On the Uniform Distribution of Numbers mod. One

Mathematische Annalen *77*, 313-352 (1916)

by Hermann Weyl

translated from the German by John McSweeney*

March 17, 2009

1 Preliminaries. The Linear Case

Given infinitely many points

$$\alpha_1, \alpha_2, \alpha_3, \ldots$$

on the real number line, one can wind this line around a circle of circumference 1 and ask whether the marks corresponding to the α_n's are uniformly distributed around the circumference. This is the case if among the first n α's, the number $n_{\mathfrak{a}}$ of them which lie in a given arc $\mathfrak{a}$ of the circle is asymptotically equal to $|\mathfrak{a}| \cdot n$:

$$\lim_{n\to\infty} \frac{n_{\mathfrak{a}}}{n} = |\mathfrak{a}| \tag{1}$$

where $|\mathfrak{a}|$ of course denotes the length of the arc $\mathfrak{a}$. Only if this limit equation is satisfied for every arc $\mathfrak{a}$ do we consider these marks to have a uniformly dense distribution on the circumference. Winding the number line around the circle corresponds to viewing the real numbers mod 1; that is, two numbers are considered equal when they differ by an integer. Among the numbers x which are congruent mod 1 to a given α, there is one and only one which satisfies the inequality $0 \leq x < 1$; this reduced number α mod 1 will be denoted by (α).

In order to obtain a criterion for uniform distribution, let us assume that we have a sequence of numbers α_n mod 1 which indeed satisfies this above law. I then claim that for any bounded, Riemann-integrable function $f(x)$ which is periodic of period 1, we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{h=1}^{n} f(\alpha_h) \quad = \quad \int_0^1 f(x)dx \text{ [1]}, \tag{2}$$

i.e. the discrete average of the values at α_n of the function f agrees with the continuous average $\int_0^1 f(x)dx$. Our condition certainly implies that (2) is satisfied for every piecewise constant function

*Corrections, comments, questions, etc. on the mathematics, language or typesetting would be greatly appreciated at mcsweeney@math.ohio-state.edu.

[1] This ergodic-type result states that the time average (left-hand side) is equal to the space average (right-hand side) [J.M. 2005].

of period 1. The core of the matter is that to any bounded, Riemann-integrable function $f(x)$ defined on the interval $0 \le x \le 1$, there correspond two piecewise constant functions f_1 and f_2, which bound it from above and below ($f_1 \le f \le f_2$), and whose integrals $\int_0^1 f_1(x)dx$, $\int_0^1 f_2(x)dx$ differ by an arbitrarily small amount. If the difference of these integrals is ε, then we have

$$\lim_{n\to\infty} \frac{1}{n}\sum_{h=1}^{n} f_1(\alpha_h) \;=\; \int_0^1 f_1(x)dx \;\ge\; \int_0^1 f(x)dx \;-\; \varepsilon.$$

For sufficiently large n we therefore have

$$\frac{1}{n}\sum_{h=1}^{n} f_1(\alpha_h) \;>\; \int_0^1 f(x)dx \;-\; 2\varepsilon,$$

and so *a fortiori*

$$\frac{1}{n}\sum_{h=1}^{n} f(\alpha_h) \;>\; \int_0^1 f(x)dx \;-\; 2\varepsilon.$$

Similarly, by considering f_2, for sufficiently large n the left hand side satisfies

$$< \;\int_0^1 f(x)dx \;+\; 2\varepsilon,$$

and so our claim is proven.

The simplest function of period 1 is

$$e^{2\pi i x} \;=\; e(x),$$

which is the fundamental *analytic invariant of the congruence classes* mod 1. Assigning the complex number $e(x)$ to the real number x corresponds to nothing other than the process of analytically winding the real number line around the circle of circumference 1. For every integer m, $e(mx)$ is also of period 1; and so under the above assumption, we have, in particular, for every integer $m \neq 0$:

$$\lim_{n\to\infty} \frac{1}{n}\sum_{h=1}^{n} e(m\alpha_h) \;=\; \int_0^1 e(mx)dx \;=\; 0.$$

The theory of Fourier series tells us that every periodic function can be written as a linear combination of the special functions $e(mx)$. From this can be derived the following converse to our result:

Theorem 1 *If for every integer $m \neq 0$ the limiting relation*

$$\sum_{h=1}^{n} e(m\alpha_h) \;=\; o(n)$$

holds, then the numbers α_n mod 1 satisfy the law of uniform distribution everywhere.

In fact, since equation (2) is obvious for the function $e(0x) \;=\; 1$, it follows from our assumption in Theorem 1 that it holds for every truncated trigonometric series:

$$f(x) \;=\; \frac{a_0}{2} + (a_1\cos 2\pi x + b_1 \sin 2\pi x) + \ldots + (a_m \cos 2\pi m x + b_m \sin 2\pi m x).$$

When $f(x)$ is an arbitrary continuous function of period 1, one can find for each positive number ε a truncated trigonometric series f_ε such that $|f - f_\varepsilon| < \varepsilon$. In that case $f_1 = f_\varepsilon - \varepsilon$ and $f_2 = f_\varepsilon + \varepsilon$ will be two truncated trigonometric series which bound f from above and below and whose integrals

$$\int_0^1 f_1 \, dx \ , \quad \int_0^1 f_2 \, dx$$

differ by 2ε. We conclude as above that equation (2) is valid for the function f. Finally, if f is a piecewise constant function of period 1, one can easily (by replacing the jumps of f with steep, straight line segments[2]) find two continuous functions f_1 and f_2 between which f lies and whose integrals differ by an arbitrarily small amount. For that reason, equation (2) is also valid for such f. This proves Theorem 1, and we now have a convenient analytic criterion for the uniform distribution of numerical sequences mod 1.

We shall give right away an application, by demonstrating:

Theorem 2 *If ξ is an irrational number, then the integer multiples of ξ*

$$1\xi, 2\xi, 3\xi, \ldots$$

reduced mod 1 occur uniformly densely everywhere.

Let m be an integer $\neq 0$, and let's put $m\xi = \eta$, so we have only to determine that

$$\sum_{h=1}^{n} e(h\eta) \ = \ o(n)$$

holds. The left-hand side can be summed as a geometric series; its absolute value is

$$= \left| \frac{e((n+1)\eta) - e(\eta)}{e(\eta) - 1} \right| \ \leq \ \frac{2}{|e(\eta) - 1|} \ = \ \frac{1}{|\sin \pi\eta|};$$

hence, since η cannot be an integer, the above is not only $= o(n)$, but is actually bounded.

Theorem 2 was proven by Bohl, Sierpiński and myself in 1909-1910 on the basis of the fact that the irrational ξ can be approximated by a fraction whose numerator is $o(n)$, with an error of $o(1/n)$.[3] A further elementary proof, not using the exponential function, is due to Bohr. It goes as follows: one chooses a large integer H, and ε is defined to be the smallest number (in absolute value) to which one of the numbers $1\xi, 2\xi, 3\xi, \ldots, H\xi$ is congruent mod 1. Let this be e.g. $J\xi \equiv \varepsilon$. Further let $[1] = [a_1, b_1]$ and $[2] = [a_2, b_2]$ be two intervals of equal length. One can determine the natural number L such that $a_1 + L\varepsilon \in (a_2 - \varepsilon, a_2]$. Denote the interval $[a_1 + L\varepsilon, b_1 + L\varepsilon]$ by $[2']$. Whenever $n\xi$ mod 1 lies in the interval $[1]$, $(n + LJ)\xi$ mod 1 lies in the interval $[2']$, and vice versa. Hence if n_1 denotes the number of the first n terms

$$1\xi, 2\xi, 3\xi, \ldots, n\xi \tag{3}$$

[2]Weyl here uses the more poetic term 'Böschungen', roughly translating to 'embankments'. (JM)

[3]P.Bohl, Crelles J. *135* (1909), pp. 189-283, in particular p. 222. W. Sierpiński, Krakow, Akad. Anz., math.-naturw. Kl[A] Jan. 1910, p.9. H. Weyl, Rend. Circ. Mat. Palermo *30*, p. 406 (1910).

which lie in [1] mod 1, and n_2 and $n_{2'}$ have analogous definitions for the intervals [2] and [2′], we have that $|n_1 - n_{2'}|$ is bounded by a finite number C for all n. The part $\mathfrak{s}$ of [2′] which overhangs off of [2] has length $< |\varepsilon|$. If two numbers $h\xi$ and $k\xi$ ($h < k$) mod 1 lie in $\mathfrak{s}$, then $(k-h)\xi$ mod 1 is smaller than ε and hence $k - h > H$. Therefore at most $[n/H] + 1$ of the numbers from (3) lie in $\mathfrak{s}$. The number of terms from (3) which lie simultaneously in [2] and [2′] (i.e. in [2′] but not in $\mathfrak{s}$) is therefore

$$\geq n_{2'} - \left[\frac{n}{H}\right] - 1.$$

Consequently we have *a fortiori*

$$n_2 \geq n_{2'} - \left[\frac{n}{H}\right] - 1 \;\geq\; n_1 - \left[\frac{n}{H}\right] - 1 - C.$$

From this we obtain

$$\limsup_{n\to\infty} \frac{n_1 - n_2}{n} \;\leq\; \frac{1}{H},$$

and since H can be taken arbitrarily large,

$$\limsup_{n\to\infty} \frac{n_1 - n_2}{n} \leq 0.$$

As a similar argument works by interchanging intervals [1] and [2], it must be the case that

$$\lim_{n\to\infty} \frac{n_1 - n_2}{n} \;=\; 0,$$

i.e. the number of terms of (3) which fall in the intervals [1] and [2], which have equal length, are asymptotically equal. If we divide the interval $[0,1]$ into finitely many parts $\mathfrak{D}$ of equal length δ, then into every $\mathfrak{D}$ fall an asymptotically equal number of the terms of (3) reduced mod 1; as every term lies in exactly one $\mathfrak{D}$, the number of terms that fall in a given $\mathfrak{D}$ must be asymptotically equal to $\delta \cdot n$. From there our claim (1) follows for every interval $\mathfrak{a}$ which is composed of a finite number of intervals of the form $\mathfrak{D}$, and therefore also for any arbitrary interval, as δ can be taken arbitrarily small.

Referring back to our general question, we shall conclude by mentioning that the limit equation (1), if it applies at all to a specific sequence α_n for every interval $\mathfrak{a}$, it is also equally valid for an arbitrary interval $\mathfrak{a}$ (of length < 1). Because we're dividing the interval $[0,1]$ into a finite number of parts $\mathfrak{D}$, whose length is δ, for all n above a certain threshold, and for all intervals $\mathfrak{D}$,

$$\delta(1-\delta)n \;\leq\; n_{\mathfrak{D}} \;\leq\; \delta(1+\delta)n.$$

For an arbitrary given interval $\mathfrak{a}$, let's look on the one hand at the intervals $\mathfrak{D}$ which lie entirely inside $\mathfrak{a}$ (they have a total length $> |\mathfrak{a}| - 2\varepsilon$), and on the other hand those that have any points at all in common with $\mathfrak{a}$ (and whose total length is $< |\mathfrak{a}| + 2\varepsilon$). One thus obtains, for any value n which exceeds the chosen threshold

$$\begin{aligned} (|\mathfrak{a}| - 2\delta)(1-\delta)n \quad &\leq \quad n_{\mathfrak{a}} \quad \leq \quad (|\mathfrak{a}| + 2\delta)(1+\delta)n, \\ \left|\frac{n_{\mathfrak{a}}}{n} - |\mathfrak{a}|\right| \quad &\leq \quad 3\delta + 2\delta^2. \end{aligned}$$

2 Generalization to Higher Dimensions. Applications

When one confers specific values

$$x_1 = \alpha_1,\ x_2 = \alpha_2,\ \ \ldots,\ x_p = \alpha_p$$

to p variables $x_1, x_2, \ldots, x_p$, one obtains a point (α) in p-dimensional space with coordinates α_i. We consider two such points to be congruent (mod 1, that is) if the corresponding coordinates of both points are congruent mod 1. Among the points which are congruent to the given (α),

$$x_1 \equiv \alpha_1, x_2 \equiv \alpha_2, \ldots, x_p \equiv \alpha_p \pmod 1,$$

there is exactly one which belongs to the unit hypercube $0 \leq x_1, x_2, \ldots, x_p < 1$; we shall call this one *reduced*. One identifies two points if they are congruent, or simply considers every system of congruent points as a single "point", which generates a closed p-dimensional manifold $\mathfrak{R}_p$ from the usual p-dimensional space. A region (a closed, bounded, point set with a well-defined volume V in the Cantor-Jordan sense) in the usual space is, as long as no two of its points are congruent, also a (simply covered) region with volume V in $\mathfrak{R}_p$. A linear transformation

$$x_i \ = \ \sum_{k=1}^{p} a_{ik} x'_k + a_i \quad (i = 1, 2, \ldots, p)$$

can be considered a transformation of $\mathfrak{R}_p$ if and only if it is integral and unimodular (i.e. when the coefficients a_{ik} are integers[4] and form a matrix of determinant ± 1).

