Harmonic Analysis as found in Analytic Number Theory

Hugh L. Montgomery

Department of Mathematics
University of Michigan
Ann Arbor, MI 48109–1109 USA
phone: 734–763–3269
`hlm@math.lsa.umich.edu`

ABSTRACT. A wide variety of questions of Harmonic Analysis arise naturally in various contexts of Analytic Number Theory; in what follows we consider a number of examples of this type.

The author is grateful to Dr. Ulrike Vorhauer for advice and assistance at all stages of preparation of this paper.

1. Uniform Distribution

The definition of uniform distribution is fairly intuitive:

DEFINITION 1. *A sequence* $\{u_n\} \in \mathbb{T}$ *is* uniformly distributed *if for any* α, $0 \leqslant \alpha < 1$, *we have*

$$\lim_{N\to\infty} \frac{1}{N} \operatorname{card}\{1 \leqslant n \leqslant N : u_n \in [0,\alpha) \pmod 1\} = \alpha.$$

Let U_N be a measure with unit masses at the points u_n for $1 \leqslant n \leqslant N$. Then the Fourier transform of U_N is the exponential sum

$$\widehat{U}_N(k) = \sum_{n=1}^{N} e(-ku_n)$$

where $e(\theta) = e^{2\pi i\theta}$. (This notation was introduced by I. M. Vinogradov.) H. Weyl [36, 37] introduced an important criterion for uniform distribution in terms of the size of the U_N, namely that the following are equivalent statements concerning a sequence $\{u_n\}$:

Research supported in part by NSF Grant DMS 0070720.

(a) The sequence $\{u_n\}$ is uniformly distributed;

(b) For each integer $k \neq 0$, $\widehat{U}_N(k) = o(N)$ as $N \to \infty$;

(c) If F is properly Riemann-integable on $\mathbb{T}$ then

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} F(u_n) = \int_{\mathbb{T}} F(x)\, dx\,.$$

From the formula for the value of a geometric series it is immediate that

$$\Big| \sum_{n=1}^{N} e(n\theta) \Big| \leqslant \frac{1}{|\sin \pi\theta|} \leqslant \frac{1}{2\|\theta\|}$$

where $\|\theta\|$ is the distance from θ to the nearest integer, $\|\theta\| = \min_{n\in\mathbb{Z}} |\theta - n|$. If θ is irrational and k is a non-zero integer then $k\theta$ is not an integer, and hence by the above with θ replaced by $k\theta$ we see that $\widehat{U}_N(k) = O_k(1)$ when $u_n = n\theta$. In this way we see easily that the sequence $n\theta$ is uniformly distributed if θ is irrational.

The proof of Weyl's Criterion depends on the existence of one-sided approximations to the characteristic function χ_I of an interval by trigonometric polynomials; these approximations should be close in the L^1 norm. Of course the existence of such trigonometric polynomials follows easily from the uniform approximation to continuous functions by trigonometric polynomials, but it is also useful to put this in a quantitative form. Erdős and Turán [9] showed that there exist trigonometric polynomials T_- and T_+ of degree at most K such that $T_-(x) \leqslant \chi_I(x) \leqslant T_+(x)$ for all x and such that $\int_{\mathbb{T}} T_\pm = \alpha + O(1/K)$, and thus that

$$\Big| \operatorname{card}\{1 \leqslant n \leqslant N : u_n \in [0,\alpha) \pmod 1\} - N\alpha \Big| \leqslant C\frac{N}{K} + C \sum_{k=1}^{K} \frac{|\widehat{U}(k)|}{k}\,.$$

In the 1970's Selberg considered how the large sieve could be refined, and in doing so discovered more natural functions that yield very sharp constants (see Selberg [32], pp. 213–226). Indeed, Beurling [5] defined the entire function

$$B(z) = \Big(\frac{\sin \pi z}{\pi}\Big)^2 \Big(\frac{2}{z} + \sum_{n=0}^{\infty} \frac{1}{(z-n)^2} - \sum_{n=1}^{\infty} \frac{1}{(z+n)^2} \Big),$$

and observed that $B(z) = O(e^{2\pi|\Im z|})$, that $B(x) \geqslant \operatorname{sgn}(x)$ for all real x, and that

$$\int_{\mathbb{R}} B(x) - \operatorname{sgn}(x)\, dx = 1\,.$$

Indeed, Beurling showed that among functions with the prior properties, this function uniquely minimizes the above integral.

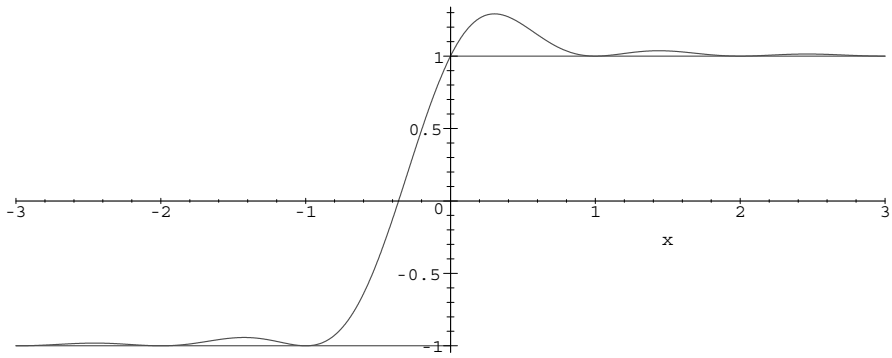

FIGURE 1. Beurling's function $B(x)$.

If $I = [a, b]$ is an interval of the real line and $\delta > 0$ then the Selberg majorant and minorant are

$$S_+(z) = \frac{1}{2}B(\delta(z-a)) + \frac{1}{2}B(\delta(b-z)),$$
$$S_-(z) = -\frac{1}{2}B(\delta(a-z)) - \frac{1}{2}B(\delta(z-b)).$$

Then $S_-(x) \leqslant \chi_I(x) \leqslant S_+(x)$ for all real x, $\widehat{S}(t) = 0$ for $|t| \geqslant \delta$, and $\int_{\mathbb{R}} S_\pm(x)\,dx = b - a \pm 1/\delta$. This approximation is optimal when $\delta(b-a)$ is an integer, and in any case is quite good. To obtain corresponding approximations for the circle group we apply the Poisson summation formula. Suppose that $b - a < 1$, so that the interval $[a, b]$ defines an arc of $\mathbb{T}$. We take $\delta = K + 1$ and set $T_\pm(x) = \sum_n S_\pm(n + x)$. Then $T_\pm$ is a trigonometric polynomial of degree not exceeding K, $T_-(x) \leqslant \chi_I(x) \leqslant T_+(x)$ for all x, and $\int_{\mathbb{T}} T_\pm(x)\,dx = b - a \pm 1/(K+1)$. This is optimal when $(b-a)(K+1)$ is an integer, and is close to optimal in any case.

Weyl's criterion has since been vastly generalized to describe the weak convergence of a sequence of measures μ_n to a limiting measure μ in terms of the convergence of the Fourier transforms of these measures. One fruitful generalization is to $\mathbb{T}^d$. Suppose that $\{\boldsymbol{u}_n\}$ is a sequence of points in $\mathbb{T}^d$ and let $\mathcal{B} = [a_1, b_1] \times \cdots \times [a_d, b_d]$ denote a box in $\mathbb{T}^d$. Then the following are equivalent:

(a) $\lim_{N\to\infty} \frac{1}{N} \operatorname{card}\{1 \leqslant n \leqslant N : \boldsymbol{u}_n \in \mathcal{B}\} = \operatorname{vol} \mathcal{B}$ for every box $\mathcal{B}$ in $\mathbb{T}^d$;

(b) If $\boldsymbol{k} \in \mathbb{Z}^d$, $\boldsymbol{k} \neq \boldsymbol{0}$, then $\sum_{n=1}^N e(\boldsymbol{k} \cdot \boldsymbol{u}_n) = o(N)$ as $N \to \infty$;

(c) If F is properly Riemann-integrable on $\mathbb{T}^d$ then $\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^N F(\boldsymbol{u}_n) = \int_{\mathbb{T}^d} F(\boldsymbol{x})\,d\boldsymbol{x}$.

Quantitative majorants in $\mathbb{T}^d$ are easily obtained by forming a product of one-dimensional majorants. Minorants are a little more elusive, but Barton, Vaaler and Montgomery [2] have given a construction that works pretty well.

