References

[1] P. Erdös and R. R. Hall, *Values of the divisor function on short intervals*, J. Number Theory 12 (1980), pp. 176–187.
[2] P. Erdös and I. Kátai, *On the growth of some additive functions on small intervals*, Acta Math. Hungar. 33 (1979), pp. 345–359.
[3] H. Halberstam and H.-E. Richert, *On a result of R. R. Hall*, J. Number Theory 11 (1979), pp. 76–89.
[4] R. R. Hall, *The maximum value of the divisor function on short intervals*, Bull. London Math. Soc. 13 (1981), pp. 121–124.
[5] G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities*, Cambridge 1934.
[6] I. Kátai, *Local growth of the number of prime divisors of consecutive integers*, Publ. Math. Debrecen 18 (1971), pp. 171–175.
[7] A. Selberg, *Note on a paper by L. G. Sathe*, J. Indian Math. Soc. 18 (1954), pp. 83–87.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF YORK
York, Y01 5DD, England
UNIVERSITÉ DE NANCY I
UER SCIENCES MATHÉMATIQUES
Boîte Postale 239
54506 Vandoeuvre les Nancy, France

Received on 21.10.1982
and in revised form on 17.1.1983 (1323)

ACTA ARITHMETICA
XLIII (1984)

The joint distribution of the binary digits of integer multiples

by

JOHANNES SCHMID (Ulm)

1. Problem and results. Each number $n \in N_0$ has a unique binary expansion

$$n = \sum_{\nu=0}^{\infty} d_\nu 2^\nu, \quad d_\nu \in \{0, 1\}.$$

We consider the "digital sum", i.e. the total

$$B(n) := \sum_{\nu=0}^{\infty} d_\nu$$

of digits 1 in this expansion.

Elementarily, $B(n)$ is binomially distributed and hence its approximate distribution is given by the central limit theorems, e.g. in the simplest form:

$$\frac{1}{2^m} \# \left\{ 0 \leqslant n < 2^m; \frac{B(n) - m/2}{\sqrt{m/4}} \leqslant \xi \right\} \underset{(m\to\infty)}{\longrightarrow} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\xi} e^{-t^2/2} dt \quad \text{for } \xi \in \boldsymbol{R}.$$

It is not difficult to show (elementarily) that the right hand term describes the limiting distribution of $B(kn)$, $k \in N$, too.

Thence, it should be interesting to know the joint distribution of $\big(B(k_1 n), B(k_2 n), \ldots, B(k_s n)\big)$ (,where, w.l.o.g., the k_ν are different and odd).

A special question of this type was asked by Stolarsky and Muskat [9] who obtained the upper estimate:

$$\# \{1 \leqslant n < x; B(kn) - B(n) = a\} < c_1(k) \frac{x}{\sqrt{\log x}} + c_2(k) \frac{x}{\log x}$$

with explicit but rather large $c_1(k)$, $c_2(k)$. Further, they conjectured tha the left side is

$$\sim \hat{c}_1(k) \frac{x}{\sqrt{\log x}} \exp\left(-\hat{c}_2(k) \frac{a^2}{\log x}\right) \quad \text{for suitable } \hat{c}_1(k), \hat{c}_2(k),$$

at least for "small" a.

For $k = 3$, a theorem of the latter type was obtained by Stolarsky [8].

The general case was treated by W. Schmidt [7], who showed that $(B(k_1 n), B(k_2 n), \ldots, B(k_s n))$ in a global sense have a limiting distribution of Gauss type.

In our main theorem we shall prove that the same is true in the local sense and moreover shall give an error term instead of a mere asymptotic.

Kátai [4] proved a global theorem of W. Schmidt's type for the differences $B(3n)-B(n)$.

MAIN THEOREM 1.1. *Let $s \in \boldsymbol{N}$; $k_1, k_2, \ldots, k_s$ different odd $\in \boldsymbol{N}$. Then, for $x \in \boldsymbol{R}$, $x > 1$, and $a := (a_\nu)_{\nu=1}^s \in \boldsymbol{Z}^s$*

$$\# \{n: 0 \leqslant n < x; \forall \nu = 1, 2, \ldots, s: B(k_\nu n) = a_\nu\}$$

$$= \frac{x}{\sqrt{2\pi \operatorname{ld} x}^s \sqrt{\det V}} e^{-\frac{1}{2\operatorname{ld} x}\left(a-\frac{\operatorname{ld} x}{2}\right)V^{-1}\left(a-\frac{\operatorname{ld} x}{2}\right)^*} + O\left(\frac{x}{\sqrt{\operatorname{ld} x}^s \sqrt{\operatorname{ld} x}}\right)$$

with the positive-definite (in particular: regular) $s \times s$-matrix V with entries

$$v_{\mu,\nu} := \frac{1}{4} \frac{\gcd^2(k_\mu, k_\nu)}{k_\mu k_\nu} \qquad (\mu, \nu = 1, 2, \ldots, s).$$

