_M(\alpha)| \ll \log M(q + M + M^2/q)^{1/2}.$$

This particular formulation of the Weyl Inequality follows from Theorem 1, Chapter 3, of [3]. To complete the re-purposing of Lemma 7, we need a nontrivial estimate on the oscillatory integral in the asymptotic formula.

Lemma 10.

$$\Big| \int_0^M e^{2\pi i x^2 \beta} \mathrm{d}x \Big| \ll |\beta|^{-1/2}.$$

Proof. By trivially bounding the integral we can assume that $|\beta|^{-1/2} \leq M$, in which case we can break up the interval and integrate by parts to see

$$\begin{aligned}
\Big| \int_0^M e^{2\pi i x^2 \beta} \mathrm{d}x \Big| &= \Big| \int_0^{|\beta|^{-1/2}} e^{2\pi i x^2 \beta} \mathrm{d}x + \int_{|\beta|^{-1/2}}^M \frac{1}{4\pi i x \beta} \frac{\mathrm{d}}{\mathrm{dx}} (e^{2\pi i x^2 \beta}) \mathrm{d}x \Big| \\
&\ll |\beta|^{-1/2} + \frac{1}{4\pi|\beta|} \left| \Big[\frac{e^{2\pi i x^2 \beta}}{x} \Big]_{|\beta|^{-1/2}}^M + \int_{|\beta|^{-1/2}}^M \frac{e^{2\pi i x^2 \beta}}{x^2} \mathrm{d}x \right| \\
&\ll |\beta|^{-1/2},
\end{aligned}$$

as required. □

A.2. **Proof of (17).** For a fixed $\alpha \in \mathfrak{m}$ we have by the pigeonhole principle that there exist

$$1 \leq q \leq M^{7/4}$$

and $(a,q) = 1$ with

$$|\alpha - a/q| < 1/qM^{7/4}.$$

If $\eta^{-2} \leq q \leq M^{1/4}$, then by reasoning identical to the proof of (16), Lemma 7 implies

$$|S_M(\alpha)| \ll q^{-1/2} M \leq \eta M.$$

If $M^{1/4} \leq q \leq M^{7/4}$, then Lemma 9 and the bound $\delta \geq N^{-1/20} \gg M^{-1/10}$ imply

$$|S_M(\alpha)| \ll M^{9/10} \ll \eta M.$$

If $1 \leq q \leq \eta^{-2}$, then, letting $\beta = \alpha - a/q$, it must be the case that

$$|\beta| > 1/\eta^2 N \gg 1/\eta^2 M^2, \tag{38}$$

as otherwise we would have $\alpha \in \mathfrak{M}$. Combining Lemma 7 and (38) with Lemma 10, the minor arc estimate is established. □

Appendix B. Exponential Sum Estimates over Shifted Primes: Proof of (23) and (24)

In this section, we return to the setting of the proof of Lemma 5, recalling all relevant notation and assumptions. We begin with an asymptotic formula for W_M near rationals with small denominator.

Lemma 11. *If $\alpha = a/q + \beta$ with $q \le (\log M)^{20}$, $(a,q) = 1$, and $|\beta| \le (\log M)^{20}/M$, then*

$$W_M(\alpha) = \frac{r(d,a,q)\phi(d)}{\phi(qd)} \int_0^M e^{2\pi i x\beta} \mathrm{d}x + O(Me^{-c\sqrt{\log M}}),$$

where

$$r(d,a,q) = \sum_{\substack{r=0 \\ (dr+1,q)=1}}^{q-1} e^{2\pi i r a/q} = \begin{cases} \mu(q)e^{2\pi i \ell a/q} & \text{if } (d,q) = 1 \text{ and } \ell \equiv -d^{-1} \bmod q \\ 0 & \text{else} \end{cases}$$

and μ is the Möbius function.

