Vol. 100, No. 2 DUKE MATHEMATICAL JOURNAL © 1999

THE DISTRIBUTION OF SPACINGS BETWEEN QUADRATIC RESIDUES

PÄR KURLBERG AND ZEÉV RUDNICK

1. Introduction. Our goal in this paper is to study the distribution of spacings (or gaps) between squares in $\mathbf{Z}/q\mathbf{Z}$, as $q \to \infty$. In the case that q is prime, a theorem of Davenport (see [3], [4], [11], and [18]) shows that the probability of two consecutive quadratic residues modulo a prime q being spaced h units apart is 2^{-h}, as $q \to \infty$. For our purposes, we may interpret this result as saying that when we normalize the spacings to have unit mean, then the distribution of spacing as $q \to \infty$ along primes is given by

$$P(s) = \sum_{h=1}^{\infty} 2^{-h} \delta\left(s - \frac{h}{2}\right),$$

that is, a sum of point masses at half-integers with exponentially decreasing weights.

In this paper, we study the spacing distribution of squares modulo q when q is square-free and *highly composite*, that is, the limiting distribution of spacings between the squares modulo q as the number of prime divisors, $\omega(q)$, tends to infinity. For odd square-free q, the number N_q of squares modulo q equals

$$N_q = \prod_{p|q} \frac{p+1}{2}.$$

This is because, if p is an odd prime, the number of squares modulo p is $(p+1)/2$ and, for q square-free, x is a square modulo q if and only if x is a square modulo p for all primes p dividing q. Thus, for odd q, the mean spacing $s_q = q/N$ equals

$$s_q = \frac{2^{\omega(q)}}{\prod_{p|q}(1+1/p)} = \frac{2^{\omega(q)}}{\sigma_{-1}(q)}.$$

For $q = 2q'$ even and square-free, it is easily seen that $s_q = s_{q'}$. It follows that $s_q \to \infty$ as $\omega(q) \to \infty$, unlike the case of prime q, where the mean spacing is essentially constant. Thus, unlike in the prime case (where the level spacing distribution was forced to be supported on a lattice), in the highly composite case, there is an a priori chance of getting a continuous distribution.

Received 3 August 1998.

1991 *Mathematics Subject Classification.* Primary 11.

Authors' work supported in part by Israel Science Foundation grant number 192/96. Kurlberg also partially supported by the European Community Training and Mobility of Researchers network "Algebraic Lie Representations," EC-contract number ERB FMRX-CT97-0100.

A relevant statistical model for the distribution of spacings is given by looking at random points in the unit interval $\mathbf{R}/\mathbf{Z}$. For independent, uniformly distributed numbers in $\mathbf{R}/\mathbf{Z}$, the spacing statistics are said to be *Poissonian*. The distribution $P(s)$ of spacings between consecutive points is that of a Poisson arrival process, that is, $P(s) = e^{-s}$ (see [6]). Moreover, the joint distribution of k-consecutive spacings is the product of k-independent exponential random variables.

It is well known (see [15]) that the spacing statistics of the superposition of several independent spectra converge to the Poisson case—the spacing statistics of uncorrelated levels. Thus, the heuristic that "primes are independent," together with Davenport's result, indicates that the spacing statistics of the squares modulo q should be, in the limit as $\omega(q) \to \infty$, Poissonian; that is, in some sense, squares modulo q behave as random numbers. It is our purpose to confirm this expectation.

In order to study the level spacings, we proceed by studying the *r-level correlation functions*. These measure clustering properties of a sequence in $\mathbf{R}/\mathbf{Z}$ on a scale of the mean spacing. Their definition and their application to computing various local spacings statistics are recalled in Appendix A. In our case, these turn out to be given by the following. For $r \geq 2$ and a bounded convex set $\mathscr{C} \subset \mathbf{R}^{r-1}$, let

$$R_r(\mathscr{C}, q) = \frac{1}{N_q} \#\big\{x_i \text{ distinct squares } \bmod q : (x_1 - x_2, \ldots, x_{r-1} - x_r) \in s\mathscr{C}\big\}.$$

This is immediately transformed into

$$R_r(\mathscr{C}, q) = \frac{1}{N_q} \sum_{h \in s\mathscr{C} \cap \mathbf{Z}^{r-1}} N(h, q), \tag{1.1}$$

where $N(h, q)$ is the number of solutions of the system of congruences $y_{i+1} - y_i = h_i \bmod q$ with $y_1, y_2, \ldots, y_r$ squares modulo q and $h = (h_1, \ldots, h_{r-1}) \in \mathbf{Z}^{r-1}$.

To compute the correlations for distinct x_i, we consider only sets $\mathscr{C}$ that a priori only contain vectors $(x_i - x_{i+1})$ with distinct coordinates. To do this, we define "roots" σ_{ij} on $\mathbf{R}^{r-1}$ for $i < j$ by $\sigma_{ij}(h) = \sum_{k=i}^{j-1} h_k$. The hyperplanes $\{\sigma_{ij} = 0\} \subset \mathbf{R}^{r-1}$ are called "walls," and $(x_i - x_{i+1})$ does not lie in any of the walls if and only if all coordinates x_i are distinct.

Our main result shows that if $\mathscr{C}$ does not intersect any wall, then $R_r(\mathscr{C}, q) \to \mathrm{vol}(\mathscr{C})$ for any sequence of square-free q with $\omega(q) \to \infty$.

THEOREM 1. *Let q be square-free, let $r \geq 2$, and let $\mathscr{C} \subset \mathbf{R}^{r-1}$ be a bounded convex set that does not intersect any of the walls. Then the r-level correlation function satisfies*

$$R_r(\mathscr{C}, q) = \mathrm{vol}(\mathscr{C}) + O\big(s^{-1/2+\epsilon}\big) \quad \text{as } s \to \infty$$

for all $\epsilon > 0$, where s is the mean spacing.

This theorem implies that all spacing statistics are Poissonian (see Appendix A). For instance, if we denote by $s_1, \dots, s_{N-1}$ the normalized differences between neighboring squares, then we have the following theorem.

THEOREM 2. *For q square-free, the limiting level spacing distribution of the squares modulo q is given by $P(t) = \exp(-t)$ as $\omega(q) \to \infty$. Moreover, under the same condition, for any $k \geq 1$, the limiting joint distribution of $(s_n, s_{n+1}, \dots, s_{n+k})$ is a product $\prod_{i=0}^{k} \exp(-t_i)$ of $k+1$ independent exponential variables.*

There are only a few known cases where the complete spacing distribution can be proved to be Poissonian as in our case. A notable example is Hooley's results (see [7], [8], [9], and [10]) that the spacings between elements coprime to q are Poissonian as the mean spacing $q/\phi(q) \to \infty$. A much more recent result is due to Cobeli and Zaharescu [2], who show that the spacings between primitive roots with respect to a prime p are Poissonian provided the mean spacing $p/\phi(p-1) \to \infty$.

The results of this paper are related to work on the level spacing distribution of the fractional parts $\{\alpha n^2\}$ (α irrational) by Rudnick, Sarnak, and Zaharescu [16], [17]. In particular, in [17], an attempt to study that problem is made by replacing α with a rational approximation b/q, and this leads to a study of the spacings of the sequence $bn^2 \bmod q$, $1 \leq n \leq N$ for N a small power of q. The available sites are exactly the set of squares with respect to q, hence our interest in the problem.

In [17], it is shown that in order for all the correlation functions of the sequence $\{\alpha n^2\}$ to have Poisson behavior, it is necessary to assume that the rational approximants b/q have denominator q that is close to square-free, hence our interest in the square-free case. For arbitrary q, it is still true that all correlations are Poissonian, but there are significant technical complications to overcome in proving this (see [13]).

We believe that the methods developed in this paper should be useful in studying similar problems, for instance, the spacing distribution of cubes modulo q, as the number of prime factors of q that are congruent to 1 modulo 3 tends to infinity. (The condition modulo 3 is necessary in order for the mean spacing to go to infinity.)

Contents of the paper. We begin with a section sketching the argument for Theorem 1 in the case of the pair correlation function. This section can be used as a guide to the rest of the paper.

In Section 3, we first reduce the problem to the case that q is odd. Then in Section 4, we analyze the behavior of $N(h, p)$, where p is prime. Squares that are distinct modulo q are not necessarily distinct modulo p; we denote by $r_{\text{eff}}(h)$ the number of squares that remains distinct after reduction modulo p. Using an inclusion-exclusion argument, we write $r_{\text{eff}}(h)$ as a linear combination of characteristic functions of certain hyperplanes over $\mathbf{Z}/p\mathbf{Z}$. Next, in Section 5, we use the multiplicative properties of the counting functions $N(h, q)$ to derive an expression for $R_r(\mathcal{C}, q)$ as a sum over divisors c of q and lattices L arising from intersections of hyperplanes modulo p for different p's (see Proposition 6).

In Section 6, we show that the main term of the sum consists of those terms for

which the product of c and the discriminant of L are small with respect to s, and an error term corresponding to terms where the product is large. In Section 7, we evaluate the main term and show that it gives us exactly $\mathrm{vol}(\mathscr{C})$, thus giving us our main result.

In Appendix A, we explain how to use Theorem 1 to derive results such as those in Theorem 2, that is, that the level spacings are Poissonian as well. Appendix B explains some background on counting lattice points in convex sets. In Appendix C, we estimate the number of divisors of q that are smaller than a fixed power of the mean spacing s.

2. The pair correlation: A sketch. In order to explain the proof of Theorem 1, we give an overview of the argument in the special case of the pair correlation function.

Let q be an odd, square-free number with $\omega(q)$ prime factors, and let I be an interval not containing the origin. As in the introduction, define the pair correlation function

$$R_2(I,q) = \frac{1}{N} \sum_{h \in sI \cap \mathbf{Z}} N(h,q),$$

where N is the number of squares modulo q, $s = q/N = 2^{\omega(q)}/\sigma_{-1}(q)$ is their mean spacing, $\sigma_{-1}(q) = \prod_{p|q}(1+(1/p))$, and $N(h,q)$ is the number of solutions in squares modulo q of the equation

$$y_1 - y_2 = h \mod q.$$

We sketch a proof that $R_2(I,q) \to |I|$ as $\omega(q) \to \infty$ ($|I|$ being the length of the interval). In fact, we have the more precise result in the following theorem.

THEOREM 3. *For q odd and square-free, we have, for all $\epsilon > 0$,*

$$R_2(I,q) = |I| + O\big(s^{-1+\epsilon}\big).$$

Proof. Here are the main steps in the argument.

Step 1. By the Chinese remainder theorem, $N(h,q) = \prod_{p|q} N(h,p)$ is a product over primes dividing q. By elementary considerations, one sees that

$$N(h,p) = \frac{p + a(h,p)}{4} \Delta(h,p) \tag{2.1}$$

with $a(h,p) = O(1)$ and

$$\Delta(h,p) = 1 + \delta(h,p), \qquad \delta(h,p) = \begin{cases} 0 & p \nmid h, \\ 1 & p \mid h. \end{cases}$$

From this, we see that

$$N(h,q) = \frac{q\,\Delta(h,q)}{4^{\omega(q)}} \sum_{c|q} \frac{a(h,c)}{c} \tag{2.2}$$

with $a(h,c) := \prod_{p|c} a(h,p) \ll c^{\epsilon}$ and $\Delta(h,q) = \prod_{p|q} \Delta(h,p)$.

