ON REPRESENTATIONS OF INTEGERS IN THIN SUBGROUPS OF $\mathrm{SL}_2(\mathbb{Z})$

JEAN BOURGAIN AND ALEX KONTOROVICH

ABSTRACT. Let $\Gamma < \mathrm{SL}(2,\mathbb{Z})$ be a free, finitely generated Fuchsian group of the second kind with no parabolics, and fix two primitive vectors $v_0, w_0 \in \mathbb{Z}^2 \setminus \{0\}$. We consider the set $\mathcal{S}$ of all integers occurring in $v_0 \gamma\, {}^t w_0$, for $\gamma \in \Gamma$. Assume that the limit set of Γ has Hausdorff dimension $\delta > 0.99995$, that is, Γ is thin but not too thin. Using a variant of the circle method, new bilinear forms estimates and Gamburd's 5/6-th spectral gap in infinite-volume, we show that $\mathcal{S}$ contains almost all of its admissible primes, that is, those not excluded by local (congruence) obstructions. Moreover, we show that the exceptional set $\mathfrak{E}(N)$ of integers $|n| < N$ which are locally admissible ($n \in \mathcal{S}(\mathrm{mod}\, q)$ for all $q \geq 1$) but fail to be globally represented, $n \notin \mathcal{S}$, has a power savings, $|\mathfrak{E}(N)| \ll N^{1-\varepsilon_0}$ for some $\varepsilon_0 > 0$.

CONTENTS

Date: January 25, 2010.
Bourgain is partially supported by NSF grant DMS-0808042.
Kontorovich is partially supported by NSF grants DMS-0802998 and DMS-0635607, and the Ellentuck Fund at IAS.

1. Introduction

Recently Bougain, Gamburd and Sarnak [BGS06, BGS08] introduced the Affine Linear Sieve, which concerns the application of various sieve methods to the setting of (possibly thin) orbits of groups of morphisms of affine n-space. Until now, the Affine Linear Sieve had produced almost-primes in great generality, in some cases giving explicit bounds for the number of factors (e.g. [Kon09, KO09]), but had not yet exhibited actual primes in thin orbits. The failure of sieve methods to produce primes stems from the well-known *parity barrier*, pinpointed by Selberg 60 years ago, that sieves alone cannot distinguish between integers having an odd or even number of prime factors. In the 1930s, I. M. Vinogradov introduced bilinear forms estimates to overcome this barrier, leading to his resolution of the ternary Goldbach problem. It is our present goal to inject bilinear forms methods into the Affine Linear Sieve to produce *primes* and not just almost primes in sets coming from thin orbits.

1.1. **Statement of the Main Theorem.**

The (multi-) set $\mathcal{S}$ of integers which we study is the following. Fix two primitive vectors $v_0, w_0 \in \mathbb{Z}^2$, and let $\Gamma < \mathrm{SL}(2,\mathbb{Z})$ be finitely-generated, free, contain no parabolic elements, and whose limit set has Hausdorff dimension $\delta > 0.99995$ (so Γ is thin but not too thin). Such Γ exist, cf. Remark 1.7. Let

$$\mathcal{S} = \langle v_0 \cdot \Gamma, w_0 \rangle = \Big\{ \langle v_0 \cdot \gamma, w_0 \rangle : \gamma \in \Gamma \Big\}, \tag{1.1}$$

where the inner product is the usual one on $\mathbb{R}^2$, and the Γ-action is usual matrix multiplication.

For ease of exposition, we focus on the example $v_0 = w_0 = (0,1)$, for which $\langle v_0\gamma, w_0\rangle = d_\gamma$, where $\gamma = \left(\begin{smallmatrix} a_\gamma & b_\gamma \\ c_\gamma & d_\gamma \end{smallmatrix}\right)$. So in this case, $\mathcal{S}$ consists of all lower right entries d_γ for $\gamma \in \Gamma$. The goal, then, is to prove that given any $\Gamma < \mathrm{SL}(2,\mathbb{Z})$ as above, as long as it is not too thin (as measured by δ), the set of lower right entries contains infinitely many prime values. (A pleasant feature of d_γ is that there is always an infinite set of admissible primes, as we see below.)

In fact, we show more: the set of integers appearing in $\mathcal{S}$ has full density, with a power savings in the exceptional set! By this we mean the following.

Theorem 1.2. *Let $\Gamma < \mathrm{SL}(2,\mathbb{Z})$ be finitely-generated, free, and have no parabolics, and let $\mathcal{S}$ be as in* (1.1)*. Let δ be the Hausdorff dimension of the limit set of Γ. Assume that δ is bounded below by the largest root of the polynomial*

$$1020 - 8897x - 5010x^2 + 12888x^3,$$

that is,

$$\delta > 0.9999493550.$$

Let $\mathfrak{E}(N)$ be the set of integers $|n| < N$ which fail the local-to-global principle, that is, n is admissible ($n \in \mathcal{S} (\mathrm{mod}\, q)$ for all integers $q \geq 1$), but nevertheless $n \notin \mathcal{S}$. Then there is some $\varepsilon_0 > 0$ such that

$$|\mathfrak{E}(N)| \ll N^{1-\varepsilon_0}, \tag{1.3}$$

as $N \to \infty$.

Remark 1.4. By the prime number theorem, the exceptional set $\mathfrak{E}(N)$ cannot contain even a positive proportion of the admissible primes, because then it would have size $\gg \frac{N}{\log N}$, contradicting the power savings in (1.3). For the choice $v_0 = w_0 = (0,1)$ giving d_γ, there are always admissible residue classes of primes, since every group contains the identity element $\gamma = I$ with $d_\gamma = 1$.

Remark 1.5. We have made no attempt at optimizing the allowed range of δ; this can surely be done with some more effort, but our proof is sufficiently involved as is. Our main point is that *some* $\delta < 1$ is allowed, and that this range can be made explicit.

Remark 1.6. One should note the parallel between (1.3) and the theorem of Montgomery and Vaughan [MV75], that the exceptional set in the Goldbach problem, that is the set $\mathfrak{E}(X)$ of even integers at most X which cannot be expressed as the sum of two primes, has a power savings, $|\mathfrak{E}(X)| \ll X^{1-\varepsilon_0}$. (Of course the full Goldbach problem is equivalent to $|\mathfrak{E}(X)| = 1$ for $X > 2$.) A big difference between the two is that Goldbach is a definite problem, in that there are only finitely many chances to find primes p_1, p_2 with $p_1 + p_2 = 2n$, whereas our present problem is indefinite: one can take larger and larger balls in Γ, whose individual entries may attain the sought-after values; see (1.12).

Remark 1.7. There exist groups Γ satisfying the above conditions. For just one example, recall the commutator group $\Gamma'(2) = [\Gamma(2), \Gamma(2)]$ of the classical congruence subgroup $\Gamma(2)$ of level 2 in $\mathrm{PSL}(2,\mathbb{Z})$. That is, let $A = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$; then $\Gamma(2) = \langle A, B \rangle$ is free, and $\Gamma'(2)$ consists of all elements of the form

$$A^{n_1}B^{m_1}A^{n_2}B^{m_2}\cdots A^{n_k}B^{m_k},$$

with $\sum_j n_j = \sum_j m_j = 0$. It is easy to see that the only trace ± 2 element of $\Gamma'(2)$ is the identity. The group $\Gamma'(2)$ is thin, having infinite index in $SL(2,\mathbb{Z})$, but is infinitely-generated, and its limit set is the entire boundary. Hence its Hausdorff dimension is $\delta_{\Gamma'(2)} = 1$. Now, take a subgroup Γ of $\Gamma'(2)$ which is generated by finitely many elements of $\Gamma'(2)$; it will still be free and have no parabolics. One can [Sul84] add more and more generators to Γ in such a way that the Hausdorff dimension δ_Γ of the limit set of Γ will be arbitrarily close to 1.

1.2. **Methods.**

Our starting point is the Hardy-Littlewood circle method; we now describe the main ingredients. One forms an exponential sum, which is essentially

$$S_N(\theta) \approx \sum_{\substack{\gamma\in\Gamma\\ \|\gamma\|<N}} e(\langle v_0\gamma, w_0\rangle\,\theta), \qquad \text{where } \theta\in[0,1] \text{ and } e(x)=e^{2\pi i x}. \tag{1.8}$$

The norm is the usual matrix norm

$$\left\| \left(\begin{smallmatrix} a & b\\ c & d\end{smallmatrix}\right)\right\| = \sqrt{a^2+b^2+c^2+d^2}.$$

Let

$$R_N(n) := \widehat{S_N}(-n) = \int_0^1 S_N(\theta)e(-n\theta)d\theta \approx \sum_{\substack{\gamma\in\Gamma\\ \|\gamma\|<N}} \mathbf{1}_{\{d_\gamma=n\}} \tag{1.9}$$

be roughly the number of representations of n, that is, the number of $\|\gamma\|<N$ having $d_\gamma=n$. Hence n appears in $\mathcal{S}$ if $R_N(n)>0$.

Divide the circle into major and minor arcs, $[0,1]=\mathfrak{M}\sqcup\mathfrak{m}$, where the major arcs $\mathfrak{M}$ consist of a union of small intervals near rationals with small denominators, and the minor arcs $\mathfrak{m}$ are the rest. Write

$$\mathcal{M}_N(n) = \int_{\mathfrak{M}} S_N(\theta)e(-n\theta)d\theta,$$

and

$$\mathcal{E}_N(n) = \int_{\mathfrak{m}} S_N(\theta)e(-n\theta)d\theta,$$

so that

$$R_N(n) = \mathcal{M}_N(n) + \mathcal{E}_N(n),$$

the first term being the "main" term and the latter being the "error."

Of course

$$R_N(n) \geq \mathcal{M}_N(n) - |\mathcal{E}_N(n)|,$$

so n is representable if

$$\mathcal{M}_N(n) > |\mathcal{E}_N(n)|.$$

Note that by Lax-Phillips [LP82], the total mass is

$$S_N(0) = \sum_{|n|<N} R_N(n) \sim c \cdot N^{2\delta},$$

and so on average, one expects numbers to appear with multiplicity roughly

$$N^{2\delta-1}.$$

In a more-or-less straightforward calculation using the spectral gap, we show, cf. Theorem 4.1, that for "almost" all $|n| < N$, the main term is

$$\mathcal{M}_N(n) \gg \frac{1}{\log\log(10+|n|)} N^{2\delta-1}, \tag{1.10}$$

if n is admissible (and of course zero otherwise).

1.3. **Local-global?**

One may *a priori* wonder whether in fact $\mathcal{E}_N$ can be controlled in $\mathcal{L}^1$, that is, uniformly in n:

$$|\mathcal{E}_N(n)| \leq \int_{\mathfrak{m}} |S_N(\theta)| d\theta \stackrel{?}{=} o(N^{2\delta-1}). \tag{1.11}$$

This (and removing "almost" before (1.10)) would lead to a genuine local-global principle – there would only be no exceptions at all (by taking N to infinity with n fixed).