In the same way that we used Weyl's Criterion to see that the sequence $\{n\theta\}$ is uniformly distributed if θ is irrational, we can use Weyl's Criterion for $\mathbb{T}^d$ to obtain a sharpening of Kronecker's Theorem. Suppose that $1, \theta_1, \ldots, \theta_d$ are linearly independent over the field $\mathbb{Q}$ of rational numbers. Then the points $n\boldsymbol{\theta} = (n\theta_1, n\theta_2, \ldots, n\theta_d)$ are not only dense in $\mathbb{T}^d$ (Kronecker's Theorem) but are actually uniformly distributed. Indeed, suppose that $\mathcal{B}$ is a box in $\mathbb{T}^d$, and let $T(\boldsymbol{x})$ be a trigonometric polynomial that minorizes $\chi_{\mathcal{B}}$ with $\int_{\mathbb{T}^d} T > 0$. Then

$$\begin{aligned}\operatorname{card}\{M+1 \leqslant n \leqslant M+N : n\boldsymbol{\theta} \in \mathcal{B}\} &\geqslant \sum_{n=M+1}^{M+N} T(n\boldsymbol{\theta}) \\ &= \sum_{\boldsymbol{k}} \widehat{T}(\boldsymbol{k}) \sum_{n=M+1}^{M+N} e(n\boldsymbol{k}\cdot\boldsymbol{\theta}) \\ &\geqslant N\widehat{T}(\mathbf{0}) - \frac{1}{2}\sum_{\boldsymbol{k}\neq\mathbf{0}} \frac{|\widehat{T}(\boldsymbol{k})|}{\|\boldsymbol{k}\cdot\boldsymbol{\theta}\|}.\end{aligned}$$

Here the last sum is finite because there are only finitely many $\boldsymbol{k}$ for which $\widehat{T}(\boldsymbol{k}) \neq 0$ and because $\boldsymbol{k}\cdot\boldsymbol{\theta}$ is never an integer. Since $\widehat{T}(\mathbf{0}) > 0$, the above is positive if N is sufficiently large. The remarkable feature here is that the expression above is independent of M. That is, if $n_1 < n_2 < n_3 < \cdots$ are the n for which $n\boldsymbol{\alpha} \in \mathcal{B}$ then the gaps $n_{i+1} - n_i$ are uniformly bounded above. This insight is critical to Bohr's definition of almost periodicity.

DEFINITION 2. *Let $f : \mathbb{R} \to \mathbb{C}$ be continuous. We say that a real number t is an ε* -almost period of f *if $|f(x+t) - f(x)| < \varepsilon$ for all real x. The function f is* almost periodic *if for every $\varepsilon > 0$ there is a number $C = C(\varepsilon)$ such that any interval of length at least C contains an ε-almost period.*

It is by the strengthened form of Kronecker's Theorem that we see that the *almost periodic polynomial*

$$T(x) = \sum_{n=1}^{N} a_n e(\lambda_n x)$$

is indeed an almost periodic function. It can also be shown that the almost periodic polynomials are dense in the space of almost periodic functions. Let $\zeta(s) = \sum_{n=1}^{\infty} n^{-s}$ be the Riemann zeta function. We write $s = \sigma + it$. If σ is fixed, $\sigma > 1$, then $\zeta(\sigma+it)$ is an almost periodic function of t. The concept of almost periodicity can be generalized to other norms. For example, when σ is fixed, $1/2 < \sigma < 1$ the function $\zeta(\sigma+it)$ is a mean-square almost periodic function, even though it is not an almost periodic function in the uniform norm. One

might expect that by almost periodicity one could show that if the zeta function has a zero with real part $> 1/2$ then it would have many such zeros, all with approximately the same real part. However, attempts to prove such an assertion have so far been unsuccessful. If one could prove such a thing then the Riemann Hypothesis would likely follow, since we have fairly good upper bounds on the number of zeros of the zeta function with large real part. Let $N(\alpha, T)$ denote the number of zeros of $\zeta(s)$ in the rectangle $\alpha \leqslant \sigma \leqslant 1$, $0 \leqslant t \leqslant T$. Then $N(\alpha, T) \ll T^{\phi(\sigma)} \log T$ where $\phi(\sigma) < 1$ if $\sigma > 1/2$. (Following Vinogradov, we say that $f \ll g$ if $f = O(g)$. This notation saves a set of parentheses when there is no main term.) Although we have not succeeded so far to use almost periodicity to produce many zeros from one, we do know that translates of the zeta function are universal among non-zero analytic functions, in the following sense: Let $\mathcal{R}$ be a rectangle, $\mathcal{R} = \{s : \sigma_1 \leqslant \sigma \leqslant \sigma_2, t_1 \leqslant t \leqslant t_2\}$ with $1/2 < \sigma_1 < \sigma_2 < 1$. Let f be analytic and non-zero on a domain containing $\mathcal{R}$. Then for any $\varepsilon > 0$ there exists a real number τ such that $|f(s) - \zeta(s + i\tau)| < \varepsilon$ uniformly for $s \in \mathcal{R}$.

almost periodicity also arises in the error term in the Prime Number Theorem. Let $\Lambda(n)$ be von Mangoldt's lambda function, which is to say that $\Lambda(n) = \log p$ when $n = p^k$ for some positive integer k, and $\Lambda(n) = 0$ otherwise. We put $\psi(x) = \sum_{n \leqslant x} \Lambda(n)$. By integration by parts we see that the Prime Number Theorem in the form $\pi(x) \sim \mathrm{li}(x)$ is equivalent to the assertion that $\psi(x) \sim x$. Since $-\zeta'(s)/\zeta(s) = \sum_{n=1}^{\infty} \Lambda(n) n^{-s}$ for $\sigma > 1$, we can recover $\psi(x)$ from the logarithmic derivative of the zeta function by an inverse Mellin transform:

$$\psi(x) = \frac{-1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{\zeta'}{\zeta}(s) \frac{x^s}{s} \, ds \, .$$

By moving the contour to the left we see that this is

$$= x - \sum_{\rho} \frac{x^\rho}{\rho} - \frac{\zeta'}{\zeta}(0) + \frac{1}{2} \sum_{k=1}^{\infty} \frac{x^{-2k}}{k} \, .$$

Here ρ runs over all the non-trivial zeros of the zeta function, which is to say all those zeros with positive real part. This explicit formula for the error term in the Prime Number Theorem is essentially one that Riemann stated and von Mangoldt later proved. Write $\rho = \beta + i\gamma$. In the quantity $x^\rho = x^\beta x^{i\gamma}$, the second factor oscillates, but more slowly as x increases. To make it periodic we make an exponential change of variables. Suppose also that the Riemann Hypothesis is true, which is to say that $\beta = 1/2$ for all ρ. Then

$$\frac{\psi(e^y) - e^y}{e^{y/2}} = -\sum_{\rho} \frac{e^{i\gamma y}}{\rho} + o(1) \, .$$

This expression is mean-square almost periodic, and the sum on the right is its Fourier expansion. It is generally believed that the $\gamma > 0$ are linearly independent over $\mathbb{Q}$, so that the terms $e^{i\gamma y}$ behave like independent random variables.

2. Exponential Sums

In his seminal work [36, 37], Hermann Weyl not only gave his criterion for uniform distribution, but also a useful method for estimating exponential sums of the form $\sum_{n=1}^{N} e(P(n))$ where P is a polynomial with real coefficients. Indeed, today such an exponential sum is called a *Weyl sum*. Weyl's basic observation was that

$$\begin{aligned}
\Big|\sum_{n=1}^{N} e(P(n))\Big|^2 &= \sum_{n=1}^{N}\sum_{m=1}^{N} e\big(P(m)-P(n)\big)\\
&= \sum_{n=1}^{N}\sum_{h=1-n}^{N-n} e\big(P(n+h)-P(n)\big)\\
&= \sum_{h=-N+1}^{N-1}\ \sum_{\substack{1\leqslant n\leqslant N\\ 1-h\leqslant n\leqslant N-h}} e\big(P(n+h)-P(n)\big)\\
&= N + 2\Re\sum_{h=1}^{N-1}\sum_{n=1}^{N-h} e\big(P(n+h)-P(n)\big).
\end{aligned}$$

This manipulation is known as 'Weyl differencing'. If $P(x)$ has degree d then $P(x+h)-P(x)$ has degree $d-1$ when $h \neq 0$. Hence if we perform this differencing $d-1$ times then we are left with a geometric series, which we know how to estimate. In this way, Weyl showed that if $P(x) = \sum_{i=0}^{d} a_i x^i$ is a polynomial with real coefficients, and if at least one of the numbers $a_1, a_2, \ldots, a_d$ is irrational, then the sequence $\{P(n)\}$ is uniformly distributed modulo 1.