Step 2. We decompose $\Delta(h,q) = \Delta(h,c)\Delta(h,q/c)$ and rewrite $\Delta(h,q/c)$ as

$$\Delta\left(h,\frac{q}{c}\right) = \prod_{p|(q/c)} \big(1+\delta(h,p)\big) = \sum_{g|(q/c)} \delta(h,g)$$

with

$$\delta(h,g) = \begin{cases} 0 & g \nmid h, \\ 1 & g \mid h. \end{cases}$$

Substituting this into the expression (2.2) for $N(h,q)$, and inserting the result into the formula for $R_2(I,q)$, we get

$$R_2(I,q) = \frac{1}{\sigma_{-1}(q)2^{\omega(q)}} \sum_{c|q} \frac{1}{c} \sum_{g|(q/c)} \sum_{h \in sI \cap g\mathbf{Z}} a(h,c)\Delta(h,c). \tag{2.3}$$

Step 3. We partition the sum into two parts, one over the pairs g,c with $gc < s$, and the leftover part over pairs with $gc \geq s$. We show this leftover part is negligible (in fact, $O(s^{-1+\epsilon})$). We first use $a(h,c)\Delta(h,c) \ll c^{\epsilon}$ and the fact that, in order for the inner sum over h to be nonempty, we need $g \ll s$ (recall that I does not contain the origin) to get that the sum over pairs with $cg > s$ is bounded by

$$\begin{aligned} s^{-1+\epsilon} \sum_{c|q} c^{-1+\epsilon} \sum_{\substack{g|(q/c) \\ g \ll s \\ cg > s}} \#(sI \cap g\mathbf{Z}) &\ll s^{-1+\epsilon} \sum_{c|q} c^{-1+\epsilon} \sum_{\substack{g|(q/c) \\ g \ll s \\ cg > s}} \frac{s}{g} \\ &\ll s^{\epsilon} \sum_{\substack{d|q \\ d > s}} d^{-1+\epsilon} \sum_{\substack{g|d \\ g \ll s}} 1. \end{aligned}$$

Now we use Lemma 18, which shows that the number of divisors $g < s$ of q is at most $O(s^{\epsilon})$, and Lemma 19 to bound the above by

$$s^{\epsilon} \sum_{\substack{d|q \\ d > s}} d^{-1+\epsilon} \ll s^{-1+\epsilon}$$

as promised.

Step 4. For each pair of c,g with $cg < s$, we first treat the inner sum over $h \in sI \cap g\mathbf{Z}$. We break it up into sums over $(s|I|/gc)+O(1)$ subintervals $[y, y+cg) \cap g\mathbf{Z}$, plus a leftover term of size at most $c^{1+\epsilon}$. For each subinterval, we use periodicity of $a(h,c)\Delta(h,c)$ under $h \mapsto h+c$ to find

$$\sum_{h \in [y,y+cg) \cap g\mathbf{Z}} a(h,c)\Delta(h,c) = \sum_{h_1=1}^{c} a\big(gh_1,c\big)\Delta\big(gh_1,c\big).$$

Because q is square-free and g divides q/c, we have that g, c are coprime. Therefore, we can change variables $h = gh_1$ to get that this last sum equals

$$\sum_{h \bmod c} a(h,c)\Delta(h,c) = \prod_{p|c} \sum_{h \bmod p} a(h,p)\Delta(h,p).$$

We evaluate the sum $\sum_{h \bmod p} a(h,p)\Delta(h,p)$ by noting that, summing (2.1) over $h \bmod p$, the sum of the left-hand side is simply the number of all pairs of squares modulo p, namely, $(p+1)^2/4$. This gives

$$\sum_{h \bmod p} a(h,p)\Delta(h,p) = p+1.$$

Thus, the inner sum over $h \in sI \cap g\mathbf{Z}$ equals

$$\begin{aligned}\sum_{h \in sI \cap g\mathbf{Z}} a(h,c)\Delta(h,c) &= \left(\frac{s|I|}{gc} + O(1)\right)\prod_{p|c}(p+1) + O\left(c^{1+\epsilon}\right) \\ &= \frac{s|I|}{g}\sigma_{-1}(c) + O\left(c^{1+\epsilon}\right).\end{aligned}$$

Step 5. Inserting this into the expression (2.3) for $R_2(I,q)$ gives

$$R_2(I,q) = \frac{1}{2^{\omega(q)}\sigma_{-1}(q)} \sum_{c|q} \frac{1}{c} \sum_{g|(q/c):gc<s} \frac{s|I|}{g}\sigma_{-1}(c) + O\left(s^{-1+\epsilon}\right).$$

Now we extend the sum to all pairs g, c, to find that, up to an error of $O(s^{-1+\epsilon})$, we have

$$\begin{aligned}R_2(I,q) &\sim |I| \frac{1}{\sigma_{-1}(q)^2} \sum_{c|q} \frac{\sigma_{-1}(c)}{c} \sum_{g|(q/c)} \frac{1}{g} \\ &= |I| \frac{1}{\sigma_{-1}(q)^2} \sum_{c|q} \frac{\sigma_{-1}(c)}{c} \sigma_{-1}\left(\frac{q}{c}\right) \\ &= |I| \frac{1}{\sigma_{-1}(q)} \sum_{c|q} \frac{1}{c} = |I|,\end{aligned}$$

which is what we need to prove our theorem. □

In the following sections, we repeat these steps with full details for the higher correlation functions, where several technical complications arise.

3. Reduction to odd q. We first show that in Theorem 1 it suffices to consider only the case of q odd. Suppose that $q = 2q'$ with q' odd and square-free. We recall

that

$$R_r(\mathscr{C}, q) = \frac{1}{N_q} \sum_{h \in s\mathscr{C} \cap \mathbf{Z}^{r-1}} N(h, q), \tag{3.1}$$

where $N(h, q)$ is the number of solutions of the system $y_{i+1} - y_i = h_i$ where $y_1, y_2, \ldots, y_r$ are squares modulo q and $h = (h_1, \ldots, h_{r-1}) \in (\mathbf{Z}/q\mathbf{Z})^{r-1}$.

By the Chinese remainder theorem, the number N_q of squares modulo q is the product

$$N_q = N_2 N_{q'} = 2N_{q'}.$$

Therefore, the mean spacing $s_q := q/N_q$ is given by

$$s_q = \frac{2q'}{2N_{q'}} = \frac{q'}{N_{q'}} = s_{q'}. \tag{3.2}$$

Moreover, again by the Chinese remainder theorem,

$$N(h, q) = N(h, 2)N(h, q'),$$

and since all residues modulo 2 are squares, we have $N(h, 2) = 2$. Thus, we find

$$\frac{N(h, q)}{N_q} = \frac{2N(h, q')}{2N_{q'}} = \frac{N(h, q')}{N_{q'}}. \tag{3.3}$$

Inserting (3.2), (3.3) into (3.1), we find that

$$R_r(\mathscr{C}, q) = R_r(\mathscr{C}, q').$$

This shows that it suffices to prove Theorem 1 for q odd, which we assume is the case in the sequel.

4. The prime case. Let $p > 2$ be a prime. For $h = (h_1, \ldots, h_{r-1}) \in (\mathbf{Z}/p\mathbf{Z})^{r-1}$, we define $N_r(h, p)$ to be the number of solutions in squares $y_i \bmod p$ (including $y_i = 0$) of the system

$$y_i - y_{i+1} = h_i \bmod p, \quad 1 \le i \le r-1. \tag{4.1}$$

This number depends crucially on the number of distinct y_j. For each $h = (h_1, \ldots, h_{r-1})$, we define $r_{\text{eff}}(h)$ to be the number of distinct y_j (not necessarily squares) satisfying the system (4.1). Since the solutions of the homogeneous system $y_i - y_{i+1} = 0 \bmod p$ are spanned by $(1, \ldots, 1)$, $r_{\text{eff}}(h)$ is well defined (independent of the particular solution y of (4.1)).

We define *roots* $\sigma_{ij}(h)$, $1 \le i < j \le r$, by

$$\sigma_{ij}(h) = \sum_{k=i}^{j-1} h_k \tag{4.2}$$

so that $\sigma_{i,i+1}(h) = h_i$, $\sigma_{ij} = \sum_{k=i}^{j-1} \sigma_{k,k+1}$. The solutions of (4.1) are all distinct if and only if $\sigma_{ij}(h) \neq 0$, for all $i < j$, since

$$y_i - y_j = \sum_{k=i}^{j-1} y_k - y_{k+1} = \sum_{k=i}^{j-1} h_k = \sigma_{ij}(h).$$

PROPOSITION 4. *Let $r_{\text{eff}}(h)$ be the number of distinct y_i in a solution of (4.1). Then*

$$N_r(h, p) = \frac{p + a(h, p)}{2^{r_{\text{eff}}}} \tag{4.3}$$

with $a(h, p) \ll_r p^{1/2}$.

Proof. The case $r_{\text{eff}}(h) = 1$ happens precisely when $h = 0$ and all y_i are equal: $y_1 = y_2 = \cdots = y_r$. In this case, the number of solutions is the number of squares modulo p, namely, $(p+1)/2$, which is of the desired form. We thus assume from now on that $r_{\text{eff}}(h) > 1$.

We first reduce the system (4.1) to a system of $r_{\text{eff}} - 1$ equations in r_{eff} variables. If $r_{\text{eff}}(h)$ is the number of distinct y_i in a solution of (4.1) (independent of y!), then we can eliminate some of the equations. Renumber the variables so that $y_1, \ldots, y_{r_{\text{eff}}}$ are the distinct coordinates of a solution, and for all $j \geq 1$, $y_{r_{\text{eff}}+j}$ equals one of these; then the system (4.1) is equivalent to the reduced system

$$y_i - y_{i+1} = h_i' \bmod p, \quad 1 \leq i \leq r_{\text{eff}} - 1 \tag{4.4}$$

(where the h_i' are renumbered h_j to give that the first r_{eff} coordinates are distinct). So we need to find the number of solutions of the reduced system (4.4).

We first eliminate those solutions where at least one of the y_j is zero. In this case, since the system (4.4) (considered as a linear system) has rank $r_{\text{eff}} - 1$ in r_{eff} variables, specifying any one of the variables determines all the others; hence the number of solutions with some coordinate zero is at most r_{eff}. Thus, we need only count solutions where all coordinates y_i are nonzero.

To every such solution in squares $y_i \neq 0 \bmod p$, write $y_i = x_i^2 \bmod p$ with $x_i \neq 0 \bmod p$. There are precisely two such solutions, namely, $\pm x_i \bmod p$. Thus, the number of possible x_i corresponding to a given solution y of (4.4) is precisely $2^{r_{\text{eff}}}$, and the number of nonzero solutions of the reduced system (4.4) with y_i squares modulo p is exactly $1/2^{r_{\text{eff}}}$ times the number of solutions of the system

$$x_i^2 - x_{i+1}^2 = h_i' \bmod p, \quad 1 \leq i \leq r_{\text{eff}} - 1 \tag{4.5}$$

with $x_i \neq 0 \bmod p$. By adding back at most r solutions, we can remove the condition $x_i \neq 0$, and then we find that

$$N_r(h, p) = \frac{1}{2^{r_{\text{eff}}}} n(h', p) + O_r(1), \tag{4.6}$$

where $n(h', p)$ is the number of solutions of

$$x_i^2 - x_{i+1}^2 = h_i' \bmod p, \quad 1 \le i \le r_{\text{eff}}(h) - 1.$$

This is just the number of solutions $(t, x_1, \dots, x_{r_{\text{eff}}})$ of the system

$$x_1^2 = t - b_1, \qquad x_2^2 = t - b_2, \dots, x_{r_{\text{eff}}}^2 = t - b_{r_{\text{eff}}} \tag{4.7}$$

with $b_1 = 0$, $b_2 = h_1'$, $b_3 = h_1' + h_2', \dots, b_{r_{\text{eff}}(h)} = h_1' + h_2' + \cdots + h_{r_{\text{eff}}-1}'$ and, in general, $b_k = \sigma_{1k}(h')$. Note that the b_i are distinct; this is equivalent to the requirement that the solutions of the reduced system (4.4) be distinct. One can now use the Riemann hypothesis for curves (see [21], and [18, Theorem 5A and Corollary 5B] for the case $b_1 = -1, b_2 = -2, \dots, b_r = -r$), to find

$$|n(h', p) - p| \ll r_{\text{eff}} 2^{r_{\text{eff}}} \sqrt{p}. \tag{4.8}$$

In addition, $|N(h, p) - n(h', p)/2^{r_{\text{eff}}}| \le r$, and so

$$N(h, p) = \frac{p + a(h, p)}{2^{r_{\text{eff}}}}$$

with

$$a(h, p) \ll 2^{r_{\text{eff}}}(r_{\text{eff}}\sqrt{p} + r) \ll_r \sqrt{p}.$$

This proves Proposition 4. □

A formula for $r_{\text{eff}}(h)$. Our next order of business is to give a formula for $r_{\text{eff}}(h)$. We begin with some combinatorial background. A *set partition* of the set $\{1, 2, \dots, r\}$ is a collection of disjoint subsets $\underline{F} = [F_1, \dots, F_t]$, $F_i \subseteq \{1, 2, \dots, r\}$, whose union is all of $\{1, 2, \dots, r\}$. We set $|\underline{F}| = t$, the number of subsets in $\underline{F}$.