Let an infinite sequence in $\mathfrak{R}_p$ be given by

$$\alpha(n): \quad x_1 \equiv \alpha_1(n), x_2 \equiv \alpha_2(n), \ldots, x_p \equiv \alpha_p(n) \pmod 1 \qquad \{n = 1, 2, 3, \ldots\}.$$

We ask as we did above for the case $p = 1$: when does this sequence lie uniformly densely in $\mathfrak{R}_p$, i.e. when is it the case that among the first n terms, the number of these which lie inside a given bounded region of volume V is asymptotically given by $V \cdot n$? The following theorem gives the criterion, whose proof is exactly as above:

Theorem 3 *The point sequence $\alpha(n)$ fills in $\mathfrak{R}_p$ uniformly densely if, for every system of integers $m_1, m_2, \ldots, m_p$ which do not all vanish, the limit equation*

$$\sum_{h=1}^{n} e(m_1\alpha_1(h) + m_2\alpha_2(h) + \ldots + m_p\alpha_p(h)) \ = \ o(n)$$

holds.

From there follows immediately the following generalization of Theorem 2:

[4] The a_i's can be arbitrary numbers.

Theorem 4 *Let $\xi_1, \xi_2, \ldots, \xi_p$ be any p numbers, among which there is no integral linear relation (i.e. no relation of the form*

$$l_1\xi_1 \ + \ l_2\xi_2 \ + \ \ldots \ + \ l_p\xi_p \ = \ l,$$

where the coefficients l_i and l are integers which are not all zero); then the point sequence

$$(n\xi_1, n\xi_2, \ldots, n\xi_p) \qquad \{n = 1, 2, 3, \ldots\}$$

is uniformly distributed mod 1.

The claim that this sequence is everywhere dense is the content of a famous approximation theorem of Kronecker[5]. The preceding much sharper theorem was initially stated by me in the summer of 1913 in a lecture at the Göttingen Mathematical Society and proven by similar methods as here. The essentially equivalent proofs of Theorem 2 by Sierpiński, Bohr and myself do not lend themselves to generalization to the higher-dimensional case, although Bohr's elementary proof does. Relative to the surprising applications which Bohr made to the theory of the Riemann ζ-function from Kronecker's theorem and this sharper one, we refer the reader to "The Riemann Zetafunction and the Theory of Prime Numbers" by Bohr and Littlewood. It will be discussed in section 5 how this statement is to be modified when there is one or more integral linear relationships between the ξ_i's.

We can replace Theorem 4's parameter n, which ran through the discrete sequence of positive integers, by a continuous parameter t which we will call *time*. From this we obtain:

Theorem 5 *If a point moves in $\mathfrak{R}_p$ with constant speed in a straight line:*

$$x_i \equiv \alpha_i + \gamma_i t \quad (\alpha_i, \gamma_i \text{ constants}), \tag{4}$$

then the time it spends in an arbitrary region is proportional to its volume. It is assumed that there is no homogeneous integral linear relation between the directional components γ_i.

During the long observation time from $t = 0$ to t, we denote by t_G the total length of the time intervals which the point spends in a given region G, so that the relative time that the point spends in a region G is understood to be $\lim_{t\to\infty} t_G/t$. The volume of a region G in $\mathfrak{R}_p$ can be considered as the *a priori* probability that a point selected uniformly at random falls in G, and therefore our theorem implies that for a straight path of constant speed, *the relative time spent is equal to the a priori probability.* The connection with statistical mechanics is evident. — Let $m_1, m_2, \ldots, m_p$ be any p integers which do not all vanish, and put

$$m_1\alpha_1 + m_2\alpha_2 + \ldots + m_p\alpha_p = \alpha, \quad m_1\gamma_1 + m_2\gamma_2 + \ldots + m_p\gamma_p = \gamma.$$

Then our principle (Theorem 3), adapted to the case of a continuous parameter t, tells us that for the proof of Theorem 5, all we are required to prove is the equality

$$\int_0^t e(\alpha + \gamma t)\ dt \ = \ o(t).$$

[5] *Die Periodensysteme von Funktionen reeller Variablen,* Ber. Preuss. Akad. Wiss. Berlin *1884,* pp. 1071-1080, and *Näherungsweise ganzzahlige Auflösung linearer Gleichungen*, ibid *1884,* pp.1179-1193, 1271-1299. (Werke III 1, pp.32-109.)

This is however obvious, as the integral on the left-hand side evaluates to

$$\frac{e(a+\gamma t)-e(\alpha)}{\gamma} \;=\; O(1).$$

One can formulate Theorem 5 in several other ways. By a *Closed Euclidean Space* we mean any closed p-dimensional manifold which has the property that to every one of its points there belongs a neighborhood in which Euclidean geometry is valid.[6] $\mathfrak{R}_p$ is such a closed Euclidean space. If Γ_0 is a group that consists of finitely many, say ν, integral unimodular linear transformations of $\mathfrak{R}_p$, and consider every system of ν points which are equivalent relative to Γ_0 to be a single point, then $\mathfrak{R}_p$ generates a closed Euclidean space $\mathfrak{R}_p^{\Gamma_0}$, over which $\mathfrak{R}_p$ spreads as an unbounded and unramified finitely-sheeted covering space, as long as none of the transformations belonging to Γ_0 possesses a fixed point in $\mathfrak{R}_p$. It can be rigorously proven (which I will do in the appendix) that every closed Euclidean space is such an $\mathfrak{R}_p^{\Gamma_0}$. Accordingly we can claim:

Theorem 6 *In a closed Euclidean space, every straight line, apart from easily characterized exceptions, spreads so that it comes arbitrarily close to every point of the space, and spends on average the same amount of time in any region of the same volume in said space.*

Let two points in $\mathfrak{R}_p$ be identified if their coordinates agree mod 1 *up to their sign*, so that any single point is determined through the congruences

$$x_1 \equiv \pm\alpha_1,\; x_2 \equiv \pm\alpha_2,\; \;\ldots,\; x_p \equiv \pm\alpha_p \pmod 1$$

with any of the 2^p possible sign combinations; thus every such point has exactly one representative in the cube of edge length 1/2:

$$0 \le x_1, x_2, \ldots, x_p \le \frac{1}{2}.$$

In this cube, the constant-speed line (4) is represented as a zigzag path, which could be described by a point mass in p-dimensional space being reflected off the walls of the cube according to the usual laws of reflection (in the case $p = 2$, as the path of a billiard ball). If there is no homogeneous integral linear relation between the velocity components relative to the coordinate axes, *a point moving according to this law spends on average the same amount of time in any area of the cube of a given size.* Kőnig and Szűcs have already shown, based on Kronecker's approximation theorem, that such a point comes arbitrarily close to every position in the cube[7]. It would not be difficult to completely characterize the possible exceptional cases which Kőnig and Szűcs also considered.

It is known that one can map $\mathfrak{R}_2$ bijectively[8] and conformally to the torus; the straight lines are mapped to the *loxodromes* of the torus, which intersect its meridians everywhere at a constant

[6]The problem of formulating the axioms of Euclidean as well as non-Euclidean geometry in such a manner that they only reveal something about the neighborhood of any one point (without invoking the dimension), and then to investigate which distinct (in the Analysis Situs [topological - JM] sense) spaces satisfy these axioms, was called the "Clifford-Klein problem of spatial forms" in honor of the originators of this question. Compare namely F. Klein, *Über Nicht-Euklidische Geometrie*, Math. Ann. *37*.

[7]D. Kőnig and A. Szűcs, *Mouvement d'un point abandonné à l'intérieur d'un cube*, Rend. Circ. Mat. Palermo *36* (1913).

[8]Weyl uses a less elegant construction, 'umkehrbar eindeutig', roughly 'invertibly well-defined' (JM)

angle. One can therefore construct a torus of any kind, if one replaces $\mathfrak{R}_2$ by the two-dimensional manifold in which two points $(x', y'), (x'', y'')$ coincide if and only if

$$x' \equiv x'' (\text{mod } a), \quad y' \equiv y'' (\text{mod } b)$$

hold; here a, b are two fixed real numbers. Every one of these manifolds can be transformed into $\mathfrak{R}_2$ by a suitable affine transformation of the coordinates xy. The torus is constructed by the rotation of a circle of radius r; let κ denote the value of the Gaussian curvature at any position on the surface. We can then say that *a loxodrome running on the torus* (except for easily characterizable exceptions, which are closed curves), *covers our torus densely everywhere; this density is not uniform, but rather is proportional to* $1 - r^2\kappa$ (therefore is larger on the parts facing towards the axis of rotation than on those facing away).

For $p = 2$ one can evidently reformulate Theorem 5 as follows: *If* $\overrightarrow{\beta\delta}$ *is a line segment*[9] *in* $\mathfrak{R}_2$, *which is not parallel to the straight path (4) of the moving point, then the moving point crosses the segment on average per unit of time as often as the area of the parallelogram spanned by the velocity vector* (γ_1, γ_2) *and the segment* $\overrightarrow{\beta\delta}$. It is assumed throughout that the ratio of the components of the velocity vector $\gamma_1 : \gamma_2$ is irrational.

This formulation is due to Bohl and was derived by him from Theorem 2. He used it to settle a question, put forth by Lagrange, about the *Superposition of Oscillations.* An oscillatory process arising from the superposition of m simple oscillations is described by the formula

$$z = \sum_{i=1}^{m} C_i \; e(\alpha_i + \gamma_i t), \tag{5}$$

in which the α_i's and the γ_i's are real constants and the C_i's are positive constants; geometrically, this can be represented by an epicycloid in the complex z-plane. We shall write $z = r \cdot e(\sigma)$, where we let r (≥ 0) denote the absolute value and σ the argument of z. Lagrange asked whether the argument σ on average per time unit approaches a constant value, i.e. whether the limit $\lim_{t\to\infty} \sigma/t$ exists. In a special case Lagrange could himself already provide an answer: if one of the C_i, say C_m, is larger than the sum of all the others, the following — even if one replaces the arguments $\alpha_i + \gamma_i t$ by independent variables — always holds:

$$|\sigma - (\alpha_m + \gamma_m t)| < \frac{1}{4}, \quad \lim_{t\to\infty} \frac{\sigma}{t} = \gamma_m.$$

In the case $m = 3$, Bohl was able to solve the non-Lagrangian case as well. He showed that: *if* $\pi \mathbf{A}_i$ *are the angles of a triangle of sides* C_i, *then*

$$\lim_{t\to\infty} \frac{\sigma}{t} = \gamma_1 \mathbf{A}_1 + \gamma_2 \mathbf{A}_2 + \gamma_3 \mathbf{A}_3. \tag{6}$$

We are assuming that there is no homogeneous integral linear relationship between the γ_i's. For the proof we may assume that $\alpha_3 = \gamma_3 = 0$. The quantity

$$z = re(\sigma) = C_1 e(x_1) + C_2 e(x_2) + C_3 \tag{7}$$

[9] A path $\vec{\beta\delta}$ in $\mathfrak{R}_p$ is naturally *not* uniquely determined by its endpoints β and δ.

as a function of x_1, x_2 vanishes at the two points

$$\beta: \; x_1 \equiv \beta_1 = \frac{1-\mathbf{A}_2}{2}, \quad x_2 \equiv \beta_2 = \frac{1+\mathbf{A}_1}{2} \pmod 1;$$
$$\delta: \; x_1 \equiv \delta_1 = \frac{1+\mathbf{A}_2}{2}, \quad x_2 \equiv \delta_2 = \frac{1-\mathbf{A}_1}{2} \pmod 1$$

in $\mathfrak{R}_2$. Cut $\mathfrak{R}_2$ along the segment

$$\overrightarrow{\beta\delta}: \quad x_1 \equiv \beta_1\tau + \delta_1(1-\tau) \pmod 1;$$
$$x_2 \equiv \beta_2\tau + \delta_2(1-\tau) \pmod 1, \quad \{0 \le \tau \le 1\}.$$

One can then show by means of very simple geometric considerations that the function $\sigma = \sigma(x_1, x_2)$, defined by (7), which has as endpoints β and δ, is well-defined and continuous in the *cut space* $\mathfrak{R}_2$, although has a jump of 1 at the $\overrightarrow{\beta\delta}$ cut. Bohl's result follows without further ado.