Herein, the O-constant does not depend on a (but may depend on s; $k_1, k_2, \ldots, k_s$).

Note. $\operatorname{ld} x$ means logarithm of x to base 2. All vectors are row vectors. $*$ denotes the transpose of a vector or matrix.

Remarks. 1. For $x = 2^m$ a second term — of order $\dfrac{2^m}{\sqrt{m}^s \sqrt{m}}$ — can be made explicit and the error term reduced to $O\left(\dfrac{2^m}{\sqrt{m}^s m}\right)$ by our Main Lemma 4.4.

2. Let

$$\hat{a} := a - \frac{\operatorname{ld} x}{2}.$$

Then, for sufficiently small C_1 (> 0)

$$|\hat{a}| \leqslant C_1 \sqrt{\operatorname{ld} x \log \operatorname{ld} x}$$

implies that $\exp\left(\dfrac{1}{2\operatorname{ld} x} \hat{a} V^{-1} \hat{a}^*\right) = o(\sqrt{\operatorname{ld} x})$, and hence for those a the main theorem gives a uniform asymptotic as $x \to \infty$.

On the other hand, for sufficiently large C_2

$$|\hat{a}| \geqslant C_2 \sqrt{\operatorname{ld} x \log \operatorname{ld} x}$$

implies that the "main term" is only of the order of the remainder term.

3. For sufficiently large $C = C(q)$, $q > 0$, it can even be proved elementarily from the properties of the binomial distribution that

$$\# \left\{0 \leqslant n < x; \exists \nu: \left|B(k_\nu n) - \frac{\operatorname{ld} x}{2}\right| \geqslant C\sqrt{\operatorname{ld} x \log \operatorname{ld} x}\right\} = O\left(\frac{x}{\sqrt{\operatorname{ld} x}^q}\right).$$

Just observe that the left hand term is

$$O\left(\# \left\{0 \leqslant n < (\max_{\nu=1,\ldots,s} k_\nu) x; \left|\frac{B(n) - \frac{\operatorname{ld} x}{2}}{\sqrt{\operatorname{ld} x}}\right| \geqslant C\sqrt{\log \operatorname{ld} x}\right\}\right).$$

With our means we can improve the $c_1(k)$ of Stolarsky and Muskat's theorem to its optimal value at the expense of being not able to specify $c_2(k)$ explicitly. Namely, as one might suspect by heuristically summing up the terms of Theorem 1.1 with $s = 2$ over those (a_1, a_2) with $a_1 - a_2 = a$:

THEOREM 1.2. *Let k_1, k_2 be different odd $\in \boldsymbol{N}$. Then, for $x \in \boldsymbol{R}$, $x > 1$, and $a \in \boldsymbol{Z}$*

$$\# \{n: 0 \leqslant n < x; B(k_1 n) - B(k_2 n) = a\} = \frac{x}{\sqrt{2\pi \operatorname{ld} x} \sqrt{\tilde{V}}} e^{-\frac{1}{2\operatorname{ld} x} \frac{a^2}{\tilde{V}}} + O\left(\frac{x}{\operatorname{ld} x}\right)$$

with

$$\tilde{V} := \frac{1}{2}\left(1 - \frac{\gcd^2(k_1, k_2)}{k_1 k_2}\right) > 0.$$

Herein, the O-constant does not depend on a (but may depend on k_1, k_2).