To each set partition $\underline{F}$, we associate a subset $V_{\underline{F}}$ of affine r-space $V = \mathbf{A}^r$ by setting

$$V_{\underline{F}} = \{s \in \mathbf{A}^r : s_i = s_j \text{ if } i, j \text{ are in some } F_k\}. \tag{4.9}$$

Correspondingly, in $H = \mathbf{A}^{r-1}$, we have a subspace

$$H_{\underline{F}} = \{h \in \mathbf{A}^{r-1} : \sigma_{ij}(h) = 0 \text{ if } i, j \text{ are in some } F_k\}. \tag{4.10}$$

Under the map $\pi : V \to H$ taking $s = (s_i) \mapsto (s_i - s_{i+1})$, we have $V_{\underline{F}} = \pi^{-1} H_{\underline{F}}$.

There is a partial ordering on the collection of all set partitions of $\{1, \dots, r\}$ with $\underline{F} \preceq \underline{G}$ if and only if every F_i is contained in some G_j. For example, $\underline{O} = [\{1, 2, \dots, r\}]$ is the maximal element of this partial ordering, with $|\underline{O}| = 1$ and $H_{\underline{O}} = (0)$. The minimal element is $\underline{r} = [\{1\}, \{2\}, \dots, \{r\}]$ with $|\underline{r}| = r$ and $H_{\underline{r}} = \mathbf{A}^{r-1}$.

The partial ordering on set partitions is inclusion-reversing on subspaces: $\underline{F} \preceq \underline{G} \Leftrightarrow V_{\underline{F}} \supseteq V_{\underline{G}} \Leftrightarrow H_{\underline{F}} \supseteq H_{\underline{G}}$. The regular part of $V_{\underline{F}}$ is

$$V_{\underline{F}}^{\times} = \left\{ s \in V_{\underline{F}} : s_i \neq s_j \text{ if } i, j \text{ are not in some } F_k \right\},$$

and we define $H_{\underline{F}}^{\times}$ likewise. Then $H_{\underline{O}}^{\times} = H_{\underline{O}} = (0)$, and every h belongs to a unique $H_{\underline{F}}^{\times}$ for some $\underline{F}$. Thus, we have

$$H = \coprod_{\underline{F}} H_{\underline{F}}^{\times},$$

and likewise,

$$H_{\underline{F}} = \coprod_{\underline{F} \preceq \underline{G}} H_{\underline{G}}^{\times}.$$

We can now give a formula for $r_{\text{eff}}(h)$:

$$r_{\text{eff}}(h) = \dim V_{\underline{F}} = \dim H_{\underline{F}} + 1 = |\underline{F}|, \tag{4.11}$$

where $\underline{F}$ is the unique set partition such that $h \in H_{\underline{F}}^{\times}$. We can write this as follows. Define

$$\delta_{\underline{F}}(h) = \begin{cases} 1 & h \in H_{\underline{F}}, \\ 0 & \text{otherwise}, \end{cases} \qquad \delta_{\underline{F}}^{\times}(h) = \begin{cases} 1 & h \in H_{\underline{F}}^{\times}, \\ 0 & \text{otherwise}. \end{cases} \tag{4.12}$$

Then

$$r_{\text{eff}}(h) = \sum_{\underline{F}} \dim(V_{\underline{F}}) \delta_{\underline{F}}^{\times}(h). \tag{4.13}$$

Similarly,

$$\Delta(h, p) := 2^{r - r_{\text{eff}}(h)} = \sum_{\underline{F}} 2^{\operatorname{codim}(V_{\underline{F}})} \delta_{\underline{F}}^{\times}(h). \tag{4.14}$$

It is convenient to express this in terms of the characteristic function $\delta_{\underline{F}}$ of the subspaces $H_{\underline{F}}$. For this, we use Möbius inversion. Since the collection of all set partitions of $\{1, \ldots, r\}$ is a partially ordered set, it has a Möbius function $\mu(\underline{F}, \underline{G})$ that is the unique function so that, for any functions ψ, ϕ on set partitions satisfying

$$\phi(\underline{F}) = \sum_{\underline{F} \preceq \underline{G}} \psi(\underline{G}), \tag{4.15}$$

we have

$$\psi(\underline{F}) = \sum_{\underline{F} \preceq \underline{G}} \mu(\underline{F}, \underline{G}) \phi(\underline{G}). \tag{4.16}$$

An explicit form of $\mu(\underline{F}, \underline{G})$ can be found in [14, §25]. We do not have any use for it.

In our case, clearly we have $H_{\underline{F}} = \coprod_{\underline{F} \preceq \underline{G}} H^{\times}_{\underline{G}}$ so that

$$\delta_{\underline{F}} = \sum_{\underline{F} \preceq \underline{G}} \delta^{\times}_{\underline{G}}. \tag{4.17}$$

Thus, we have

$$\delta^{\times}_{\underline{F}} = \sum_{\underline{F} \preceq \underline{G}} \mu(\underline{F}, \underline{G}) \delta_{\underline{G}}. \tag{4.18}$$

This gives us the formula for $\Delta(h, p) = 2^{r - r_{\mathrm{eff}}(h)}$. From (4.14) and (4.18), we find

$$\Delta(h, p) = \sum_{\underline{G}} \lambda(\underline{G}) \delta_{\underline{G}}(h) \tag{4.19}$$

with

$$\lambda(\underline{G}) = \sum_{\underline{F} \preceq \underline{G}} \mu(\underline{F}, \underline{G}) 2^{\operatorname{codim} V_{\underline{F}}}. \tag{4.20}$$

For use in Section 7, we need to know the sum of the product of $\Delta(h, p)$ with the error term $a(h, p)$ in (4.3) over all vectors h.

LEMMA 5. *We have the following relation:*

$$\sum_{h \bmod p} a(h, p) \Delta(h, p) = (p+1)^r - p^r \sum_{\underline{G}} \lambda(\underline{G}) p^{-\operatorname{codim} H_{\underline{G}}}.$$

Proof. We have, by definition,

$$N(h, p) = \frac{p + a(h, p)}{2^r} \Delta(h, p)$$

so that

$$a(h, p) \Delta(h, p) = 2^r N(h, p) - p \Delta(h, p).$$

Now sum over all $h \bmod p$. The sum of $N(h, p)$ is just the total number of r-tuples of squares modulo p, namely, $((p+1)/2)^r$. To sum $\Delta(h, p)$ over h, we use (4.19). Since the sum over all h of $\delta_{\underline{G}}(h)$ is just the number of vectors in the subspace $H_{\underline{G}}$, namely, $p^{\dim H_{\underline{G}}} = p^{r-1-\operatorname{codim} H_{\underline{G}}}$, we find

$$\begin{aligned} \sum_{h \bmod p} a(h, p) \Delta(h, p) &= (p+1)^r - p \sum_{h \bmod p} \Delta(h, p) \\ &= (p+1)^r - p^r \sum_{\underline{G}} \lambda(\underline{G}) p^{-\operatorname{codim} H_{\underline{G}}} \end{aligned}$$

as required. □

5. A formula for $R_r(\mathscr{C}, q)$. In order to prove Theorem 1, we give an expression (5.2) for the r-level correlation $R_r(\mathscr{C}, q)$ that involves summing over the intersection of the dilated set $s\mathscr{C}$ with various lattices.

Recall that, for each set partition $\underline{G}$ of $\{1, \dots, r\}$, we associate a subspace $H_{\underline{G}} \subseteq (\mathbf{Z}/p\mathbf{Z})^{r-1}$. Now, given a divisor $d \bmod q$, let $\mathscr{G} = \otimes_{p|d} \underline{G}^{(p)}$ be a tuple of such set partitions, one for each prime p dividing d (recall that q, and hence d, is square-free). Let $L(\mathscr{G}) \subset \mathbf{Z}^{r-1}$ be the preimage of $\prod_{p|d} H_{\underline{G}^{(p)}}$ under the reduction map $\mathbf{Z}^{r-1} \to \prod_{p|d} (\mathbf{Z}/p\mathbf{Z})^{r-1} \simeq (\mathbf{Z}/d\mathbf{Z})^{r-1}$. $L(\mathscr{G})$ is a lattice whose discriminant (that is, the index in $\mathbf{Z}^{r-1}$) is

$$\operatorname{disc}(\mathscr{G}) = \prod_{p|d} p^{\operatorname{codim}(H_{\underline{G}^{(p)}})}.$$

The support $\operatorname{supp}(\mathscr{G})$ of $L(\mathscr{G})$ is the product of all primes p for which $H_{\underline{G}^{(p)}} \neq (\mathbf{Z}/p\mathbf{Z})^{r-1}$:

$$\operatorname{supp}(\mathscr{G}) = \prod_{p : \underline{G}^{(p)} \neq [\{1\}, \dots, \{r-1\}]} p.$$

Since $\operatorname{codim}(H_{\underline{G}^{(p)}}) \leq r-1$, we get

$$\operatorname{supp}(\mathscr{G}) \mid \operatorname{disc}(\mathscr{G}) \mid \operatorname{supp}(\mathscr{G})^{r-1}.$$

We set

$$\lambda(\mathscr{G}) = \prod_{p|d} \lambda\big(\underline{G}^{(p)}\big),$$

where $\lambda(\underline{G})$ is given by (4.20). For a divisor $c \mid q$, we also set

$$a(h, c) := \prod_{p|c} a(h, p), \qquad \Delta(h, c) := \prod_{p|c} \Delta(h, p).$$

Note that by Proposition 4,

$$a(h, c) \ll c^{1/2+\epsilon}, \qquad \Delta(h, c) \ll c^{\epsilon} \tag{5.1}$$

for all $\epsilon > 0$.

Our formula for $R_r(\mathscr{C}, q)$ is the following proposition.

PROPOSITION 6. *The r-level correlation function is given by*

$$R_r(\mathscr{C}, q) = \frac{s}{2^{r\omega(q)}} \sum_{c|q} \frac{1}{c} \sum_{\operatorname{supp}(\mathscr{G}) | (q/c)} \lambda(\mathscr{G}) \sum_{h \in s\mathscr{C} \cap L(\mathscr{G})} a(h, c) \Delta(h, c). \tag{5.2}$$

Proof. We have that

$$R_r(\mathscr{C}, q) = \frac{1}{N} \sum_{h \in s\mathscr{C} \cap \mathbf{Z}^{r-1}} N(h, q).$$

By the Chinese remainder theorem,

$$N(h,q) = \prod_{p|q} N(h,p).$$

We rewrite formula (4.3) in the form

$$N(h,p) = \frac{p + a(h,p)}{2^r} \Delta(h,p),$$

where

$$\Delta(h,p) = 2^{r - r_{\mathrm{eff}}(h)}.$$

Thus, we find

$$N(h,q) = \frac{q\Delta(h,q)}{2^{r\omega(q)}} \sum_{c|q} \frac{a(h,c)}{c} = \frac{q}{2^{r\omega(q)}} \sum_{c|q} \Delta\left(h, \frac{q}{c}\right) \frac{a(h,c)\Delta(h,c)}{c}. \tag{5.3}$$

Inserting (5.3), we get a formula for $R_r(\mathscr{C}, q)$. Recalling that $N = q/s$,

$$R_r(\mathscr{C}, q) = \frac{s}{2^{r\omega(q)}} \sum_{c|q} \frac{1}{c} \sum_{h \in s^{\mathscr{C}}} \Delta\left(h, \frac{q}{c}\right) a(h,c)\Delta(h,c). \tag{5.4}$$

Next we use the expression (4.19) for $\Delta(h,p)$ to write $\Delta(h, q/c) = \prod_{p|(q/c)} \Delta(h,p)$ in the form

$$\Delta\left(h, \frac{q}{c}\right) = \prod_{p|(q/c)} \sum_{\underline{G}^{(p)}} \lambda\left(\underline{G}^{(p)}\right) \delta\left(h, \underline{G}^{(p)}\right) = \sum_{\mathscr{G} = \otimes_{p|(q/c)} \underline{G}^{(p)}} \lambda(\mathscr{G}) \delta(h, \mathscr{G}), \tag{5.5}$$

where the sum is over all tuples of set partitions $\mathscr{G} = \otimes_{p|(q/c)} \underline{G}^{(p)}$, one for each prime dividing q/c, and we put for each such tuple $\mathscr{G}$,

$$\lambda(\mathscr{G}) := \prod_{p|(q/c)} \lambda\left(\underline{G}^{(p)}\right)$$

and

$$\delta(h, \mathscr{G}) := \prod_{p|(q/c)} \delta\left(h, \underline{G}^{(p)}\right) = \begin{cases} 1 & h \in H_{\underline{G}^{(p)}} \bmod p \text{ for all } p \mid \dfrac{q}{c}, \\ 0 & \text{otherwise.} \end{cases}$$

This is the characteristic function of the lattice $L(\mathscr{G})$ whose support $\mathrm{supp}(\mathscr{G})$ divides q/c. Thus, we get the desired expression

$$R_r(\mathscr{C}, q) = \frac{s}{2^{r\omega(q)}} \sum_{c|q} \frac{1}{c} \sum_{\mathrm{supp}(\mathscr{G})|(q/c)} \lambda(\mathscr{G}) \sum_{h \in s^{\mathscr{C}} \cap L(\mathscr{G})} a(h,c)\Delta(h,c). \qquad \square$$

6. Evaluating the r-level correlations. In order to estimate the correlations using Proposition 6, we partition the sum (5.2) into two parts: The first consisting of pairs c and $\mathcal{G}$ such that $c\,\mathrm{disc}(\mathcal{G}) < s$, and the second consisting of the pairs for which $c\,\mathrm{disc}(\mathcal{G}) > s$. We show that the first part gives the main term and the second part is negligible.