As Theorem 5 is valid for all p, it is not difficult to completely answer Lagrange's question not only for the case $m = 3$ but also for all m (at which Bohl was not successful). For $m = 4$, the non-Lagrangian case splits into two further subcases. I shall present here the result for one of these cases [10]:

Theorem 7 *Let four positive numbers $C_1 < C_2 < C_3 < C_4$ be given, for which $C_1 + C_4 > C_2 + C_3$ and $C_4 < C_1 + C_2 + C_3$ hold. It is then possible for a planar linkage of sides C_i to run through in a single cycle all the incongruent states which can be formed by a planar quadrilateral of sides C_i. If $2\pi\xi_i$ denote the angles which the sides C_i of such a quadrilateral in its plane make against any fixed straight line in the plane, then the angle differences $2\pi(\xi_i - \xi_k)$ all return together to their initial values upon the completion of such a cycle. One can form the integral ranging over a cycle*

$$\mathbf{A}_1 = \frac{1}{2}\int (\xi_2 d\xi_3 - \xi_3 d\xi_2) + (\xi_3 d\xi_4 - \xi_4 d\xi_3) + (\xi_4 d\xi_2 - \xi_2 d\xi_4)$$

and analogously for $\mathbf{A}_2$, $\mathbf{A}_3$, $\mathbf{A}_4$. These integrals are independent of the way in which the planar linkage runs through its incongruent states, and are also independent of the fixed line against which the angles $2\pi\xi_i$ are measured. For that reason I will call these the integral invariants of the linkage. Their sum is equal to 1. By superposing four simple oscillations with amplitude C_i, we obtain the oscillation

$$z = re(\sigma) = \sum_{i=1}^{4} C_i e(\alpha_i + \gamma_i t);$$

its argument σ grows as $\sum_{i=1}^{4} \gamma_i \mathbf{A_i}$ on average per unit of time, as long as there is no homogeneous integral linear relationship among the γ_i's.

The relation between the integral invariants $\mathbf{A}_i$ and the sides C_i of the linkage seems analogous to the relation between the angles and sides of a triangle in planar geometry.

[10] Compare a lecture of mine given at the Spring 1914 meeting of the Swiss Mathematical Society: *Une application de la théorie des nombres à la mécanique statistique et à la théorie des perturbations*, Enseignement Mathématique *16*, No. 6 (1914).

Lagrange's question is of particular interest in *astronomy*, where r denotes the numerical eccentricity of the orbit of a planet, and σ the perihelion longitude (or r, the sine of the orbit's inclination relative to the fixed plane of the planet system, and σ, the knot longitude). Formulas of the form (5) represent therefore in particular the *secular evolution of perihelion and knot longitudes.* The non-Lagrangian case is only valid for *Venus* and *Earth* in our solar system.

3 The Series $\sum e(\varphi(n))$ for a Polynomial φ

Until now we have assumed that the points in $\mathfrak{R}_p$ which we have been considering depend in a *linear* fashion on a parameter, be it continuous or discrete. We now move to the more general case where the dependence is given by *polynomials of higher degree.* Let the parameter t (time) run through its values *continuously*, so that the corresponding questions can be dealt with by considering the formula

$$\int_0^t e(\varphi(t))\ dt \ = \ o(t), \tag{8}$$

which is valid whenever $\varphi(t)$ is any polynomial in t which does not reduce to a mere constant. Since the linear case has been considered, we may assume the degree of $\varphi(t)$ to be greater than 1. Let's fix a t_0 such that for $t \geq t_0$ the derivative $\varphi'(t)$ has constant sign, let's say positive, and make the substitution $\varphi(t) = x$ in the integral

$$2\pi i \int_{t_0}^t e(\varphi(t))\ dt \qquad (t > t_0).$$

It therefore becomes

$$2\pi i \int_{t=t_0}^t e(x)\frac{dx}{\varphi'(t)} \ = \ \int_{t_0}^t \frac{de(x)}{\varphi'(t)}.$$

Integration by parts yields

$$\left[\frac{e(x)}{\varphi'(t)}\right]_{t_0}^t \ + \ \int_{t_0}^t \frac{\varphi''(t)}{(\varphi'(t))^2}\ e(\varphi(t))\ dt.$$

Since $\int_{t_0}^\infty \frac{|\varphi''|}{\varphi'^2} dt$ converges, the convergence of

$$\int_0^\infty e(\varphi(t))\ dt$$

follows. We have therefore proven a sharper form of (8), namely

$$\int_0^t e(\varphi(t))\ dt \ = \ O(1).$$

With the help of our principle we can conclude from this:

Theorem 8 *If $\varphi_1(t),\ \varphi_2(t),\ \ldots,\ \varphi_p(t)$ are p polynomials among which there is no linear combination with integer coefficients (which do not all vanish) which reduces to a mere constant, then a point moving in $\mathfrak{R}_p$ according to the law*

$$x_i \ \equiv \ \varphi_i(t) \quad (i = 1, 2, \ldots, p)$$

spends on average an equal amount of time (given an infinite observation time) in regions of equal volume.

The analysis becomes more intricate when we replace the continuous parameter t by a discrete n, which runs through the natural numbers. In this case everything depends on showing that:

Theorem 9 *If $\varphi(z)$ is a polynomial of the q^{th} degree:*

$$\varphi(z) \ = \ \alpha_q z^q + \alpha_{q-1} z^{q-1} + \ldots + \alpha_0$$

in which one of the coefficients $\alpha_q, \alpha_{q-1}, \ldots, \alpha_1$ is irrational, then the limit equation

$$\sum_{h=0}^{n} e(\varphi(h)) \ = \ o(n) \tag{9}$$

holds.

{**Addendum** *If α_l is the irrational coefficient of highest index, then if we fix α_l, (9) holds uniformly for all values of the coefficients $\alpha_{l-1}, \ldots, \alpha_1, \alpha_0$.*}

For $\varphi(z) = \alpha z^q$ this theorem was first articulated by Hardy and Littlewood at Cambridge (1912)[11]. For polynomials $\varphi(z)$ of the second degree, these authors have since published in Acta Mathematica a proof based on an application of Cauchy's Integral Theorem[12]; as I gather from a friendly communiqué from Hardy, they obtained Theorem 9 to the same extent that I did, and their proof, which differs significantly from mine, will shortly appear in Acta Mathematica.

Consider in the usual q-dimensional space all the lattice points $\mathfrak{r} = (r_1, r_2, \ldots r_q)$, i.e. all the points with integer coefficients r_i, which belong to the "octahedral" region

$$|\mathfrak{r}| \ = \ |r_1| + |r_2| + \ldots + |r_q| \ \leq \ n.$$

Denote the number of all such points n_q; it is equal to

$$2^q \binom{n}{q} \ + \ 2^{q-1} \binom{n}{q-1} \binom{q}{1} \ + \ 2^{q-2} \binom{n}{q-2} \binom{q}{2} \ + \ \ldots \ + \ 1.$$

We do not need to know the exact value of n_q; we merely need, for large enough n, the (obvious from a geometric point of view) asymptotic formula

$$n_q \ \sim \ \frac{(2n)^q}{q!}.$$

As a preamble to Theorem 9 we will need the following lemma:

Lemma *Let ξ be an irrational number. Among the n_q lattice points*

$$\mathfrak{r} \ = \ (r_1, r_2, \ldots, r_q)$$

[11] In their lecture *Some Problems of Diophantine Approximation*; cf. the Conference Proceedings (*Proc. Int'l. Conf.Math. Cambridge* (1912), pp. 223-229)

[12] G.H. Hardy and J.E. Littlewood, Acta Math. *37*, pp. 193-238, Theorem 2.14 on p. 213. This work is the continuation of the essay *Some Problems of Diophantine Approximations*, which begins on p. 155, whose 3rd part is still to come. [This 3rd part never appeared (addendum 1955).]

which belong to the octahedral region $|\mathfrak{r}| \le n$*, consider those for which the quantity*

$$r_1 r_2 \ldots r_q \xi \ \mathit{mod}\ 1$$

lies in the arbitrary given interval (a, b) *of length* $c = b - a < 1$*. We claim that the number of these is asymptotic to* $c \cdot n_q$*.*

Pursuant to the principle set forth in Theorem 1, for the proof we need only deduce the limiting equality

$$\lim_{n\to\infty} \frac{1}{n_q} \sum_{|\mathfrak{r}|\le n} e(r_1 r_2 \ldots r_q \xi) \;=\; 0. \tag{10}$$

If indeed this is true for every irrational number ξ, then it will hold as well for any integer multiple $m\xi$ of ξ ($m \neq 0$). As the claimed equality is correct for $q = 1$, we can propagate the proof to q by means of the conclusion for $q-1$[13]. In the sum on the left-hand side of (10), I will first of all evaluate the summation over r_q, and thereafter over the remaining r's. $\mathfrak{r}'$ will denote the "projected" lattice point $(r_1, r_2, \ldots, r_{q-1})$ in the $(q-1)$-dimensional coordinate space ($x_q = 0$), and I will therefore set

$$\sum_{\mathfrak{r}} \;=\; \sum_{\mathfrak{r}'} \sum_{r_q}$$

The outside summation indexed $\mathfrak{r}'$ is taken over the octahedral region $|\mathfrak{r}'| \le n$, and for every such $\mathfrak{r}'$, the inner one over all the integers which satisfy the condition

$$|r_q| \;\le\; n - |\mathfrak{r}'|.$$

The inner summation I can evaluate (it's a geometric series); if we denote

$$r_1 r_2 \ldots r_{q-1} \;=\; R,$$

it yields

$$\left| \sum_{r_q} e(r_q R \xi) \right| \;\le\; \frac{1}{|\sin(\pi R \xi)|}.$$

As this sum consists of at most $2n + 1$ terms, we also have

$$\left| \sum_{r_q} e(r_q R \xi) \right| \;\le\; 2n + 1.$$

Choose a positive number ε ($< 1/2$). We assume that our theorem is already established for $q - 1$, so that we know that the number of the n_{q-1} lattice points $\mathfrak{r}'$, for which $R\xi$ mod 1 lies between $-\varepsilon$ and $+\varepsilon$ is asymptotically $2\varepsilon \cdot n_{q-1}$, so for large enough n is certainly $< 3\varepsilon \cdot n_{q-1}$. For these $\mathfrak{r}'$ we will use the second of the estimates just given for the inner sum; for the rest, the first suffices:

$$\left| \sum_{r_q} e(r_q R \xi) \right| \;\le\; \frac{1}{\sin \pi\varepsilon}.$$

[13] Of course this is an inductive proof, although Weyl never actually uses the word. –J.M.

So altogether we have

$$\left|\sum_{\mathfrak{r}} e(r_1 r_2 \dots r_q \xi)\right| \;=\; \left|\sum_{\mathfrak{r}}\right| \;\leq\; \sum_{\mathfrak{r}'}\left|\sum_{r_q}\right| \;\leq\; n_{q-1}\left\{3\varepsilon(2n+1)+\frac{1}{\sin\pi\varepsilon}\right\}.$$

Since

$$\lim_{n\to\infty}\frac{n_{q-1}(2n+1)}{n_q} \;=\; q,$$

we consequently have, for sufficiently large n:

$$\frac{1}{n_q}\left|\sum_{\mathfrak{r}} e(r_1 r_2 \dots r_q \xi)\right| \;<\; \varepsilon(3q+1),$$

and that was the content of our claim.