In the Weyl differencing, it is somewhat a disadvantage that the parameter h runs all the way from 1 to $N-1$. Later, van der Corput found that h can be restricted. For $1 \leqslant n \leqslant N$ let y_n be a complex number, and suppose that $y_n = 0$ if $n < 1$ or $n > N$. Then

$$\begin{aligned}
(H+1)^2\Big|\sum_{n} y_n\Big|^2 &= \Big|\sum_{h=0}^{H}\sum_{n} y_{n+h}\Big|^2\\
&= \Big|\sum_{n}\sum_{h=0}^{H} y_{n+h}\Big|^2
\end{aligned}$$

By Cauchy's inequality this is

$$\begin{aligned}
&\leqslant (N+H)\sum_n \Big|\sum_{h=0}^{H} y_{n+h}\Big|^2 \\
&= (N+H)\sum_{h=0}^{H}\sum_{k=0}^{H}\sum_n y_{n+h}\overline{y_{n+k}} \\
&= (N+H)(H+1)\sum_n |y_n|^2 \\
&\qquad + 2(N+H)\Re\sum_{r=1}^{H}(H+1-r)\sum_n y_{n+r}\overline{y_n}\,.
\end{aligned}$$

Thus we obtain van der Corput's inequality, which asserts that

$$\Big|\sum_{n=1}^{N} y_n\Big|^2 \leqslant \frac{N+H}{H+1}\sum_{n=1}^{N}|y_n|^2 + \frac{2(N+H)}{H+1}\sum_{h=1}^{H}\Big(1-\frac{h}{H+1}\Big)\Big|\sum_{n=1}^{N-h} y_{n+h}\overline{y_n}\Big|\,.$$

On taking $y_n = e(ku_n)$ and applying Weyl's Criterion twice, we obtain

THEOREM 1. (van der Corput) *If for each positive integer* h *the sequence* $\{u_{n+h}-u_n\}$ *is uniformly distributed* (mod 1), *then the sequence* $\{u_n\}$ *is uniformly distributed* (mod 1).

More recently it has been recognized that the above remains true even when h is restricted to lie in certain subsets of positive integers; we say that $\mathcal{H}$ is a *van der Corput set* if the above is true when h is restricted to lie in $\mathcal{H}$. This is equivalent to the existence of non-negative cosine polynomials

$$T(x) = a_0 + \sum_{\substack{1\leqslant h\leqslant H\\ h\in\mathcal{H}}} a_h \cos 2\pi hx$$

with $T(0)=1$ and a_0 arbitrarily small. In this context it is no accident that we see the coefficients of the Fejér kernel in van der Corput's inequality. Since the set of perfect squares constitute a van der Corput set, there exist non-negative cosine polynomials of the form

$$T(x) = a_0 + \sum_{h=1}^{H} a_h \cos 2\pi h^2 x$$

with $T(0)=1$ and a_0 arbitrarily small, but it is not known how rapidly a_0 tends to 0 as $H\to\infty$.

van der Corput devised a general method for estimating exponential sums of the form $\sum_{a\leqslant n\leqslant b} e(f(n))$ where f is a sufficiently smooth real valued function. As an example of a simple first result of this type, we mention that if $0 < \lambda_2 \leqslant f''(x) \leqslant A\lambda_2$ for $a \leqslant x \leqslant b$ then

$$\sum_{a\leqslant n\leqslant b} e(f(n)) \ll_A \lambda_2^{1/2}(b-a) + \lambda_2^{-1/2} .$$

In Process A of van der Corput, one exponential sum is made to give rise to another by a suitable application of the van der Corput inequality. This method is destructive in the sense that usually some cancellation is lost. In van der Corput's Process B, one takes $r(x) = e(f(x))$ for $a \leqslant x \leqslant b$, and $r(x) = 0$ otherwise. Then by the Poisson summation formula,

$$\sum_{a\leqslant n\leqslant b} e(f(n)) = \sum_n r(n) = \sum_\nu \widehat{r}(\nu) = \sum_\nu \int_a^b e(f(x) - \nu x)\, dx .$$

If f' is strictly increasing with $f'(a) = \alpha$, $f'(b) = \beta$, then for $\alpha \leqslant \nu \leqslant \beta$ we obtain a stationary phase at x_ν where $f'(x_\nu) = \nu$. If $f''(x) > 0$ for $a \leqslant x \leqslant b$ then the above is approximately

$$\sum_{\alpha\leqslant\nu\leqslant\beta} \frac{e(f(x_\nu) - \nu x_\nu + 1/8)}{\sqrt{f''(x_\nu)}} .$$

Thus the problem of estimating one sum is reduced to that of estimating another. Process B is non-destructive, since a second application of it takes us back to our initial sum. For a certain class of functions f, these two processes lead to estimates of the form

$$\sum_{a\leqslant n\leqslant b} e(f(n)) \ll \big(\max|f'|\big)^k (b-a)^\ell$$

for certain pairs (k,ℓ) in the square $0 \leqslant k \leqslant 1/2 \leqslant \ell \leqslant 1$. If (k,ℓ) is an exponent pair then Process A gives the exponent pair $\big(\frac{k}{2k+2}, \frac{k+\ell+1}{2k+2}\big)$, and Process B gives the exponent pair $(\ell - 1/2, k + 1/2)$. It is trivial that $(0,1)$ is an exponent pair. By Process B it follows that $(1/2, 1/2)$ is an exponent pair, and by Process A this yields the further pair $(1/6, 2/3)$. The collection of exponent pairs that can be obtained in this way is indicated in Figure 2 below.

Recently, Huxley [17], building on work of Bombieri and Iwaniec [6], has slightly enlarged the region of known exponent pairs, but we are still far from proving the conjecture that (k,ℓ) is an exponent pair if $k > 0$ and $\ell > 1/2$. This is a quite deep conjecture, since the special case $f(x) = t\log x$ yields the Lindelöf Hypothesis, which asserts that $\zeta(1/2 + it) \ll t^\varepsilon$ for every $\varepsilon > 0$. Some useful exponent pairs are given in Table 1.

Quantitative estimates can also be derived for the Weyl sum $\sum_{n=1}^N e\big(P(n)\big)$ in terms of the rational approximations to the coefficients of P. For example, by Weyl's method we find

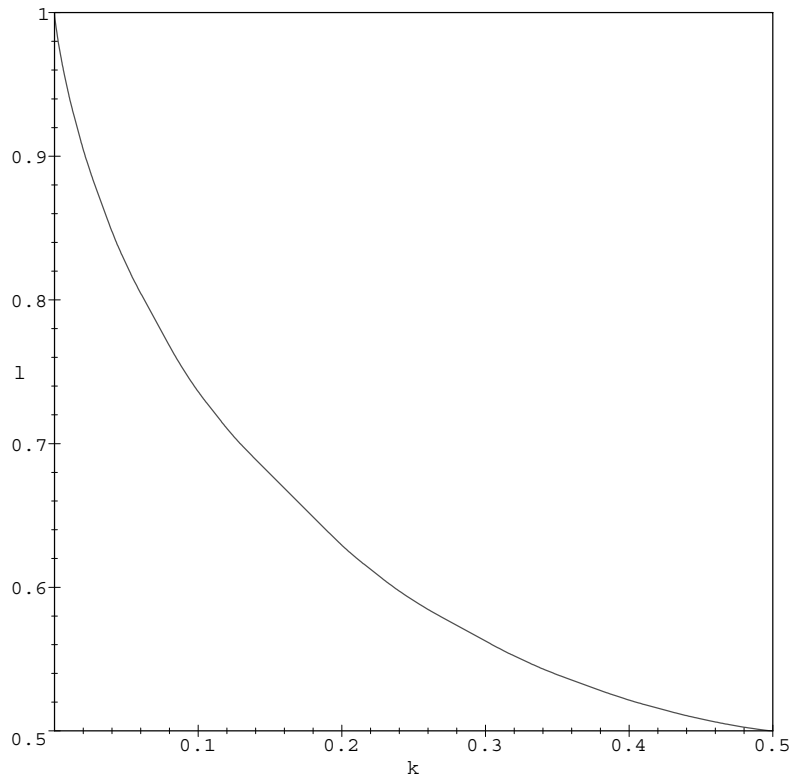

FIGURE 2. Exponent Pairs derived by van der Corput's processes.

that if $P(x) = \sum_{i=0}^{d} \alpha_i x^i$ and $|\alpha_d - a/q| \leqslant 1/q^2$ where $(a,q) = 1$, then

$$\sum_{n=1}^{N} e\big(P(n)\big) \ll_d N^{1+\varepsilon}\Big(\frac{1}{q} + \frac{1}{N} + \frac{q}{N^d}\Big)^{2^{1-d}} .$$

Here the most favorable circumstance is when $N \leqslant q \leqslant N^{d-1}$, in which case the upper bound is of the order $N^{1-2^{1-d}+\varepsilon}$. This is only slightly better than the trivial bound N if d is large, and it falls far short of what we conjecture, which is that

$$\sum_{n=1}^{N} e\big(P(n)\big) \ll_d N^{1+\varepsilon}\Big(\frac{1}{q} + \frac{q}{N^d}\Big)^{1/d} .$$

Here the most favorable situation arises when $q \approx N^{d/2}$, and then the upper bound is of the order $N^{1/2+\varepsilon}$.