6.1. The case $c\,\mathrm{disc}(\mathcal{G}) > s$. We use $a(h,c) \ll c^{1/2+\epsilon}$ (5.1) and $\Delta(h,c) \ll c^{\epsilon}$ to see that this term is bounded by

$$\frac{s}{2^{r\omega(q)}} \sum_{c|q} \frac{1}{c} \sum_{\substack{\mathrm{supp}(\mathcal{G})|q/c \\ c\,\mathrm{disc}(\mathcal{G})>s}} |\lambda(\mathcal{G})| \#\left\{s\mathcal{C} \cap L(\mathcal{G})\right\} c^{1/2+\epsilon}. \tag{6.1}$$

By the Lipschitz principle (Lemma 16),

$$\#\left\{s\mathcal{C} \cap L(\mathcal{G})\right\} \ll \frac{\mathrm{vol}(s\mathcal{C})}{\mathrm{disc}(\mathcal{G})} + s^{r-2},$$

and since $\mathrm{vol}(s\mathcal{C}) = s^{r-1}\,\mathrm{vol}(\mathcal{C})$, we find that

$$\#\left\{s\mathcal{C} \cap L(\mathcal{G})\right\} \ll \frac{s^{r-1}}{\mathrm{disc}(\mathcal{G})} + s^{r-2}. \tag{6.2}$$

Moreover, in order that $s\mathcal{C} \cap L(\mathcal{G}) \neq \emptyset$, we see that we need $\mathrm{supp}(\mathcal{G}) \ll s^{r(r-1)/2}$, since $\mathcal{C}$ does not intersect the walls. This is a consequence of the following observation. Let $\mathcal{C} \subset \mathbf{R}^{r-1}$ be a bounded convex set. Define

$$\mathrm{diam}_1(\mathcal{C}) = \max\left\{\sum_{k=1}^{r-1} |x_k| : x \in \mathcal{C}\right\}.$$

Note that diam_1 scales linearly: $\mathrm{diam}_1(s\mathcal{C}) = s\,\mathrm{diam}_1(\mathcal{C})$ for all $s > 0$.

LEMMA 7. *If $\mathrm{supp}(\mathcal{G}) > \mathrm{diam}_1(s\mathcal{C})^{r(r-1)/2}$, then $s\mathcal{C} \cap L(\mathcal{G})$ is contained in the walls $\{h \in \mathbf{R}^{r-1} : \sigma_{ij}(h) = 0$ for some $i < j\}$.*

Proof. Let $d_{ij}(\mathcal{G})$ be the product of the primes p such that σ_{ij} vanishes on $H_{\underline{G}^{(p)}}$, that is, so that

$$\sigma_{ij}(x) = 0 \bmod p \quad \text{for all } x \in L(\mathcal{G}).$$

Then $d_{ij}(\mathcal{G}) \mid \mathrm{supp}(\mathcal{G})$, and moreover, we claim that

$$\mathrm{disc}(\mathcal{G}) \mid \prod_{i<j} d_{ij}(\mathcal{G}).$$

It is enough to check this one prime at a time and is equivalent to saying that

$$\mathrm{codim}\left(H_{\underline{G}^{(p)}}\right) \leq \#\left\{i < j : \sigma_{ij} = 0 \text{ on } H_{\underline{G}^{(p)}}\right\},$$

which follows since $H_{\underline{G}^{(p)}}$ is given by vanishing of some of the σ_{ij}.

Now note that if $\mathrm{supp}(\mathcal{G}) > d^{r(r-1)/2}$, then for some $i < j$, $d_{ij}(\mathcal{G}) > d$ because $\mathrm{supp}(\mathcal{G}) \le \mathrm{disc}(\mathcal{G}) \le \prod_{i<j} d_{ij}(\mathcal{G})$, and the last product consists of $r(r-1)/2$ factors. If we take $d = \mathrm{diam}_1(s\mathcal{C}) = s\,\mathrm{diam}_1(\mathcal{C})$, then one has $d_{ij}(\mathcal{G}) > \mathrm{diam}_1(s\mathcal{C})$ for some $i < j$. However, $\sigma_{ij}(h) = 0 \bmod d_{ij}(\mathcal{G})$ and so $\sigma_{ij}(h) = m d_{ij}(\mathcal{G})$ for some integer m. If $m = 0$, then h lies in a wall. If $m \neq 0$, then being an integer, $|m| \ge 1$, and so

$$|\sigma_{ij}(h)| \ge d_{ij}(\mathcal{G}) > \mathrm{diam}_1(s\mathcal{C}).$$

Since

$$\sigma_{ij}(h) = \left|\sum_{k=i}^{j-1} h_k\right| \le \sum_{k=i}^{j-1} |h_k| \le \sum_{k=i}^{r-1} |h_k|,$$

we find that

$$\sum_{k=i}^{r-1} |h_k| > \mathrm{diam}_1(s\mathcal{C}).$$

Thus, $h \notin s\mathcal{C}$ by definition of $\mathrm{diam}_1(s\mathcal{C})$. □

By Lemma 7, together with $|\lambda(\mathcal{G})| \ll \mathrm{supp}(\mathcal{G})^{\epsilon}$, (6.1) is bounded by

$$\frac{s}{2^{r\omega(q)}} \sum_{c|q} c^{-1/2+\epsilon} \sum_{\substack{\mathrm{supp}(\mathcal{G})|(q/c) \\ c\,\mathrm{disc}(\mathcal{G})>s \\ \mathrm{supp}(\mathcal{G})\ll s^{r(r-1)/2}}} \mathrm{supp}(\mathcal{G})^{\epsilon} \left(\frac{s^{r-1}}{\mathrm{disc}(\mathcal{G})} + s^{r-2}\right). \tag{6.3}$$

We split the sum into two parts and use $s < 2^{\omega(q)}$ to bound (6.3) by the sum of

$$\frac{1}{s} \sum_{c|q} c^{-1/2+\epsilon} \sum_{\substack{\mathrm{supp}(\mathcal{G})|(q/c) \\ c\,\mathrm{disc}(\mathcal{G})>s}} \mathrm{supp}(\mathcal{G})^{\epsilon} \frac{s}{\mathrm{disc}(\mathcal{G})} \tag{6.4}$$

and

$$\frac{1}{s} \sum_{c|q} c^{-1/2+\epsilon} \sum_{\substack{\mathrm{supp}(\mathcal{G})|(q/c) \\ c\,\mathrm{disc}(\mathcal{G})>s \\ \mathrm{supp}(\mathcal{G})\ll s^{r(r-1)/2}}} \mathrm{supp}(\mathcal{G})^{\epsilon}. \tag{6.5}$$

We begin by noting that the number of $\mathcal{G}$ with $\mathrm{supp}(\mathcal{G}) = g$ is $O(g^{\epsilon})$, that is,

$$\sum_{\mathrm{supp}(\mathcal{G})=g} 1 \ll g^{\epsilon}. \tag{6.6}$$

Since we sum over $\mathrm{supp}(\mathcal{G}) \ll s^{r(r-1)/2}$ in (6.5), we have $\mathrm{supp}(\mathcal{G})^\epsilon \ll s^{\epsilon'}$, and thus (6.5) is bounded by

$$\frac{1}{s}\sum_{c|q} c^{-1/2+\epsilon} \sum_{\substack{g|(q/c)\\ g\ll s^{r(r-1)/2}}} g^\epsilon \ll s^{-1+\epsilon}\sum_{c|q} c^{-1/2+\epsilon} \sum_{\substack{g|(q/c)\\ g\ll s^{r(r-1)/2}}} 1.$$

By Lemma 18, the number of divisors of q/c that are less than $s^{r(r-1)/2}$ is at most s^ϵ, so this term is bounded by

$$s^{-1+\epsilon}\sum_{c|q} c^{-1/2+\epsilon}.$$

Since

$$\sum_{c|q} c^{-1/2+\epsilon} = \prod_{p|q}\left(1+\frac{1}{p^{1/2-\epsilon}}\right) \ll \prod_{p|q}(1+1)^{\epsilon'} \ll s^{\epsilon''},$$

the contribution of (6.5) is at most $O(s^{-1+\epsilon})$.

It now remains to bound (6.4). We first consider the terms for which $c\,\mathrm{supp}(\mathcal{G}) > s$. Now, $\mathrm{disc}(\mathcal{G}) \geq \mathrm{supp}(\mathcal{G})$, so if $c\,\mathrm{supp}(\mathcal{G}) > s$, then certainly $c\,\mathrm{disc}(\mathcal{G}) > s$, and the sum of the corresponding terms in (6.4) is bounded by

$$\begin{aligned}\frac{1}{s}\sum_{c|q} c^{-1/2+\epsilon} &\sum_{\substack{\mathrm{supp}(\mathcal{G})|(q/c)\\ c\,\mathrm{supp}(\mathcal{G})>s}} \mathrm{supp}(\mathcal{G})^\epsilon \frac{s}{\mathrm{supp}(\mathcal{G})}\\ &= \sum_{c|q} c^{-1/2+\epsilon} \sum_{\substack{g|(q/c)\\ cg>s}} \frac{1}{g^{1-\epsilon}} \sum_{\mathrm{supp}(\mathcal{G})=g} 1 \ll \sum_{c|q} c^{-1/2+\epsilon} \sum_{\substack{g|(q/c)\\ cg>s}} \frac{1}{g^{1-\epsilon}},\end{aligned}$$

by (6.6). Changing the variable to $d = cg$, which is a divisor of q satisfying $d > s$, this is bounded by

$$\sum_{\substack{d|q\\ d>s}}\sum_{c|d} \frac{c^{-1/2+\epsilon}}{(d/c)^{1-\epsilon}} = \sum_{\substack{d|q\\ d>s}} \frac{1}{d^{1-\epsilon}} \sum_{c|d} c^{1/2+\epsilon}.$$

Now the sum $\sum_{c|d} c^{1/2+\epsilon}$ is bounded by $\tau(d)d^{1/2+\epsilon} \ll d^{1/2+\epsilon'}$, so the above is bounded by

$$\sum_{\substack{d|q\\ d>s}} d^{-1/2+\epsilon} \ll s^{-1/2+\epsilon}$$

by Lemma 19. This bounds the contribution of c, $\mathcal{G}$ with $c\,\mathrm{supp}(\mathcal{G}) > s$.

If $c\,\mathrm{disc}(\mathcal{G}) > s$ then $s/\mathrm{disc}(\mathcal{G}) \le c$. This, together with (6.6) implies that

$$\frac{1}{s}\sum_{c|q} c^{-1/2+\epsilon} \sum_{\substack{\mathrm{supp}(\mathcal{G})|(q/c)\\ c\,\mathrm{disc}(\mathcal{G})>s\\ c\,\mathrm{supp}(\mathcal{G})<s}} \mathrm{supp}(\mathcal{G})^{\epsilon}\frac{s}{\mathrm{disc}(\mathcal{G})}$$

$$\ll \frac{1}{s}\sum_{c|q} c^{1/2+\epsilon} \sum_{\substack{g|(q/c)\\ cg<s}} g^{\epsilon} \ll s^{-1/2+\epsilon}\sum_{\substack{c|q\\ c<s}}\sum_{\substack{g|q\\ g<s}} 1$$

$$\ll s^{-1/2+\epsilon}\left(\sum_{\substack{c|q\\ c<s}} 1\right)^2 \ll s^{-1/2+\epsilon},$$

since $\sum_{\substack{c|q\\ c<s}} 1 \ll s^{\epsilon}$ by Lemma 19. Consequently, (6.4) is $O(s^{-1/2+\epsilon})$. (Note that we only used $\mathrm{supp}(\mathcal{G}) \ll s^{r(r-1)/2}$ to bound (6.5).)