For the proof of Theorem 9 we will make use of the following consequence of our lemma: The number of systems of integers $r_1, r_2, \dots, r_{q-1}$ for which

$$|\mathfrak{r}'| \;=\; |r_1|+|r_2|+\dots+|r_{q-1}| \;\leq\; n$$

holds, and for which $R\xi = r_1 r_2 \dots r_{q-1}\xi$ mod 1 lies between $-\varepsilon$ and $+\varepsilon$, is, for large enough n, less than $3\varepsilon \cdot n_{q-1}$.

After this preliminary preparation, we may undertake the study of the sum

$$\sigma_n \;=\; \sum_{h=0}^{n} e(\varphi(h)).$$

First, we assume that the highest coefficient α_q of $\varphi(z)$ is irrational.

First Step. The conjugate of σ_n is

$$\overline{\sigma_n} \;=\; \sum_{h=0}^{n} e(-\varphi(h));$$

therefore

$$|\sigma_n|^2 \;=\; \sigma_n\overline{\sigma_n} \;=\; \sum_{h=0}^{n}\sum_{k=0}^{n} e(\varphi(h))\cdot e(-\varphi(k)) \;=\; \sum_{h,k} e(\varphi(h)-\varphi(k)).$$

Let us put $h = k + r$; then we will have

$$\varphi(h) \;=\; \varphi(k+r) \;=\; \varphi(k) + r\varphi(r,k).$$

$\varphi(r,k)$ is a polynomial[14] in k and r, which only contains terms of order $q-1$ or lower; in its expansion the coefficient α_0 no longer appears, and the expansion (ordered by decreasing powers of k) begins with the term $q\alpha_q k^{q-1}$. We therefore now have

$$|\sigma_n|^2 \;=\; \sum_r\sum_k e(r\varphi(r,k)).$$

[14] Weyl refers to such functions as 'ganze rationale Funktionen' ('entire rational functions'), which seems to be his way of denoting polynomials in several variables.

The range of summation is described by

$$0 \ \le\ k\ \le\ n, \qquad 0\ \le\ k+r\ \le\ n.$$

Here r runs through the entire interval from $-n$ to $+n$; for every such r in the inner sum, we have that k runs through all the integers in the interval from 0 to $n-|r|$, or from $|r|$ to n, depending on whether $r \geq 0$ or $r \leq 0$.

Second Step. We shall use the symbol $n_q,\ q = 1, 2, \ldots$ in the same sense as in the lemma, so from the last equation one obtains, with the help of the Schwarz inequality,

$$|\sigma_n|^4\ \le\ n_1 \sum_r \left| \sum_k e(r\varphi(r,k)) \right|^2 .$$

In what follows, I shall now repeat the procedure from Step 1. We have

$$\left| \sum_k e(r\varphi(r,k)) \right|^2 \ =\ \sum_{k,l} e(r\varphi(r,k) - r\varphi(r,l)).$$

Again I shall write

$$k = l+s, \quad \varphi(r,k) = \varphi(r, l+s) = \varphi(r,l) + s\varphi(r,s,l).$$

The polynomial function $\varphi(r,s,l)$ of r, s and l contains only terms of order $\leq q-2$ and its expansion, ordered by decreasing powers of l, begins with the term $q(q-1)\alpha_q l^{q-2}$; the coefficients $\alpha_0,\ \alpha_1$ no longer appear. At the present stage we have

$$|\sigma_n|^4\ \le\ n_1 \sum_{r,s} \sum_l e(rs\varphi(r,s,l)).$$

Here (r,s) runs through the two-dimensional "octahedron" $|r|+|s| \leq n$, and l through the interval which arises from the summation interval for k described above, if one removes $|s|$ numbers from above or from below, depending on whether $s \geq 0$ or $s \leq 0$.

The *Third Step* begins with one more application of the Schwarz inequality, which yields

$$|\sigma_n|^8\ \le\ n_1^2 n_2 \sum_{r,s} \left| \sum_l e(rs\varphi(r,s,l)) \right|^2 ,$$

and then proceeds in a similar fashion as above.

We must continue the process until the $(q-1)^{\text{st}}$ step. To this end, we will distinguish the summation letters $h, k, l, \ldots; r, s, \ldots$ with the help of indices, as follows:

$$\begin{aligned}
h_1 &= r_1 + h_2; & \varphi(h_1) &= \varphi(h_2) & &+ r_1\varphi(r_1, h_2),\\
h_2 &= r_2 + h_3; & \varphi(r_1, h_2) &= \varphi(r_1, h_3) & &+ r_2\varphi(r_1, r_2, h_3),\\
&\ \ \cdots & &\ \ \cdots & &\ \ \cdots\\
h_{q-1} &= r_{q-1} + h; & \varphi(r_1, \ldots, r_{q-2}, h_{q-1}) &= \varphi(r_1, \ldots, r_{q-2}, h) & &+ r_{q-1}\varphi(r_1, r_2, \ldots, r_{q-1}, h).
\end{aligned}$$

The last of these functions φ, a function of the q arguments $r_1, r_2, \ldots, r_{q-1}, h$, only contains terms of degree 0 and 1, and the coefficient by which h is multiplied is $q!\alpha_q$:

$$\varphi(r_1, r_2, \ldots, r_{q-1}, h) \;=\; q!\alpha_q h + (\beta_0 + \beta_1 r_1 + \beta_2 r_2 + \ldots + \beta_{q-1} r_{q-1}).$$

The coefficients β can be computed from the first two coefficients of $\varphi(z)$, α_q and α_{q-1}, in a manner which is of no interest to us at this moment. I shall introduce the following abbreviations:

$$R = r_1 r_2 \ldots r_{q-1},\;\; \rho = R\;(\beta_0 + \beta_1 r_1 + \ldots + \beta_{q-1} r_{q-1}),\;\; \xi = q!\;\alpha_q$$

$$Q = 2^{q-1}, \quad N = (n_1)^{2^{q-3}} (n_2)^{2^{q-4}} \cdots n_{q-3}^2 n_{q-2},$$

(for $q = 2$ we set $N = 1$), so finally the inequality

$$|\sigma_n|^Q \;\leq\; N \sum_{\mathfrak{r}'} \left\{ e(\rho) \sum_h e(R\xi h) \right\} \tag{11}$$

comes about, in which $\mathfrak{r}' = (r_1, r_2, \ldots, r_{q-1})$ runs through the octahedron $|\mathfrak{r}'| \leq n$, and h through an interval (depending on $\mathfrak{r}'$) of $n + 1 - |\mathfrak{r}'|$ integers; the latter is obtained from $\{0, 1, \ldots, n\}$ by removing $|r_i|$ terms from below or above, depending on whether $r_i \geq 0$ or $r_i \leq 0$. In the expression for N let us replace every n_i with n^i, so that N becomes a power of n, whose exponent is

$$1 \cdot 2^{q-3} + 2 \cdot 2^{q-4} + 3 \cdot 2^{q-5} + \ldots + (q-2) \cdot 2^0 \;=\; Q - q.$$

Therefore for N we have the asymptotic formula

$$N \;\sim\; \kappa \cdot n^{Q-q},$$

in which κ is a constant which only depends on q.

In (11) we may evaluate the inner summation, indexed by h, as it is a geometric series. We obtain, for all $\mathfrak{r}'$ for which $R\xi$ mod 1 does not lie between $-\varepsilon$ and $+\varepsilon$,

$$\left| \sum_h \right| \;\leq\; \frac{1}{\sin \pi\varepsilon};$$

for the remaining $\mathfrak{r}'$, the number of which is less than $3\varepsilon \cdot n_{q-1}$ for large enough n, we shall once again use the crude estimate

$$\left| \sum_h \right| \;\leq\; n + 1.$$

Then from our formula (11) follows the inequality

$$|\sigma_n|^Q \;\leq\; N \cdot n_{q-1} \left\{ 3\varepsilon(n+1) + \frac{1}{\sin \pi\varepsilon} \right\}.$$

As soon as n is sufficiently large, we get

$$\left| \frac{\sigma_n}{n} \right|^Q \;\leq\; 3\varepsilon \left(\frac{\kappa \cdot 2^{q-1}}{(q-1)!} + 1 \right),$$

and our proof is finished.

If the highest coefficient is not irrational, but rather, say, $\alpha_q, \alpha_{q-1}, \ldots, \alpha_{l+1}$ are rational and α_l is irrational, let G be the common denominator of the fractions $\alpha_q, \alpha_{q-1}, \ldots \alpha_{l+1}$. We then partition $\sum_n$ into G parts according to the remainder n mod G. We shall therefore replace n by $Gn + r$ and have

$$\sum_{h=0}^{Gn-1} e(\varphi(h)) \;=\; \sum_{r=0}^{G-1} \sum_{h=0}^{n-1} e(\varphi(Gh + r)).$$

But for every $r = 0, 1, \ldots, G-1$, $\varphi(Gz + r)$ is congruent mod 1 to a polynomial $\psi_r(z)$ of the l^{th} degree[15], whose highest coefficient $\alpha_l G^l$ is irrational. Therefore each of the G sums into which we broke up the original sum is $o(n)$.

In the same vein one can consider sums of the form

$$\sum ne(\varphi(n)), \qquad \sum \frac{1}{n} e(\varphi(n))$$

or, more generally,

$$\sum_n a_n e(\varphi(n))$$

in which $a_1, a_2, a_3, \ldots$ are positive numbers whose sum diverges.

Theorem 10 *The limiting relation*

$$\sum_{h=0}^{n} a_h e(\varphi(h)) \;=\; o\left(\sum_{h=0}^{n} a_h\right)$$

is valid for every divergent series $a_0 + a_1 + a_2 + \ldots$ of positive terms which decrease monotonically. If the terms increase monotonically, the same claim still holds under the condition

$$na_n \;=\; O\left(\sum_{h=0}^{n} a_h\right);$$

for example this is always true if a_n grows as a power of n.

This result is obtained by an application of the well-known method of partial summation:

$$\sum_{h=0}^{n} a_h e(\varphi(h)) \;=\; a_n \sigma_n \;+\; \sum_{h=0}^{n-1} \sigma_h (a_h - a_{h+1}).$$

Let $\varepsilon > 0$ be given, so that starting from a certain h we have $|\sigma_h| < \varepsilon \cdot h$, and therefore

$$\left|\sum_{h=0}^{n} a_h e(\varphi(h))\right| \;\leq\; C_\varepsilon \;+\; \varepsilon n a_n \;+\; \varepsilon \sum_{h=0}^{n-1} h|a_h - a_{h+1}|,$$

where C_ε is a constant which may depend on ε, but not on n. If the a_n are monotonically decreasing, then the right-hand summation is

$$\sum_{h=0}^{n-1} h(a_h - a_{h+1}) \;=\; \sum_{h=1}^{n} a_h - na_n,$$

[15]Two polynomials are said to be congruent mod 1 if their difference is a polynomial with integer coefficients.

and we find that
$$\left|\sum_{h=0}^{n} a_h e(\varphi(h))\right| \;\le\; C_\varepsilon \;+\; \varepsilon \sum_{h=0}^{n} a_h.$$
Thus this case has been dealt with. If however the a_n increase monotonically, the right-hand summation
$$= \sum_{h=0}^{n-1} h(a_{h+1} - a_h) \;=\; na_n - \sum_{h=1}^{n} a_h \;<\; na_n,$$
from which we obtain
$$\left|\sum_{h=0}^{n} a_h e(\varphi(h))\right| \;\le\; C_\varepsilon \;+\; 2\varepsilon n a_n.$$

I would like to alert the reader's attention to the equation [16]
$$\sum_{h=1}^{n} \frac{e(\varphi(h))}{h} \;=\; o(\lg n).$$
There are also cases where $a_0 + a_1 + a_2 + \ldots$ converges which are important. For example, consider the simplest ϑ-function:
$$\vartheta(v; z) \;=\; \sum_{n=-\infty}^{+\infty} z^{n^2} e^{2\pi i n v} \quad (|z| < 1)$$
for a real argument v (z is usually denoted by q). Let's write $z = re^{2\pi i \alpha}$, where we set $|z| = r$, and retain from the sum only the part in which n runs from 0 to $+\infty$, so the ϑ-series becomes
$$\sum_{n=0}^{\infty} r^{n^2} e(\varphi(n)) \quad \{\varphi(n) = \alpha n^2 + vn\}.$$
We want to investigate the behavior of the ϑ-function, when z converges along a radius to a point on the unit circle, i.e. we want r (keeping α and v fixed) to approach 1 from below. We may write our series as
$$\sum_{n=0}^{\infty} c_n r^n,$$
in which $c_n = 0$, as long as n is not a perfect square, and $c_n = e(\varphi(m))$, if n is a perfect square $= m^2$. Now partial summation yields
$$\sum_{n=0}^{\infty} c_n r^n \;=\; \sum_{n=0}^{\infty} C_n(r^n - r^{n+1}) \;=\; (1-r) \cdot \sum_{n=0}^{\infty} C_n r^n,$$