But algebra intervenes, and (1.11) is **false**! Consider changing w_0 from $(0,1)$ to $(1,0)$, which now picks up the lower left entry c_γ instead of d_γ. The above analysis all goes through, and setting $n = 0$, there are no local obstructions, so the main term is

$$\mathcal{M}_N(0) \gg N^{2\delta-1}. \tag{1.12}$$

This count is a manifestation of the fact that, as our setup is indefinite (cf. Remark 1.6), there are more and more chances to hit $n = 0$ by letting N grow. Hence $R_N(0)$ now counts the set of matrices $\gamma \in \Gamma$ with $c_\gamma = 0$, which are necessarily parabolic. But we assumed that Γ has no parabolics (other than I and possibly $-I$)! So $R_N(0) \leq 2$, and we have

$$N^{2\delta-1} \ll |R_N(0) - \mathcal{M}_N(0)| = |\mathcal{E}_N(0)| = \left| \int_{\mathfrak{m}} S_N(\theta) d\theta \right| \leq \int_{\mathfrak{m}} |S_N(\theta)| d\theta,$$

thereby disproving (1.11). Hence the exceptional set need not be empty.

Remark 1.13. It may still be the case that the exceptional set is finite. One example of such is the question of curvatures in an integral Apollonian gasket, cf. [GLM$^+$03, Sar07, KO08, BF10], where it is expected that every sufficiently large admissible number appears.

Remark 1.14. One faces a similar obstacle upon trying to use the circle method to count sums of four squares – one cannot introduce minor arcs because the $\mathcal{L}^1$ norm is as big as the main term. This issue was overcome by Kloosterman [Klo27], but his refinement is impossible in our context (as again, it would imply a complete local-to-global phenomenon, which is false).

Instead we average over n, which bypasses the $\mathcal{L}^1$ norm and places the $\mathcal{L}^2$ norm in the spotlight. We prove in Theorem 8.18 that for some $\eta > 0$,

$$\sum_{|n|<N} |\mathcal{E}_N(n)|^2 = \int_{\mathfrak{m}} |S_N(\theta)|^2 d\theta \ll N^{4\delta-1-\eta}. \tag{1.15}$$

A straightforward argument gives (1.3) from (1.10) and (1.15), cf. Theorem 8.19.

1.4. **Ingredients.**

It is the bilinear (in fact multilinear) structure of the set $\mathcal{S} = \langle v_0\Gamma, w_0\rangle$ which we exploit. A key ingredient in our approach is the observation that, instead of (1.8), one can consider the exponential sum $S_N(\theta)$ of the form

$$\sum_{\substack{\gamma_1\in\Gamma \\ \|\gamma_1\|<N^{1/2}}} \sum_{\substack{\gamma_2\in\Gamma \\ \|\gamma_2\|<N^{1/2}}} e(\langle v_0\gamma_1\gamma_2, w_0\rangle\, \theta), \tag{1.16}$$

say, whose transform $R_N(n) = \widehat{S_N}(-n)$ is neither a lower nor upper bound for the true number of representatives, but certainly still has the property that if $R_N(n) > 0$, then n is representable. (Even (1.16) is an oversimplification of our actual exponential sum, for which see §3.)

The advantage of the above formulation is that, since $\langle v_0\gamma_1\gamma_2, w_0\rangle = \langle v_0\gamma_1, {w_0}^t\gamma_2\rangle$, one can rewrite $S_N(\theta)$ as

$$\sum_{x\in\mathbb{Z}^2}\sum_{y\in\mathbb{Z}^2} \mu(x)\nu(y)e(\langle x, y\rangle\, \theta),$$

for appropriate measures μ and ν. We then develop several new bounds for bilinear forms of the above type, cf. Theorems 5.1, 6.1, and 7.1.

Another key ingredient in the above technology is the ability to count effectively (with power savings error terms, uniformly over congruence subgroups and their cosets) the number of elements of a group of isometries of an infinite-volume hyperbolic manifold lying in certain restricted regions. This analysis is carried out together with Peter Sarnak in the companion paper [BKS09]. It is our pleasure to thank him for many illuminating conversations during this work.

1.5. **Organization.**

The paper is organized as follows. In §2, we state several estimates on the number of orbital points in various restricted regions, which will be necessary in the sequel. Their proofs are given in [BKS09]. In §3, S_N is properly defined and the major and minor arcs are introduced (even these require a slight deviation from the traditional construction). The major arcs are controlled in §4, the minor arcs are disposed of in Sections 5, 6, and 7, and all of the ingredients are assembled in the final §8.

2. Estimates of Orbital Regions

In this section, we state some estimates which are used in the sequel. Their proofs appear in [BKS09].

Let $\Gamma < \mathrm{SL}(2,\mathbb{Z})$ be a free, finitely generated Fuchsian group of the second kind with no parabolics, and let δ be the Hausdorff dimension of the limit set of Γ. Assume throughout that $\delta > 5/6$. Let $0 < \sigma < 1/4$ be another parameter to be chosen later, cf. (3.1). Let $q \geq 1$, and define a "congruence" subgroup of Γ of level q to be a group which contains the principal congruence subgroup

$$\Gamma(q) := \{\gamma \in \Gamma : \gamma \equiv I (\mathrm{mod}\, q)\}.$$

We require the following Sobolev-type norm. Fix $T \geq 1$ and let $\{X_1, X_2, X_3\}$ be a basis for the Lie algebra $\mathfrak{g}$. Then define the $\mathcal{S}_{\infty,T}$ norm by

$$\mathcal{S}_{\infty,T} f = \max_{X \in \{0, X_1, X_2, X_3\}} \sup_{g \in G, \|g\| < T} |d\pi(X).f(g)|,$$

that is, the supremal value of first order derivatives of f in a ball of radius T in G.

The following is a form of "dualization" (cf. [NS09]), an analogue of Poisson summation.

Theorem 2.1. *Fix any $\gamma_0 \in \Gamma$ and a congruence subgroup $\Gamma_1(q) < \Gamma$ of level $q \geq 1$. Let $f : G \to \mathbb{C}$ be a smooth function with $|f| \leq 1$ There is a fixed "bad" integer $\mathfrak{B}$ which depends only on Γ such that for $q = q'q''$, $q' \mid \mathfrak{B}$,*

$$\sum_{\substack{\gamma \in \gamma_0 \cdot \Gamma(q) \\ \|\gamma\| < T}} f(\gamma) = \frac{1}{[\Gamma : \Gamma(q)]} \left(\sum_{\substack{\gamma \in \Gamma \\ \|\gamma\| < T}} f(\gamma) + \mathcal{E}_{q'} \right) + O\left(T^{\frac{6}{7}2\delta + \frac{5}{21}} (1 + \mathcal{S}_{\infty,T} f)^{6/7} \right).$$

Here $\mathcal{E}_{q'} \ll T^{2\delta - \alpha_0}$, with $\alpha_0 > 0$, and all implied constants are independent of q'' and γ_0.

Proof. This is Theorem 1.13 in [BKS09], using Gamburd's [Gam02] spectral gap $\Theta = 5/6$. □

Theorem 2.2. *Let $v_0, w \in \mathbb{Z}^2$ and assume that $n \in \mathbb{Z}$, $\frac{N}{K_0} < |n| < N$, $|w| < N^{1-\sigma}$, $|v_0| \leq 1$, and $|n| < |v_0||w|N^{\sigma}$. Then*

$$\sum_{\substack{\gamma \in \Gamma \\ \|\gamma\| < N^{\sigma}}} \mathbf{1}\left\{ |\langle v_0 \gamma, w \rangle - n| < \frac{N}{2K_0} \right\} \gg \frac{N^{2\delta\sigma}}{K_0} + O\left(N^{\sigma(\frac{3}{4} + \frac{1}{4}2\delta)} (\log N)^{1/4} \right).$$

Proof. This is an application of Theorem 1.14 in [BKS09]. □

Theorem 2.3. *Fix (c,d) and $y = (y_1, y_2)$ in $\mathbb{Z}^2$ with $|y| < N^{1/2}$, $|(c,d)| < N^{\sigma}$ and $|y| < N^{1/2-\sigma}|(c,d)|$. Then*

$$\sum_{\substack{\gamma \in \Gamma \\ \|\gamma\| < N^{1/2-\sigma}}} \mathbf{1}\left\{ |(c,d)\gamma - y| < \frac{N^{1/2}}{K} \right\} \mathbf{1}\left\{ (c,d)\gamma \equiv y (\bmod\, q) \right\}$$

$$\ll \frac{N^{\delta(1-2\sigma)}}{K^{1+\delta} q^2} + N^{(1/2-\sigma)(\frac{6}{7}2\delta + \frac{5}{21})},$$

as $N \to \infty$.

Proof. This follows from Theorem 1.15 in [BKS09]. □

3. Setup of the Exponential Sum and Major/Minor Arcs

3.1. The Exponential Sum S_N.

Let $\Gamma < \mathrm{SL}(2,\mathbb{Z})$ be finitely-generated, free, contain no parabolics, and have limit set with Hausdorff dimension $\delta > 1/2$. Fix two primitive vectors $v_0, w_0 \in \mathbb{Z}^2 \setminus \{0\}$. Recall the shape of the exponential sum function from (1.16), $S_N(\theta)$ is approximately

$$\sum_{\substack{\gamma_1\in\Gamma \\ \|\gamma_1\|<N^{1/2}}} \sum_{\substack{\gamma_2\in\Gamma \\ \|\gamma_2\|<N^{1/2}}} e(\langle v_0\gamma_1\gamma_2, w_0\rangle\,\theta),$$

and this is used to develop certain bilinear forms estimates.

Assuming Γ has no parabolics, the vectors $v_0\gamma_1$ are all unique. But of course the products $\gamma_1\gamma_2$, though capable of reaching elements in Γ of norm N, can also back-track (e.g. if $\gamma_2 = \gamma_1^{-1}$). To prevent this, use that Γ is free and restrict the range of γ_1 and γ_2 further by writing each as a word in the generators of Γ, and controlling the concluding letter of γ_1 and beginning letter of γ_2. Then it is not possible to back-track, and each element arising as the product $\gamma = \gamma_1\gamma_2$ is unique.

Unfortunately, the above tweak is problematic for another reason. Namely, one would like to perform certain estimates involving spectral theory, in particular, appealing to the spectral gap. Conditions such as $\|\gamma\| < N$ can be encoded spectrally, whereas restrictions on letters appearing in representations of γ as a word in the generators cannot.

So we add another element γ_3 of small norm, say N^σ with

$$\sigma < 1/4. \tag{3.1}$$

(See (6.5) for the place where this is used.) This has the effect of ruining unique representations of $\gamma = \gamma_1\gamma_2\gamma_3$, but not by too much, while still allowing for estimates using spectral theory. To distinguish their roles, we will call $\xi = \gamma_1$, $\varpi = \gamma_2$, and $\gamma = \gamma_3$. The exact definition follows.

Fix a subset $\Xi = \Xi_{N^{1/2}} \subset \Gamma$ consisting of elements $\xi \in \Gamma$ with $\|\xi\| < N^{1/2}$, which when written (uniquely, since Γ is free) as a reduced word in the generators of Γ, all end with the same letter. By the pigeonhole principle, one can choose Ξ so that the number of elements in $\Xi_{N^{1/2}}$ is $\gg N^\delta$ (the implied constant depending only on Γ and the number of generators of Γ).

Also fix a subset $\Pi = \Pi_{N^{1/2-\sigma}} \subset \Gamma$ consisting of elements $\varpi \in \Gamma$ with $\|\varpi\| < N^{1/2-\sigma}$, which are written as a reduced word in the generators of Γ, all starting with the same letter. (If this letter happens to be the inverse of the ending letter in Ξ, add a different fixed letter to the beginning of every element in Π.) Again the number of elements in $\Pi_{N^{1/2-\sigma}}$ is $\gg N^{\delta(1-2\sigma)}$.