6.2. The case $c\,\mathrm{disc}(\mathcal{G}) \le s$. Fix $c \ge 1$ and $\mathcal{G}$, and partition the lattice points in $s\mathcal{C} \cap L(\mathcal{G})$ into two subsets as follows. Fix a reduced fundamental cell (see B.1) $P = P(\mathcal{G})$ for the lattice $L = L(\mathcal{G})$. Then cP is a reduced fundamental cell for the dilated lattice cL. We can tile $\mathbf{R}^{r-1}$ by the translates $h_c + cP$, $h_c \in cL$.

Definition. We say that $x \in L \cap s\mathcal{C}$ is *c-interior* if there is some $y \in cL$ so that $x \in y + cP \subseteq s\mathcal{C}$. We say that $x \in L \cap s\mathcal{C}$ is a *c-boundary* point otherwise.

Note that the notion depends on c and on the choice of a fundamental cell P for L. An important fact is that if $\mathrm{dist}(x, \partial(s\mathcal{C})) \gg_r c\,\mathrm{disc}(L)$, then x is c-interior. This follows from Lemma 15 since $\mathrm{diam}(cP) \ll_r c\,\mathrm{disc}(L)$.

LEMMA 8. *Let P be a fundamental cell for the lattice $L \subseteq \mathbf{Z}^{r-1}$, $c \ge 1$ so that $\gcd(c, \mathrm{disc}(L)) = 1$. Then, for $y \in cL$, the intersection $L \cap (y + cP)$ with L of the translate of the dilated cell $y + cP$ consists of a full set of representatives of $\mathbf{Z}^{r-1}/c\mathbf{Z}^{r-1}$.*

Proof. If $P = \{\sum_{j=1}^{r-1} x_j \vec{\ell}_j : 0 \le x_j < 1\}$, then the c^{r-1} lattice points $y + \sum_{j=1}^{r-1} n_j \vec{\ell}_j, n_j = 0, 1, \ldots, c-1$ in $L \cap y + cP$ are clearly inequivalent modulo cL and are the only points of L in this intersection. We show that if $\gcd(c, \mathrm{disc}(L)) = 1$, then they are inequivalent modulo $c\mathbf{Z}^{r-1}$. To see this, it suffices to show that $L \cap c\mathbf{Z}^{r-1} = cL$. By the theorem on elementary divisors, there is a basis $\{\vec{e}_j\}$ of $\mathbf{Z}^{r-1}$ and integers $d_j \ge 1$ so that $\{d_j \vec{e}_j\}$ is a basis of L, and $\mathrm{disc}(L) = \prod_{j=1}^{r-1} d_j$. If $x \in L \cap c\mathbf{Z}^{r-1}$, then $x = \sum_{j=1}^{r-1} m_j d_j \vec{e}_j \in L$, and also $x = c\sum_{j=1}^{r-1} n_j \vec{e}_j \in c\mathbf{Z}^{r-1}$. Comparing coefficients, we find

$$m_j d_j = c n_j, \quad j = 1, \ldots, r-1. \tag{6.7}$$

Now, since $d_j \mid \mathrm{disc}(L)$ and $\gcd(c, \mathrm{disc}(L)) = 1$, we have that $\gcd(c, d_j) = 1$, and so

(6.7) shows that $m_j = 0 \bmod c$ and $x \in cL$. □

LEMMA 9. (a) *The number of points y of cL so that $y + cP \subset s\mathcal{C}$ is*

$$\frac{\mathrm{vol}(s\mathcal{C})}{\mathrm{disc}(cL)} + O\left(\left(\frac{s}{c}\right)^{r-2}\right).$$

(b) *The number of c-boundary points of L is $\ll cs^{r-2}$.*

Proof. (a) If $y = cz \in cL$, then $y + cP \subseteq s\mathcal{C}$ if and only if $z \in L \cap (s/c)\mathcal{C}$ and $z + P \subseteq (s/c)\mathcal{C}$. Thus, we need to count $N := \#\{z \in L \cap (s/c)\mathcal{C} : z + P \subseteq (s/c)\mathcal{C}\}$. An upper bound is obtained by a packing argument: Since the translates $z + P$ are disjoint and contained in $(s/c)\mathcal{C}$, we get

$$N \,\mathrm{vol}(P) \le \mathrm{vol}\left(\frac{s}{c}\mathcal{C}\right),$$

and so

$$N \le \frac{\mathrm{vol}\left((s/c)\mathcal{C}\right)}{\mathrm{disc}(L)} = \frac{s^{r-1}\,\mathrm{vol}(\mathcal{C})}{c^{r-1}\,\mathrm{disc}(L)}. \tag{6.8}$$

For a lower bound, note that if $z \in L \cap (s/c)\mathcal{C}$ satisfies $\mathrm{dist}(z, \partial((s/c)\mathcal{C})) > \mathrm{diam}(P)$, then $z + P \subseteq (s/c)\mathcal{C}$. By the Lipschitz principle (Lemma 16) and Lemma 17, the number $\tilde{N}$ of such points is

$$\tilde{N} = \frac{\mathrm{vol}\left\{x \in (s/c)\mathcal{C} : \mathrm{dist}\left(x, \partial\left((s/c)\mathcal{C}\right)\right) \ge \mathrm{diam}(P)\right\}}{\mathrm{disc}(L)} + O\left(\left(\frac{s}{c}\right)^{r-2}\right).$$

Further,

$$\mathrm{vol}\left\{x \in \frac{s}{c}\mathcal{C} : \mathrm{dist}\left(x, \partial\left(\frac{s}{c}\mathcal{C}\right)\right) \ge \mathrm{diam}(P)\right\} = \mathrm{vol}\left(\frac{s}{c}\mathcal{C}\right) + O\left(\mathrm{diam}(P)\left(\frac{s}{c}\right)^{r-2}\right),$$

and so

$$\begin{aligned}\tilde{N} &= \frac{\mathrm{vol}\left((s/c)\mathcal{C}\right)}{\mathrm{disc}(L)} + O\left(\frac{\mathrm{diam}(P)(s/c)^{r-2}}{\mathrm{disc}(L)} + \left(\frac{s}{c}\right)^{r-2}\right) \\ &= \frac{\mathrm{vol}\left((s/c)\mathcal{C}\right)}{\mathrm{disc}(L)} + O\left(\left(\frac{s}{c}\right)^{r-2}\right)\end{aligned}$$

because $\mathrm{diam}(P) \ll_r \mathrm{disc}(L)$.

Since $N \ge \tilde{N}$, together with the upper bound (6.8), we find

$$N = \frac{\mathrm{vol}\left((s/c)\mathcal{C}\right)}{\mathrm{disc}(L)} + O\left(\left(\frac{s}{c}\right)^{r-2}\right).$$

(b) For the number of c-boundary points, we subtract the number of c-interior points from the total number of points of $L \cap s\mathcal{C}$. The total number of points in $L \cap s\mathcal{C}$ is given by the Lipschitz principle (Lemma 16):

$$L \cap s\mathcal{C} = \frac{\mathrm{vol}(s\mathcal{C})}{\mathrm{disc}(L)} + O\big(s^{r-2}\big). \tag{6.9}$$

To count the number of c-interior points, we can write each uniquely as $y + p$, with y as in part (a) and $p \in L \cap cP$. Now $\#(L \cap cP) = c^{r-1}$ (see Lemma 8), and so, by part (a), the number of c-interior points is

$$Nc^{r-1} = \frac{\mathrm{vol}(s\mathcal{C})}{\mathrm{disc}(L)} + O\big(cs^{r-2}\big). \tag{6.10}$$

Subtracting (6.10) from (6.9) gives us part (b). □

Fix $\mathcal{G}$, $c \geq 1$ with $c\,\mathrm{disc}(\mathcal{G}) \leq s$. Note that since q is square-free and $\mathrm{supp}(\mathcal{G}) \mid (q/c)$, we have $\gcd(c, \mathrm{disc}(\mathcal{G})) = 1$. We now estimate the sum

$$\sum_{h \in L(\mathcal{G}) \cap s\mathcal{C}} a(h,c)\Delta(h,c).$$

We divide this into two sums, Σ_{int} over the c-interior points and Σ_{bd} over the c-boundary points. We use $a(h,c)\Delta(h,c) \ll c^{1/2+\epsilon}$ to bound Σ_{bd} by

$$\#\{c\text{-boundary points}\}c^{1/2+\epsilon} \ll cs^{r-2}c^{1/2+\epsilon} = c^{3/2+\epsilon}s^{r-2}.$$

The contribution of the c-interior points is computed by writing each such h as $h = y + h_0$ with $h_0 \in cP \cap L$ and $y \in cL \cap s\mathcal{C}$. For each y, we get all possible h_0 that run over a full set of representatives of $\mathbf{Z}^{r-1}/c\mathbf{Z}^{r-1}$ since $\gcd(c, \mathrm{disc}(\mathcal{G})) = 1$ (Lemma 8). Denote the number of such y by N; by Lemma 9(a), $N = (\mathrm{vol}((s/c)\mathcal{C})/\mathrm{disc}(L)) + O((s/c)^{r-2})$. Moreover,

$$a\big(y + h_0, c\big)\Delta\big(y + h_0, c\big) = a\big(h_0, c\big)\Delta\big(h_0, c\big),$$

since $y \in cL(\mathcal{G}) \subset c\mathbf{Z}^{r-1}$. Thus,

$$\begin{aligned}\Sigma_{\text{int}} &= N \sum_{h_0 \bmod c} a(h_0,c)\Delta(h_0,c)\\ &= \left(\frac{\operatorname{vol}\left((s/c)\mathscr{C}\right)}{\operatorname{disc}(L)} + O\left(\left(\frac{s}{c}\right)^{r-2}\right)\right) \sum_{h_0 \bmod c} a(h_0,c)\Delta(h_0,c)\\ &= \frac{\operatorname{vol}\left((s/c)\mathscr{C}\right)}{\operatorname{disc}(L)} \sum_{h_0 \bmod c} a(h_0,c)\Delta(h_0,c) + O\left(\left(\frac{s}{c}\right)^{r-2} c^{r-1} c^{1/2+\epsilon}\right)\\ &= \frac{\operatorname{vol}\left((s/c)\mathscr{C}\right)}{\operatorname{disc}(L)} \sum_{h_0 \bmod c} a(h_0,c)\Delta(h_0,c) + O\left(c^{3/2+\epsilon} s^{r-2}\right).\end{aligned}$$

Thus, the total contribution of the pairs with $c\operatorname{disc}(\mathscr{G}) \le s$ is

(6.11)

$$\begin{aligned}&\frac{s}{2^{r\omega(q)}} \sum_{c|q} \frac{1}{c} \sum_{\substack{\operatorname{supp}(\mathscr{G})|(q/c)\\ c\operatorname{disc}(\mathscr{G})\le s}} \lambda(\mathscr{G}) \sum_{h \in s\mathscr{C}\cap L(\mathscr{G})} a(h,c)\Delta(h,c)\\ &\qquad = \frac{s}{2^{r\omega(q)}} \sum_{c|q} \frac{1}{c} \sum_{\substack{\operatorname{supp}(\mathscr{G})|(q/c)\\ c\operatorname{disc}(\mathscr{G})\le s}} \lambda(\mathscr{G}) \frac{\operatorname{vol}(s\mathscr{C})}{c^{r-1}\operatorname{disc}(\mathscr{G})} \sum_{h_0 \bmod c} a(h_0,c)\Delta(h_0,c)\\ &\qquad\quad + O\left(\frac{s}{2^{r\omega(q)}} \sum_{c|q} \frac{1}{c} \sum_{\substack{\operatorname{supp}(\mathscr{G})|(q/c)\\ c\operatorname{disc}(\mathscr{G})\le s}} |\lambda(\mathscr{G})| c^{3/2+\epsilon} s^{r-2}\right).\end{aligned}$$