We can also sharpen W. Schmidt's result by giving an error term. Heuristically, this is done by summing up the terms of Theorem 1.1 over those a with

$$\frac{a_\nu - \frac{\operatorname{ld} x}{2}}{\sqrt{\operatorname{ld} x}} \leqslant \xi_\nu \quad \text{for all } \nu.$$

THEOREM 1.3. *Let* s, k_ν, V *be as in Theorem* 1.1. *Then, for* $x \in \mathbf{R}$, $x > 1$, *and* $\xi := (\xi_\nu)_{\nu=1}^s \in \mathbf{R}^s$

$$\#\left\{n: 0 \leqslant n < x;\ \forall \nu = 1, 2, \ldots, s:\ \frac{B(k_\nu n) - \frac{\operatorname{ld} x}{2}}{\sqrt{\operatorname{ld} x}} \leqslant \xi_\nu\right\}$$

$$= \frac{x}{\sqrt{2\pi}^s \sqrt{\det V}} \int_{t \leqslant \xi} \cdots \int e^{-\frac{1}{2} t V^{-1} t^*} dt + O\left(\frac{x}{\sqrt{\operatorname{ld} x}}\right).$$

Herein, the O-constant does not depend on ξ *(but may depend on* s; $k_1, k_2, \ldots, k_s$*).*

Actually, in W. Schmidt's version the $\operatorname{ld} x$ on the left side has to be replaced by $\operatorname{ld} n$. We omit the straightforward proof that this does not change the right side.

For some interesting corollaries to this theorem the reader is referred to W. Schmidt's paper.

If desired, one could obtain by our means a global theorem for the differences $B(k_1 n) - B(k_2 n)$ related to Theorem 1.2 as Theorem 1.3 is related to Theorem 1.1 and thus generalize Kátai's result.

Starting point for our proof is the recursion formula in Lemma 2.1, which is analogous to the one used by Stolarsky and Muskat [9], (1.7) & (1.8).

By introducing Fourier-transforms the problem is essentially translated to the estimation of the mth power of a matrix valued function $\Phi(t)$, $t \in [-\pi, \pi]^s$, for large m.

$\Phi^m(t)$ is estimated by employing the maximum modulus eigenvalue $\lambda(t)$ of $\Phi(t)$ and the corresponding projection operator. Hereby we follow the main lines of Höglund's treatment of Markov chains [3], especially Theorem 3.1, while most details are specific to our problem.

The notion of a set like $\mathscr{A}$ (Lemma 2.3) and the properties of V (Lemma 4.3) are contained in W. Schmidt's paper [7]; for Lemma 4.3(b) we can also use the more general result in J. Schmid [6].

2. A recursion and an equivalent matrix problem. Remember that throughout this paper $k_1, k_2, \ldots, k_s$ are different odd $\in \mathbf{N}$.

We are interested in the number of $n \in [0, x)$ with $B(k_\nu n) = a_\nu$. In order to obtain a one-step recursion, however, it is necessary to consider also nonzero residue classes mod k_ν, that is to count the n with $B(k_\nu n + h_\nu) = a_\nu$. Furthermore, since the recursion works only for intervals of length a power of 2, we decompose $[0, x)$ into intervals of the form $[l2^m, (l+1)2^m)$. This will motivate the parameters l and h in the following definition:

For $a := (a_\nu)_{\nu=1}^s \in \mathbf{Z}^s$; $h := (h_\nu)_{\nu=1}^s \in \mathbf{Z}^s$, $0 \leqslant h_\nu \leqslant k_\nu - 1$; and $l, m \in \mathbf{N}_0$ let

$$S_l(h, a; 2^m) := \#\{l2^m \leqslant n < (l+1)2^m;\ \forall \nu = 1, 2, \ldots, s: B(k_\nu n + h_\nu) = a_\nu\}. \tag{2.1}$$

Then the recursion is given by

LEMMA 2.1. *For* $m > 0$

$$S_l(h, a; 2^m) = S_l(h', a - r'; 2^{m-1}) + S_l(h'', a - r''; 2^{m-1}),$$

where

$$h' = h'(h) := \left(\left[\frac{h_\nu}{2}\right]\right)_{\nu=1}^s, \quad h'' = h''(h) := \left(\left[\frac{h_\nu + k_\nu}{2}\right]\right)_{\nu=1}^s,$$

and

$$r' = r'(h) := (p(h_\nu))_{\nu=1}^s, \quad r'' = r''(h) := (p(h_\nu + k_\nu))_{\nu=1}^s$$

with

$$p(u) := \begin{cases} 1, & u \text{ odd}, \\ 0, & u \text{ even}. \end{cases}$$

(Of course, $p(h_\nu + k_\nu) = 1 - p(h_\nu)$, *because* k_ν *is odd.)*

Proof. Let

$$l2^m \leqslant n < (l+1)2^m, \quad n = 2n' + d, \quad d \in \{0, 1\}.$$

Then, for all $\nu = 1, 2, \ldots, s$, the following are equivalent:

$$B(k_\nu n + h_\nu) = a_\nu,$$

$$B\left(2k_\nu n' + 2\left[\frac{h_\nu + k_\nu d}{2}\right] + p(h_\nu + k_\nu d)\right) = a_\nu,$$

$$B\left(k_\nu n' + \left[\frac{h_\nu + k_\nu d}{2}\right]\right) = a_\nu - p(h_\nu + k_\nu d).$$

$d = 0$ yields the first, $d = 1$ the second right hand term.