[16] The original goal of Hardy and Littlewood's research was the obtention of the limiting relation
$$\zeta(1+ti) \;=\; o(\lg t)$$
for the Riemann ζ-function, i.e.
$$\sum_{n=1}^{t} \frac{e^{ti \lg n}}{n} \;=\; o(\lg t). \qquad (*)$$
(Cf. Zur Abschätzung von $\zeta(1+ti)$ *Math. Zeitschrift* (1921) Vol.10 pp. 88-101.) This can be proven with the method used here on $\sum e(\varphi(n))$; as far as I can tell, this statement is not as deep as Theorem 9, as the derivation of (*) does not need to rely on our Criterion (Theorem 1). I will publish my proof later, (perhaps as an addendum to Hardy and Littlewood's work on this topic).

$$C_n = c_0 + c_1 + \ldots + c_n = \sum_{m=0}^{\sqrt{n}} e(\varphi(m)).$$

If α is irrational, we will have

$$C_n = o(\sqrt{n}) \quad \text{or} \quad C_n = o\left(\frac{3 \cdot 5 \cdots (2n+1)}{2 \cdot 4 \cdots 2n}\right).$$

Since

$$\sum_{n=0}^{\infty} \frac{3 \cdot 5 \cdots (2n+1)}{2 \cdot 4 \cdots 2n} r^n = (1-r)^{-\frac{3}{2}}$$

holds, we therefore have

$$\lim_{r \to 1} \left\{ \sqrt{1-r} \cdot \sum_{n=0}^{\infty} c_n r^n \right\} = 0.$$

We thus obtain for the ϑ-function[17]:

$$\vartheta(z; v) = o\left(\frac{1}{\sqrt{1-|z|}}\right).$$

An analogous result holds for the more general function

$$\vartheta(z; v_1, v_2, \ldots, v_{q-1}) = \sum_{n=-\infty}^{+\infty} z^{n^q} e(v_1 n + v_2 n^2 + \ldots + v_{q-1} n^{q-1}).$$

Theorem 11 *Let z approach a point on the unit circle along a constant angle which is incommensurable with 2π. Then*

$$\vartheta(z; v_1, v_2, \ldots, v_{q-1}) = o\left(\frac{1}{\sqrt[q]{1-|z|}}\right)$$

uniformly in the real arguments v.

4 Conclusions for the Uniform Distribution of Points in $\mathfrak{R}_p$

With the help of the criterion upon which our whole investigation is based, we obtain the following consequences of Theorem 9:

Theorem 12 *If $\varphi(z)$ is a polynomial with constant term α_0 and if not all coefficients of $\varphi(z) - \alpha_0$ are rational, then the sequence of numbers*

$$\varphi(1), \varphi(2), \varphi(3), \ldots$$

is uniformly distributed mod 1.

In particular,

[17]cf. G.H. Hardy and J.E. Littlewood, Theorem 7 of the Cambridge lecture. (*Proc. Inter. Math. Camb.* (1912), pp. 223-229. –JM)

Theorem 13 *If ξ is an irrational number, then the sequence of points*

$$1\xi, 4\xi, 9\xi, 16\xi, 25\xi, \ldots,$$

when one winds the real number line around a circle of circumference 1, covers the circumference of this circle uniformly densely. An analogous result naturally holds if we replace the squares with cubes, fourth powers, etc.

More generally,

Theorem 14 *If*

$$\varphi_1(z),\ \varphi_2(z),\ \ldots,\ \varphi_p(z)$$

are any p polynomials such that no integral linear combination of the $\varphi_i(z)$ is congruent to a constant mod 1, then the sequence of points

$$x_1 \equiv \varphi_1(n),\ x_2 \equiv \varphi_2(n),\ \ldots,\ x_p \equiv \varphi_p(n) \quad (\text{mod } 1), \qquad (n = 1, 2, 3, \ldots)$$

lies uniformly densely in $\mathfrak{R}_p$.

Indeed, for any system of integers $m_1, m_2, \ldots, m_p$, which do not all vanish,

$$\varphi(z) = m_1\varphi_1(z) + m_2\varphi_2(z) + \ldots + m_p\varphi_p(z)$$

is a polynomial which satisfies the conditions of Theorem 9.

Theorem 14 contains as special cases:

Theorem 15 *Let ξ be an irrational number, and consider q subintervals $[a_i, b_i]$ $(i = 1, 2, \ldots, q)$ of the unit interval $[0, 1]$. Then the number of the first n integers $h = 1, 2, \ldots, n$ for which the reduced mod 1 numbers $(h^i\xi)$ lie in $[a_i, b_i]$ for all i, $1 \le i \le q$, is asymptotically equal to*

$$(b_1 - a_1)(b_2 - a_2)\ldots(b_q - a_q)n$$

as $n \to \infty$.

Theorem 16 *Let $\xi_1, \xi_2, \ldots, \xi_p$ be any p numbers, among which there is no integral linear relationship, and consider pq arbitrary subintervals $[a_{ij}, b_{ij}]$ of the unit interval. The number of h among $1, \ldots, n$ satisfying all the conditions*

$$\begin{array}{l} a_{11} \le (h\xi_1) \le b_{11},\ a_{21} \le (h^2\xi_1) \le b_{21}, \ldots,\ a_{q1} \le (h^q\xi_1) \le b_{q1}, \\ a_{12} \le (h\xi_2) \le b_{12},\ a_{22} \le (h^2\xi_2) \le b_{22}, \ldots,\ a_{q2} \le (h^q\xi_2) \le b_{q2}, \\ \quad \cdot \quad \cdot \quad \cdot \\ a_{1p} \le (h\xi_p) \le b_{1p},\ a_{2p} \le (h^2\xi_p) \le b_{2p}, \ldots,\ a_{qp} \le (h^q\xi_p) \le b_{qp} \end{array}$$

is asymptotically equal to

$$n \cdot \prod_{i=1}^{q}\prod_{j=1}^{p}(b_{ij} - a_{ij}).$$

In their first paper in Acta Mathematica[18] Hardy and Littlewood showed (by elementary means) that numbers h of the form required in Theorems 15 and 16 actually exist; they require this fact for their proof of Theorem 9. The theorems established here, which we deduced in the reverse order by using Theorem 9 (which we proved in a different but direct way) were merely conjectured by them.

A further interesting special case of Theorem 14 is the following:

Theorem 17 *Let $\varphi(z)$ be a polynomial in which z^l is the highest power which has an irrational coefficient; we want to know when and how often it happens that some l consecutive terms of the sequence*

$$\varphi(1), \varphi(2), \varphi(3), \ldots$$

mod 1 belong to l given intervals $[a_i, b_i]$. It turns out that the relative frequency of such an event is equal to the a priori probability

$$\prod_{i=1}^{l}(b_i - a_i).$$

One must check whether there can be l integers $m_0, m_1, \ldots, m_{l-1}$ so that

$$\sum_{r=0}^{l-1} m_r \varphi(z+r)$$

is a polynomial all of whose coefficients, except for the constant term, are rational numbers. When checking this condition, one can replace $\varphi(z)$ by a truncated polynomial obtained from φ by removing all terms of power greater than that of z^l. If the truncated $\varphi(z)$ is

$$\alpha z^l + \alpha_1 z^{l-1} + \ldots + \alpha_l \ , \qquad (\alpha \text{ irrational}),$$

then

$$\varphi(z+r) \;=\; \alpha z^l + \left[\alpha\binom{l}{1}r + \alpha_1\right] z^{l-1} \;+\; \left[\alpha\binom{l}{2}r^2 + \alpha_1\binom{l-1}{1}r + \alpha_2\right] z^{l-2} \;+\; \ldots$$

If the numbers m_r are to have the required property, then they must satisfy the equations

$$\begin{aligned}
\sum_{r=0}^{l-1} m_r &= 0,\\
\sum_{r=0}^{l-1} r m_r &= 0,\\
\ldots\quad &\quad \ldots\\
\sum_{r=0}^{l-1} r^{l-1} m_r &= 0,
\end{aligned}$$

[18]Some Problems of Diophantine Approximation I: The Fractional Part of $n^k\theta$. *Acta Math.*, 1914, Vol 37, pp. 155-191.

which gives

$$m_0 = m_1 = \ldots = m_{l-1} = 0.$$

We also immediately realize that the theorem can no longer be valid for more than l consecutive terms of the sequence $\varphi(n)$.

5 The Exceptional Cases

Let once again

$$\varphi_1(z),\ \varphi_2(z),\ \ldots,\ \varphi_p(z)$$

be any p polynomials with real coefficients without constant term.[19] I now assume that there exist integers l_i so that

$$l_1\varphi_1(z)\ +\ l_2(z)\ +\ \ldots\ +\ l_p\varphi_p(z)$$

is a polynomial with nothing but rational coefficients. I will denote a system of integers $(l_1, l_2, \ldots, l_p)$ which satisfies this condition by an $\mathfrak{l}$-point. Such $\mathfrak{l}$-points form a *lattice*, i.e. for each $\mathfrak{l}$-point we have a $(-\mathfrak{l})$-point, and the sum $\mathfrak{l}' + \mathfrak{l}''$ is also such a point.[20] According to a well-known procedure, which was described by Minkowski[21], one can define $q \leq p$ $\mathfrak{l}$-points

$$\mathfrak{l}\ =\ (l_{k1}, l_{k2}, \ldots, l_{kp}), \quad (k = 1, 2, \ldots, q)$$

in such a way that among them there is no relation $\sum_{k=1}^{q} y_k \mathfrak{l}_k = 0$ with arbitrary, not all vanishing coefficients y_k, but where all $\mathfrak{l}$-points can be obtained of the $\mathfrak{l}_k$ with *integer* coefficients:

$$\mathfrak{l}\ =\ \sum_{k=1}^{q} h_k \mathfrak{l}_k.$$

The points $\mathfrak{x}\ =\ (x_1, x_2, \ldots, x_p)$ which by means of arbitrary *real* coefficients y_k can be represented in the form

$$\mathfrak{x}\ =\ \sum_{k=1}^{q} y_k \mathfrak{l}_k$$

constitute a q-dimensional linear manifold $\mathfrak{Y}$ in the usual p-dimensional space. By a lattice point I mean any point $\mathfrak{x}$ with integer coordinates x_i; I then claim that the $\mathfrak{l}$-points constitute all of the lattice points lying on the linear manifold $\mathfrak{Y}$. Since no linear relation holds between the points $\mathfrak{l}_k$, there is in the following matrix

$$\begin{matrix} l_{11}, & l_{12}, & \ldots, & l_{1p} \\ \ldots & \ldots & \ldots & \ldots \\ l_{q1}, & l_{q2}, & \ldots, & l_{qp} \end{matrix}$$

[19] The last assumption is for convenience only.

[20] The way in which points are added, and multiplied by scalars, hardly needs explanation.