For large fixed N and $\theta \in [0,1]$, define the exponential sum function by

$$S_N(\theta) := \sum_{\xi\in\Xi} \sum_{\varpi\in\Pi} \sum_{\substack{\gamma\in\Gamma \\ \|\gamma\|<N^\sigma}} e(\langle v_0 \cdot \gamma\xi\varpi, w_0\rangle\, \theta). \tag{3.2}$$

For $n \in \mathbb{Z}$, let

$$\begin{aligned} R_N(n) \;&:=\; \widehat{S_N}(n) = \int_0^1 S_N(\theta) e(-n\theta) d\theta \\ &=\; \sum_{\xi\in\Xi} \sum_{\varpi\in\Pi} \sum_{\substack{\gamma\in\Gamma \\ \|\gamma\|<N^\sigma}} \mathbf{1}_{\{\langle v_0\cdot\gamma\xi\varpi, w_0\rangle = n\}} \end{aligned}$$

be the representation function.

3.2. **Major/Minor Arcs Decomposition.**

Our decomposition into major and minor arcs is made as follows. By Dirichlet's theorem, for every irrational number $\theta \in [0,1]$ (and hence for almost every, with respect to Lebesgue measure), there is a $q < N^{1/2}$ and $(a,q) = 1$ such that $\theta = \frac{a}{q} + \beta$, with

$$\left|\theta - \frac{a}{q}\right| = |\beta| < \frac{1}{qN^{1/2}}.$$

Define the major arcs as the set of $\theta = \frac{a}{q} + \beta$ with $q < Q_0$ and $|\beta| < K_0/N$, where we set

$$Q_0 = N^{\alpha_0}, \qquad \text{and} \qquad K_0 = N^{\kappa_0}.$$

These are taken as large as possible while still controlling the main term, cf. Theorem 4.1.

For technical reasons of harmonic analysis, the main term is not simply the integral of the exponential sums over the major arcs. We mollify the sharp cutoff by introducing certain weights. To this end, let ψ be the triangle function

$$\psi(x) := \begin{cases} 0 & \text{if } |x| \geq 1, \\ 1-x & \text{if } 0 < x < 1, \\ 1+x & \text{if } -1 < x < 0. \end{cases} \tag{3.3}$$

Then its Fourier transform is

$$\widehat{\psi}(y) = \left(\frac{\sin(\pi y)}{\pi y}\right)^2;$$

in particular it is positive.[1]

Let $\Psi_{N,K_0}(\beta)$ be defined by

$$\Psi_{N,K_0}(\beta) := \sum_{m\in\mathbb{Z}} \psi\bigg((\beta+m)N/K_0\bigg). \tag{3.4}$$

In particular Ψ is well-defined on the circle $[0,1]$ and is a triangle function about the interval $[-K_0/N, K_0/N]$.

Let the mollified major arcs "indicator" function be

$$\mathfrak{M}(\theta) := \sum_{1\le q<Q_0} \sum_{(a,q)=1} \Psi_{N,K_0}\left(\theta - \frac{a}{q}\right). \tag{3.5}$$

Then define the main term by

$$\mathcal{M}_N(n) := \int_0^1 \mathfrak{M}(\theta) S_N(\theta) e(-n\theta) d\theta. \tag{3.6}$$

The minor arcs function is then given by

$$\mathfrak{m}(\theta) := 1 - \mathfrak{M}(\theta),$$

and the error term is

$$\mathcal{E}_N(n) := \int_0^1 \mathfrak{m}(\theta) S_N(\theta) e(-n\theta) d\theta.$$

Recall that we intend to bound the $\mathcal{L}^2$ norm of $\mathcal{E}_N$, that is, control the integral

$$\int_0^1 |\mathfrak{m}(\theta) S_N(\theta)|^2 d\theta. \tag{3.7}$$

For parameters $Q < N^{1/2}$ and $K < N^{1/2}$, we decompose the circle into dyadic pieces of the form

$$W_{Q,K} := \left\{ \theta = \frac{a}{q} + \beta : (a,q) = 1, q \sim Q, |\beta| \sim \frac{K}{N} \right\},$$

where $x \sim X$ means $\frac{1}{2}X \le x < X$. Then a bound for (3.7) will follow from controlling

$$\int_{W_{Q,K}} |S_N(\theta)|^2 d\theta,$$

with either $Q > Q_0$ or $K > K_0$, or both. This is the task undertaken in Sections 5, 6, and 7.

[1] While of greatest importance to us is the positivity of the Fourier transform, we would also like ψ to approximate the indicator function of the interval $[-1,1]$. Being positive makes $\widehat{\psi}$ the square of some function, and so ψ is a convolution of a function with itself. The convolution of the indicator function with itself is a triangle function, hence our choice of ψ.

4. The Major Arcs

This section is devoted to the proof of

Theorem 4.1. *There is a set* $\mathfrak{E}(N) \subset [-N, N]$ *of size* $|\mathfrak{E}(N)| \ll N^{1-\varepsilon_0}$ *such that the following holds. For* $|n| < N$ *and* $n \notin \mathfrak{E}(N)$*, the main term is*

$$\mathcal{M}_N(n) = \begin{cases} \gg \frac{1}{\log\log(10+|n|)} N^{2\delta-1} & \text{if } n \in \mathcal{S}(\bmod q) \text{ for all } q \geq 1 \\ 0 & \text{otherwise,} \end{cases}$$

provided that

$$K_0 = N^{\kappa_0} \qquad \text{and} \qquad Q_0 = N^{\alpha_0},$$

with

$$\kappa_0 < \frac{3}{2}\sigma(\delta - \frac{1}{2}), \tag{4.2}$$

and

$$21\alpha_0 + 13\kappa_0 < (2\delta - \frac{5}{3})\sigma. \tag{4.3}$$

That is, we need to augment the exceptional set a bit to allow the main term estimate to fail for a small number of n.

4.1. Breaking into Modular and Archimedean Components.

Recall that the main term is

$$\mathcal{M}_N(n) = \int_0^1 \mathfrak{M}(\theta) S_N(\theta) e(-n\theta) d\theta,$$

where the exponential sum is

$$S_N(\theta) = \sum_{\xi\in\Xi} \sum_{\varpi\in\Pi} \sum_{\substack{\gamma\in\Gamma \\ \|\gamma\|<N^\sigma}} e(\langle v_0 \cdot \gamma, w_0 {}^t\varpi {}^t\xi \rangle \theta),$$

and the major arc weights $\mathfrak{M}(\theta)$ are given in terms of the triangle function ψ by:

$$\mathfrak{M}(\theta) = \sum_{1\leq q<Q_0} \sum_{(a,q)=1} \Psi_{N,K_0}\left(\theta - \frac{a}{q}\right).$$

For fixed ϖ and ξ, write

$$w = w_0 {}^t\varpi {}^t\xi.$$

Let $\Gamma_1(q)$ denote the group $\Gamma_1(q) := \{\gamma \in \Gamma : v_0\gamma \equiv v_0(q)\}$, which is clearly a "congruence" subgroup of Γ of level q. Note that any $\gamma \in \Gamma$

can we written as $\gamma = \gamma_1\gamma_2$, where $\gamma_1 \in \Gamma_1(q)$, and γ_2 is a representative chosen from the quotient group $\Gamma_1(q)\backslash\Gamma$. Note further that

$$\langle v_0 \cdot \gamma_1\gamma_2, w\rangle \equiv \langle v_0 \cdot \gamma_2, w\rangle \pmod q.$$

Therefore

$$\sum_{\substack{\gamma\in\Gamma \\ \|\gamma\|<N^\sigma}} e\left(\langle v_0 \cdot \gamma, w\rangle \left(\frac{a}{q} + \beta\right)\right) \tag{4.4}$$

$$= \sum_{\gamma_2\in\Gamma_1(q)\backslash\Gamma} e\left(\langle v_0 \cdot \gamma_2, w\rangle \frac{a}{q}\right) \sum_{\substack{\gamma_1\in\Gamma_1(q) \\ \|\gamma_1\gamma_2\|<N^\sigma}} e\left(\langle v_0 \cdot \gamma_1\gamma_2, w\rangle \beta\right)$$

Assume for simplicity that Γ has spectral gap $(\Theta, \mathfrak{B})$ with $\Theta = 5/6$ and $\mathfrak{B} = 1$. (The general case of $\mathfrak{B} > 1$ is handled similarly.)

Lemma 4.5. *For $|\beta| < K_0/N$,*

$$\sum_{\substack{\gamma_1\in\Gamma_1(q) \\ \|\gamma_1\gamma_2\|<N^\sigma}} e\left(\langle v_0 \cdot \gamma_1\gamma_2, w\rangle \beta\right) = \frac{1}{[\Gamma : \Gamma_1(q)]} \sum_{\substack{\gamma\in\Gamma \\ \|\gamma\|<N^\sigma}} e\left(\langle v_0 \cdot \gamma, w\rangle \beta\right) + O\left(K_0^{6/7} N^{\sigma(\frac{6}{7}2\delta + \frac{5}{21})}\right).$$

Proof. This follows easily from Theorem 2.1. □

Putting everything together, the main term is

$$\begin{aligned} \mathcal{M}_N(n) &= \sum_{\xi,\varpi} \sum_{q<Q_0} \sum_{(a,q)=1} \sum_{\gamma_2\in\Gamma_1(q)\backslash\Gamma} \frac{e\left((\langle v_0 \cdot \gamma_2, w\rangle - n)\frac{a}{q}\right)}{[\Gamma : \Gamma_1(q)]} \\ &\quad \times \sum_{\substack{\gamma\in\Gamma \\ \|\gamma\|<N^\sigma}} \int_0^1 \Psi_{N,K_0}(\beta)\, e\Big((\langle v_0\gamma, w\rangle - n)\beta\Big) d\beta \\ &\quad + O\left(N^{2\delta(1-\sigma)} Q_0^{3+\varepsilon} \frac{K_0}{N} K_0^{6/7} N^{\sigma(\frac{6}{7}2\delta + \frac{5}{21})}\right), \end{aligned}$$

since $[\Gamma : \Gamma_1(q)] \ll q^{1+\varepsilon}$. Define the Ramanujan sum

$$c_q(x) := \sum_{(a,q)=1} e(ax/q).$$

Let the singular series be

$$\mathfrak{S}_{N,\xi,\varpi}(n) := \sum_{q<Q_0} \frac{1}{[\Gamma : \Gamma_1(q)]} \sum_{\gamma_2\in\Gamma_1(q)\backslash\Gamma} c_q(\langle v_0\gamma_2, w\rangle - n),$$

and the singular integral be

$$\tau_{N,\xi,\varpi}(n) := \sum_{\substack{\gamma\in\Gamma \\ \|\gamma\|<N^\sigma}} \int_0^1 \Psi_{N,K_0}(\beta) e\Big((\langle v_0\gamma, w\rangle - n)\beta\Big) d\beta,$$

so that

$$\mathcal{M}_N(n) = \sum_{\xi,\varpi} \mathfrak{S}_{N,\xi,\varpi}(n)\tau_{N,\xi,\varpi}(n) + O\left(N^{2\delta(1-\sigma)} Q_0^{3+\varepsilon}\frac{K_0}{N} K_0^{6/7} N^{\sigma(\frac{6}{7}2\delta+\frac{5}{21})}\right). \tag{4.6}$$

We have divided the main term into modular and archimedean components.