This leads to a non-trivial estimate for $|S(\boldsymbol{\alpha})|$.

TABLE 1. Some Exponent Pairs.

(k,ℓ)	**Operation**
$(0,1)$	
$(1/254,247/254)$	$AAAAAAB$
$(1/126,20/21)$	$AAAAAB$
$(1/86,161/172)$	$AAAABAAB$
$(1/62,57/62)$	$AAAAAB$
$(1/50,181/200)$	$AAABABAAB$
$(1/42,25/28)$	$AAABAAB$
$(2/53,181/212)$	$AABABAAAB$
$(1/24,27/32)$	$AABABAAB$
$(1/22,101/121)$	$AABABABAAB$
$(1/20,33/40)$	$AABAAB$
$(13/238,97/119)$	$AABAABAAB$
$(11/186,25/31)$	$AABAAAB$
$(4/49,75/98)$	$ABABAAAB$
$(11/128,97/128)$	$ABABAABAAB$
$(1/11,3/4)$	$ABABAAB$
$(1/10,81/110)$	$ABABABAAB$
$(13/106,75/106)$	$ABAABAAB$
$(11/86,181/258)$	$ABAABABAAB$
$(11/78,161/234)$	$ABAAABAAB$
$(22/117,25/39)$	$BABAAABAAB$
$(26/129,27/43)$	$BABAABABAAB$
$(11/53,33/53)$	$BABAABAAB$
$(13/55,3/5)$	$BABABABAAB$
$(1/4,13/22)$	$BABABAAB$
$(33/128,75/128)$	$BABABAABAAB$
$(13/49,57/98)$	$BABABAAAB$
$(19/62,52/93)$	$BAABAAAB$
$(75/238,66/119)$	$BAABAABAAB$
$(13/40,11/20)$	$BAABAAB$
$(81/242,6/11)$	$BAABABABAAB$
$(11/32,13/24)$	$BAABABAAB$
$(75/212,57/106)$	$BAABABAAAB$
$(11/28,11/21)$	$BAAABAAB$
$(81/200,13/25)$	$BAAABABAAB$
$(13/31,16/31)$	$BAAAAB$
$(75/172,22/43)$	$BAAAABAAB$
$(19/42,32/63)$	$BAAAAAB$
$(60/127,64/127)$	$BAAAAAAB$
$(1/2,1/2)$	B

3. Power Sums

Although we spend a lot of effort to estimate exponential sums, in the opposite direction it is sometimes useful to show that a cancelling sum is not always too cancelling. Turán's method of power sums provides tools of exactly this sort. Since error terms in analytic number theory are often expressed as a sum of oscillatory terms (as we have already seen in the case of the error term in the Prime Number Theorem), Turán's method assists us in proving that such error terms are sometimes large. To exemplify the method, we describe Turán's First Main Theorem. Let

$$s_\nu = \sum_{n=1}^{N} b_n z_n^\nu$$

and suppose that $|z_n| \geqslant 1$ for all n. Suppose that M is a given non-negative integer. We wish to show that there is a ν, $M+1 \leqslant \nu \leqslant M+N$, such that $|s_\nu|$ is not too small compared with $|s_0|$. To this end we employ a simple duality argument, which is typical of Turán's method. Suppose that numbers a_ν have been determined so that

$$s_0 = \sum_{\nu=0}^{N-1} a_\nu s_{M+1+\nu}. \tag{3.1}$$

Then

$$|s_0| \leqslant \Big(\sum_{\nu=0}^{N-1} |a_\nu|\Big) \max_{0\leqslant\nu\leqslant N-1} |s_{M+1+\nu}|\,.$$

By the definition of s_ν we see that (3.1) asserts that

$$\sum_{n=1}^{N} b_n = \sum_{n=1}^{N} b_n z_n^{M+1} \sum_{\nu=0}^{N-1} a_\nu z_n^\nu\,.$$

This identity certainly holds for arbitrary b_n provided that

$$1 = z_n^{M+1} \sum_{\nu=0}^{N-1} a_\nu z_n^\nu$$

for $1 \leqslant n \leqslant N$. That is, $P(z) = \sum_{\nu=0}^{N-1} a_\nu z^\nu$ should be a polynomial of degree at most $N-1$ that satisfies the N conditions $P(z_n) = z_n^{-M-1}$. Without loss of generality the z_n are distinct, and hence $P(z)$ is uniquely determined. It can be shown that $\sum_{\nu=0}^{N-1} |a_\nu| \leqslant \sum_{k=0}^{N-1} \binom{M+k}{k} 2^k$, and thus we find that

$$\max_{M+1\leqslant\nu\leqslant M+N} |s_\nu| \geqslant c(M,N)|s_0|$$

where

$$c(M,N) = \Big(\sum_{k=0}^{N-1} \Big(\frac{M+k}{k}\Big) 2^k\Big)^{-1}.$$

The constant here is best-possible, but it is disappointingly small, since it is only a little larger than

$$\Big(\frac{N}{2e(M+N)}\Big)^{N-1}.$$

Suppose that $T(x)$ is an exponential polynomial of N terms and period 1, say

$$T(x) = \sum_{n=1}^{N} b_n e(\lambda_n x)$$

where the λ_n are integers. Let I be a closed arc of the circle group $\mathbb{T}$, and let L denote the length of I. Then by Turán's First Main Theorem, it is easy to show that

$$\max_{x\in I} |T(x)| \geqslant \Big(\frac{L}{2e}\Big)^{N-1} \max_{x\in\mathbb{T}} |T(x)|.$$

Although the constant here depends on the number N of terms in $T(x)$, it is noteworthy that it is independent of the size of the frequencies λ_n. This inequality makes it possible to give a simple and motivated proof of the Fabry Gap Theorem.

The small constant $c(M,N)$ can be replaced by a larger constant if one is prepared to allow ν to run over a range longer than N. In more restricted situations the lower bound can be very good indeed. For example, suppose that

$$s_\nu = \sum_{n=1}^{N} e(\nu\theta_n). \tag{3.2}$$

Then

$$\sum_{\nu=1}^{K} \Big(1 - \frac{\nu}{K+1}\Big) |s_\nu|^2 = \sum_{m=1}^{N}\sum_{n=1}^{N}\sum_{\nu=1}^{K} \Big(1 - \frac{\nu}{K+1}\Big) e(\nu(\theta_m - \theta_n)).$$

Since the expression is real, we may take real parts to see that the above is

$$= \sum_{m=1}^{N}\sum_{n=1}^{N}\sum_{\nu=1}^{K} \Big(1 - \frac{\nu}{K+1}\Big) \cos 2\pi\nu(\theta_m - \theta_n) = \sum_{m=1}^{N}\sum_{n=1}^{N} \frac{1}{2}(\Delta_{K+1}(\theta_m - \theta_n) - 1)$$

where $\Delta_{K+1}(\theta)$ denotes the Fejér kernel. Since $\Delta_{K+1}(\theta) \geqslant 0$ for all θ, and $\Delta_{K+1}(0) = K+1$, it follows that the above is

$$\geqslant \frac{1}{2}(K+1)N - \frac{1}{2}N^2.$$

Thus if $K \geqslant (1+\varepsilon)N$ then $\max_{1\leqslant\nu\leqslant K}|s_\nu| \geqslant C(\varepsilon)\sqrt{N}$, and in particular

$$\max_{1\leqslant\nu\leqslant 2N}|s_\nu| \geqslant \sqrt{N/2}\,.$$

By working similarly with higher moments one can obtain still better lower bounds over longer ranges of ν: If s_ν is given by (3.2) and $1 \leqslant m \leqslant N/2$ then there is a ν, $1 \leqslant \nu \leqslant (12N/m)^m$ such that $|s_\nu| \geqslant \frac{1}{4}\sqrt{mN}$. It would be useful to have examples to show that this is close to best-possible.