To estimate the error in (6.11), note that the condition $c\operatorname{disc}(\mathscr{G}) \le s$ implies $c\operatorname{supp}(\mathscr{G}) \le s$ since $\operatorname{supp}(\mathscr{G}) \le \operatorname{disc}(\mathscr{G})$, so for an upper bound, we may replace the summation over pairs satisfying the former condition by the sum over pairs satisfying the latter. Noting that $2^{\omega(q)} \ge s$, this gives

$$\begin{aligned}\frac{s}{2^{r\omega(q)}} \sum_{c|q} \frac{1}{c} \sum_{\substack{\operatorname{supp}(\mathscr{G})|(q/c)\\ c\operatorname{disc}(\mathscr{G})\le s}} |\lambda(\mathscr{G})| c^{3/2+\epsilon} s^{r-2} &\ll s^{-1+\epsilon} \sum_{c|q} c^{1/2+\epsilon} \sum_{\substack{\operatorname{supp}(\mathscr{G})|q/c\\ c\operatorname{supp}(\mathscr{G})\le s}} |\lambda(\mathscr{G})|\\ &\ll s^{-1+\epsilon} \sum_{c|q} c^{1/2+\epsilon} \sum_{\substack{g|(q/c)\\ cg\le s}} \sum_{\operatorname{supp}(\mathscr{G})=g} |\lambda(\mathscr{G})|.\end{aligned}$$

Now $|\lambda(\mathscr{G})| \ll \operatorname{supp}(\mathscr{G})^{\epsilon}$ and the number of $\mathscr{G}$ with $\operatorname{supp}(\mathscr{G}) = g$ is $O(g^{\epsilon})$, which is $O(s^{\epsilon})$ since $g \le cg \le s$, so that the above is bounded by

$$s^{-1+\epsilon} \sum_{c|q} c^{1/2+\epsilon} \sum_{\substack{g|(q/c)\\ cg\le s}} 1.$$

The number of small divisors g of q/c with $g \le s/c \le s$ is at most s^{ϵ}, so the above

is at most

$$s^{-1+\epsilon}\sum_{\substack{c|q\\ c\le s}} c^{1/2+\epsilon} \ll s^{-1+\epsilon}s^{1/2+\epsilon}\#\{c \mid q : c \le s\} \ll s^{-1/2+\epsilon'},$$

which gives that the error term in (6.11) is $O(s^{-1/2+\epsilon})$.

We now extend the sum of the first term in (6.11) to all the pairs c, $\mathcal{G}$, introducing an error which was bounded in Section 6.1 by $O(s^{-1/2+\epsilon})$. (This is the term (6.4) that was bounded without using the condition $\operatorname{supp}(\mathcal{G}) \ll s^{r(r-1)/2}$.)

In summary, we find that the following proposition holds.

PROPOSITION 10. *For $r \ge 2$, we have*

(6.12)

$$R_r(\mathcal{C},q) = \frac{s}{2^{r\omega(q)}}\sum_{c|q}\frac{1}{c}\sum_{\operatorname{supp}(\mathcal{G})|(q/c)}\lambda(\mathcal{G})\frac{\operatorname{vol}(s\mathcal{C})}{c^{r-1}\operatorname{disc}(L)}\sum_{h_0 \bmod c} a(h_0,c)\Delta(h_0,c) + O\left(s^{-1/2+\epsilon}\right).$$

7. The main term. We now treat the main term of (6.12). Define

$$\mathcal{M} = \frac{s}{2^{r\omega(q)}}\sum_{c|q}\frac{1}{c}\sum_{\operatorname{supp}(\mathcal{G})|(q/c)}\lambda(\mathcal{G})\frac{\operatorname{vol}(s\mathcal{C})}{c^{r-1}\operatorname{disc}(\mathcal{G})}\sum_{h_0 \bmod c} a(h_0,c)\Delta(h_0,c).$$

We show that

$$\mathcal{M} = \operatorname{vol}(\mathcal{C}),$$

which, with (6.12), proves Theorem 1.

The sum over $h \bmod c$ is multiplicative:

$$\sum_{h \bmod c} a(h,c)\Delta(h,c) = \prod_{p|c}\sum_{h \bmod p} a(h,p)\Delta(h,p).$$

Furthermore, by Lemma 5,

$$\sum_{h \bmod p} a(h,p)\Delta(h,p) = (p+1)^r - p^r\sum_{\underline{G}^{(p)}}\lambda\left(\underline{G}^{(p)}\right)p^{-\operatorname{codim} H_{\underline{G}^{(p)}}}.$$

Now note that since $p^{\operatorname{codim} H_{\underline{G}^{(p)}}} = \operatorname{disc}(\underline{G}^{(p)})$, we get

$$\mathcal{M} = \frac{s}{2^{r\omega(q)}}\sum_{c|q}\frac{1}{c}\sum_{\operatorname{supp}(\mathcal{G})|(q/c)}\lambda(\mathcal{G})\frac{s^{r-1}\operatorname{vol}(\mathcal{C})}{c^{r-1}\operatorname{disc}(\mathcal{G})}\prod_{p|c}\left((p+1)^r - p^r\sum_{\underline{G}^{(p)}}\frac{\lambda\left(\underline{G}^{(p)}\right)}{\operatorname{disc}\left(\underline{G}^{(p)}\right)}\right)$$

$$= \frac{\operatorname{vol}(\mathcal{C})s^r}{2^{r\omega(q)}}\sum_{c|q}\frac{1}{c^r}\sum_{\operatorname{supp}(\mathcal{G})|(q/c)}\frac{\lambda(\mathcal{G})}{\operatorname{disc}(\mathcal{G})}\prod_{p|c}\left((p+1)^r - p^r\sum_{\underline{G}^{(p)}}\frac{\lambda\left(\underline{G}^{(p)}\right)}{\operatorname{disc}\left(\underline{G}^{(p)}\right)}\right).$$

Furthermore,

$$\sum_{\mathrm{supp}(\mathcal{G})|(q/c)} \frac{\lambda(\mathcal{G})}{\mathrm{disc}(\mathcal{G})} = \prod_{p|(q/c)} \sum_{\underline{G}^{(p)}} \frac{\lambda\left(\underline{G}^{(p)}\right)}{\mathrm{disc}\left(\underline{G}^{(p)}\right)}.$$

Therefore, we find that

$$\begin{aligned}\mathcal{M} &= \mathrm{vol}(\mathcal{C}) \frac{1}{\sigma_{-1}(q)^r} \sum_{c|q} \prod_{p|(q/c)} \sum_{\underline{G}^{(p)}} \frac{\lambda\left(\underline{G}^{(p)}\right)}{\mathrm{disc}\left(\underline{G}^{(p)}\right)} \prod_{p|c} \left(\left(1+\frac{1}{p}\right)^r - \sum_{\underline{G}^{(p)}} \frac{\lambda\left(\underline{G}^{(p)}\right)}{\mathrm{disc}\left(\underline{G}^{(p)}\right)} \right) \\ &= \mathrm{vol}(\mathcal{C}) \frac{1}{\sigma_{-1}(q)^r} \sum_{c|q} A\left(\frac{q}{c}\right) B(c).\end{aligned}$$

Thus, $\mathcal{M}$ is a multiple of the Dirichlet convolution of the multiplicative functions A, B, with $A(1) = B(1) = 1$,

$$A(p) = \sum_{\underline{G}^{(p)}} \frac{\lambda\left(\underline{G}^{(p)}\right)}{\mathrm{disc}\left(\underline{G}^{(p)}\right)},$$

and (since $(1+1/p)^r = \sigma_{-1}(p)^r$)

$$B(p) = \sigma_{-1}(p)^r - A(p). \tag{7.1}$$

Now, by (7.1), we have

$$\begin{aligned}(A * B)(q) &:= \sum_{c|q} A\left(\frac{q}{c}\right) B(c) = \prod_{p|q} \left(A(1)B(p) + A(p)B(1)\right) \\ &= \prod_{p|q} \sigma_{-1}(p)^r = \sigma_{-1}(q)^r.\end{aligned}$$

Finally, this gives the main term of $R_r(\mathcal{C}, q)$:

$$\mathcal{M} = \mathrm{vol}(\mathcal{C}) \frac{1}{\sigma_{-1}(q)^r} (A * B)(q) = \mathrm{vol}(\mathcal{C}) \frac{1}{\sigma_{-1}(q)^r} \sigma_{-1}(q)^r = \mathrm{vol}(\mathcal{C}).$$

Appendices

Appendix A. Recovering the level spacing from the correlations. In this appendix, we explain how to recover the various spacing distributions from the correlation functions. This is well known in the physics literature (e.g., [15]) and is certainly implicit in Hooley's work [8], [9], [10], but we do not know of a good source for it in the mathematical literature. A very detailed treatment of this and more will appear in a forthcoming book by Katz and Sarnak [12].

We begin with $\mathbf{R}/\mathbf{Z}$, which we think of as the circle with unit circumference. We denote by $\{x\}$ the fractional part of x. If $n \le x < n+1$, n integer, then $\{x\} = x - n$. We set

$$((x)) = \begin{cases} \{x\} & 0 \le \{x\} < \dfrac{1}{2}, \\ \{x\} - 1 & \dfrac{1}{2} \le \{x\} < 1. \end{cases}$$

We order the points in $\mathbf{R}/\mathbf{Z}$ counterclockwise and write $x \succ y$ if the points lie in a segment of length $< 1/2$ on $\mathbf{R}/\mathbf{Z}$ and x follows y. The *signed distance* on $\mathbf{R}/\mathbf{Z}$ is given by $((x-y))$; thus, $-1/2 \le ((x-y)) < 1/2$. In terms of the signed distance, $x \succ y$ if and only if $((x-y)) > 0$.

Given a finite set S of N points on $\mathbf{R}/\mathbf{Z}$, and $k \ge 2$, the k-level correlation functions measure clustering properties of the sequence $S \subset \mathbf{R}/\mathbf{Z}$ on a scale of the mean spacing $1/N$. For a k-tuple of points $x = (x_1, \dots, x_k)$ of S, the oriented distance vector is

$$D(x) = \big(((x_1 - x_2)), \dots, ((x_{k-1} - x_k))\big). \tag{A.1}$$

Given a bounded set $\mathscr{C} \subset \mathbf{R}^{k-1}$, we define the k-level correlation as

$$R_k(\mathscr{C}, S) = \frac{1}{N} \# \left\{ x \in S^k : D(x) \in \frac{1}{N}\mathscr{C} \right\}.$$

As an example, let $\Delta^{k-1} \subset \mathbf{R}^{k-1}$ be the standard open simplex

$$\Delta^{k-1} = \left\{ \big(y_1, \dots, y_{k-1}\big) \mid y_i > 0, \sum_{i=1}^{k-1} y_i < 1 \right\},$$

and for $t > 0$, set $\mathscr{C} = t\Delta^{k-1}$. Then if $N > 2t$, $D(x) \in (1/N)\mathscr{C} = (t/N)\Delta^{k-1}$ means that the following are true:

(1) $((x_i - x_{i+1})) > 0$, that is, $x_1 \succ x_2 \succ \cdots \succ x_k$;

(2) the points all lie in an arc of length at most t/N.

As another example, write $k - 1 = i + j$, and for $t_1, t_2 > 0$, set $\mathscr{C} = t_1\Delta^i \times t_2\Delta^j$, which we can write as

$$\mathscr{C} = \big\{ \big(y_1, \dots, y_k\big) : y_m > 0, y_1 + y_2 + \cdots + y_i < t_1, y_{i+1} + \cdots + y_{i+j} < t_2 \big\}.$$

Then $D(x) \in (1/N)\mathscr{C}$ if and only if $x_1 \succ x_2 \succ \cdots \succ x_k$ and $x_1, \dots, x_{i+1}$ lie in an arc of length $< t_1/N$, and $x_{i+1}, \dots, x_{i+j+1} = x_k$ lie in an arc of length $< t_2/N$.

Given any subset $T \subseteq S$ that is contained in a semicircle, the ordering gives us unique initial and final elements of T, and we can write $T = \{x_{\text{init}} = x_1 \prec x_2 \prec \cdots \prec x_{\text{fin}}\}$. We denote by $|T|$ the number of elements of T, and by $\operatorname{diam}(T)$ the distance $\operatorname{dist}(x_{\text{init}}, x_{\text{fin}})$ between the initial and final points of T. If T consists of just the initial and final points, we say that T is a consecutive pair. A *consecutive k-tuple* of S is

a k-tuple of elements $x_1 = x_{\text{init}} \prec \cdots \prec x_k = x_{\text{fin}}$ so that there are no points of S between x_j and x_{j+1}, for $1 \le j < k$.