In the next step, to make the recursion more manageable, we combine the infinitely many a's into Fourier-transforms

$$(\varphi_{l,m}(t))_h := \sum_{a \in \mathbf{Z}^s} S_l(h, a; 2^m) e^{iat^*}, \tag{2.2a}$$

where $t = (t_\nu)_{\nu=1}^s$, $t_\nu \in [-\pi, \pi]$, say, at the same time interpreting these functions as components of a $(k_1 k_2 \ldots k_s)$-dimensional vector $\varphi_{l,m}(t)$ indexed by the set of s-tuples $\{h; 0 \leqslant h_\nu \leqslant k_\nu - 1\}$.

The recursion takes the form

$$\begin{aligned}\big(\varphi_{l,m}(t)\big)_h &= \sum_{a\in\mathbf{Z}^s}\big(S_l(h', a-r'; 2^{m-1})+S_l(h'', a-r''; 2^{m-1})\big)e^{iat^*}\\ &= \sum_{b\in\mathbf{Z}^s}\big(S_l(h', b; 2^{m-1})e^{ir't^*}+S_l(h'', b; 2^{m-1})e^{ir''t^*}\big)e^{ibt^*}\\ &= \big(\varphi_{l,m-1}(t)\big)_{h'}e^{ir't^*}+\big(\varphi_{l,m-1}(t)\big)_{h''}e^{ir''t^*},\end{aligned}$$

which can, obviously, be written as

$$\varphi_{l,m}(t) = \varphi_{l,m-1}(t)\,\Phi(t), \tag{2.3}$$

where $\Phi(t)$ is the $(k_1k_2\ldots k_s)\times(k_1k_2\ldots k_s)$-matrix with the entries

$$\big(\Phi(t)\big)_{g,h} := \begin{cases} e^{ir't^*} & \text{if } g = h', \\ e^{ir''t^*} & \text{if } g = h'', \\ 0 & \text{otherwise.}\end{cases} \tag{2.2b}$$

Apart from the case $s = 1$, $k_1 = 1$ we always have $h' \neq h''$, so that each column of $\Phi(t)$ contains exactly two nonzero entries. In the exceptional case the two terms must be added. Since this exception is just the well-known binomial distribution case, we shall not bother about the evident changes necessary in the following to prove it along the present lines.

By iterating (2.3) we find

$$\varphi_{l,m}(t) = \varphi_{l,0}(t)\,\Phi^m(t), \tag{2.4a}$$

where $\varphi_{l,0}(t)$ is easily determined from (2.2a) and (2.1):

$$\big(\varphi_{l,0}(t)\big)_h = \exp\Big(i\big(B(k_\nu l+h_\nu)\big)_{\nu=1}^s t^*\Big). \tag{2.4b}$$

Finally, the $S_l(h, a; 2^m)$ can be obtained explicitly from the $\varphi_{l,m}(t)$ by the Fourier inversion formula. For convenience, we shall state this with a matrix $\Psi(t)$ instead of $\Phi(t)$, where $\Psi^m(t)$ is obtained from $\Phi^m(t)$ by extracting the factor which corresponds to the approximate mean value $m/2$ of the desired distribution. Namely, let

$$\Psi(t) := e^{-\frac{i}{2}(t_1+t_2+\ldots+t_s)}\Phi(t). \tag{2.5}$$

$\Psi(t)$ now fulfills a symmetry property which we shall state in Lemma 3.4.

The Fourier inversion formula is:

LEMMA 2.2

$$S_l(h, a; 2^m) = \sum_{\substack{g=(g_\nu)_{\nu=1}^s\\ 0\leqslant g_\nu\leqslant k_\nu-1}} \frac{1}{(2\pi)^s}\int\limits_{[-\pi,\pi]^s}\!\!\cdots\!\int e^{-i\left(a-\frac{m}{2}-b_l(g)\right)t^*}\big(\Psi^m(t)\big)_{g,h}\,dt, \tag{2.6}$$

where

$$b_l(g) := \big(B(k_\nu l+g_\nu)\big)_{\nu=1}^s.$$

Proof. The proof is obvious by (2.4a), (2.4b), and (2.5) above.