For a more extensive survey of Turán's method one may consult Montgomery [21], pp. 85–107. For a detailed account of the subject and important applications, one should see the book of Turán [33].

4. Irregularities of Distribution

We now consider how well-distributed N points can be. If the points are to fall in $[0,1]$ then we could take $u_n = n/N$. These points are extremely well-distributed in the sense that the number of them in a subinterval $[0,\alpha]$ is $N\alpha + O(1)$. However, we find that it is not so easy to distribute points well in $\mathbb{T}^2$. Suppose that our points are $\boldsymbol{u}_n = (u_1, u_2)$, let $\mathcal{R}(\boldsymbol{\alpha})$ be the rectangle $\mathcal{R}(\boldsymbol{\alpha}) = [0,\alpha_1]\times[0,\alpha_2]$, and let $D(\boldsymbol{\alpha})$ be the discrepancy function

$$D(\boldsymbol{\alpha}) = \operatorname{card}\{1 \leqslant n \leqslant N : \boldsymbol{u}_n \in \mathcal{R}(\boldsymbol{\alpha})\} - N\alpha_1\alpha_2\,.$$

Roth [30] used a construction suggestive of wavelets to show that

$$\int_{\mathbb{T}^2} D(\boldsymbol{\alpha})^2\, d\boldsymbol{\alpha} \gg \log N\,.$$

Since it is also possible to construct points that are this well-distributed, this solves the problem as to distribution in mean-square. Ostrowski [27] had observed that $D(\boldsymbol{\alpha}) \ll \log N$ when $\boldsymbol{u}_n = (n/N, n\sqrt{2})$. The problem of showing that in any case $\|D\|_\infty \gg \log N$ was solved by Schmidt [31] by means of a complicated induction. However, a curious difference arises here. Roth's argument generalizes easily to $\mathbb{T}^k$ to show that $\|D\|_2 \gg_k (\log N)^{(k-1)/2}$ for any $k > 1$. However, Schmidt's approach has not been extended to $k > 2$. It has been conjectured that $\|D\|_\infty \gg_k (\log N)^{k-1}$. This would be best possible, in view of constructions of Halton [14]. On the other hand, Pollington has recently mounted a wavelet approach to this problem that has led him to conclude that one should be able to show that $\|D\|_\infty \gg_k (\log N)^{k/2}$, and that this sup norm need not be larger. Hence one should regard the question of how large $\|D\|_\infty$ need be to be a wide open unsolved problem when $k > 2$. Halász [13] devised a variant of Roth's method that gives Schmidt's Theorem concerning $\|D\|_\infty$ when $k = 2$, and also gives the lower bound $\|D\|_1 \gg \sqrt{\log N}$ when $k = 2$. This is best possible, since Chen [8] has shown that if $0 < p < \infty$ and k is given, $k \geqslant 2$, then there exists a configuration of N points in $\mathbb{T}^k$ for which $\|D\|_p \ll_{p,k} (\log N)^{(k-1)/2}$.

Roth obtained his lower bound by the Cauchy–Schwarz inequality,

$$\int_{\mathbb{T}^k} D(\boldsymbol{\alpha})F(\boldsymbol{\alpha})\,d\boldsymbol{\alpha} \leqslant \left(\int_{\mathbb{T}^k} D(\boldsymbol{\alpha})^2\,d\boldsymbol{\alpha}\right)^{1/2}\left(\int_{\mathbb{T}^k} F(\boldsymbol{\alpha})^2\,d\boldsymbol{\alpha}\right)^{1/2}$$

where F is a test function defined to be a sum, $F = \sum F_r$, of more basic orthogonal functions. Halász similarly used the inequality

$$\int_{\mathbb{T}^2} D(\boldsymbol{\alpha})F(\boldsymbol{\alpha})\,d\boldsymbol{\alpha} \leqslant \|D\|_\infty \|F\|_1$$

where $F = \prod_r (1 + cF_r)$. This bears a strong resemblance to a Riesz product, as occurs in the theory of lacunary trigonometric series. For the lower bound of $\|D\|_1$, he wrote

$$\int_{\mathbb{T}^2} D(\boldsymbol{\alpha})F(\boldsymbol{\alpha})\,d\boldsymbol{\alpha} \leqslant \|D\|_1 \|F\|_\infty$$

where

$$F = \prod_{r=1}^{R}\left(1 + \frac{i}{\sqrt{R}}F_r\right).$$

Since $-1 \leqslant F_r \leqslant 1$, it follows that $|F| \leqslant (1 + 1/R)^R \ll 1$.

Let $e(x_1), e(x_2), \ldots$ be an infinite sequence of unimodular complex numbers, and put $P_N(z) = \prod_{n=1}^{N}(z - e(x_n))$. Erdős asked whether it is possible to choose the x_n in such a way that the numbers $\max_{|z|\leqslant 1}|P_N(z)|$ are bounded as $N \to \infty$. Note that $\Im \log P_N(e(x))$ is just the discrepancy of the first N points, while the problem now being considered involves the harmonic conjugate $\Re \log P_N(e(x))$, but with the important difference that we need this quantity to be large and positive, not just large in absolute value. That this sequence can not remain bounded was first proved by Wagner [35], by means of a modified form of Schmidt's method. Later Beck [3] used Halász's modified form of Roth's method to obtain this in the following sharp quantitative form: There is an absolute constant $\delta > 0$ such that

$$\max_{|z|\leqslant 1}|P_N(z)| > N^\delta$$

for infinitely many N.

Measuring the distribution of points in $\mathbb{T}^k$ relative to rectangles with sides parallel to the coordinate axes is only one of many possibilities. If $\mathcal{S}$ is a measurable set then the quantity

$$D(\mathcal{S}) = \operatorname{card}\{1 \leqslant n \leqslant N : \boldsymbol{u}_n \in \mathcal{S}\} - N \operatorname{vol} \mathcal{S}$$

provides a measure of the distribution of the $\boldsymbol{u}_n$. But when we consider $\mathcal{S}$ we would also include its translates, so we put $d(\boldsymbol{\alpha}) = D(\mathcal{S} + \boldsymbol{\alpha})$. Then $\widehat{d}(\mathbf{0}) = 0$, but for $\boldsymbol{k} \neq \mathbf{0}$ we have $\widehat{d}(\boldsymbol{k}) = \widehat{\chi}_{\mathcal{S}}(-\boldsymbol{k})\widehat{U}_N(\boldsymbol{k})$ where $\widehat{U}_N(\boldsymbol{k}) = \sum_{n=1}^{N} e(-\boldsymbol{k}\cdot\boldsymbol{u}_n)$. Hence by Parseval's identity,

$$\int_{\mathbb{T}^k} d(\boldsymbol{\alpha})^2\,d\boldsymbol{\alpha} = \sum_{\boldsymbol{k}\neq\mathbf{0}} |\widehat{\chi}_{\mathcal{S}}(\boldsymbol{k})|^2 |\widehat{U}_N(\boldsymbol{k})|^2.$$

The rate that $\widehat{\chi}_{\mathcal{S}}(\boldsymbol{k})$ tends to 0 as $\boldsymbol{k}$ tends to infinity in a particular direction depends on how much of the boundary of $\mathcal{S}$ is orthogonal to the direction in question. Thus in the case of a rectangle the Fourier coefficients decay slowly in the direction of the coordinate axes, but comparatively rapidly in other directions. Thus one may expect that points may be found so that $\widehat{U}_N(\boldsymbol{k})$ is small when $\widehat{\chi}_{\mathcal{S}}(\boldsymbol{k})$ is large, and vice versa. By using the Fejér kernel as in our discussion of power sums we can show that

$$\sum_{\substack{|k_1|\leqslant X_1\\ |k_2|\leqslant X_2\\ \boldsymbol{k}\neq\boldsymbol{0}}} |\widehat{U}_N(\boldsymbol{k})|^2 \geqslant N X_1 X_2 - N^2$$

for any positive real numbers X_1, X_2. In this way we can recover Roth's lower bound for $\|D\|_2$. By averaging over disks (and averaging over the radius as well) we find that there is a disk $\mathcal{D}$ for which $D(\mathcal{D}) \gg N^{1/4}$. More generally, if we start with a set $\mathcal{S}$, and are allowed to shrink, translate, and rotate $\mathcal{S}$, then we obtain this larger order of magnitude, in view of a general principle governing the mean square decay of the Fourier transform of the characteristic function of a set: If $\mathcal{C}$ is a simple, closed, piecewise C^1 curve in $\mathbb{R}^2$, and $\mathcal{S}$ is its interior, then

$$\int_{|\boldsymbol{t}|\geqslant R} |\widehat{\chi}_{\mathcal{S}}(\boldsymbol{t})|^2\, d\boldsymbol{t} \sim \frac{|\mathcal{C}|}{2\pi^2 R}$$

as $R\to\infty$. This is due to Montgomery [19], [20] pp. 114–119; see also Herz [15, 16].