[21] *Diophantische Approximationen* (Leipzig 1907), sec. 14; the procedure is more succinctly described, e.g. on page 78 of my book *Die Idee der Riemannschen Fläche* (Leipzig 1913). (In German this procedure is called 'Adaption eines Zahlgitters an ein enthaltenes', meaning something like 'reduction of a number lattice to one contained in it'. (JM))

a nonzero determinant of order q; say

$$L = \|l_{ik}\|_{i=1,2,\dots,q,\ k=1,2,\dots,q}.$$

From the first q of the equations

$$x_i \;=\; \sum_{k=1}^{q} y_k l_{ki} \quad (i = 1, 2, \dots, p),$$

we conclude that if $\mathfrak{x} = (x_i)$ is a lattice point lying on $\mathfrak{Y}$, and if we solve the system with respect to y, the y_k are fractions with denominator L. Therefore $L\mathfrak{x}$ is an $\mathfrak{l}$-point, but hence $\mathfrak{x}$ is also (going back to the definition of $\mathfrak{l}$-points), and thus in accordance with the choice of the points $\mathfrak{l}_k$ the coefficients y_k must have been integers. As all lattice points lying in $\mathfrak{Y}$ can be represented in the form $\sum_{k=1}^{q} y_k \mathfrak{l}_k$ with integers y_k, one finds, by continuing this procedure, that one can append still $r = p - q$ more

$$\mathfrak{m}_h \;=\; (m_{h1}, m_{h2}, \dots, m_{hp}) \quad (h = 1, 2, \dots, r)$$

so that every lattice point $\mathfrak{x}$ allows a representation of the form

$$\mathfrak{x} \;=\; (y_1\mathfrak{l}_1 + \dots + y_q\mathfrak{l}_q) \;+\; (z_1\mathfrak{m}_1 + \dots + z_r\mathfrak{m}_r) \tag{12}$$

with integer coefficients y_k, z_h. Hence equation (12) or

$$x_i \;=\; (l_{1i}y_1 + \dots + l_{qi}y_q) \;+\; (m_{1i}z_1 + .. + z_{ri}z_r)$$

yields an integral, unimodular linear transformation of the coordinates x into the new coordinates y, z, under which the system of lattice points is preserved. The same also holds therefore for the contragradient substitution

$$y_k \;=\; \sum_{i=1}^{p} l_{ki}x_i \quad (k = 1, 2, \dots, q) \tag{13}$$
$$z_h \;=\; \sum_{i=1}^{p} m_{hi}x_i \quad (h = 1, 2, \dots, r)$$

These formulas thus represent a linear transformation of $\mathfrak{R}_p$.

In particular, we form the polynomials

$$f_k(z) \;=\; \sum_{i=1}^{p} l_{ki}\varphi_i(z) \quad (k = 1, 2, \dots, q),$$
$$\psi_h(z) \;=\; \sum_{i=1}^{p} m_{hi}\varphi_i(z) \quad (h = 1, 2, \dots, r).$$

The first q have nothing but rational coefficients, whose common denominator will be denoted by G. Among the last r however, there is no combination $\sum_{h=1}^{r} t_h\psi_h(z)$ with integers $t_1, t_2, \dots, t_r$ which is a polynomial with rational coefficients, other than the trivial case $t_h = 0$ for all h. We have the task of studying the points in $\mathfrak{R}_p$ defined by

$$x_i \;\equiv\; \varphi_i(n) \pmod 1 \quad (i = 1, 2, \dots, p),$$

as n runs through the positive integers. By means of our integral unimodular transformation, the problem is reduced to studying the points (y, z):

$$y_k \equiv f_k(n) \qquad z_h \equiv \psi_h(n) \qquad (\text{mod } 1)$$
$$(k = 1, 2, \ldots, q), \qquad (h = 1, 2, \ldots, r).$$

For every integer n, define an *r-dimensional linear manifold* $\mathfrak{C}_n$ in $\mathfrak{R}_p$ by the congruences

$$y_1 \equiv f_1(n),\ y_2 \equiv f_2(n),\ \ldots,\ y_q \equiv f_q(n) \quad (\text{mod } 1)$$

(passing to the y, z — coordinates). The $\mathfrak{C}_n$'s are all mutually parallel because they're all parallel to the r-dimensional manifold

$$y_1 \equiv 0,\ y_2 \equiv 0,\ \ldots,\ y_q \equiv 0 \quad (\text{mod } 1).$$

Moreover, $\mathfrak{C}_n$ and $\mathfrak{C}_m$ are identical if the numbers n and m are congruent mod G, so that we may restrict our attention to a single period

$$\mathfrak{C}_1,\ \mathfrak{C}_2,\ \ldots,\ \mathfrak{C}_G. \tag{14}$$

Among the $\mathfrak{C}_n$ there are therefore only finitely many distinct ones: $\mathfrak{C}'$, $\mathfrak{C}''$, …; let these occur in a single period (by (14)) m', m'', … times respectively, $(m' + m'' + \ldots = G)$. We claim that the points

$$y_k \equiv f_k(n) \qquad z_h \equiv \psi_h(n) \qquad (\text{mod } 1) \tag{15}$$
$$(k = 1, 2, \ldots, q), \qquad (h = 1, 2, \ldots, r)$$

are distributed in $\mathfrak{R}_p$ (as n runs through the positive integers) with completely uniform density among the mutually parallel r-dimensional linear manifolds $\mathfrak{C}_1,\ \mathfrak{C}_2,\ \ldots,\ \mathfrak{C}_G$. It is to be understood that the density on any single manifold is uniform, and on any two distinct manifolds, say $\mathfrak{C}'$, $\mathfrak{C}''$, the respective densities are proportional to the multiplicities m', m''.

The fact that the points in (15) only appear on the $\mathfrak{C}_n$'s is self-evident. In order to prove the claim of uniformly dense distribution, we repeat our procedure from the previous paragraphs, in which we show that for every bounded function

$$F(y_1, y_2, \ldots, y_q; z_1, z_2, \ldots, z_r)$$

defined on $\mathfrak{C}_n$ which is Riemann-integrable with respect to the variables z_n, and of period 1 in all variables, the following limit equation holds:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} F(y_k = f_k(m), z_h = \psi_h(m)) \tag{16}$$
$$= \frac{1}{G} \sum_{n=1}^{G} \int_0^1 \int_0^1 \cdots \int_0^1 F(y_k = f_k(n); z_1, z_2, \ldots, z_r) dz_1 dz_2 \ldots dz_r.$$

It suffices to show this for

$$F = \Phi(y_1, y_2, \ldots, y_q) \cdot e(t_1 z_1 + \ldots + t_r z_r)$$

where the t_h are arbitrary integers and Φ a function only defined for the G[22] systems of values (y) which repeat periodically according to the formula $y_k = f_k(n) \bmod 1$, as n runs through the positive integers. If one of the positive integers t_h is not equal to 0, then by Theorem 9, the left-hand side of the claimed equation (16) evaluates to zero. This holds for the right-hand side as well, as every term of the G-term summation there $= 0$. If on the other hand all $t_h = 0$, then F is a function of the y's only; for such functions, the validity of our equation is immediately obvious.

The formulation of the result can be made independently of the linear transformation used in the proof.

Theorem 18 *Given any p polynomials $\varphi_1(z), \ldots, \varphi_p(z)$ with real coefficients, let α_i be the constant term in $\varphi_i(z)$. By an '$\mathfrak{l}$-point' we mean any system of integers $\mathfrak{l} = (l_1, l_2, \ldots, l_p)$ for which the linear combination*

$$l_1\{\varphi_1(z) - \alpha_1\} + l_2\{\varphi_2(z) - \alpha_2\} + \ldots + l_p\{\varphi_p(z) - \alpha_p\} = f_{\mathfrak{l}}(z)$$

becomes a polynomial with rational coefficients. The coefficients of the polynomials corresponding to the different $\mathfrak{l}$-points have common denominator G. For any integer n, the congruences which hold simultaneously for all $\mathfrak{l}$-points

$$l_1(x_1 - \alpha_1) + \ldots + l_p(x_p - \alpha_p) \equiv f_{\mathfrak{l}}(n) \quad (mod\ 1)$$

define a linear r-dimensional manifold $\mathfrak{C}_n$ in $\mathfrak{R}_p$. The $\mathfrak{C}_n$ $(n = 1, 2, 3, \ldots)$ are mutually parallel and repeat in a cycle of period G; among them, let the distinct ones be $\mathfrak{C}', \mathfrak{C}'', \ldots$ (there are only finitely many of them); let them appear in the cycle with multiplicities $m', m'', \ldots$ respectively. As n runs through the positive integers, the points in $\mathfrak{R}_p$ defined by

$$x_1 \equiv \varphi_1(n), x_2 \equiv \varphi_2(n), \ldots, x_p \equiv \varphi_p(n) \quad (mod\ 1)$$

are distributed among the finitely many linear manifolds $\mathfrak{C}', \mathfrak{C}'', \ldots$ so that each one is not only densely but indeed uniformly densely covered; for two distinct $\mathfrak{C}$'s, the densities are proportional to their multiplicities m.

Armed with this theorem, one is in position to state how Theorems 15 and 16 would be modified if there existed integral linear relationships between the ξ_i's, as well as an analogue of Theorem 17[23] in the case where more than l consecutive $\varphi(n)$'s are considered.

6 Extension to Several Parameters

Up until this point we have examined polynomial functions of a single variable z, and have let z run through the positive integers. We may however generalize our investigation to the case where

[22] In the original text there is a typo and there is a C where I assume G is meant. –J.M.

[23] in the original text this is incorrectly referred to as Theorem 16. –J.M.

we use several variables instead of only z. I shall restrict myself to the case of two variables u, v. Given once again p polynomial functions

$$\varphi_1(u,v), \ldots, \varphi_p(u,v),$$

we wish to make the simplifying assumption that no integral linear combination of the above yields a polynomial which is congruent to a constant mod 1. In the (u,v)-plane, let $\mathfrak{K}$ be a bounded region (a closed, bounded point set with well-defined area $J > 0$). Dilate $\mathfrak{K}$ from the origin by a ratio of $t : 1$ into a surface piece $t\mathfrak{K}$ — where t should be a large real number so that the number — n_t of lattice points (points with integer coordinates u, v) lying in $t\mathfrak{K}$ is asymptotically given by Jt^2 as $t \to \infty$. To each one of these lattice points (u,v) associate in $\mathfrak{R}_p$ the point

$$x_1 \equiv \varphi_1(u,v),\ x_2 \equiv \varphi_2(u,v),\ \ldots,\ x_p \equiv \varphi_p(u,v) \pmod 1. \tag{17}$$

In $\mathfrak{R}_p$, take an arbitrary volume V, so that our claim of uniform distribution now reads as follows:

Theorem 19 *Among the n_t points (17) which one obtains when one evaluates at all the lattice points (u,v) of $t\mathfrak{K}$, single out those which in $\mathfrak{R}_p$ belong to the region V; in the limit as $t \to \infty$, the ratio of the number of these $n_t^{(V)}$ to n_t is $V : 1$.*

Should the condition disallowing an integral linear relationship between the functions $\varphi_i(u,v)$ not be satisfied, the assertion is modified in the same manner as in the previous paragraphs. If the φ_i's are linear functions of u and v, we have thus sharpened the *most general* approximation theorem of the type (due to Kronecker) — the proposition "everywhere dense" is replaced by "everywhere uniformly dense"[24]. The proof will be completed once we show

Theorem 20 *Let $\varphi(u,v)$ be a polynomial function of u and v, whose coefficients, except for the constant term, are not all rational, and form the sum*

$$\sigma_t = \sum e(\varphi(u,v))$$

which extends over all n_t lattice points (u,v) which belong to the region $t\mathfrak{K}$; it is then the case that

$$\lim_{t\to\infty} \frac{\sigma_t}{n_t} = 0.$$

An important case of theorems of this type arises for example when one takes polynomial functions $\Phi(z)$ with arbitrary *complex* coefficients, and lets the argument z take values in the Gaussian integers; regarding the distribution of the values of this function for these given arguments, one stipulates that two complex numbers (values of the function Φ) which differ by by a Gaussian integer are to be considered the same. One would then restrict z by the condition $|z| \leq t$, and then let the positive number t grow arbitrarily large.

One could also impose some other more general conditions, instead of the assumption made here, that the region $t\mathfrak{K}$, which grows infinite in all directions, little by little encompassing all the

[24] Compare Werke III 1, pp. 104–105.

lattice points, grows *homothetically*; I will however restrict myself to carrying out the proof for a region growing according to the law of homothecy. By considering an exhaustion[25] of $\mathfrak{K}$, it suffices to assume $\mathfrak{K}$ to be a rectangle

$$a_1 \le u \le a_2, \quad b_1 \le v \le b_2.$$

The number of integers u which satisfy the inequality

$$a_1 t \le u \le a_2 t$$

is $= m_t \{\sim (a_2 - a_1)t\}$, the number of integers v for which

$$b_1 t \le v \le b_2 t$$

holds amounts to $n_t\{\sim (b_2 - b_1)t\}$; therefore in $t\mathfrak{K}$ there lie $m_t n_t$ lattice points (u, v).