Our problem is thus reduced to estimating $\Psi^m(t)$. Because this depends largely on the maximum modulus eigenvalue(s) of $\Psi(t)$, classical results due to Perron, Frobenius, and Wielandt (see our Lemma 3.1) make it desirable that $\Psi(t)$ is irreducible.

Hereby, a matrix Ψ is called *irreducible* if it does *not* have the form

$$\begin{bmatrix}\Psi_{11} & \Psi_{12}\\ 0 & \Psi_{22}\end{bmatrix}$$

with quadratic submatrices Ψ_{11}, Ψ_{22}, and cannot be brought into this form by a consistent relabelling of the rows and columns (otherwise: reducible).

An interpretation in Markov chain terminology (see Chung [1]) may be illuminating: Call the transpose of $\Psi(0)/2$ the transition matrix of a Markov chain with states h, the transition probabilities being $1/2$ for both the transition $h\to h'$ and $h\to h''$ (and 0 otherwise). As is well-known, a limiting distribution of $\big(\Psi(0)/2\big)^m$ is obtained most easily for an irreducible chain or, equivalently, an irreducible transition matrix.

For the relation of $\big(\Phi(t)/2\big)^m$ to Markov chain problems the reader may consult Höglund [3], introduction.

In general, our matrix Ψ is reducible. Namely, note that all $g_\nu \leqslant (k_\nu-1)/2$ if $g = h'$, and all $g_\nu \geqslant (k_\nu-1)/2$ if $g = h''$. If therefore some $g_\nu < (k_\nu-1)/2$ and some $g_\nu > (k_\nu-1)/2$, then neither $g = h'$ nor $g = h''$, and hence the gth row consists of 0's only.

In view of this, let $\mathscr{A}$ denote the subset of the index set $\{h;\ 0\leqslant h_\nu \leqslant k_\nu-1\}$ belonging to Ψ_{11}. If $\mathscr{A}$ contains $h = 0$, the only value of h eventually interesting us, we can reformulate Lemma 2.2 with Ψ_{11} instead of Ψ and g and h restricted to $\mathscr{A}$ accordingly. Again it is desirable that Ψ_{11} is irreducible.

In the next lemma we shall give a set $\mathscr{A}$ of this kind. It can be described by the linearly ordered set $[0, 1)$, which facilitates the computations involving h' and h'' in the following. (Cf. the notion of "admissible vector" in W. Schmidt [7], formula (5.1).) The proof of the irreducibility of Ψ_{11} will then be given in the next section after stating the eigenvalue theorems for irreducible matrices.

LEMMA 2.3. *Let*

$$\mathscr{A} := \{h = (h_\nu)_{\nu=1}^s \in \mathbf{Z}^s;\ \exists x\in[0, 1)\colon \forall \nu = 1, 2, \ldots, s\colon h_\nu = [k_\nu x]\};$$

further let I_h be the set of all $x\in[0, 1)$ representing $h\in\mathscr{A}$ by $h = ([k_\nu x])_{\nu=1}^s$, and let L_h be its measure (length).

Then:

(a) (i) $0 \in \mathscr{A}$,

(ii) *For all* $h = ([k_\nu x])_{\nu=1}^s \in \mathscr{A}$

$$h' = \left(\left[k_\nu \frac{x}{2}\right]\right)_{\nu=1}^s \in \mathscr{A} \tag{2.7a}$$

and

$$h'' = \left(\left[k_\nu \frac{x+1}{2}\right]\right)_{\nu=1}^s \in \mathscr{A}. \tag{2.7b}$$

(b) *For all* $h \in \mathscr{A}$ I_h *is a left closed, right open subinterval of* $[0,1)$. *The endpoints of such intervals are exactly those rationals in* $[0,1)$ *that can be written with one of the* $k_1, k_2, \ldots, k_s$ *as denominator.*

Proof. $h' = \left[\frac{[k_\nu x]}{2}\right] = \left[\frac{k_\nu x}{2}\right]$, hence (2.7a). (2.7b) analogously. The rest of the proof is obvious.