Proof. First we see that for any $x \ge 0$ we have

$$W_x(a/q) = \sum_{m=1}^{x} \nu(x) e^{2\pi i m a/q} = \frac{\phi(d)}{d} \sum_{r=0}^{q-1} e^{2\pi i r a/q} \psi(dx+1, dr+1, qd). \tag{39}$$

Noting that $(dr+1, qd) = 1$ if and only if $(dr+1, q) = 1$, we have by (39) and Lemma 3 that

$$W_x(a/q) = \frac{r(d,a,q)\phi(d)}{\phi(qd)} x + O(qMe^{-c\sqrt{\log M}})$$

for all $x \le M$. Then, fixing α as in Lemma 11, we apply integration by parts twice to yield

$$\begin{aligned} W_M(\alpha) &= \sum_{m=1}^{M} \nu(m) e^{2\pi i m a/q} e^{2\pi i m\beta} \\ &= W_M(a/q)e^{2\pi i M\beta} - \int_0^M W_x(a/q)(2\pi i\beta)e^{2\pi i x\beta}\mathrm{d}x \\ &= \frac{r(d,a,q)\phi(d)}{\phi(qd)}\Big(Me^{2\pi i M\beta} - \int_0^M x(2\pi i\beta)e^{2\pi i x\beta}\mathrm{d}x\Big) + O(q(M+M^2\beta)e^{-c\sqrt{\log M}}) \\ &= \frac{r(d,a,q)\phi(d)}{\phi(qd)} \int_0^M e^{2\pi i x\beta}\mathrm{d}x + O(Me^{-c\sqrt{\log M}}), \end{aligned}$$

as required. In the remaining proofs, we only use that $|r(d,a,q)| \le 1$ when $(a,q) = 1$, and we leave the evaluation claimed in Lemma 11 to the interested reader. □

B.1. **Proof of (23).** Since $\delta^{-1} \ll \log M$, the hypotheses of Lemma 11 are satisfied when $\alpha \in \mathbf{M}_q \subseteq \mathfrak{M}$. Recalling that $|r(d,a,q)| \le 1$ and $\phi(qd) \ge \phi(q)\phi(d)$, the result follows. □

For the minor arcs, we make further use of Lemma 11 as well as the following estimate of Vinogradov, a suitable analog to the Weyl Inequality used famously in his solution to the ternary Goldbach problem, which can be found in Theorem 3.1 of [10].

Lemma 12 (Vinogradov)**.** *If $|\alpha - a/q| < q^{-2}$ and $(a,q) = 1$, then*

$$|V_x(\alpha)| \ll (\log x)^4(\sqrt{qx} + x^{4/5} + x/\sqrt{q}),$$

where

$$V_x(\alpha) = \sum_{p \in \mathcal{P}_x} \log p \; e^{2\pi i p\alpha}.$$

Corollary 13. *If $|\alpha - a/q| < q^{-2}$ and $(a,q) = 1$, then*

$$|W_M(\alpha)| \ll d(\log M)^4(\sqrt{qM} + M^{4/5} + M/\sqrt{q}).$$

Proof. Exploiting (37), we see

$$\frac{d}{\phi(d)}|W_M(\alpha)| = \Big|\sum_{\substack{p\in\mathcal{P}_{dM+1}\\ p\equiv 1 \bmod d}} \log p\; e^{2\pi i(p-1)\alpha/d}\Big| = \Big|\sum_{p\in\mathcal{P}_{dM+1}} \log p\; e^{2\pi i(p-1)\alpha/d}\;\frac{1}{d}\sum_{r=0}^{d-1} e^{2\pi i(p-1)r/d}\Big|$$
$$\le \frac{1}{d}\sum_{r=0}^{d-1}\Big|\sum_{p\in\mathcal{P}_{dM+1}} \log p\; e^{2\pi i(p-1)(\alpha+r)/d}\Big| = \frac{1}{d}\sum_{r=0}^{d-1}\Big|V_{dM+1}\Big(\frac{\alpha+r}{d}\Big)\Big|.$$

If $|\alpha - a/q| < q^{-2}$ and $(a,q) = 1$, then for any fixed $0 \le r \le d-1$, the pigeonhole pricniple yields $1 \le q' \le 2dq$ and $(a',q') = 1$ with $|(\alpha+r)/d - a'/q'| < 1/2dqq'$. We also know that $|(\alpha+r)/d - (a+rq)/qd| < 1/dq^2$, so in particular we have

$$\Big|\frac{a'}{q'} - \frac{a+rq}{qd}\Big| < \frac{1}{2dqq'} + \frac{1}{dq^2}. \tag{40}$$