We let $\varphi(u, v)$ be ordered according to decreasing powers of v, and the coefficients of these powers of v, which are polynomials in u, shall be ordered by decreasing powers of u. Assuming that in this ordering the highest member possesses as coefficient an *irrational* number entails no restriction. We shall write

$$\varphi(u, v) \;=\; v^q \psi(u) \;+\; v^{q-1}\psi_1(u) \;+\; \ldots.$$

In order not to regress to the case of one variable we will need to take $q \ge 1$. To

$$\sigma_t \;=\; \sum_{a_1 t \le u \le a_2 t} \left\{ \sum_{b_1 t \le v \le b_2 t} e(\varphi(u, v)) \right\}$$

we apply the Schwarz inequality, when $q > 1$:

$$|\sigma_t|^2 \;\le\; m_t \sum_u \left| \sum_v e(\varphi(u, v)) \right|^2$$

and arrange $|\sum_v|^2$ by the same method as in section 3. We therefore obtain, using the notation introduced there,

$$|\sigma_t|^Q \;\le\; m_t^{Q-1} N_t \cdot \sum_{u; r_1, \ldots, r_{q-1}} \left\{ e(\rho) \sum_h e(q!\; \psi(u)\; r_1 r_2 \ldots r_{q-1} h) \right\}, \tag{18}$$

$$N_t \sim \kappa \cdot n_t^{Q-q}.$$

The region of the outer summation is given by

$$a_1 t \le u \le a_2 t, \quad |\mathfrak{r}| \;=\; |r_1| \;+\; |r_2| \;+\; \ldots \;+\; |r_{q-1}| \;\le\; n_t - 1. \tag{19}$$

The inner summation extends over a connected interval of $n_t - |\mathfrak{r}|$ integers. We need now show that

$$\lim_{t \to \infty} \frac{\sigma_t}{m_t n_t} \;=\; 0.$$

[25] Weyl seems to be suggesting that we can approximate our region by disjoint rectangles, prove the result for these, and disregard the error. He seems here unusually imprecise in that he doesn't elaborate on the details involved in passing to the limit in such an 'Exhaustion'. (JM)

If $\psi(u)$ does not actually contain the variable u, and thus is simply an irrational constant α, then the conclusion follows as for one variable. If however ψ contains the variable u, we can rely on the following lemma:
Among all the $m_t n_t^{(q-1)}$ lattice points $(u; r_1, r_2, \ldots, r_{q-1})$ which satisfy the conditions (19), consider those for which the quantity

$$r_1 r_2 \cdots r_{q-1} \cdot \varphi(u)$$

mod 1 lies between $-\varepsilon$ and $+\varepsilon$; the ratio of the number of these to $m_t n_t^{(q-1)}$ as $t \to \infty$ is $2\varepsilon : 1$. Here $\varphi(u)$ is any polynomial in u whose highest coefficient is irrational.

This lemma must be brought to bear on the polynomial $\varphi(u) = q!\psi(u)$. This can be achieved through an inductive process. We know that it holds for $q = 1$, by Theorem 9. Assuming that it holds for q, we'll show that it holds for $q + 1$ as well. It suffices to prove that

$$\lim_{t\to\infty} \frac{1}{m_t n_t^{(q)}} \sum_{u; r_1, \ldots, r_q} e(r_1 r_2 \cdots r_q \; \varphi(u)) \; = \; 0 \tag{20}$$

holds, when the sum is taken over

$$a_1 t \; \leq \; u \; \leq \; a_2 t, \quad |r_1| \; + \; |r_2| \; + \; \ldots \; + \; |r_q| \; \leq \; n_t - 1.$$

For fixed $u; r_1, r_2, \ldots, r_{q-1}$ we extend the sum to r_q and and employ, as earlier, for this simple sum a double estimate, depending on whether $r_1 r_2 \ldots r_{q-1} \varphi(u)$ mod 1 lies between $-\varepsilon$ and $+\varepsilon$ or not. By realizing the fact that the number of systems of values $(u; r_1, r_2, \ldots, r_{q-1})$ satisfying the first property is asymptotically $2\varepsilon m_t n_t^{(q-1)}$, we obtain the limit equation (20) in a fashion similar to the one in the proof of of the lemma in section 3.

The *two* variables u, v can be replaced by three or more. Our lemma, on which is based an essential part of our investigation, is a special case of the generalized Theorem 19.

7 On the Uniform Distribution of Arbitrary Numerical Sequences

Hardy and Littlewood asked the general question: When does a sequence of increasing integers

$$l_1 \; < \; l_2 \; < \; l_3 \; < \; \ldots$$

possess the property that for every irrational number ξ, the sequence of values

$$l_1\xi, \; l_2\xi, \; l_3\xi, \; \ldots \tag{21}$$

mod 1 comes arbitrarily close to any value?[26] We could also ask: when is the sequence everywhere uniformly dense? A thorough answer is not forthcoming. We may however claim that for any such given sequence l_n the sequence (21) is uniformly densely distributed everywhere for every value of ξ, if we disregard ceratin values of ξ which form a set of measure 0. This exceptional set naturally contains all rational numbers, however it is not known whether there are any further elements.

[26]G.H. Hardy and J.E. Littlewood, Acta Math. *37*, p.156.

Although I believe that such theorems (in which an undetermined exceptional set of measure 0 appears) should not be too highly esteemed, I would still like to briefly justify this claim here. My proof is based on the following lemma of integral calculus[27].

If $f_n(x)$ are continuous functions on the interval $[0,1]$, for which the sum of the integrals

$$\sum_{n=1}^{\infty} \int_0^1 |f_n(x)|^2 dx$$

converges, then for all x except perhaps a set of measure 0, $f_n(x)$ converges to 0 as n gets arbitrarily large.

Proof. Let ε be an arbitrarily small positive number, and $\mathfrak{A}_n$ the subset of $[0,1]$ in which $|f_n(x)| \geq \varepsilon$; thus if A_n is the Lebesgue measure of these sets $\mathfrak{A}_n$, the inequality

$$\varepsilon^2 A_n \ \leq \ \int_0^1 |f_n(x)|^2 dx$$

holds.

Form the union set[28]

$$\mathfrak{C}_n \ = \ \mathfrak{A}_{n+1} \ + \ \mathfrak{A}_{n+2} \ + \ \ldots$$

so that the points x belonging to $\mathfrak{C}_n$ are those x for which at least one of the inequalities

$$|f_{n+1}(x)|, \ |f_{n+2}(x)|, \ \ldots \ \geq \ \varepsilon$$

holds. The measure C_n of $\mathfrak{C}_n$ satisfies the inequality

$$C_n \ \leq \ \frac{1}{\varepsilon^2} \cdot \sum_{\nu=n+1}^{\infty} |f_\nu(x)|^2 dx \ = \ \frac{J_n}{\varepsilon^2}. \tag{22}$$

The sets $\mathfrak{C}_1, \mathfrak{C}_2, \mathfrak{C}_3, \ldots$ have the property that each one contains the next in the sequence; their intersection

$$\lim_{n\to\infty} \mathfrak{C}_n \ = \ \mathfrak{C} \text{ [29]}$$

has, since

$$\lim_{n\to\infty} J_n \ = \ 0,$$

measure 0, by (22). For every point x of its complement there is an index after which all the $|f_n(x)|$'s are less than ε. For such x we therefore have that

$$\limsup_{n\to\infty} |f_n(x)| \ \leq \ \varepsilon.$$

And now if one lets ε take the values $1/2, 1/3, 1/4, \ldots$, say, the proof of the lemma is complete.

[27] On the strength of this very lemma I previously proved the so-called Riesz-Fischer Theorem. Math. Ann. *67*, p.243f. (1909).

[28] Of course in modern notation this would read $\mathfrak{C}_n = \bigcup_{k=n+1}^{\infty} \mathfrak{A}_k$. (J.M. 2005)

[29] again nowadays this would be written $\bigcap_{n=1}^{\infty} \mathfrak{C}_n = \mathfrak{C}$.

We now investigate the functions

$$f_n(x) \ = \ \frac{1}{n}\sum_{h=1}^{n} e(l_h x),$$

for which we have

$$\int_0^1 |f_n(x)|^2 dx \ = \ \frac{1}{n^2}\sum_{h,k=1}^{n}\int_0^1 e(l_h x - l_k x)dx \ = \ \frac{1}{n}.$$

A direct application of our lemma is not possible due to the divergence of the harmonic series $\sum 1/n$. However, if we choose from the integers the perfect squares, we obtain

$$\lim_{n\to\infty} f_{n^2}(x) \ = \ 0$$

for all x except a set of measure 0. Let n be an arbitrary integer, and then define the integer ν through the condition

$$\nu^2 \ \le \ n \ < \ (\nu+1)^2.$$

Thus we have

$$\left|n f_n(x) \ - \ \nu^2 f_{\nu^2}(x)\right| \ \le \ 2\nu, \qquad \left|f_n(x) \ - \ \frac{\nu^2}{n} f_{\nu^2}(x)\right| \ \le \ \frac{2}{\sqrt{n}},$$

therefore we also have

$$\lim_{n\to\infty} f_n(x) \ = \ 0$$

for all x not belonging to $\mathfrak{A}$. Neither x, nor $2x$, nor $3x$, … belongs to the set $\mathfrak{A}$ (which repeats with period 1), thus

$$\lim_{n\to\infty} f_n(mx) \ = \ 0$$

for every integer $m \neq 0$. For such a number x the values

$$l_1 x, \ l_2 x, \ l_3 x, \ \ldots$$

mod 1 satisfy the law of uniformly dense distribution.

With regard to the integers l_n, let's only stipulate that

$$l_1 \ \le \ l_2 \ \le \ l_3 \ \le \ \ldots$$

whereby

$$\int_0^1 |f_n(x)|^2 dx \ = \ \frac{h_1^2 + h_2^2 + \ldots + h_m^2}{n^2},$$

if among the $l_1, \ldots, l_n$ the first h_1 of them are equal, and then the following h_2 are, and so forth, and finally the final h_m agree. Denoting by $h^{(n)}$ the largest of the numbers $h_1, h_2, \ldots, h_m$, we get

$$h_1^2 + h_2^2 + \ldots + h_m^2 \ \le \ h^{(n)} \ (h_1 + h_2 + \ldots + h_m) \ = \ h^{(n)} \cdot n,$$

hence

$$\int_0^1 |f_n(x)|^2 dx \ \le \ \frac{h^{(n)}}{n}.$$

I shall now replace the earlier assumption $h^{(n)} = 1$ by the much more general condition that two positive numbers ε and c should exist, so that

$$h^{(n)} \;\leq\; \frac{c \cdot n}{(\lg n)^{1+\varepsilon}}$$

holds. Then our above method becomes applicable. We select from the sequence of positive integers the following:

$$n_\nu \;=\; \left[e^{\nu^{1-b}}\right], \quad b \;=\; \frac{\varepsilon}{2+\varepsilon}, \quad (\nu = 1, 2, 3, \ldots)$$

and we obtain

$$\int_0^1 |f_{n_\nu}(x)|^2 dx \;\leq\; \frac{c}{\nu^{1+b}}.$$

Consequently, except for certain values of x in the interval $[0, 1]$ which form a set $\mathfrak{A}$ of measure 0,

$$\lim_{\nu \to \infty} f_{n_\nu}(x) \;=\; 0.$$

Let n be an arbitrary integer, and define ν through the condition

$$n_\nu \;\leq\; n \;<\; n_{\nu+1}$$

whence we find

$$\left| f_n(x) \;-\; \frac{n_\nu}{n} f_{n_\nu}(x) \right| \;\leq\; \frac{n_{\nu+1} - n_\nu}{n} \;\leq\; \frac{n_{\nu+1}}{n_\nu} \;-\; 1.$$

Since

$$(\nu+1)^{1-b} \;-\; \nu^{1-b} \;<\; \nu^{-b} \quad (0 < b < 1)$$

holds, the following

$$\frac{n_{\nu+1}}{n_\nu} \;-\; 1 \;<\; e^{\nu^{-b}} \;-\; 1, \quad \lim_{\nu \to \infty} \left\{ \frac{n_{\nu+1}}{n_\nu} \;-\; 1 \right\} \;=\; 0$$

is valid. So everywhere but in $\mathfrak{A}$,

$$\lim_{n \to \infty} f_n(x) \;=\; 0.$$

In the present case the law of uniform distribution thus holds for "almost all" x.