For $x < 1/2$, let $N_k(x)$ be the number of k-tuples of diameter smaller than x; this is 0 if $k \gg 1$. It is clear from the definitions and the discussion above that we can describe these functions in terms of the correlation function of the simplex $x\Delta^{k-1}$ by

$$R_k\left(x\Delta^{k-1}, S\right) = \frac{1}{N} N_k\left(\frac{x}{N}\right). \tag{A.2}$$

Furthermore, let $g(x)$ be the number of consecutive pairs of diameter less than x, that is, the number of spacings between consecutive elements of S of length less than x. We may express g in terms of an alternating sum of N_k's as follows.

LEMMA 11. *With g and N_k as above, we have for $x < 1/2$,*

$$g(x) = \sum_{k\ge 2} (-1)^k N_k(x).$$

Moreover, for all $n \ge 1$, we have the inequalities

$$\sum_{k=2}^{2n+1} (-1)^k N_k(x) \le g(x) \le \sum_{k=2}^{2n} (-1)^k N_k(x).$$

Before giving the proof, we need the following elementary lemma on sums of binomial coefficients.

LEMMA 12. *Let $m \ge 0$ be an integer. Then $\sum_{i=0}^{m} (-1)^i \binom{m}{i} = 0$ unless $m = 0$, in which case the sum equals* 1. *Moreover,*

$$\sum_{i=0}^{2n+1} (-1)^i \binom{m}{i} \le \sum_{i=0}^{m} (-1)^i \binom{m}{i} \le \sum_{i=0}^{2n} (-1)^i \binom{m}{i}.$$

Proof. The first part is just the binomial expansion of $(1-1)^m$. As for the second part, if $m \ge 1$, use the identity $\binom{m}{i} = \binom{m-1}{i} + \binom{m-1}{i-1}$ to find $\sum_{i=0}^{k} (-1)^i \binom{m}{i} = (-1)^k \binom{m-1}{k}$, from which the claim follows. □

We can now prove Lemma 11.

Proof of Lemma 11. For each pair $T = \{a \succ b\}$ of diameter less than $1/2$, we associate X_T, the set of all i-tuples $x_1 \succ \cdots \succ x_i$ in S such that $(x_1, x_i) = (a, b)$. The set of all tuples of diameter less than x is thus expressed as a *disjoint union* of the X_T's as T ranges over all pairs of diameter less than x. If we let N_i^T be the number of i-tuples in X_T, then $N_i = \sum_T N_i^T$. But $N_i^T = \binom{|T|-2}{i}$, so by Lemma 12, $\sum_{i\ge 2} (-1)^i N_i^T$ is 0 unless T is a consecutive pair, in which case the alternating sum

is 1. Summing over all consecutive pairs, we get that $g(x) = \sum_{k\geq 2}(-1)^k N_k(x)$. Lemma 12 also gives that for $n > 0$,

$$\sum_{i=2}^{2n+1}(-1)^i N_i^T \leq \sum_{i\geq 2}(-1)^i N_i^T \leq \sum_{i=2}^{2n}(-1)^i N_i^T.$$

Summing over all T, we get the second assertion. □

A.1. The joint level spacing. An (i, j)-tuple of diameter (x, y) is an $(i+j)$-tuple $x_1 \succ \cdots \succ x_i \succ x_{i+1} \succ \cdots \succ x_{i+j}$ (all lying in an arc of length $< 1/2$) such that $\operatorname{dist}(x_i, x_1) = x$ and $\operatorname{dist}(x_{i+j}, x_i) = y$.

For $i \geq 2$, $j \geq 1$, and $x + y < 1/2$, we let $N_{i,j}(x, y)$ be the number of (i, j)-tuples of diameter at most (x, y). Let $g(x, y)$ be the number of consecutive triples $x_1 \succ x_2 \succ x_3$ of diameter smaller than (x, y). Analogously to Lemma 11 we have the following lemma.

LEMMA 13. *If we let* $A_k(x, y) = \sum_{i+j=k} N_{i,j}(x, y)$, *then*

$$g(x, y) = \sum_{k\geq 3}(-1)^{k+1} A_k(x, y).$$

Moreover, for $n \geq 0$, *we have the inequalities*

$$\sum_{k=3}^{3+2n+1}(-1)^{k+1} A_k(x, y) \leq g(x, y) \leq \sum_{k=3}^{3+2n}(-1)^{k+1} A_k(x, y).$$

Proof. For each triple $T = \{a \succ b \succ c\}$ of diameter at most (x, y), let X_T be the set of (i, j)-tuples $x_1 \succ \cdots \succ x_i \succ x_{i+1} \succ \cdots \succ x_{i+j}$ such that $(x_1, x_i, x_{i+j}) = (a, b, c)$, and let $N_{i,j}^T$ be the number of (i, j)-tuples in X_T. We may write the set of (i, j)-tuples of diameter smaller than (x, y) as a disjoint union of X_T's, as T ranges over all $(2, 1)$-tuples with diameter at most (x, y). Given T, we may count tuples of type (i, j) in X_T as follows. Let M, N be the number of elements of S between a, b and b, c, respectively (we allow both M and N to be 0). Then $N_{i,j}^T = \binom{M}{j-2}\binom{N}{i-1}$. Moreover, $A_k^T = \sum_{i+j=k} N_{i,j}^T = \binom{M+N}{k-3}$ since there are $\binom{M+N}{k-3}$ ways of choosing $k-3$ objects out of M "blue" and N "red" objects. By Lemma 12, we see that $\sum_{k\geq 3}(-1)^{k+1} A_k^T = \sum_{k\geq 3}(-1)^{k+1}\binom{M+N}{k-3}$ is 0 unless T is a consecutive $(2, 1)$-tuple, in which case it is 1. Now Lemma 12, together with $A_k^T = \binom{M+N}{k-3}$, shows that

$$\sum_{k=3}^{3+2n-1}(-1)^{k+1} A_k^T \leq \sum_{k\geq 3}(-1)^{k+1} A_k^T \leq \sum_{k=3}^{3+2n}(-1)^{k+1} A_k^T.$$

Summing over all triples T of diameter at most (x, y), we are done. □

A.2. Applications to squares $\bmod q$. We let

$$S_q = \left\{ \frac{n}{q} : 0 \le n \le q-1, n \text{ a square modulo } q \right\} \subset \frac{\mathbf{R}}{\mathbf{Z}}$$

be the image in $\mathbf{R}/\mathbf{Z}$ of the set of squares in $\mathbf{Z}/q\mathbf{Z}$. The mean spacing between elements of S_q is $1/N_q$, where N_q is the number of squares modulo q. For $x > 0$, $g_q(x/N_q)$ is the number of consecutive pairs in S_q of diameter at most x/N_q, that is, the number of normalized consecutive spacings of length $< x$. We set

$$\tilde{P}(x) = \lim_{q\to\infty} \frac{1}{N_q} g_q\left(\frac{x}{N_q}\right).$$

This is the limiting proportion of normalized consecutive spacings in S_q of length at most x (this normalization sets the mean spacing to be unity). $\tilde{P}(x)$ is the cumulant of the level spacing distribution $P(s)$ of the introduction. Likewise, we set for $x, y > 0$,

$$\tilde{P}(x,y) = \lim_{q\to\infty} \frac{1}{N_q} g_q\left(\frac{x}{N_q}, \frac{y}{N_q}\right)$$

the cumulant of the joint level spacing distribution.

For a bounded convex set $\mathscr{C} \subset \mathbf{R}^{k-1}$ not intersecting the walls, and $N \gg 1$, $(1/N_q)\mathscr{C}$ will be contained in the cube $(-1/2, 1/2)^{k-1}$. For $x = (n/q) \in S_q^k$, $(0 \le n_i < q$ are squares modulo $q)$ the oriented distance vector $D(x)$ (see (A.1)) lies in $(1/N_q)\mathscr{C}$ if and only if there is an integer vector $h \in (q/N_q)\mathscr{C} \cap \mathbf{Z}^{k-1}$ so that

$$x_i - x_{i+1} = h_i \bmod q, \quad 1 \le i \le k-1.$$

Denoting by $N(h,q)$ the number of solutions of the above system in squares n_i modulo q, we find that the correlation function $R_k(\mathscr{C}, q) := R_k(\mathscr{C}, S_q)$ satisfies

$$R_k(\mathscr{C}, q) = \frac{1}{N_q} \sum_{h \in s\mathscr{C} \cap \mathbf{Z}^{k-1}} N(h,q) \tag{A.3}$$

with $s = q/N_q$.

LEMMA 14. *If* $x, y > 0$, *then*

$$\tilde{P}(x) = 1 - e^{-x}$$

and

$$\tilde{P}(x,y) = \left(1-e^{-x}\right)\left(1-e^{-y}\right).$$

Proof. As noted above (see (A.2)), we can express the functions $N_k(x)$ in terms of the correlation functions associated to the simplex $x\Delta^{k-1}$, whose volume is $(x^{k-1}/(k-1)!)$:

$$R_k\left(x\Delta^{k-1}; q\right) = \frac{1}{N_q} N_k\left(\frac{x}{N_q}\right).$$

From Theorem 1, we know that

$$R_k\left(x\Delta^{k-1};q\right) = x^{k-1}\operatorname{vol}\left(\Delta^{k-1}\right) + O_k\left(s^{-1/2+\epsilon}\right)$$
$$= \frac{x^{k-1}}{(k-1)!} + O_k\left(s^{-1/2+\epsilon}\right).$$

By Lemma 11, we see that for $n > 0$,

$$\sum_{i=1}^{1+2n+1} (-1)^{i+1}\frac{x^i}{i!} \le \liminf_{q\to\infty} \frac{g_q(x/N_q)}{N_q}$$

and

$$\limsup_{q\to\infty} \frac{g_q(x/N_q)}{N_q} \le \sum_{i=1}^{1+2n} (-1)^{i+1}\frac{x^i}{i!}.$$

Letting $n \to \infty$ and noting that the above polynomials are truncations of the Taylor series of $1-e^{-x}$, we are done.

For the second part of the lemma, recall that $N_{i,j}(x,y)$ is the number of ordered $i+j$-tuples of elements of S_q such that the first i are contained in an interval of length x, and the last j elements lie in an interval of length y. Thus, analogously to (A.2), $N_{i,j}(x,y)$ is a scaled version of the $(i+j-1)$-correlation with respect to the convex set $x\Delta^{i-1} \times y\Delta^j$:

$$\frac{N_{i,j}\left((x/N_q),(y/N_q)\right)}{N_q} = R_{i+j}\left(x\Delta^{i-1}\times y\Delta^j;q\right).$$

By Theorem 1,

$$R_{i+j}\left(x\Delta^{i-1}\times y\Delta^j;q\right) = \frac{x^{i-1}y^j}{(i-1)!j!} + O_{i,j}\left(s^{-1/2+\epsilon}\right)$$

since

$$\operatorname{vol}\left(x\Delta^{i-1}\times y\Delta^j\right) = \frac{x^{i-1}y^j}{(i-1)!j!}.$$

Letting $A_k(x,y) = \sum_{i+j=k} N_{i,j}(x,y)$ and using Lemma 12, we get

$$\limsup_{q\to\infty} \frac{1}{N_q} g_q\left(\frac{x}{N_q},\frac{y}{N_q}\right) \le \sum_{k=3}^{3+2n} (-1)^{k+1} \sum_{\substack{i+j=k\\ i>1\\ j>0}} \frac{x^{i-1}y^j}{(i-1)!j!}$$

and

$$\sum_{k=3}^{3+2n} (-1)^{k+1} \sum_{\substack{i+j=k\\ i>1\\ j>0}} \frac{x^{i-1}y^j}{(i-1)!j!} \le \liminf_{q\to\infty} \frac{1}{N_q} g_q\left(\frac{x}{N_q},\frac{y}{N_q}\right).$$

Since the above polynomials are truncations of the Taylor series for $(1-e^{-x})(1-e^{-y})$, we are done. □

Appendix B. Some geometry of numbers

B.1. Given a basis $\vec{\ell}_1, \dots, \vec{\ell}_n$ of a lattice L in $\mathbf{R}^n$, the fundamental cell is the half open set

$$P\big(\{\vec{\ell}_i\}\big) := \big\{x_1\vec{\ell}_1 + \cdots + x_n\vec{\ell}_n : 0 \le x_i < 1\big\}.$$

It serves as a fundamental domain for the action of L on $\mathbf{R}^n$ by translations. The volume of $P(\{\vec{\ell}_i\})$ is the discriminant $\operatorname{disc}(L)$ of the lattice L:

$$\operatorname{vol}\big(P\big(\{\vec{\ell}_i\}\big)\big) = \big|\det\big(\vec{\ell}_1, \dots, \vec{\ell}_n\big)\big| = \operatorname{disc}(L).$$

B.2. We need the following basic fact (due to Mahler and Weyl) from reduction theory. In any dimension $n \ge 1$, there are constants $0 < c_n' < c_n''$ so that any lattice $L \subset \mathbf{R}^n$ has a basis $\vec{\ell}_1, \dots, \vec{\ell}_n$ which is reduced in the sense that

$$c_n' \le \frac{|\vec{\ell}_1| \cdots |\vec{\ell}_n|}{\operatorname{disc}(L)} \le c_n''. \tag{B.1}$$

This is a consequence of Minkowski's second theorem on successive minima (see [1, Lemma V.8] or [20, §6]). This basis is not unique in general.