4.2. **The Archimedean Component.**

Lemma 4.7. *For an integer x,*

$$\int_0^1 \Psi_{N,K_0}(\beta)e(x\beta)d\beta \geq \frac{2K_0}{5N}\cdot \mathbf{1}_{\{|x|<\frac{N}{2K_0}\}}.$$

Proof. Inserting (3.4) gives

$$\begin{aligned}\int_0^1 \Psi_{N,K_0}(\beta)e(x\beta)d\beta &= \int_{\mathbb{R}} \psi\left(\beta\frac{N}{K_0}\right) e(x\beta)d\beta \\ &= \frac{K_0}{N}\widehat{\psi}\left(x\frac{K_0}{N}\right).\end{aligned}$$

Using $\widehat{\psi}(y) = \left(\frac{\sin(\pi y)}{\pi y}\right)^2$, one checks elementarily that $\widehat{\psi}(y) > 0.4$ for $|y| < 1/2$. □

Applied to $\tau_{N,\xi,\varpi}(n)$, the above gives

$$\tau_{N,\xi,\varpi}(n) \geq \frac{2K_0}{5N} \sum_{\substack{\gamma\in\Gamma \\ \|\gamma\|<N^\sigma}} \mathbf{1}_{\left\{|\langle v_0\gamma,w\rangle - n|<\frac{N}{2K_0}\right\}}$$

Applying Theorem 2.2 gives

$$\begin{aligned}\tau_{N,\xi,\varpi}(n) &\gg \frac{2K_0}{5N}\frac{N^{2\delta\sigma}}{K_0} + O\left(\frac{K_0}{N}N^{(2\delta+3)\sigma/4}(\log N)^{1/4}\right) \\ &\gg \frac{N^{2\delta\sigma}}{N} + O\left(\frac{K_0}{N}N^{(2\delta+3)\sigma/4}(\log N)^{1/4}\right),\end{aligned} \tag{4.8}$$

as long as $w = w_0{}^t\varpi\,{}^t\xi$ satisfies $|w| \asymp N^{1-\sigma}$. But if $|w| \ll N^{1-\sigma-\varepsilon}$, then $|\langle v_0\gamma, w\rangle| \ll N^{1-\varepsilon}$, and these values of n may be discarded into the exceptional set $\mathfrak{E}(N)$.

With $K_0 = N^{\kappa_0}$, the bound (4.8) is significant as long as

$$\kappa_0 < \frac{3}{2}\sigma(\delta - \frac{1}{2}). \tag{4.9}$$

4.3. **Modular Component.**

Assume that $\Gamma(q)\backslash\Gamma \cong \mathrm{SL}(2,q)$ for all q; minor changes are need to accommodate the more general case. With ξ and ϖ fixed and $w = w_0{}^t\varpi\,{}^t\xi$, we evaluate the singular series:

$$\mathfrak{S}_{N,\xi,\varpi}(n) = \sum_{q<Q_0} \frac{1}{\mathrm{SL}(2,q)} \sum_{\gamma\in\mathrm{SL}(2,q)} c_q(\langle v_0\gamma, w\rangle - n).$$

Extend the q sum to infinity with a negligible error. The factors are multiplicative, and the main contribution comes from the primes, not prime powers. Hence we estimate just the prime contribution, which is

$$\prod_p \left(1 + \frac{1}{|\,\mathrm{SL}(2,p)|} \sum_{\gamma\in\mathrm{SL}(2,p)} c_p(\langle v_0\gamma, w\rangle - n)\right).$$

By changing representatives, we may assume without loss of generality that $v_0 = w = (0,1)$, so that $\langle v_0\gamma, w\rangle = d_\gamma$, the lower right entry. The Ramanujan sum $c_p(x)$ on primes is either $p-1$ if $x \equiv 0(p)$ or -1 othewise. There are two cases.

(1) Case $n \equiv 0(p)$: Suppose $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ has $d \equiv n \equiv 0$. Then b is invertible ($p-1$ choices) and uniquely determines c; a is free (p choices). The number of such γ is thus $p^2 - p$.

The number of γ with $d \neq n(p)$ is $p^3 - p - (p^2 - p) = p^3 - p^2$ because $|\,\mathrm{SL}(2,p)| = p(p^2-1)$. Hence

$$\begin{aligned}\frac{1}{|\,\mathrm{SL}(2,p)|} \sum_{\gamma\in\mathrm{SL}(2,p)} c_p(d-n) &= \frac{1}{p^3-p}\Big[(p-1)(p^2-p) + (-1)(p^3-p^2)\Big] \\ &= -\frac{1}{p+1}.\end{aligned}$$

(2) Case $n \neq 0(p)$: Suppose γ has $d \equiv n \neq 0(p)$. Then b and c are free (p^2 choices) and a uniquely determined by $a = (1+bc)d^{-1}$.

The number of γ with $d \neq n(p)$ is $p^3 - p - (p^2) = p^3 - p^2 - p$. Hence

$$\begin{aligned}\frac{1}{|\,\mathrm{SL}(2,p)|} \sum_{\gamma\in\mathrm{SL}(2,p)} c_p(d-n) &= \frac{1}{p^3-p}\Big[(p-1)(p^2) + (-1)(p^3-p^2-p)\Big] \\ &= \frac{1}{p^2-1}.\end{aligned}$$

To leading order, the singular series is thus

$$\mathfrak{S}_{N,\xi,\varpi}(n) \gg \prod_{p\nmid n}\left(1+\frac{1}{p^2-1}\right)\prod_{p|n}\left(1-\frac{1}{p+1}\right) \gg \frac{1}{\log\log n},$$

as desired.

4.4. **Conclusion.**

Returning to (4.6), we have shown that for $n \notin \mathfrak{E}(N)$,

$$\mathcal{M}_N(n) \gg \frac{1}{\log\log n}N^{2\delta-1} + O\left(N^{2\delta(1-\sigma)}Q_0^{3+\varepsilon}\frac{K_0}{N}K_0^{6/7}N^{\sigma(\frac{6}{7}2\delta+\frac{5}{21})}\right).$$

With $Q_0 = N^{\alpha_0}$, this is conclusive if

$$3\alpha_0 + \frac{13}{7}\kappa_0 < (2\delta - \frac{5}{3})\sigma/7. \tag{4.10}$$

This completes the proof of Theorem 4.1.

5. Minor Arcs I: L^∞ norm of $S_N(\theta)$

We establish in this section an L^∞ bound for $S_N\Big|_{W_{Q,K}}$. By itself, this bound is useful only if both K and Q are quite small, but it will be combined later with other estimates.

Theorem 5.1. *Write* $\theta = \frac{a}{q} + \beta$ *with* $q < N^{1/2}$, $|\beta| < \frac{1}{qN^{1/2}}$, *and* $|\beta| \sim \frac{K}{N}$. *Then*

$$|S_N(\theta)| \ll N^{(3\delta+1)/2}\left(\frac{1}{K^{(1+\delta)/2}q} + N^{-\frac{1}{84}(6\delta-5)(1-2\sigma)}\right). \tag{5.2}$$

5.1. **Rewriting $S_N(\theta)$.**

Write

$$\begin{aligned} S_N(\theta) &= \sum_{\xi\in\Xi}\sum_{\substack{\varpi\in\Pi,\gamma\in\Gamma\\ \|\gamma\|<N^\sigma}} e(\langle v_0\gamma\varpi, w_0\,{}^t\xi\rangle\,\theta) \\ &= \sum_x\sum_y \mu(x)\nu(y)e(\langle x,y\rangle\,\theta), \end{aligned} \tag{5.3}$$

where

$$\mu(x) = \sum_{\xi \in \Xi} \mathbf{1}(x = w_0 \, {}^t\xi),$$

and

$$\nu(y) = \sum_{\substack{\varpi \in \Pi, \gamma \in \Gamma \\ \|\gamma\| < N^{\sigma}}} \mathbf{1}(y = v_0 \gamma \varpi).$$

Note that x and y are primitive vectors in $\mathbb{Z}^2$. The assumption that Γ contains no parabolics implies that the values $w_0 \, {}^t\xi$ are unique, and hence

$$\mu \leq 1. \tag{5.4}$$

Note also that

$$\operatorname{supp} \mu, \ \operatorname{supp} \nu \subset B_{N^{1/2}}.$$

On the other hand, ν does not have a unique decomposition. For γ fixed, the value of $v_0 \gamma \varpi$ is unique, and there are $N^{\delta\sigma}$ choices for γ. Hence we have, crudely,

$$\nu \ll N^{\delta\sigma}. \tag{5.5}$$

The number of elements captured by each measure is

$$\sum_x \mu(x) \asymp \sum_y \nu(y) \asymp N^{\delta}. \tag{5.6}$$

5.2. **Bounding $S_N(\theta)$.**

As the support of ν is in a ball in $\mathbb{Z}^2$ of radius $N^{1/2}$, we can break ν into 64 pieces $\nu = \sum_\alpha \nu_\alpha$, where the support of each piece ν_α is in a square of side length

$$\frac{1}{4} N^{1/2}.$$

Each piece obviously retains the bound $\nu_\alpha \ll N^{\delta\sigma}$, and we have the bound

$$|S_N(\theta)| \leq \sum_\alpha |S_\alpha(\theta)|,$$

where

$$S_\alpha(\theta) := \sum_x \sum_y \mu(x) \nu_\alpha(y) e(\langle x, y \rangle \, \theta).$$

Let $\Upsilon : \mathbb{R}^2 \to \mathbb{R}$ be is a smooth, non-negative function which is at least 1 in the unit square $[-1,1]^2$, and $\operatorname{supp} \widehat{\Upsilon} \in B_{\frac{1}{10}}$, say.

Apply Cauchy-Schwarz, insert $\Upsilon(x/N^{1/2})$ (since μ has support in a ball of radius $N^{1/2}$), and extend the x sum to all of $\mathbb{Z}^2$ (effectively replacing the thin group Γ by the full modular group $\mathrm{SL}(2,\mathbb{Z})$):

$$\begin{aligned}|S_\alpha(\theta)| &\ll \left(\sum_{x\in\mathbb{Z}^2}|\mu(x)|^2\right)^{1/2}\left(\sum_{x\in\mathbb{Z}^2}\left|\sum_{y\in\mathbb{Z}^2}\nu_\alpha(y)e(\langle x,y\rangle\,\theta)\right|^2\Upsilon\left(\frac{x}{N^{1/2}}\right)\right)^{1/2}\\ &\ll N^{\delta/2}\left(\sum_{x\in\mathbb{Z}^2}\left|\sum_{y\in\mathbb{Z}^2}\nu_\alpha(y)e(\langle x,y\rangle\,\theta)\right|^2\Upsilon\left(\frac{x}{N^{1/2}}\right)\right)^{1/2},\end{aligned}$$

where we used (5.4) and (5.6).