If we replace x by x/m (m a positive integer), we get to the case in which the l_n are fractions with a finite common denominator m.

Finally let

$$\lambda_1 \;\leq\; \lambda_2 \;\leq\; \lambda_3 \;\leq\; \ldots$$

be any sequence of real numbers increasing to infinity; we assume that the growth of λ is not too slow as a function of the index, specifically that two positive numbers ε and c of the following type exist: if the index grows from n by $n/(\lg n)^{1+\varepsilon}$, then λ must grow by at least c. We will restrict the real variable x to a finite interval $|x| \leq A$. We take an arbitrary positive integer m and assign to every λ_n a fraction l_n with denominator m, which differs from λ_n by at most $1/(2m)$. Among the n first numbers l_n there can certainly not be a set of more than $n/(\lg n)^{1+\varepsilon}$ which are all equal (as long as $m > 1/c$). Consequently, for all x up to a set $\mathfrak{A}_m$ of measure 0:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{h=1}^{n} e(l_h x) \;=\; 0.$$

Since

$$|e(x_1) - e(x_2)| \;\leq\; 2\pi\, |x_1 - x_2|$$

holds, we have that

$$\left| \frac{1}{n} \sum_{h=1}^{n} e(\lambda_h x) \;-\; \frac{1}{n} \sum_{h=1}^{n} e(l_h x) \right| \;\leq\; \frac{A\pi}{m},$$

and so for all x not belonging to the set $\mathfrak{A}_m$,

$$\limsup_{n\to\infty} \frac{1}{n} \left| \sum_{h=1}^{n} e(\lambda_h x) \right| \;\leq\; \frac{A\pi}{m}.$$

Form the set $\mathfrak{A}$ by including the points x in the interval $[-A, A]$ which are in all but finitely many of the sets $\mathfrak{A}_m$ $(m = 1, 2, 3, \ldots)$; $\mathfrak{A}$ then has measure 0, and for all values x not in $\mathfrak{A}$, we have the limit equation

$$\lim_{n\to\infty} \frac{1}{n} \sum_{h=1}^{n} e(\lambda_h x) \;=\; 0.$$

We have thus proven:

Theorem 21 *Let $\lambda_1, \lambda_2, \lambda_3, \ldots$ be any sequence of real numbers, where to the function λ of the indices, we can associate two positive constants ε and c such that whenever the index increases from n by more than $n/(\lg n)^{1+\varepsilon}$, λ will have increased by at least c. If x is then a real number not belonging to a certain set of measure 0, then the sequence of values*

$$\lambda_1 x,\;\; \lambda_2 x,\;\; \lambda_3 x,\;\; \ldots$$

mod 1 is uniformly densely distributed[30].

8 Appendix. On Closed Euclidean Spaces

By a closed Euclidean space we mean a closed p-dimensional manifold[31] with the property that Euclidean geometry is valid in the neighborhood of every point. In the usual p-dimensional space,

[30] For the case $\lambda_n = a^n$ (a a positive integer) this was proven in a sharper form by G.H. Hardy and J.E. Littlewood, Acta Math. *37*, p.183 ff. The general question was developed independently of my work by Towler, a student of Hardy's and Littlewood's, in the Londoner Proceedings.

[31] Relative to the analysis-situs (topological -JM) idea used in this appendix, I refer the reader to my book *Die Idee der Riemannschen Fläche*, (namely sections 4 and 9). [Nowadays one should refer to the polished third edition from 1955, especially sections 4 and 8, instead of the first edition of 1913 (addendum 1955).] Relative to the connection between the problem of non-Euclidean space shapes with Riemannian function theory see P. Koebe, Ann. Mat. [3], *21* (Lagrange-Festband), p.57 ff.

in which any system of real numbers $(x_1, x_2, \dots x_p)$ denotes a point, let Γ be a discrete group of Euclidean motions; suppose Γ possesses a finite fundamental domain and contains no pure rotations. If one groups together all points of the usual space which are equivalent with respect to Γ into a single "point" of a new manifold $\mathfrak{K}_\Gamma$, then $\mathfrak{K}_\Gamma$ is, in our sense of the term, manifestly a closed Euclidean space; I shall call this a *Crystal* (relative to the group Γ). In this appendix we wish to prove that

Theorem 22 *Every closed Euclidean space is a crystal.*

According to a general theorem proven by Bieberbach[32], every group of motions Γ of the above form contains p independent translations, from which any translation in Γ can be composed. Through an affine transformation of the coordinates $(x_1, x_2, \dots, x_p)$ we may take these translations to be

$$(1, 0, \dots, 0);\ (0, 1, \dots, 0); \dots ; (0, 0, \dots, 1)$$

so that the points x that coincide in $\mathfrak{R}_p$ are also identical in $\mathfrak{K}_\Gamma$: in $\mathfrak{R}_p$, Γ becomes a group consisting of finitely many integral linear unimodular transformations[33], call it Γ_0, and one can produce $\mathfrak{K}_\Gamma$ from $\mathfrak{R}_p$ if one collapses all the points which are equivalent relative to Γ_0. This is the claim which we made in section 2.

To recognize that every closed Euclidean space $\mathfrak{R}$ is a crystal, we proceed as follows: if $\mathfrak{p}$ is any point of $\mathfrak{R}$, then for small enough r the interior of the sphere about $\mathfrak{p}$ of radius r (i.e. the totality of the points which can be linked to $\mathfrak{p}$ by a straight line of length $< r$) is a neighborhood of $\mathfrak{p}$ which can be mapped bijectively and congruently to the interior of a sphere of radius r in the usual Euclidean space. This will not be the case for arbitrarily large r, as $\mathfrak{R}$ is a closed Euclidean space. There is, however, a certain positive number $r(\mathfrak{p})$, which separates the r's for which this map is possible $(r \leq r(\mathfrak{p}))$ from those for which that is not the case. $r(\mathfrak{p})$ is a continuous function of $\mathfrak{p}$: as long as a point $\mathfrak{q}$ lies close enough to $\mathfrak{p}$, the quantity $|r(\mathfrak{q}) - r(\mathfrak{p})|$ is at most as large as the distance between the points $\mathfrak{p}, \mathfrak{q}$. Therefore $r(\mathfrak{p})$ has (being a continuous, positive function on a closed manifold) a *positive* minimum r_0. Therefore we have shown: in the interior of a ball of radius r_0 in $\mathfrak{R}$, Euclidean geometry is always valid, wherever the center of the ball may lie. Another consequence is that each of the halves into which a straight line is divided in $\mathfrak{R}$ has infinite length (which of course does not exclude the case that it may be a closed curve that is being traversed infinitely often). For, having traversed the line up to some point, the line will have to continue from there by a segment of length at least r_0 in the given direction.

We now choose a fixed point $\mathfrak{p}_0$ in $\mathfrak{R}$ and a system of rectangular coordinates at $\mathfrak{p}_0$. If γ is an arbitrary curve originating from $\mathfrak{p}_0$ and ending at $\mathfrak{p}$ in $\mathfrak{R}$, then we can translate the coordinate-frame in its origin along γ so that the axes always remain parallel. Let $x_1, x_2, \dots, x_p$ be the components of the total translation, which the origin undergoes when moving from $\mathfrak{p}_0$ to $\mathfrak{p}$, its "translation

[32] *Über die Bewegungsgruppen der Euklidischen Räume*, Math. Ann. *70*, p. 333 (1911). See also G. Frobenius, Ber. Kgl. Preuss. Akad. Wiss. *1911*, p. 663.

[33] The group of translations T contained in Γ is an invariant subgroup of Γ. In the notation of group theory, one would write $\Gamma_0 = \Gamma : T$.

components along γ". (They by no means need to be =0 when the curve γ returns to its initial point.) I claim: every curve γ originating at $\mathfrak{p}_0$ defines a "point" $\overline{\mathfrak{p}}$ of the new manifold $\overline{\mathfrak{R}}$, which lies "over the endpoint $\mathfrak{p}$ of γ"; two curves γ, γ' beginning at $\mathfrak{p}$ and ending at $\mathfrak{p}_0$ shall only define the same point $\overline{\mathfrak{p}}$ "over" $\mathfrak{p}$ when the translation components have the same values along γ as they have along γ'. So let $\overline{\mathfrak{p}}$ be a point defined by γ and lying above $\mathfrak{p}$, and K the interior of a ball centered at $\mathfrak{p}$, whose radius is $\leq r_0$; append to γ all possible curves γ_1 leaving $\mathfrak{p}$ and running through K; the points defined by $\gamma + \gamma_1$ for all such curves γ_1 form a "neighborhood" $\overline{K}$ of $\overline{\mathfrak{p}}$. For such an assignment, for every point of K there exists one and only one point of $\overline{K}$, so we have thus constructed an unramified covering space $\overline{\mathfrak{R}}$ over $\mathfrak{R}$. This covering space is "regular", i.e. it never happens that for two curves in $\overline{\mathfrak{R}}$ which have the same "trace curve" in $\mathfrak{R}$, one is open and the other closed. Now let $\overline{\mathfrak{p}_1}$, $\overline{\mathfrak{p}_1}'$ be two points in $\overline{\mathfrak{R}}$ which "lie on top of each other", i.e. lie over the same point $\mathfrak{p}_1$ of $\mathfrak{R}$; then there is a unique invertible well-defined continuous mapping from $\overline{\mathfrak{R}}$ into itself, where every point of $\overline{\mathfrak{R}}$ maps to a point lying on top of it ("decktransformation"), in particular $\overline{\mathfrak{p}_1}$ into $\overline{\mathfrak{p}_1}'$. These transformations form a discrete group Γ_R. The measure of the length is preserved from $\mathfrak{R}$ to $\overline{\mathfrak{R}}$; the decktransformations are congruent maps from $\overline{\mathfrak{R}}$ into itself.

To every point $\overline{\mathfrak{p}}$ in $\overline{\mathfrak{R}}$ correspond p unique numbers $x_1, x_2, \ldots, x_p$, namely the translation components along the curve which defines $\overline{\mathfrak{p}}$, and thus a point $(x) = (x_1, x_2, \ldots, x_p)$ in the Euclidean space R with rectangular coordinates x_i. The mapping $\overline{\mathfrak{p}} \mapsto (x)$ is well-defined, continuous, and length-preserving. We now change our point of view, and from now let's agree to consider $\overline{\mathfrak{p}}$ as lying above a point (x) in R. Then $\overline{\mathfrak{R}}$ changes into an unramified covering space over R, which is *unbounded.* Above the origin in R there certainly lies a point $\overline{\mathfrak{p}}_0$ of $\overline{\mathfrak{R}}$. We trace from the origin an arbitrary half-line g in R and trace in $\overline{\mathfrak{R}}$ a continuously changing point $\overline{\mathfrak{p}}$, which originates at $\overline{\mathfrak{p}}_0$ and whose trace point runs through this line in R (Weierstrass' Principle of Analytic Continuation); from this we obtain a unique curve $\overline{\gamma}$ in $\overline{\mathfrak{R}}$ above g. It is not possible on $\overline{\mathfrak{R}}$ to reach a "boundary" or a "critical point"; this is because from every point that we may have reached, we can still go a distance of at least r_0 along g. As $\overline{\mathfrak{R}}$ is, relative to R, unramified and unbounded, *and as the usual Euclidean Space is simply connected*, $\overline{\mathfrak{R}}$ must be everywhere exactly one-sheeted over R, i.e. $\overline{\mathfrak{R}}$ *is mapped, through the map* $\overline{\mathfrak{p}} \mapsto (x)$, *bijectively and congruently to the Euclidean space* R. By virtue of this map the group of transformations $\Gamma_{\mathfrak{R}}$ of $\overline{\mathfrak{R}}$ (relative to $\mathfrak{R}$) appears as a discrete group of motions Γ in R, and $\mathfrak{R}$ itself is mapped bijectively in a length-preserving and continuous manner to the crystal $\mathfrak{K}_\Gamma$. If $\mathfrak{R}$ is closed, this must also hold for $\mathfrak{K}_\Gamma$, i.e. the group Γ must possess a finite fundamental domain. And with that our proof has come to an end.

Notes

Do we want to refer to Theorem 3 as the "principle" or the "criterion'?

When to use "homogeneous" with "linear combination"?

Should I put the gothic r's in boldface instead?