B.3. We define the diameter $\operatorname{diam}(L)$ of the lattice L to be the minimum of the diameters of all fundamental cells for L.

LEMMA 15. *The diameter of an integer lattice $L \subseteq \mathbf{Z}^n$ is bounded by the discriminant of L:*

$$\operatorname{diam}(L) \ll_n \operatorname{disc}(L), \tag{B.2}$$

the implied constant depending only on the dimension n.

Proof. It suffices to show that if $P(\{\vec{\ell}_i\})$ is the fundamental cell of an integer lattice $L \subseteq \mathbf{Z}^n$ with respect to a reduced basis $\{\vec{\ell}_i\}$, then the diameter of $P(\{\vec{\ell}_i\})$ is bounded by the discriminant of L:

$$\operatorname{diam}\big(P\big(\{\vec{\ell}_i\}\big)\big) \ll_n \operatorname{disc}(L). \tag{B.3}$$

To see this, note that, since $L \subseteq \mathbf{Z}^n$ is an integer lattice, the length of any nonzero vector in L is at least 1; then this implies that a reduced basis has *bounded eccentricity*:

$$1 \le |\vec{\ell}_1| \le |\vec{\ell}_2| \le \cdots \le |\vec{\ell}_n| \le c_n'' \operatorname{disc}(L) \tag{B.4}$$

(assuming we ordered the basis vectors according to their length). Indeed, using (B.1) together with $|\vec{\ell}_i| \ge 1$, we get an upper bound for the longest basis vector $\vec{\ell}_n$,

$$|\vec{\ell}_n| = 1 \cdot |\vec{\ell}_n| \le |\vec{\ell}_1| \cdot |\vec{\ell}_2| \cdot \cdots \cdot |\vec{\ell}_n| \le c_n'' \operatorname{disc}(L).$$

Thus, the diameter of the fundamental cell $P(\{\vec{\ell}_i\})$ is at most

$$\sum_{i=1}^{n} |\vec{\ell}_i| \le n|\vec{\ell}_n| \le c_n'' \operatorname{disc}(L)$$

as required. □

B.4. It is useful to note that for integer dilates cL of a lattice L, $c \ge 1$, the diameter scales linearly: $\operatorname{diam}(cL) = c \operatorname{diam}(L)$, while the discriminant scales with c^n: $\operatorname{disc}(cL) = c^n \operatorname{disc}(L)$. Thus, to bound the diameter of a dilate of an integer lattice, we use

$$\operatorname{diam}(cL) \ll_n c \operatorname{disc}(L). \tag{B.5}$$

B.5. The Lipschitz principle

Definition. A set $\mathscr{C} \subset \mathbf{R}^n$ is of *class m* if the intersection of every line with $\mathscr{C}$ consists of at most m intervals (including the degenerate case when some of the intervals are points) and if the same is true for the projection of $\mathscr{C}$ on every linear subspace.

Thus, for instance, a convex set is of class 1.

We use the following form of the "Lipschitz principle" from the geometry of numbers to estimate the number of lattice points in a region of $\mathbf{R}^n$.

LEMMA 16. *Let $L \subset \mathbf{Z}^n$ be an integer lattice of discriminant* $\operatorname{disc}(L)$*, and let $\mathscr{C} \subset \mathbf{R}^n$ be a set of class m (e.g., a convex set). Suppose that $\mathscr{C}$ lies in a ball of radius R around the origin. Then*

$$\#(L \cap \mathscr{C}) = \frac{\operatorname{vol}(\mathscr{C})}{\operatorname{disc}(L)} + O\big(R^{n-1}\big). \tag{B.6}$$

This follows from the Lipschitz principle for the integer lattice proven by Davenport [5], as adapted by W. Schmidt (see [19, Lemma 1]).

We apply the Lipschitz principle to certain subsets of convex sets. For this purpose, we need the following lemma.

LEMMA 17. *Let $\mathscr{C} \subset \mathbf{R}^n$ be a convex set, let $d > 0$, and define*

$$\mathscr{C}_d := \big\{x \in \mathscr{C} : \operatorname{dist}(x, \partial\mathscr{C}) \ge d\big\}$$

to be the set of points of $\mathscr{C}$ of distance at least d from the boundary $\partial\mathscr{C}$ of $\mathscr{C}$. Then $\mathscr{C}_d$ is convex.

Proof. What we need to show is that for any $x_1, x_2 \in \mathscr{C}_d$ and $\lambda \in [0,1]$, the point $x_3 = x_1 + \lambda(x_2 - x_1)$ also lies in $\mathscr{C}_d$, that is, if $|y| \le d$, then $x_3 + y \in \mathscr{C}$. But $x_3 + y = (x_1 + y) + \lambda((x_2 + y) - (x_1 + y))$, that is, $x_3 + y$ lies on a line between $x_1 + y$

and x_2+y. These two points lie in $\mathcal{C}$ since $x_1, x_2 \in \mathcal{C}_d$. By convexity, so does x_3+y. □

Appendix C. Counting small divisors. In the paper, we need to use some estimates for the number of divisors of q that are smaller than a fixed power of the mean spacing s. As is well known, the number of all divisors of q is $O(q^\epsilon)$ for all $\epsilon > 0$. This is not enough for our purposes, as we need a bound that is $O(s^\epsilon)$. This is provided by the following lemmas.

LEMMA 18. *Let q be square-free, and let $s = 2^{\omega(q)}/\sigma_{-1}(q)$. Fix $\alpha > 0$. Then as $s \to \infty$,*

$$\#\{d \mid q : d < s^\alpha\} = O(s^\epsilon)$$

for all $\epsilon > 0$.

Proof. We start by bounding products of k distinct primes below by k^k; we may assume that the primes are the first k primes. Then by the prime number theorem,

$$\log \prod_{i=1}^{k} p_i = \sum_{i=1}^{k} \log p_i \sim p_k \sim k \log k.$$

Exponentiating, we see that the product is bounded below by k^k. Now,

$$\#\{d \mid q : d < s^\alpha\} = \sum_j a_j,$$

where $a_j = a(j, s^\alpha, q)$ is the number of divisors of q that are smaller than s^α and have precisely j prime factors. But if $j > N$, where N is the smallest integer such that $N^N \geq s^\alpha$, then $a_j = 0$. Moreover, setting $w = \omega(q)$, we see that $a_j \leq \binom{w}{j}$. Hence,

$$\sum_{\substack{d \mid q \\ d < s^\alpha}} 1 \leq \sum_{j \leq N} \binom{w}{j} \leq N \binom{w}{N}.$$

By Stirling's formula, $\binom{w}{N} \ll (w^N/(N/e)^N)$. Thus,

$$\sum_i a_j \leq N \binom{w}{N} \ll N \left(\frac{we}{N}\right)^N \ll N \left(\frac{N \log(N) e}{\alpha N \log(2)}\right)^N$$

since $N^N \geq s^\alpha \gg 2^{w\alpha(1-\epsilon)}$ implies that $w \leq (N \log(N)/\alpha \log(2))$. Thus,

$$\{d \mid q : d < s^\alpha\} \ll N \left(\frac{\log(N) e}{\alpha \log(2)}\right)^N \ll (C \log N)^N,$$

but the last term is clearly $O(s^\epsilon)$. □

LEMMA 19. *If $\alpha > 0$, then* $\sum_{\substack{d|q \\ d>s}} d^{-\alpha} \ll s^{-\alpha+\epsilon}$.

Proof. We divide the sum into two parts: one over $s < d < s^R$ and the other over $d > s^R$ (R is a parameter chosen later). For the first, we use the fact that there are few (namely, $O(s^\epsilon)$) divisors d of q with $d < s^R$ to bound that contribution by

$$\sum_{\substack{d|q \\ s<d<s^R}} d^{-\alpha} \ll \sum_{\substack{d|q \\ s<d<s^R}} s^{-\alpha} \ll s^{-\alpha+\epsilon}.$$

For the summands with $d > s^R$, use $d^{-\alpha} < s^{-R\alpha}$ and $\tau(q) = 2^{\omega(q)} \ll s^{1+\epsilon}$ to get

$$\sum_{\substack{d|q \\ d>s^R}} d^{-\alpha} \ll s^{-R\alpha}\tau(q) \ll s^{1-R\alpha+\epsilon}.$$

Now choose $R > 0$ so that $1 - R\alpha < -\alpha$ to conclude the lemma. □

REFERENCES

[1] J. W. S. CASSELS, *An Introduction to the Geometry of Numbers*, Grundlehren Math. Wiss. **99**, Springer-Verlag, Berlin, 1959.

[2] C. COBELI AND A. ZAHARESCU, *On the distribution of primitive roots* mod p, Acta Arith. **83** (1998), 143–153.

[3] H. DAVENPORT, *On the distribution of quadratic residues (*mod p*)*, J. London Math. Soc. **6** (1931), 49–54; *On the distribution of quadratic residues (*mod p*)*, J. London Math. Soc. **8** (1933), 46–52.

[4] ———, *On character sums in finite fields*, Acta Math. **71** (1939), 99–121.

[5] ———, *On a principle of Lipschitz*, J. London Math. Soc. **26** (1951), 179–183; *Corrigendum*, J. London Math. Soc. **39** (1964), 580.

[6] W. FELLER, *An Introduction to Probability Theory and Its Applications, Vol. 2*, Wiley, New York, 1966.

[7] C. HOOLEY, *On the difference of consecutive numbers prime to n*, Acta Arith. **8** (1962–63), 343–347.

[8] ———, *On the difference between consecutive numbers prime to n, II*, Publ. Math. Debrecen **12** (1965), 39–49.

[9] ———, *On the difference between consecutive numbers prime to n, III*, Math. Z. **90** (1965), 355–364.

[10] ———, "On the intervals between consecutive terms of sequences" in *Analytic Number Theory (St. Louis, 1972)*, Proc. Sympos. Pure Math. **24**, Amer. Math. Soc., Providence, 1973, 129–140.

[11] N. KATZ, *Sommes exponentielles: cours à Orsay, automne, 1979*, Astérisque **79**, Soc. Math. France, Paris, 1979.

[12] N. KATZ AND P. SARNAK, *Random Matrices, Frobenius Eigenvalues, and Monodromy*, Amer. Math. Soc. Colloq. Publ. **45**, Amer. Math. Soc., Providence, 1999.

[13] P. KURLBERG, *The distribution of spacings between quadratic residues, II*, submitted for publication.

[14] J. H. VAN LINT AND R. M. WILSON, *A Course in Combinatorics*, Cambridge Univ. Press, Cambridge, 1992.

[15] M. L. MEHTA, *Random Matrices*, 2d ed., Academic Press, Boston, 1991.
[16] Z. RUDNICK AND P. SARNAK, *The pair correlation function of fractional parts of polynomials*, Comm. Math. Phys. **194** (1998), 61–70.
[17] Z. RUDNICK, P. SARNAK, AND A. ZAHARESCU, in preparation.
[18] W. M. SCHMIDT, *Equations over Finite Fields: An Elementary Approach*, Lecture Notes in Math. **536**, Springer-Verlag, Berlin, 1976.
[19] ———, *Northcott's theorem on heights, II: The quadratic case*, Acta Arith. **70** (1995), 343–375.
[20] C. L. SIEGEL, *Lectures on the Geometry of Numbers*, rewritten by K. Chandrasekharan, Springer-Verlag, Berlin, 1989.
[21] A. WEIL, *Sur les courbes algébriques et les variétés qui s'en déduisent*, Actualités Sci. Indust. **1041**, Hermann, Paris, 1948.

RAYMOND AND BEVERLY SACKLER SCHOOL OF MATHEMATICAL SCIENCES, TEL AVIV UNIVERSITY, TEL AVIV 69978, ISRAEL