5. The Large Sieve

The large sieve was originated by Linnik [18] in a somewhat obscure form. It gained new life in the hands of Rényi [28], who viewed it as a statement about almost independent vectors. Today we usually think of this as an extension of Bessel's inequality for vectors in an inner product space: Let $\phi_1, \ldots, \phi_R$ be arbitrary vectors in an inner product space. Then the following three assertions concerning the constant C are equivalent:

(a) For any vector ξ in the space,

$$\sum_{r=1}^{R} |(\xi, \phi_r)|^2 \leqslant C\|\xi\|^2;$$

(b) For any complex numbers u_r we have

$$\Big|\sum_{1\leqslant r,s\leqslant R} u_r\overline{u_s}(\phi_r, \phi_s)\Big| \leqslant C\sum_{r=1}^{R} |c_r|^2;$$

(c) $C = \rho\big([(\phi_r, \phi_s)]\big)$.

(Here $\rho(A)$ denotes the spectral radius of the matrix A.) Rényi took the coordinates of his vectors to depend on arithmetic progressions or on Dirichlet characters, but the vectors he obtained in doing so were not very close to orthogonal, so the estimates he obtained were imperfect. Roth [29] had the excellent idea of taking $\phi_r = (e(n\alpha_r))$ where the α_r are well-spaced in $\mathbb{T}$. These vectors are quite close to being orthogonal, and we now think of the most basic form of the large sieve as being a statement about the mean square of a trigonometric polynomial at well-spaced points. That is, we take

$$S(\alpha) = \sum_{n=M+1}^{M+N} a_n e(n\alpha),$$

and we consider inequalities of the shape

$$\sum_{r=1}^{R} |S(\alpha_r)|^2 \leqslant \Delta \sum_{n=M+1}^{M+N} |a_n|^2.$$

We suppose that $\|\alpha_r - \alpha_s\| \geqslant \delta$ for $r \neq s$, and want Δ to depend on N and δ. If $a_n = e(-n\alpha_1)$ then $S(\alpha_1) = N$, and thus we must have $\Delta \geqslant N$. By averaging over translations of the α_r we can also show that $\Delta \geqslant \delta^{-1} - 1$. We find that Δ does not have to be much larger than is required by these considerations.

Gallagher [10] used an inequality of the Sobolev type,

$$|f(\alpha)| \leqslant \frac{1}{\delta} \int_{\alpha-\delta/2}^{\alpha+\delta/2} |f(x)|\, dx + \frac{1}{2} \int_{\alpha-\delta/2}^{\alpha+\delta/2} |f'(x)|\, dx,$$

to show that one can take $\Delta = 1/\delta + \pi N$. This is the best constant with respect to δ, but in arithmetic settings the coefficient of N is more important. The main advantage of Gallagher's approach is that it generalizes readily to other families of functions. To obtain good dependence on N we note that by duality the stated inequality is equivalent to the inequality

$$\sum_{n=M+1}^{M+N} \Big| \sum_{r=1}^{R} y_r e(n\alpha_r) \Big|^2 \leqslant \Delta \sum_{r=1}^{R} |y_r|^2.$$

On the left hand side we square out and take the sum over n inside. The diagonal terms give $N \sum_r |y_r|^2$, so it remains to consider the non-diagonal terms. This brings us to Hilbert's Inequality, which asserts that

$$\Big| \sum_{\substack{r,s \\ r \neq s}} \frac{y_r \overline{y_s}}{r-s} \Big| \leqslant \pi \sum_r |y_r|^2.$$

Montgomery and Vaughan [23], with some assistance from Selberg, found that this can be generalized, so that

$$\Big|\sum_{\substack{r,s\\ r\neq s}} \frac{y_r\overline{y_s}}{\lambda_r-\lambda_s}\Big| \leqslant \frac{\pi}{\delta}\sum_r |y_r|^2$$

provided that $|\lambda_r-\lambda_s| \geqslant \delta$ whenever $r\neq s$. Moreover for the circle group we have, correspondingly, the inequality

$$\Big|\sum_{\substack{r,s\\ r\neq s}} \frac{y_r\overline{y_s}}{\sin\pi(\alpha_r-\alpha_s)}\Big| \leqslant \frac{1}{\delta}\sum_r |y_r|^2$$

where $\|\alpha_r-\alpha_s\| \geqslant \delta$ for $r\neq s$. This gives the large sieve with the factor $\Delta = N+1/\delta$. With a little more care one can obtain $\Delta = N+1/\delta-1$, and with this constant there are situations in which equality can occur.

In arithmetic situations the α_r are usually taken to be the Farey fractions a/q, in which $(a,q)=1$ and $q\leqslant Q$. Since $\|a/q-a'/q'\| \geqslant 1/(qq') \geqslant 1/Q^2$ when a/q and a'/q' are distinct modulo 1, we find that

$$\sum_{q=1}^{Q}\sum_{\substack{a=1\\ (a,q)=1}} |S(a/q)|^2 \leqslant (N+Q^2)\sum_{n=M+1}^{M+N}|a_n|^2.$$

The generalized Hilbert Inequality can also be established in a weighted form, in which we find that

$$\Big|\sum_{\substack{r,s\\ r\neq s}} \frac{y_r\overline{y_s}}{\lambda_r-\lambda_s}\Big| \leqslant \frac{3}{2}\pi\sum_r \frac{|y_r|^2}{\delta_r}$$

where $|\lambda_r-\lambda_s| \geqslant \delta_r$ when $s\neq r$. Here the constant $\frac{3}{2}\pi$ is certainly not best possible. The above also has a counterpart for the circle group, and hence we have a weighted form of the large sieve,

$$\sum_{r=1}^{R}\frac{|S(\alpha_r)|^2}{N+\frac{3}{2\delta_r}} \leqslant \sum_{n=M+1}^{M+N}|a_n|^2\,.$$

Many other variants of the large sieve have been derived, involving, for example, maximal partial sums (via the Carleson–Hunt Theorem), or the Hardy–Littlewood Maximal Theorem. As for further generalizations of Hilbert's Inequality, Montgomery and Vaaler [22] have shown that if $\rho_r=\beta_r+i\gamma_r$ with $\beta_r\geqslant 0$ for all r and $|\gamma_r-\gamma_s|\geqslant\delta_r$ for $s\neq r$, then

$$\Big|\sum_{\substack{r,s\\ r\neq s}} \frac{y_r\overline{y_s}}{\rho_r+\overline{\rho_s}}\Big| \leqslant 84\sum_r \frac{|y_r|^2}{\delta_r}\,.$$

The proof depends on the theory of H^2 functions that are analytic in a half-plane.

6. Dirichlet Series

Let $D(s) = \sum_{n=1}^{N} a_n n^{-s}$. By Hilbert's Inequality we see that

$$\int_0^T |D(it)|^2\,dt = (T + O(N)) \sum_{n=1}^{N} |a_n|^2,$$

and the weighted Hilbert Inequality gives

$$\int_0^T |D(it)|^2\,dt = \sum_{n=1}^{N} |a_n|^2 (T + O(n))\,.$$

Thus we have some limitation on the amount of time that $|D|$ is large. Suppose that $|a_n| \leqslant 1$ for all n. Then by applying the above to D^2 we find that

$$\int_0^T |D(it)|^4\,dt \ll (T + N^2) N^{2+\varepsilon}\,.$$

It would be very useful if we could interpolate between these estimates, in the sense that

$$\int_0^T |D(it)|^q\,dt \ll (T + N^{q/2}) N^{q/2+\varepsilon}$$

for real q, $2 \leqslant q \leqslant 4$. Indeed, this implies the Density Hypothesis concerning the Riemann zeta function. Many years ago, Hardy and Littlewood had conjectured that if $|\widehat{f}(k)| \leqslant \widehat{F}(k)$ for all k and $q \geqslant 2$ then $\|f\|_q \ll_q \|F\|_q$, and it can be shown that this majorant conjecture would imply the conjecture above. However, Bachelis [1] used a method of Katznelson to show that this is true only when q is an even integer (in which case it holds with constant 1).