Open the square, reverse orders, and apply Poisson summation:

$$\begin{aligned}|S_\alpha(\theta)| &\ll N^{\delta/2}\left(\sum_{y\in\mathbb{Z}^2}\sum_{y'\in\mathbb{Z}^2}\nu_\alpha(y)\nu_\alpha(y')\sum_{x\in\mathbb{Z}^2}e(\langle x,y-y'\rangle\,\theta)\Upsilon\left(\frac{x}{N^{1/2}}\right)\right)^{1/2}\\ &\ll N^{\delta/2}\left(\sum_{y\in\mathbb{Z}^2}\sum_{y'\in\mathbb{Z}^2}\nu_\alpha(y)\nu_\alpha(y')\sum_{k\in\mathbb{Z}^2}\int_{x\in\mathbb{R}^2}e(\langle x,y-y'\rangle\,\theta)\Upsilon\left(\frac{x}{N^{1/2}}\right)e(-\langle x,k\rangle)dx\right)^{1/2}\\ &\ll N^{\delta/2}N^{1/2}\left(\sum_{y\in\mathbb{Z}^2}\sum_{y'\in\mathbb{Z}^2}\nu_\alpha(y)\nu_\alpha(y')\sum_{k\in\mathbb{Z}^2}\widehat{\Upsilon}\left(N^{1/2}(\theta(y-y')-k))\right)\right)^{1/2}.\end{aligned}$$

Since $\operatorname{supp}\widehat{\Upsilon}\in B_{\frac{1}{10}}$, there is at most one contribution in the k sum, which is of size $\ll|\widehat{\Upsilon}(0)|\ll 1$, and only occurs if

$$\|\theta(y-y'))\|\le\frac{1}{10N^{1/2}}.^2$$

As $\theta=\frac{a}{q}+\beta$, we have

$$\left|\frac{a}{q}(y-y')\right|\le|\theta(y-y')|+|\beta(y-y')|.$$

Note that since ν_α has support in a rectangle whose side lengths are $\le\frac{1}{4}N^{1/2}$, we can bound

$$|y-y'|\le\frac{\sqrt{2}}{4}N^{1/2}.$$

Since

$$|\beta|\le\frac{1}{qN^{1/2}},$$

[2] Here $\|\cdot\|$ is the distance to the nearest lattice point in $\mathbb{Z}^2$.

we have

$$|\beta(y-y')| \le \frac{1}{qN^{1/2}}\frac{\sqrt{2}}{4}N^{1/2} \le \frac{\sqrt{2}}{4q}.$$

Using $\|\theta(y-y')\| \le \frac{1}{10N^{1/2}}$ and $q < N^{1/2}$ gives

$$\|\frac{a}{q}(y-y')\| \le \frac{1}{10N^{1/2}} + \frac{\sqrt{2}}{4q},$$

or

$$\|a(y-y')\| \le \frac{1}{10}\frac{q}{N^{1/2}} + \frac{\sqrt{2}}{4} < 0.5,$$

and hence

$$y \equiv y' (\operatorname{mod} q).$$

Moreover,

$$\|\theta(y-y')\| = |\beta(y-y')| \le \frac{1}{10N^{1/2}}.$$

Putting everything together gives

$$|S_N(\theta)| \ll N^{\delta/2}N^{1/2}\left(\sum_{y\in\mathbb{Z}^2}\nu(y)\sum_{\substack{y'\in\mathbb{Z}^2\\ y\equiv y'(q),|y-y'|\le\frac{1}{10|\beta|N^{1/2}}}}\nu(y')\right)^{1/2}. \quad (5.7)$$

Let $|\beta| = \frac{K}{N}$, and assume that $K \gg 1$ (the opposite case is handled similarly). We will analyze the innermost sum as follows.

Increase ν_α to all of ν. Fix y and recall the definition of ν. We wish to bound

$$\begin{aligned}\sum_{\substack{y'\in\mathbb{Z}^2\\ y\equiv y'(q),|y-y'|\le\frac{1}{10|\beta|N^{1/2}}}}\nu(y') &= \sum_{\varpi\in\Pi}\sum_{\substack{\gamma\in\Gamma\\ \|\gamma\|<N^\sigma}}\mathbf{1}\left\{v_0\gamma\varpi \equiv y(q), |v_0\gamma\varpi - y| \le \frac{1}{10|\beta|N^{1/2}}\right\}\\ &\le \sum_{\substack{\gamma\in\Gamma\\ \|\gamma\|<N^\sigma}}\sum_{\substack{\varpi\in\Gamma\\ \|\varpi\|<N^{1/2-\sigma}}}\mathbf{1}\left\{(c,d)\varpi \equiv y(q), |(c,d)\varpi - y| \le \frac{1}{10|\beta|N^{1/2}}\right\},\end{aligned}$$

where we have written $(c,d) = v_0\gamma$ and relaxed the condition $\varpi \in \Pi$ (which constrains the starting letter of ϖ in addition to its norm) to just $\varpi \in \Gamma$ with $\|\varpi\| < N^{1/2-\sigma}$.

Continuing to hold y fixed, we also fix (c,d), that is, fix γ. By throwing away small n's into an exceptional set $\mathfrak{E}(N)$, we may restrict to those y for which $|y| \asymp N^{1/2}$ and those γ for which $|(c,d)| \asymp N^{\sigma}$. Now we are in position to apply Theorem 2.3 to the inner sum over ϖ.

$$\sum_{\substack{\varpi\in\Gamma \\ \|\varpi\|<N^{1/2-\sigma}}} \mathbf{1}\left\{(c,d)\varpi \equiv y(q), |(c,d)\varpi - y| \le \frac{N^{1/2}}{10K}\right\} \ll \frac{N^{\delta-2\delta\sigma}}{K^{1+\delta}q^2} + N^{(1/2-\sigma)(\frac{6}{7}2\delta+\frac{5}{21})}.$$

The sum over γ contributes $N^{2\delta\sigma}$ and the sum over y contributes N^{δ}. Inserting everything into (5.7) gives

$$\begin{aligned} S_N(\theta) &\ll N^{\delta/2}N^{1/2}\left(N^{\delta}N^{2\delta\sigma}\left(\frac{N^{\delta-2\delta\sigma}}{K^{1+\delta}q^2} + N^{(1-2\sigma)(\frac{6}{7}\delta+\frac{5}{42})}\right)\right)^{1/2} \\ &\ll N^{\delta/2}N^{1/2}\left(\frac{N^{\delta}}{K^{(1+\delta)/2}q} + N^{\delta/2}N^{\delta\sigma}N^{(1/2-\sigma)(\frac{6}{7}\delta+\frac{5}{42})}\right) \\ &\ll \frac{N^{(3\delta+1)/2}}{K^{(1+\delta)/2}q} + N^{\delta+1/2+\delta\sigma+(1/2-\sigma)(\frac{6}{7}\delta+\frac{5}{42})}. \end{aligned}$$

This proves (5.2).

6. Minor Arcs II: Average $|S_N(\theta)|$ over P_β

Now we fix β and average $S_N(\theta)$.

Theorem 6.1. *Recall that* $\theta = \frac{a}{q} + \beta$, *with* $\frac{1}{2}Q \le q < Q < N^{1/2}$ *and* $|\beta| < \frac{1}{qN^{1/2}}$. *Fix* β *with* $|\beta| < \frac{2}{QN^{1/2}}$, *and let*

$$P_{Q,\beta} := \left\{\theta = \frac{a}{q} + \beta : q \sim Q, (a,q) = 1\right\},$$

so $|P| \asymp Q^2$. *Then*

$$\sum_{\theta\in P_{Q,\beta}} |S_N(\theta)| \ll_{\varepsilon} N^{\delta+1+\varepsilon}Q\left(Q^{-1/2} + N^{-\sigma} + N^{-\sigma/2-1/2}Q\right). \quad (6.2)$$

Write

$$\begin{aligned} S_N(\theta) &= \sum_{\xi\in\Xi}\sum_{\varpi\in\Pi}\sum_{\substack{\gamma\in\Gamma\\ \|\gamma\|<N^\sigma}} e(\langle v_0\cdot\gamma\xi\varpi, w_0\rangle\,\theta) \\ &= \sum_{\xi\in\Xi,\varpi\in\Pi}\sum_{\substack{\gamma\in\Gamma\\ \|\gamma\|<N^\sigma}} e(\left\langle v_0\gamma\ , w_0\,{}^t(\xi\varpi)\right\rangle\ \theta) \\ &= \sum_{x\in B_{N^{1-\sigma}}}\sum_{y\in B_{N^\sigma}} \mu(x)\nu(y)e(\langle x,y\rangle\,\theta), \end{aligned}$$

where μ and ν are now measures with $\operatorname{supp}\mu\subset B_{N^{1-\sigma}}$ and $\operatorname{supp}\nu\subset B_{N^\sigma}$ defined by:

$$\mu(x) = \sum_{\xi\in\Xi,\varpi\in\Pi}\mathbf{1}(x = w_0(\xi\varpi)^*),$$

$$\nu(y) = \sum_{\substack{\gamma\in\Gamma\\ \|\gamma\|<N^\sigma}}\mathbf{1}(y = v_0\gamma).$$

Since products of the form $\xi\varpi$ are unique, we have

$$\mu,\ \nu\le 1.$$

Write $|S_N(\theta)| = \zeta(\theta)S(\theta)$, where $|\zeta(\theta)|=1$. Then for any $\Omega\subset[0,1]$,

$$\int_\Omega |S_N(\theta)|d\theta = \int_\Omega \zeta(\theta)S_N(\theta)d\theta = \sum_x\mu(x)\sum_y\int_\Omega\zeta(\theta)\nu(y)e(\langle x,y\rangle\,\theta)d\theta.$$

Recall that $\Upsilon\in C^\infty(\mathbb{R}^2)$ is a smooth, non-negative function which is at least 1 in the unit square $[-1,1]^2$ with $\operatorname{supp}\widehat{\Upsilon}\in B_{\frac{1}{10}}$. Apply Cauchy-Schwarz, and insert Υ to retain the condition that $\operatorname{supp}\mu\subset B_{N^{1-\sigma}}$.

$$\begin{aligned} \int_\Omega |S_N(\theta)|d\theta &\ll \left(\sum_x\mu(x)^2\right)^{1/2}\left(\sum_x\left|\sum_y\int_\Omega\nu(y)\zeta(\theta)e(\langle x,y\rangle\,\theta)d\theta\right|^2\Upsilon\left(\frac{x}{N^{1-\sigma}}\right)\right)^{1/2} \\ &\ll N^{\delta(1-\sigma)}\left(\sum_y\sum_{y'}\int_{\theta\in\Omega}\int_{\theta'\in\Omega}\zeta(\theta)\overline{\zeta(\theta')}\nu(y)\nu(y')\right. \\ &\qquad\left.\times\sum_x e(\langle x, y\theta - y'\theta'\rangle)\Upsilon\left(\frac{x}{N^{1-\sigma}}\right)d\theta d\theta'\right)^{1/2} \end{aligned} \tag{6.3}$$

In the above we used the bound

$$\sum_x\mu(x)^2\le\sum_x\mu(x)\ll N^{2\delta(1-\sigma)}.$$

Write (6.3) with $\Omega = P$ (and the integral as a sum). Apply Poisson summation in the x sum and use supp $\widehat{\Upsilon} \in B_{1/10}$, together with $|\zeta| \le 1$:

$$\sum_{\theta \in P} |S_N(\theta)| \ll N^{\delta(1-\sigma)} N^{1-\sigma} \left(\sum_y \sum_{y'} \nu(y)\nu(y') \sum_{\theta \in P} \sum_{\theta' \in P} \mathbf{1}_{\{\|y\theta - y'\theta'\| < \frac{1}{10N^{1-\sigma}}\}} \right)^{1/2}. \tag{6.4}$$

As $\nu \le 1$ and supported on primitive vectors in B_{N^σ}, our task is then to count the number, say A, of points in the parentheses.