If $q' < q$, then since $(a,q) = 1$ the two fractions above cannot be equal, hence

$$\frac{1}{2dqq'} + \frac{1}{dq^2} > \frac{1}{dqq'},$$

which implies $q' \ge q/2$. In any case $q/2 \le q' \le 2dq$, so by Lemma 12 we have

$$\Big|V_{dM+1}\Big(\frac{\alpha+r}{d}\Big)\Big| \ll (\log(dM+1))^4(\sqrt{q'(dM+1)} + (dM+1)^{4/5} + (dM+1)/\sqrt{q'})$$
$$\ll d(\log M)^4(\sqrt{qM} + M^{4/5} + M/\sqrt{q}),$$

and the corollary follows. □

B.2. **Proof of (24).** For a fixed $\alpha \in \mathfrak{m}$ we have by the pigeonhole principle that there exist

$$1 \le q \le M/(\log M)^{20}$$

and $(a,q) = 1$ with

$$|\alpha - a/q| < (\log M)^{20}/qM.$$

If $\eta^{-2} \le q \le (\log M)^{20}$, then by reasoning identical to the proof of (23), Lemma 11 implies

$$|W_M(\alpha)| \ll M/\phi(q) \ll \eta M,$$

where the last inequality follows from the fact that $\phi(q) \gg \sqrt{q}$.

If $(\log M)^{20} \le q \le M/(\log M)^{20}$, then Corollary 13 and the bound $d, \delta^{-1} \ll \log M$ imply

$$|W_M(\alpha)| \ll M/(\log M)^5 \le \eta M.$$

If $1 \le q \le \eta^{-2}$, then, letting $\beta = \alpha - a/q$, it must be the case that

$$|\beta| > 1/\eta^2 N \gg 1/\eta^2 M, \tag{41}$$

as otherwise we would have $\alpha \in \mathfrak{M}$. Combining Lemma 11 and (41) with the bound

$$\Big|\int_0^M e^{2\pi i x\beta}\mathrm{d}x\Big| = \Big|\frac{e^{2\pi i M\beta} - 1}{2\pi i\beta}\Big| < |\beta|^{-1},$$

the minor arc estimate is established. □

References

[1] H. Furstenberg, *Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions*, J. d'Analyse Math, 71 (1977), 204-256.

[2] N. Lyall, Á. Magyar, *Polynomial configurations in difference sets*, J. Number Theory 129 (2009), pp. 439-450.

[3] H. L. Montgomery, *Ten lectures on the interface between analytic number theory and harmonic analysis*, CBMS Regional Conference Series in Mathematics, 84.

[4] H. L. Montgomery, R. C. Vaughan, *Multiplicative Number Theory I. Classical Theory*, Cambridge Studies in Advanced Mathematics 97, 2007.

[5] J. Pintz, W. L. Steiger, E. Szemerédi, *On sets of natural numbers whose difference set contains no squares*, J. London Math. Soc. 37 (1988), pp. 219-231.

[6] K. F. Roth, *On certain sets of integers*, J. London Math. Soc. 28 (1953), pp. 104-109.

[7] I. Ruzsa, T. Sanders, *Difference sets and the primes*, Acta. Arith. 131, no. 3 (2008), pp.281-201.

[8] A. Sárközy, *On difference sets of sequences of integers I*, Acta. Math. Hungar. 31 (1-2) (1978), pp. 125-149.

[9] A. Sárközy, *On difference sets of sequences of integers III*, Acta. Math. Hungar. 31(3-4) (1978), pp. 355-386.

[10] R. C. Vaughan, *The Hardy-Littlewood method*, Cambridge University Press, Second Edition, 1997.

Department of Mathematics, The University of Georgia, Athens, GA 30602, USA

E-mail address: lyall@math.uga.edu

Department of Mathematics, The University of Georgia, Athens, GA 30602, USA

E-mail address: arice@math.uga.edu