In any case there is more that can be said about the number of times a Dirichlet polynomial can be large than follows from moment estimates. For example, if $0 \leqslant t_1 < t_2 < \ldots < t_R \leqslant T$ and $t_{r+1} - t_r \geqslant 1$ for all r, if $|D(it_r)| \geqslant V$ for all r, and if $|a_n| \leqslant 1$ for all n, then it is known that $R \ll N^2 V^{-2} T^{\varepsilon}$ provided that $V^2 > N T^{1/2+\varepsilon}$. It has been conjectured that this estimate for R holds when the last condition is weakened to read $V^2 \geqslant N T^{\varepsilon}$. However, Bourgain [7] has shown that if this is so then every Kakeya set in $\mathbb{R}^d$, $d \geqslant 2$, has Hausdorff dimension d. Thus there are those that doubt such a strong conjecture.

7. Spectral Characteristics of Zeros of the Zeta Function

Let $h(d)$ denote the class number of the quadratic number field with discriminant d. In an effort to derive a useful lower bound for $h(d)$ when d is negative, it was recognized that it would suffice to have a good supply of pairs of zeros of the Riemann zeta function that are $< c\frac{2\pi}{\log t}$ apart where $c < 1/2$. With this motivation, an attempt was made in 1971 to determine the distribution of $\gamma - \gamma'$ as γ and γ' run over nearby ordinates of zeros of the

zeta function. One begins with a generalization of the explicit formula noted earlier. We observe that

$$\sum_{n\leqslant x}\Lambda(n)n^{-s} = \frac{-1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\frac{\zeta'}{\zeta}(s+w)\frac{x^w}{w}\,dw$$
$$= -\frac{\zeta'}{\zeta}(s) + \frac{x^{1-s}}{1-s} - \sum_{\rho}\frac{x^{\rho-s}}{\rho-s} + \sum_{n=1}^{\infty}\frac{x^{-2n-s}}{2n+s}\,.$$

We assume the Riemann Hypothesis and combine two such explicit formulæ to see that

$$2\sum_{\gamma}\frac{x^{i\gamma}}{1+(t-\gamma)^2} = -x^{-1/2}\Bigl(\sum_{n\leqslant x}\Lambda(n)\Bigl(\frac{x}{n}\Bigr)^{-1/2+it} + \sum_{n>x}\Lambda(n)\Bigl(\frac{x}{n}\Bigr)^{3/2+it}\Bigr)$$
$$+ x^{-1+it}\bigl(\log t + O(1)\bigr) + O(x^{1/2}/t)\,.$$

We take the modulus squared of both sides, and integrate over $0\leqslant t\leqslant T$. We write $x=T^{\alpha}$ for $0\leqslant\alpha\leqslant 1$, and note that the expression on the right hand side can be asymptotically evaluated by using our mean square estimate for Dirichlet polynomials. Let

$$F(\alpha) = \Bigl(\frac{T}{2\pi}\log T\Bigr)^{-1}\sum_{\substack{0<\gamma\leqslant T\\ 0<\gamma'\leqslant T}} T^{i\alpha(\gamma-\gamma')}w(\gamma-\gamma')$$

where $w(u)=4/(4+u^2)$. Then we find that F is real, even, non-negative, and $F(\alpha) = (1+o(1))T^{-2\alpha}\log T + \alpha + o(1)$ as $T\to\infty$, uniformly for $0\leqslant\alpha\leqslant 1$. When $\alpha>1$ the method fails because the Dirichlet polynomial is too long. But only the terms near the diagonal contribute, and one can use the Hardy–Littlewood quantitative form of the Twin Prime Conjecture to estimate that those terms contribute. In this way one is led to guess that $F(\alpha)=1+o(1)$ uniformly for $1\leqslant\alpha\leqslant A$, for any $A>1$. This is known as the Strong Pair Correlation Conjecture. By taking Fourier transforms, we are led to a conjecture concerning the distribution of the frequencies $\gamma-\gamma'$: The number of pairs γ,γ' of ordinates of zeros, $0\leqslant\gamma\leqslant T$, $0\leqslant\gamma'\leqslant T$, for which $2\pi\alpha/\log T\leqslant\gamma-\gamma'\leqslant 2\pi\beta/\log T$ is asymptotic to

$$\Bigl(\delta + \int_{\alpha}^{\beta} 1 - \Bigl(\frac{\sin\pi u}{\pi u}\Bigr)^2\,du\Bigr)\frac{T}{2\pi}\log T\,.$$

Here $\delta=1$ if $0\in[\alpha,\beta]$, and $\delta=0$ otherwise. That is, δ is a Dirac point mass at 0. This arises because of the possibility that $\gamma=\gamma'$ when $\alpha<0<\beta$. This is the Weak Pair Correlation Conjecture. Freeman Dyson observed that the density function here is that of a random hermitian matrix of unitary type, and thus we take the above, although only a conjecture, as evidence that the zeros of the zeta function are spectral in nature.

Goldston and Montgomery [12] showed that if the Riemann Hypothesis is true then the Strong Pair Correlation Conjecture is equivalent to the estimate

$$\int_0^X (\psi(x+h)-\psi(x)-h)^2\,dx \sim hX\log X/h$$

for $x^\varepsilon \leqslant h \leqslant x^{1-\varepsilon}$. A heuristic argument in favor of this can be obtained by expanding out and using the Hardy–Littlewood Conjecture concerning the number of d-twin primes not exceeding x.

Recently Montgomery and Soundararajan [26] considered the higher moments

$$\mu_k(X,h) = \frac{1}{X}\int_0^X (\psi(x+h)-\psi(x)-h)^k\,dx\,.$$

On expanding, one encounters enumerations of prime k-tuples. The Hardy–Littlewood Prime k-tuple Conjecture asserts that if $d_1, d_2, \ldots d_k$ are distinct integers then

$$\sum_{n\leqslant X}\prod_{i=1}^k \Lambda(n+d_i) \;=\; \mathfrak{S}(\mathcal{D})X + E(X,\mathcal{D})$$

where $\mathfrak{S}(\mathcal{D})$ is the 'singular series'

$$\mathfrak{S}(\mathcal{D}) \;=\; \sum_{\substack{1\leqslant q_i<\infty\\(1\leqslant i\leqslant k)}}\prod_{i=1}^k \frac{\mu(q_i)}{\phi(q_i)}\sum_{\substack{1\leqslant a_i\leqslant q_i\\(a_i,q_i)=1\\\sum a_i/q_i\in\mathbb{Z}}} e\Big(\sum_{i=1}^k \frac{a_i d_i}{q_i}\Big)\,.$$

Gallagher [11] considered the moments $\mu_k(X,h)$ for smaller h of the form $h = c\log X$, and for that purpose showed that

$$\sum_{\substack{0\leqslant d_i\leqslant X\\d_i\ \mathrm{distinct}\\(1\leqslant i\leqslant k)}} \mathfrak{S}(\mathcal{D}) \;\sim\; X^k$$

as $X \to \infty$. For larger h, the mean value h is much larger than the usual size of the difference $\psi(x+h)-\psi(x)-h$, and so it is useful to consider the arithmetic function $\Lambda_0(n) = \Lambda(n)-1$, whose mean value is asymptotically zero. For this function we have an alternative formulation of the prime k-tuple Conjecture, which asserts that

$$\sum_{n\leqslant X}\prod_{i=1}^k \Lambda_0(n+d_i) \;=\; \mathfrak{S}_0(\mathcal{D})X + E_0(X,\mathcal{D})$$

where now

$$\mathfrak{S}_0(\mathcal{D}) \;=\; \sum_{\substack{1<q_i<\infty \\ (1\leqslant i\leqslant k)}} \prod_{i=1}^{k} \frac{\mu(q_i)}{\phi(q_i)} \sum_{\substack{1\leqslant a_i\leqslant q_i \\ (a_i,q_i)=1 \\ \sum a_i/q_i\in\mathbb{Z}}} e\Big(\sum_{i=1}^{k} \frac{a_i d_i}{q_i}\Big)\,.$$