The set A contains

(1) lattice points $y = (y_1, y_2) \in B_{N^\sigma}$, and $y' = (y_1', y_2') \in B_{N^\sigma}$; and
(2) points on the circle $\theta = \frac{a}{q} + \beta$ and $\theta' = \frac{a'}{q'} + \beta$ (same β – this is the key!).

which satisfy:

(1) $|y|, |y'| < N^\sigma$, primitive vectors,
(2) $y, y' \ne (0,0)$; and
(3) $\|y_1\theta - y_1'\theta'\| < \frac{1}{10N^{1-\sigma}}$, and $\|y_2\theta - y_2'\theta'\| < \frac{1}{10N^{1-\sigma}}$, where $\|\cdot\|$ is the distance to nearest integer.

We note first that

$$\begin{aligned} \left\| (y_2' y_1 - y_1' y_2) \frac{a}{q} \right\| &= \left\| y_2' \left(y_1 \frac{a}{q} - y_1' \frac{a'}{q'} \right) - y_1' \left(y_2 \frac{a}{q} - y_2' \frac{a'}{q'} \right) \right\| \\ &\le \left\| y_2' \left(y_1 \frac{a}{q} - y_1' \frac{a'}{q'} \right) \right\| + \left\| y_1' \left(y_2 \frac{a}{q} - y_2' \frac{a'}{q'} \right) \right\| \end{aligned}$$

and that, since $|\beta| < \frac{2}{QN^{1/2}}$,

$$\begin{aligned} \left\| y_1 \frac{a}{q} - y_1' \frac{a'}{q'} \right\| &\le \|y_1\theta - y_1'\theta'\| + |\beta(y_1 - y_1')| \\ &\le \frac{1}{10N^{1-\sigma}} + \frac{2}{QN^{1/2}} \cdot 2N^\sigma. \end{aligned}$$

Therefore

$$\left\| y_2' \left(y_1 \frac{a}{q} - y_1' \frac{a'}{q'} \right) \right\| \le \frac{1}{10N} + \frac{2}{QN^{1/2}} \cdot 2N^{2\sigma},$$

and hence

$$\left\| (y_2' y_1 - y_1' y_2) \frac{a}{q} \right\| \le 2 \left(\frac{1}{10N} + \frac{2}{QN^{1/2}} \cdot 2N^{2\sigma} \right),$$

or

$$\|(y_2' y_1 - y_1' y_2)\, a\| \le 2 \left(\frac{q}{10N} + \frac{2q}{QN^{1/2}} \cdot 2N^{2\sigma} \right) \le \frac{1}{5N^{1/2}} + 8 \frac{1}{N^{1/2}} N^{2\sigma}.$$

Choose σ so that $N^\sigma < \frac{1}{4}N^{1/4}$, say. (this is where the condition $\sigma \approx \frac{1}{4}$ arises, cf. (3.1)). That is,

$$\sigma < 1/4 - \frac{\log 4}{\log N}. \tag{6.5}$$

Then the right hand side is < 1, and so

$$y_2' y_1 - y_1' y_2 \equiv 0(q).$$

In the same way, we deduce that

$$y_2' y_1 - y_1' y_2 \equiv 0(q'),$$

and hence

$$y_2' y_1 - y_1' y_2 \equiv 0(\tilde{q}),$$

where $\frac{1}{2}Q \leq \tilde{q} \leq Q^2$ is the least common multiple of q and q'.

The rest of the analysis breaks down into three regions: Either

(i) $y_1 y_2' - y_2 y_1' \neq 0$ (but is $\equiv 0(\tilde{q})$); or
(ii) $y_1 y_2' - y_2 y_1' = 0$ but $y_1 y_2 y_1' y_2' \neq 0$; or
(iii) $y_1 y_2' - y_2 y_1' = 0$ and $y_1 y_2 y_1' y_2' = 0$.

We handle these separately.

6.1. **Region (i).**

Proposition 6.6. *The contribution to A from Region (i) is*

$$\ll N^{(1+\delta)2\sigma} Q$$

The proof is as follows. Write

$$\tilde{q} \mid (y_2' y_1 - y_1' y_2),$$

and

$$y_1 y_2' - y_2 y_1' \neq 0,$$

so in particular $\tilde{q} \leq 2N^{2\sigma}$. Recall also that $\tilde{q} < Q^2$, and hence

$$\tilde{q} \leq \min(Q^2, 2N^{2\sigma}) \leq Q\sqrt{2}N^\sigma.$$

Lemma 6.7.

$$\|y_1 \frac{a}{q} - y_1' \frac{a'}{q'}\| = 0.$$

Proof. Assume not, then $\|y_1 \frac{a}{q} - y_1' \frac{a'}{q'}\|$ is at least $1/\tilde{q}$. But then (using $|\beta| < \frac{2}{QN^{1/2}}$),

$$\begin{aligned} \frac{1}{Q\sqrt{2}N^\sigma} &\leq \frac{1}{\tilde{q}} \leq \left\| y_1 \frac{a}{q} - y_1' \frac{a'}{q'} \right\| \leq \|y_1\theta - y_1'\theta'\| + |\beta(y_1 - y_1')| \\ &\leq \frac{1}{10N^{1-\sigma}} + \frac{2}{QN^{1/2}} 2N^\sigma, \end{aligned}$$

or (using $Q < N^{1/2}$ and $N^{2\sigma} < \frac{1}{16}N^{1/2}$),

$$\frac{1}{\sqrt{2}} \leq \frac{QN^{2\sigma}}{10N} + \frac{2}{N^{1/2}} 2N^{2\sigma} \leq \frac{Q\frac{1}{16}N^{1/2}}{10N} + \frac{2}{N^{1/2}} 2\frac{1}{16}N^{1/2} \leq \frac{1}{160} + \frac{1}{4},$$

which is obviously a contradiction. □

Now we have that

$$y_1 \frac{a}{q} \equiv y_1' \frac{a'}{q'} (\operatorname{mod} 1).$$

The same argument of course applies to y_2, y_2', that is,

$$y_2 \frac{a}{q} \equiv y_2' \frac{a'}{q'} (\operatorname{mod} 1)$$

Some more notation: Let $d = (q, q')$ and write $q = dq_1$, $q' = dq_1'$, with $(q_1, q_1') = 1$. Recall that $(a, q) = 1$. Hence

$$y_1 a q_1' \equiv y_1' a' q_1 (\operatorname{mod} d q_1 q_1'). \tag{6.8}$$

Looking mod q_1 gives

$$y_1 a q_1' \equiv 0 (\operatorname{mod} q_1),$$

which forces

$$q_1 \mid y_1,$$

since $(q_1, aq_1') = 1$. The same argument applies to show that

$$q_1 \mid y_2, \qquad q_1' \mid y_1', \qquad \text{and} \qquad q_1' \mid y_2'.$$

But since y is a primitive vector, $(y_1, y_2) = 1$, and hence

$$q_1 = 1 \qquad \text{and} \qquad d = q.$$

By the same token, $q_1' = 1$ and $d = q'$, so in fact (!)

$$q = q'.$$

Then (6.8) and its companion become

$$y_1 a \equiv y_1' a'(q), \qquad \text{and} \qquad y_2 a \equiv y_2' a'(q).$$

We count the contribution to A as follows. There are $\ll Q$ choices for q, then $\ll q^2$ choices for a, a'. There are $\ll N^{2\sigma\delta}$ choices for primitive pairs (y_1, y_2). Then y_1' and y_2' are determined mod q, and hence there are $\ll N^{2\sigma}q^{-2}$ choices for them, crudely (we are not using any spectral theory here!). Altogether, the contribution is

$$\ll \sum_{q\sim Q} q^2 N^{2\sigma\delta}\frac{N^{2\sigma}}{q^2} \ll N^{2\sigma(1+\delta)}Q. \tag{6.9}$$

This proves Proposition 6.6.

6.2. **Region (ii).**

Proposition 6.10. *The contribution to A from Region (ii) is*

$$\ll_\varepsilon N^{2\sigma\delta+\varepsilon}Q^2 + N^{2\sigma\delta+\sigma-1+\varepsilon}Q^4,$$

for any $\varepsilon > 0$.

Recall that in this region,

$$y_1y_2' - y_2y_1' = 0,$$

and so

$$\tilde{q} \mid (y_2'y_1 - y_1'y_2)$$

is vacuous. We also have in this region that y_1, y_2, y_1', y_2', are all non-zero. Moreover, the vectors y and y' are primitive. By unique factorization, $y_1y_2' = y_2y_1'$ forces

$$y_1 = \pm y_1', \qquad\qquad y_2 = \pm y_2'.$$

Hence there are $\ll N^{2\sigma\delta}$ choices for y, y'.

Let $q_1 = (y_1, q)$ and $q_1' = (y_1', q') = (y_1, q')$. As $q_1, q_1' \mid y_1$, there are $\ll_\varepsilon N^\varepsilon$ choices for q_1 and q_1'. Assume without loss of generality that $q_1 \le q_1'$. Fix a' and q', for which there are

$$\ll Q \cdot \frac{Q}{q_1'} \tag{6.11}$$

choices.

Write $y_1 = q_1z_1$ and $q = q_1q_2$. Then $\|y_1\frac{a}{q} - y_1'\frac{a'}{q'} + \beta(y_1 - y_1')\| < \frac{1}{10N^{1-\sigma}}$ becomes

$$\|z_1\frac{a}{q_2} - \psi\| < \frac{1}{10N^{1-\sigma}},$$

where $\psi = y_1'\frac{a'}{q'} + \beta(y_1 - y_1')$ is already fixed.

The grid in the unit interval of possible values of $z_1 \frac{a}{q_2}$ as a and q_2 vary has mesh of size at least

$$\frac{4q_1^2}{Q^2}.$$

Hence the set of values of $z_1 \frac{a}{q_2}$ satisfying the above proximity to ψ is

$$\ll \frac{Q^2}{4q_1^2} \frac{1}{N^{1-\sigma}} + 1 \tag{6.12}$$

Let $z_1 \frac{a}{q_2} \equiv \tilde{\psi} (\mathrm{mod}\, 1)$ for some fixed grid point $\tilde{\psi}$. Since

$$(q_2, a, z_1) = 1,$$

this determines q_2 uniquely. Then a is determined $(\mathrm{mod}\, q_2)$, so has

$$\frac{q}{q_2} = q_1 \tag{6.13}$$

possible values.

Combining (6.11), (6.12), and (6.13), we have that the contribution to A from Region (ii) is at most:

$$\begin{aligned} \sum_y \nu(y) \sum_{y'=\pm y} \sum_{\substack{q_1 | y_1, q_1' | y_1 \\ q_1 \le q_1' \le \min(Q, N^\sigma)}} & Q \cdot \frac{Q}{q_1'} \left(1 + \frac{Q^2}{4q_1^2} \frac{1}{N^{1-\sigma}} \right) q_1 \\ \ll_\varepsilon \quad & N^{2\sigma\delta+\varepsilon} Q^2 + N^{2\sigma\delta+\sigma-1+\varepsilon} Q^4, \end{aligned}$$

for any $\varepsilon > 0$, as claimed.