Thus $\mathfrak{S}_0$ is the same as $\mathfrak{S}$ except that the possibility that $q_i = 1$ is now excluded. The mean value of $\mathfrak{S}_0$ is of course smaller, and hence more difficult to determine, but by elaborating on work of Montgomery and Vaughan [24] concerning the distribution of reduced residues modulo q in short intervals it can be shown that

$$\sum_{\substack{0\leqslant d_i\leqslant X \\ d_i \text{ distinct} \\ (1\leqslant i\leqslant k)}} \mathfrak{S}_0(\mathcal{D}) \;=\; \begin{cases} \dfrac{k!}{(k/2)!2^{k/2}}\,(-X\log X)^{k/2} + O\big(X^{k/2}(\log X)^{k/4}\big) & \text{if } k \text{ is even,} \\ O\big(X^{k/2-1/(7k)+\varepsilon}\big) & \text{if } k \text{ is odd.} \end{cases}$$

This is established unconditionally; when combined with plausible hypotheses concerning the size and behavior of the error terms $E(X,h)$ we are led to expect that

$$\mu_k(X,h) = \big(c_k + o(1)\big) X h^{k/2} (\log X/h)^{k/2}$$

where the c_k are the moments of the normalized normal variable,

$$c_k \;=\; \begin{cases} \dfrac{k!}{(k/2)!2^{k/2}} & \text{if } k \text{ is even,} \\ 0 & \text{if } k \text{ is odd.} \end{cases}$$

Since these moments occur uniquely in the case of normal distribution, we are led to expect that the distribution of $\psi(x+h) - \psi(x)$, for $0 \leqslant x \leqslant X$, is approximately normal with mean h and variance $h\log X/h$.

Suppose that $X > T$. In terms of zeros of the zeta function, the Strong Pair Correlation Conjecture seems to be telling us that the mean square size of the sum

$$\sum_{0<\gamma\leqslant T} \cos\gamma\log x \qquad (0 \leqslant x \leqslant X)$$

is the same as if the terms were uncorrelated random variables. (We recall from the theory of probability that if X_i are uncorrelated variables then $\mathrm{Var}\,(\sum X_i) = \sum \mathrm{Var}(X_i)$.) It seems that our new speculations concerning the μ_k can similarly be interpreted as asserting that the above sum is distributed as if the terms are independent random variables (as in the Central Limit Theorem). How this relates to the spectral nature of the zeros, or any possible underlying operators remains to be seen.

References

[1] G. F. Bachelis, *On the upper and lower majorant properties in* $L^p(G)$, Quart. J. Math. (Oxford) (2) **24** (1973), 119–128.

[2] J. T. Barton, H. L. Montgomery and J. D. Vaaler, *Note on a Diophantine inequality in several variables*, to appear.

[3] J. Beck, *The modulus of polynomials with zeros on the unit circle: a problem of Erdős*, Ann. of Math. (2) **134** (1991), 609–651.

[4] J. Beck and W. W. L. Chen, *Irregularities of distribution*, Cambridge Tract 89, Cambridge University Press, Cambridge, 1987.

[5] A. Beurling, *Sur les intégrales de Fourier absolument convergentes et leur application à une transformation fonctionelle*, Neuvième congrès des mathématiciens scandinaves, Helsingfors, 1938

[6] E. Bombieri and H. Iwaniec, *On the order of* $\zeta(\frac{1}{2}+it)$, Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4) **13** (1986), 449–472.

[7] J. Bourgain, *On the distribution of Dirichlet sums*, J. Anal. Math. **60** (1993), 21–32.

[8] W. W. L. Chen, *On irregularities of distribution, I*, Mathematika **27** (1980), 153–170.

[9] P. Erdős and P. Turán, *On a problem in the theory of uniform distribution* I, Nederl. Akad. Wetensch. Proc. **51** (1948), 1146–1154; II, 1262–1269, (= Indag. Math. **10**, 370–378; 406–413).

[10] P. X. Gallagher, *The large sieve*, Mathematika **14** (1967), 14–20.

[11] ———, *On the distribution of primes in short intervals*, Mathematika **23** (1976), 4–9.

[12] D. A. Goldston and H. L. Montgomery, *On pair correlations of zeros and primes in short intervals*, Analytic number theory and Diophantine problems (Stillwater, 1984), Birkhäuser Verlag, Boston–Basel–Berlin, 1987, 183–203.

[13] G. Halász, *On Roth's method in the theory of irregularities of point distributions*, Recent Progress in Analytic Number Theory (Durham, 1979), Vol. 2, Academic Press, London, 1981, 79–84.

[14] J. H. Halton, *On the efficiency of certain quasirandom sequences of points in evaluating multidimensional integrals*, Num. Math. **2** (1960), 84–90.

[15] C. S. Herz, *Fourier transforms related to convex sets*, Ann. of Math. (2) **75** (1961), 81–92.

[16] C. S. Herz, *On the number of lattice points in a convex set*, Amer. J. Math. **84** (1962), 126–133.

[17] M. N. Huxley, *Area, Lattice Points and Exponential Sums*, Clarendon Press, Oxford, 1996.

[18] Ju. V. Linnik, *The large sieve*, Dokl. Akad. Nauk SSSR **30** (1941), 292–294.

[19] H. L. Montgomery, *The analytic principle of the large sieve*, Bull. Amer. Math. Soc. **84** (1978), 547–567.

[20] H. L. Montgomery, *Irregularities of distribution*, Congress of Number Theory (Zarautz, 1984), Universidad del Pais Vasco, Bilbao, 1989, 11–27.

[21] H. L. Montgomery, *Ten lectures on the interface between analytic number theory and harmonic analysis*, CBMS No. 84, Amer. Math. Soc., Providence, 1994.

[22] H. L. Montgomery and J. D. Vaaler, *A further generalization of Hilbert's inequality*, Mathematika **45** (1999), 35–39.

[23] H. L. Montgomery and R. C. Vaughan, *Hilbert's inequality*, J. London Math. Soc. (2) **8** (1974), 73–82.

[24] ———, *On the distribution of reduced residues*, Annals of Math. **123** (1986), 311–333.

[25] H. L. Montgomery and K. Soundararajan, *Beyond pair correlation*, to appear.

[26] H. L. Montgomery and K. Soundararajan, *Primes in short intervals*, to appear.

[27] A. Ostrowski, *Bemerkungen zur Theorie der Diophantischen Approximationen* I, Abh. Math. Sem. Hamburg **1** (1922), 77–98; II, **1** (1922), 250–251; III, **41** (1926), 224.

[28] A. Rényi, *Un nouveau théorème concernant les fonctions indépendantes et ses applications à la théorie des nombres*, J. Math. Pures Appl. **28** (1949), 137–149.

[29] K. F. Roth, *On irregularities of distribution*, Mathematika **1** (1954), 73–79.

[30] K. F. Roth, *On the large sieves of Linnik and Rényi*, Mathematika **12** (1965), 1–9.
[31] W. M. Schmidt, *Irregularities of distribution, VII*, Acta Arith. **21** (1972), 49–50.
[32] A. Selberg, *Collected Papers*, Volume II, Springer-Verlag, Berlin, 1991.
[33] P. Turán, *On a new method of analysis and its applications*, Wiley-Interscience, New York, 1984.
[34] J. D. Vaaler, *Some extremal functions in Fourier analysis*, Bull. Amer. Math. Soc. **12** (1985), 183–216.
[35] G. S. Wagner, *On a problem of Erdős in Diophantine approximation*, Bull. London Math. Soc. **12** (1980), 81–88.
[36] H. Weyl, *Über ein Problem aus dem Gebiete der diophantischen Approximationen*, Nachr. Ges. Wiss. Göttingen, Math.-phys. Kl. (1914), 234–244; Gesammelte Abhandlungen, Band I, Springer-Verlag, Berlin–Heidelberg–New York, 1968,487–497.
[37] *Über die Gleichverteilung von Zahlen mod. Eins*, Math. Ann. **77** (1916), 313–352; Gesammelte Abhandlungen, Band I, Springer-Verlag, Berlin–Heidelberg–New York, 1968, 563–599; Selecta Hermann Weyl, Birkhäuser Verlag, Basel–Stuttgart, 1956, 111–147.