6.3. **Region (iii):**

Proposition 6.14. *The contribution to A from Region (iii) is*

$$\ll N^\sigma Q^2.$$

Recall that in this region,

$$y_1 y_2' = y_2 y_1',$$

and $y_1 = 0$, say. By primitivity, $y_2 = \pm 1$, and by the above, $y_1' = 0$ and again $y_2' = \pm 1$.

The analysis is now the same as in Region (ii) except $q_1 = q_1' = 1$, so there are no ε's. The contribution is

$$\begin{aligned} &\ll \quad Q \cdot Q\left(1 + \frac{Q^2}{4}\frac{1}{N^{1-\sigma}}\right) \ll Q^2 + Q^4 N^{\sigma-1} \\ &\ll \quad Q^2 + Q^2 N^{\sigma} \ll Q^2 N^{\sigma}. \end{aligned}$$

6.4. **Conclusion.**

Combining (6.4) with Propositions 6.6, 6.10, and 6.14, gives

$$\begin{aligned} \sum_{\theta\in P} |S(\theta)| \quad &\ll \quad N^{\delta(1-\sigma)} N^{1-\sigma}\ A^{1/2} \\ &\ll_\varepsilon \quad N^{\delta(1-\sigma)} N^{1-\sigma}\ \left(N^{2\sigma(1+\delta)}Q + N^{2\sigma\delta+\varepsilon}Q^2 + N^{2\sigma\delta+\sigma-1+\varepsilon}Q^4 + N^{\sigma}Q^2\right)^{1/2} \\ &\ll_\varepsilon \quad N^{\delta(1-\sigma)} N^{1-\sigma}\ QN^{\sigma+\sigma\delta+\varepsilon}\left(Q^{-1/2} + N^{-\sigma} + N^{-\sigma/2-1/2}Q\right), \end{aligned}$$

as claimed. This completes the proof of Theorem 6.1.

7. Minor Arcs III: Average $|S_N|^2$ over $W_{Q,K}$

The main goal of this section is to prove

Theorem 7.1.

$$\int_{W_{Q,K}} |S_N(\theta)|^2 d\theta \ll \log N\ \left(N^{\delta+\delta\sigma}\left\|\left.S_N(\theta)\right|_{W_{Q,K}}\right\|_{L^\infty} + N^{2\delta+1-\sigma}\right). \tag{7.2}$$

We begin with a lemma.

Lemma 7.3. *Let*

$$\Omega \subset [0,1]$$

be a finite union of open intervals. Then

$$\int_{\Omega} |S_N(\theta)| d\theta \ll \max\left(N^{\delta+\delta\sigma}\ ,\ N^{\delta+(1-\sigma)/2}|\Omega|^{1/2}\right). \tag{7.4}$$

Proof. Returning to (6.3), apply Poisson summation in the x sum and use supp $\widehat{\Upsilon} \in B_{1/10}$, together with $|\zeta| \leq 1$:

$$\int_\Omega |S_N(\theta)| d\theta \ll N^{\delta(1-\sigma)} N^{1-\sigma} \left(\sum_y \sum_{y'} \nu(y)\nu(y') \, meas \left\{ (\theta, \theta') : \begin{array}{l} \|\theta y_1 - \theta' y_1'\| < \frac{1}{N^{1-\sigma}} \\ \|\theta y_2 - \theta' y_2'\| < \frac{1}{N^{1-\sigma}} \end{array} \right\} \right)^{1/2}.$$

Let $Y := \begin{pmatrix} y_1 & y_2 \\ -y_1' & -y_2' \end{pmatrix}$ and consider two regions: either $\det Y = 0$ or not.

If not, then $(\theta, \theta') \mapsto (\theta, \theta')Y = (\theta y_1 - \theta' y_1', \theta y_2 - \theta' y_2')$ is a map which is measure preserving $(\text{mod}\, 1 \times 1)$. Hence the preimage has

$$meas \left\{ (\theta, \theta') : \begin{array}{l} \|\theta y_1 - \theta' y_1'\| < \frac{1}{N^{1-\sigma}} \\ \|\theta y_2 - \theta' y_2'\| < \frac{1}{N^{1-\sigma}} \end{array} \right\} \ll N^{2\sigma - 2},$$

and there are $N^{4\sigma\delta}$ choices for y, y'.

If on the other hand the determinant of Y is zero, then by primitivity, $y = \pm y'$, so there are only $N^{2\sigma\delta}$ choices. Assume $y_1 \neq 0$. Fix $\theta' \in \Omega$ (contributing at most $|\Omega|$); then θ satisfies $\|y_1\theta - \theta_0\| < \frac{1}{N^{1-\sigma}}$ for some fixed θ_0. Hence the contribution is at most

$$N^{2\sigma\delta} |\Omega| N^{\sigma - 1}.$$

Combining the two regions gives

$$\begin{aligned} \int_\Omega |S_N(\theta)| d\theta &\ll N^{\delta(1-\sigma)} N^{1-\sigma} \left(N^{4\sigma\delta} N^{2\sigma-2} + N^{2\sigma\delta} |\Omega| N^{\sigma-1} \right)^{1/2} \\ &\ll N^{\delta + \delta\sigma} + N^{\delta + (1-\sigma)/2} |\Omega|^{1/2}, \end{aligned}$$

as claimed. □

Proof of Theorem 7.1. Let $W_{Q,K} = \Omega$. As $|S_N(\theta)| \ll N^{2\delta}$, we can take a dyadic subdivision $M \ll N^{2\delta}$ of $\ll \log N$ terms, and decompose Ω into level sets

$$\Omega = \bigsqcup_\alpha \Omega_\alpha,$$

according to the size of $|S_N(\theta)|$. So if $\theta \in \Omega_\alpha$, then $\frac{M}{2} \leq |S_N(\theta)| < M$ with $M \ll N^{2\delta}$.

On any such level set, we have

$$\frac{1}{|\Omega_\alpha|} \int_{\Omega_\alpha} |S_N(\theta)| d\theta \asymp \sup_{\theta \in \Omega_\alpha} |S_N(\theta)|,$$

so

$$\begin{aligned}\int_{\Omega}|S_N(\theta)|^2d\theta &\ll \log N \sup_{\alpha} \int_{\Omega_\alpha}|S_N(\theta)|^2d\theta\\ &\ll \log N \sup_{\alpha} \sup_{\theta\in\Omega_\alpha}|S_N(\theta)|\int_{\Omega_\alpha}|S_N(\theta)|d\theta\\ &\ll \log N \sup_{\alpha} \sup_{\theta\in\Omega_\alpha}|S_N(\theta)|\max\left(N^{\delta+\delta\sigma}, N^{\delta+(1-\sigma)/2}|\Omega_\alpha|^{1/2}\right)\\ &\ll \log N \max\left(\sup_{\theta\in\Omega}|S_N(\theta)|N^{\delta+\delta\sigma},\right.\\ &\qquad \left.\sup_{\alpha} N^{\delta+(1-\sigma)/2}|\Omega_\alpha|^{-1/2}\int_{\Omega_\alpha}|S_N(\theta)|d\theta\right)\\ &\ll \log N \max\left(\sup_{\theta\in\Omega}|S_N(\theta)|N^{\delta+\delta\sigma}\ ,\ N^{2\delta+(1-\sigma)}\right)\\ &\ll \log N \left(N^{\delta+\delta\sigma}\left\|S_N(\theta)\Big|_{\Omega}\right\|_{L^\infty}+N^{2\delta+1-\sigma}\right),\end{aligned}$$

where we used (7.4) twice (the second time under the assumption $N^{\delta+\delta\sigma} \ll N^{\delta+(1-\sigma)/2}|\Omega_\alpha|^{1/2}$). This completes the proof. □

8. Putting It All Together

Now we combine the previous estimates to show that the minor arcs are small. We will write

$$Q = N^\alpha \qquad \text{and} \qquad K = N^\kappa,$$

with $\alpha, \kappa \in [0, 1/2]$. For the vast majority of θ, the weight $\mathfrak{m}(\theta)$ is identically one. We deal with these θ's first, via the following three Lemmata.

Lemma 8.1. *As $N \to \infty$,*

$$\int_{W_{Q,K}} |S_N(\theta)|^2 d\theta \ll N^{4\delta-1-\eta},$$

if

$$\alpha + \frac{1+\delta}{2}\kappa > \frac{3}{2}(1-\delta) + \delta\sigma, \tag{8.2}$$

$$\sigma > 2(1-\delta), \tag{8.3}$$

and

$$\sigma < \frac{132\delta - 131}{96\delta - 10}. \tag{8.4}$$

This Lemma is conclusive if either K or Q is large.

Proof. Putting (5.2) into (7.2) and ignoring log's gives:

$$\int_{W_{Q,K}} |S_N(\theta)|^2 d\theta \ll N^{\delta+\delta\sigma} N^{(3\delta+1)/2} \left(\frac{1}{K^{(1+\delta)/2} Q} + N^{-\frac{1}{84}(1-2\sigma)(6\delta-5)} \right) + N^{2\delta+1-\sigma}.$$

For the last term of the right-hand side to be bounded by $N^{4\delta-1-\eta}$, we need

$$\sigma > 2(1-\delta).$$

For the middle term above to be bounded by $N^{4\delta-1-\eta}$, we need

$$\sigma < \frac{132\delta - 131}{96\delta - 10}.$$

Lastly, the first term above is controlled if

$$\alpha + \frac{1+\delta}{2}\kappa > \frac{3}{2}(1-\delta) + \delta\sigma.$$

As δ is very near 1, this essentially requires that $\alpha + \kappa > \sigma$ in order to bound the minor arcs outright. ☐

If both K and Q are too small to apply the above, but not small enough to be in the major arcs, we try the next lemma, which is conclusive if K is almost as small as the major arcs.

Lemma 8.5. *As $N \to \infty$,*

$$\int_{W_{Q,K}} |S_N(\theta)|^2 d\theta \ll N^{4\delta-1-\eta},$$

if

$$\kappa > \frac{1-\delta}{\delta}, \tag{8.6}$$

and

$$1 - \delta + \kappa + 2\alpha < \frac{1}{42}(6\delta - 5)(1 - 2\sigma). \tag{8.7}$$

Proof. Using (5.2) twice leads to the bound

$$\begin{aligned} \int_{W_{Q,K}} |S_N(\theta)|^2 d\theta &\ll Q^2 \frac{K}{N} \left(\frac{N^{3\delta+1}}{K^{1+\delta} Q^2} + N^{-\frac{1}{42}(6\delta-5)(1-2\sigma)} \right) \\ &= \frac{N^{3\delta}}{K^{\delta}} + Q^2 K N^{3\delta} N^{-\frac{1}{42}(6\delta-5)(1-2\sigma)}. \end{aligned}$$

In order for this to be sufficient, we need both

$$\kappa > \frac{1-\delta}{\delta},$$

and

$$1-\delta+\kappa+2\alpha < \frac{1}{42}(6\delta-5)(1-2\sigma),$$

as claimed. □

The remaining case is when K and Q are too small to apply Lemma 8.1, and K is too small for Lemma 8.5 to suffice, that is, when Q is small, but not so small as to be in the major arcs.

Lemma 8.8. *As $N \to \infty$,*

$$\int_{W_{Q,K}} |S_N(\theta)|^2 d\theta \ll N^{4\delta-1-\eta},$$

if the conditions (8.9) *through* (8.14) *are satisfied.*

Proof. Now add (5.2) to (6.2):

$$\begin{aligned}\int_{W_{Q,K}} |S_N(\theta)|^2 d\theta &\ll \sup |S_N(\theta)| \frac{K}{N} \sum_P |S_N(\theta)| \\ &\ll N^{(3\delta+1)/2}\left(\frac{1}{K^{(1+\delta)/2}Q} + N^{-\frac{1}{84}(6\delta-5)(1-2\sigma)}\right)\frac{K}{N} \\ &\quad \times N^{\delta+1+\varepsilon}Q\left(Q^{-1/2}+N^{-\sigma}+N^{-\sigma/2-1/2}Q\right).\end{aligned}$$

For these to all be conclusive, that is, $\ll N^{4\delta-1-\eta}$, we need the following six conditions.

$$(1-\delta)\kappa + 3(1-\delta) < \alpha \tag{8.9}$$

and

$$(1-\delta)\kappa/2 + 3(1-\delta)/2 < \sigma \tag{8.10}$$

and

$$(1-\delta)\kappa/2 + 3(1-\delta)/2 + \alpha < (1+\sigma)/2 \tag{8.11}$$

and

$$3(1-\delta)/2 + \kappa + \alpha/2 < \frac{1}{84}(6\delta-5)(1-2\sigma) \tag{8.12}$$

and

$$3(1-\delta)/2 + \kappa + \alpha < \frac{1}{84}(6\delta-5)(1-2\sigma) + \sigma \tag{8.13}$$

and

$$3(1-\delta)/2 + \kappa + 2\alpha < \frac{1}{84}(6\delta-5)(1-2\sigma) + \sigma/2 + 1/2. \tag{8.14}$$

□

Lastly, dispose of those θ with non-trivial weights in the minor arcs function $\mathfrak{m}(\theta)$. Recall from (3.3) the triangle function ψ used to form $\mathfrak{M}(\theta)$ and $\mathfrak{m}(\theta)$. Note that in $[-1,1]$ the function $1-\psi(x)$ is just $|x|$. These values of θ (which should be contained in the major arcs but receive some small weights here) are controlled as follows:

$$\begin{aligned}\int_{\mathfrak{m}(\theta)\neq 1}|\mathfrak{m}(\theta)|^2|S_N(\theta)|^2d\theta &\ll \sum_{q<Q}\sum_{(a,q)=1}\int_{|\beta|<K/N}(\beta N/K)^2\frac{N^{3\delta+1}}{(N\beta)^{1+\delta}Q^2}d\beta \\ &\ll \frac{N^{3\delta}}{K^{\delta}},\end{aligned}$$

using the first term in the bound (5.2) ignoring the second term, which was handled in (8.7). This is estimated in the same way as in Lemma 8.5.

8.1. **Conclusion.**

We must now collect all of the inequalities needed above and try to make sense of them.

We need to find values of δ and σ such that every pair (α,κ) either lies in the major arcs, that is, satisfies (4.2) and (4.3), or lies in one of the three regions described in Lemmata 8.1, 8.5, and 8.8.

The main roles in α and κ are played by (4.3) in the major arcs, and (8.6) and (8.9) in the minor arcs. Putting these together with the main condition (8.2) on δ and σ gives the system of inequalities:

$$\begin{cases}\sigma > 2(1-\delta)\\ 21\left((1-\delta)\left(\frac{1-\delta}{\delta}\right)+3(1-\delta)\right)+13\left(\frac{1-\delta}{\delta}\right) < \left(2\delta-\frac{5}{3}\right)\sigma.\end{cases} \tag{8.15}$$

The region in the (δ,σ) plane for which (8.15) and (8.4) are satisfied is depicted in Figure 1. Solving for the minimal value of δ in this region, one finds that this value is The minimal value for δ is the largest root of the cubic polynomial:

$$1020-8897x-5010x^2+12888x^3, \tag{8.16}$$

whose plot is shown in Fig. 2. The approximate value is

$$\delta\approx 0.9999493550. \tag{8.17}$$

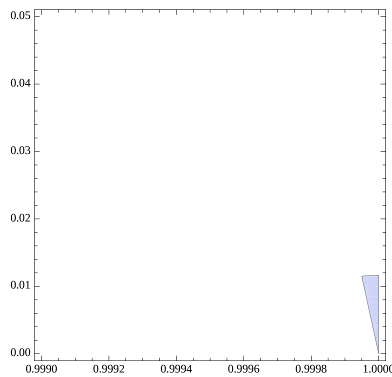

FIGURE 1. Values in the (δ, σ) plane satisfying (8.15) and (8.4).

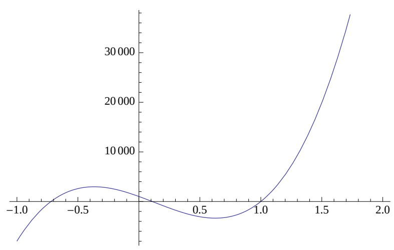

FIGURE 2. The plot of the polynomial (8.16).

The corresponding value of σ is the root nearest the origin of the polynomial

$$4995 - 434163x + 149452x^2 + 700x^3,$$

which takes the approximate value

$$\sigma \approx 0.011550825843.$$

With these values of δ and σ, Fig. 3 shows the overlapping regions, corresponding to the major arcs, and Lemmata 8.1, 8.5, and 8.8.

The largest region corresponds to Lemma 8.1, where we see the negatively sloped line corresponding to the condition (8.4). The triangle to the right on the bottom corresponds to Lemma 8.5, since it contains the vertical line showing the inequality $\kappa > \frac{1-\delta}{\delta}$. The triangle on the left side corresponds to Lemma 8.8. On the magnified image, the vertical line is again Lemma 8.5 and the horizontal line is Lemma 8.8. The triangle based at the origin corresponds to the major arcs, with the negatively sloped line being the condition (4.3).

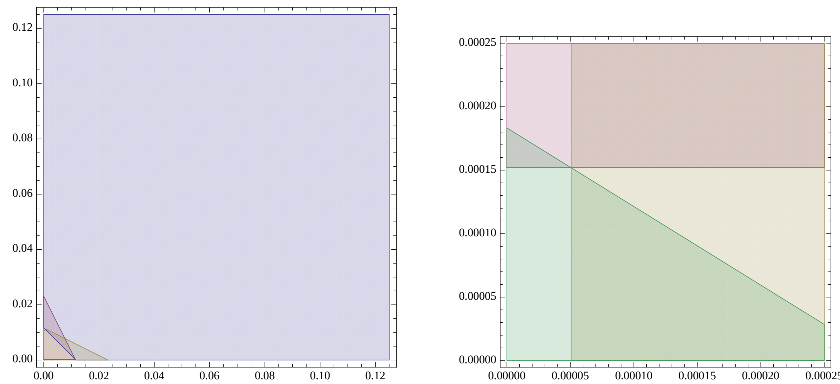

FIGURE 3. Overlapping regions at different magnifications in the (κ, α)-plane corresponding to the major arcs and Lemmata 8.1, 8.5, and 8.8.

We have proved:

Theorem 8.18. *Assume that*

$$\delta > 0.9999493550.$$

Then there is a choice for σ *such that as* $N \to \infty$, *there is some* $\eta > 0$ *with*

$$\sum_{|n|<N} |\mathcal{E}_N(n)|^2 = \int_0^1 |\mathfrak{m}(\theta)|^2 |S_N(\theta)|^2 d\theta \ll N^{4\delta-1-\eta}.$$

We conclude with the standard argument below.

Theorem 8.19. *Theorems 4.1 and 8.18 imply Theorem 1.2.*

Proof. Let $\mathfrak{E}(N)$ be the set of exceptions up to N. Let $\mathcal{Z}$ denote those integers passing local obstructions. Then

$$\begin{aligned} |\mathfrak{E}(N)| = & \sum_{\substack{|n|<N \\ n \in \mathcal{Z}, |\mathcal{E}_N(n)| > \mathcal{M}_N(n)}} 1 \\ \ll & \sum_{\substack{|n|<N \\ |\mathcal{E}_N(n)| \gg \frac{1}{\log\log n} N^{2\delta-1}}} 1 \\ \ll & \sum_{|n|<N} |\mathcal{E}_N(n)|^2 (\log\log N)^2 N^{2-4\delta} \\ \ll & N^{1-\eta}(\log\log N)^2. \end{aligned}$$

This completes the proof. $\square$

References

[BF10] J. Bourgain and E. Fuchs, 2010. In preparation.

[BGS06] Jean Bourgain, Alex Gamburd, and Peter Sarnak. Sieving and expanders. *C. R. Math. Acad. Sci. Paris*, 343(3):155–159, 2006.

[BGS08] Jean Bourgain, Alex Gamburd, and Peter Sarnak. Affine linear sieve, expanders, and sum-product, 2008.

[BKS09] J. Bourgain, A. Kontorovich, and P. Sarnak. Sector estimates for hyperbolic isometries, 2009. Preprint.

[Gam02] Alex Gamburd. On the spectral gap for infinite index "congruence" subgroups of $SL_2(\mathbf{Z})$. *Israel J. Math.*, 127:157–200, 2002.

[GLM+03] R.L. Graham, J.C. Lagarias, C.L. Mallows, A.R. Wilks, and C.H. Yan. Apollonian circle packings:number theory. *J. Numb. Th.*, 100:1–45, 2003.

[Klo27] H. D. Kloosterman. On the representation of numbers in the form $ax^2 + by^2 + cz^2 + dt^2$. *Acta Math.*, 49(3-4):407–464, 1927.

[KO08] A. Kontorovich and H. Oh. Apollonian circle packings and closed horospheres on hyperbolic 3-manifolds, 2008. Preprint, *http://arxiv.org/abs/0811.2236*.

[KO09] A. Kontorovich and H. Oh. Almost prime Pythagorean triples in thin orbits, 2009. Preprint, *http://arxiv.org/abs/1001.0370*.

[Kon09] A. V. Kontorovich. The hyperbolic lattice point count in infinite volume with applications to sieves. *Duke J. Math.*, 149(1):1–36, 2009. *http://arxiv.org/abs/0712.1391*.

[LP82] P.D. Lax and R.S. Phillips. The asymptotic distribution of lattice points in Euclidean and non-Euclidean space. *Journal of Functional Analysis*, 46:280–350, 1982.

[MV75] H. L. Montgomery and R. C. Vaughan. The exceptional set in Goldbach's problem. *Acta Arith.*, 27:353–370, 1975.

[NS09] A. Nevo and P. Sarnak. Prime and almost prime integral points on principal homogeneous spaces, 2009.

[Sar07] P. Sarnak. Letter to J. Lagarias, 2007. Online at *http://www.math.princeton.edu/sarnak*.

[Sul84] D. Sullivan. Entropy, Hausdorff measures old and new, and limit sets of geometrically finite Kleinian groups. *Acta Math.*, 153(3-4):259–277, 1984.

E-mail address: `bourgain@ias.edu`

IAS, Princeton, NJ

E-mail address: `avk@math.ias.edu`

IAS and Brown, Princeton, NJ