Now, we restrict all indices h, g to the set $\mathscr{A}$, but, for convenience, continue to write Ψ instead of Ψ_{11}:

Reread the definitions of $\varphi_{l,m}$, Φ, and Ψ and Lemma 2.2 with the additional restriction $h \in \mathscr{A}$, $g \in \mathscr{A}$. ($\varphi_{l,m}$ becomes a $(\#\mathscr{A})$-vector and Φ, Ψ a $(\#\mathscr{A})\times(\#\mathscr{A})$-matrix. The sum in Lemma 2.2 is over $g \in \mathscr{A}$.)

3. Spectral properties of $\Psi(t)$. An estimate of $\Psi^m(t)$ is closely related to the spectral properties of $\Psi(t)$. These shall be treated in the present section.

The classical eigenvalue theorems about non-negative matrices — like $\Psi(0)$ — due to Perron and Frobenius are (quoted here from Wielandt [10], for an English reference see, e.g., Gantmacher [2], vol. II, chapter XIII):

LEMMA 3.1. *Let A be a $k\times k$-matrix with non-negative entries and let A be irreducible. Then*:

(a) *A has a simple, positive maximum modulus eigenvalue λ. The eigenvector corresponding to λ can be chosen to be positive (componentwise), and λ is the only eigenvalue with this property. λ is the only maximum modulus eigenvalue if $A^m > 0$ for some $m \in \mathbf{N}$.*

(b) *Let B be a complex $k\times k$-matrix with*

$$|b_{\mu,\nu}| \leqslant a_{\mu,\nu} \quad \forall \mu, \nu\colon 1 \leqslant \mu, \nu \leqslant k$$

($a_{\mu,\nu}$, $b_{\mu,\nu}$ being the entries of A, B, respectively), and let β be a maximum modulus eigenvalue of B. Then

$$|\beta| \leqslant \lambda,$$

and equality, i.e. $\beta = e^{i\eta}\lambda$ for an $\eta \in \mathbf{R}$, holds if and only if there is a diagonal matrix D whose diagonal elements all have modulus 1 such that

$$B = e^{i\eta} DAD^{-1}.$$

Part (b)—useful for us in relating the spectrum of $\Psi(t)$ to that of $\Psi(0)$—is Wielandt's contribution.

We still have to prove the irreducibility of $\Psi(0)$ and then can apply Perron/Frobenius:

LEMMA 3.2. (a) *For $m \in \mathbf{N}$; $g, h \in \mathscr{A}$, $h = ([k_\nu x])_{\nu=1}^s$, $x \in [0,1)$,*

$$\frac{(\Psi^m(0))_{g,h}}{2^m} = \frac{1}{2^m}\#\left\{0 \leqslant u < 2^m;\ \frac{x}{2^m} + \frac{u}{2^m} \in I_g\right\} \xrightarrow[(m\to\infty)]{} L_g.$$

(b) *Consequently: $\Psi^m(0) > 0$ for all sufficiently large m; in particular, $\Psi(0)$ is irreducible.*

$\lambda(0) := 2$ is a simple, positive eigenvalue of $\Psi(0)$, exceeding all other eigenvalues in modulus. $(1, 1, \ldots,)$ is a left, $(L_g)^*_{g\in\mathscr{A}}$ a right eigenvector of $\Psi(0)$ to $\lambda(0)$.

Proof. (a) The Markov chain interpretation of Section 2 simplifies the formulation of the proof:

For a Markov chain with states in $\mathscr{A}$ and transposed transition matrix $\frac{\Psi(0)}{2}$ $\frac{(\Psi^m(0))_{g,h}}{2^m}$ is the probability of reaching state g from state h after m steps. Let us write, for short, "parameter x" instead of "state $([k_\nu x])_{\nu=1}^s$". Then, starting with parameter x, in one step exactly the parameters $x/2$ and $(x+1)/2$ are reached each with probability $1/2$; in two steps $x/4$, $(x+1)/4$, $(x+2)/4$, $(x+3)/4$ each with probability $1/4$; and so on. Finally, in m steps exactly the parameters $(x+u)/2^m$, $0 \leqslant u < 2^m$, are reached each with probability $1/2^m$. The probabilities belonging to the same state g add up to $v/2^m$, where v is the number of u's such that $(x+u)/2^m \in I_g$, that is $v = 2^m L_g + O(1)$; whence $v/2^m \xrightarrow[(m\to\infty)]{} L_g$.

(b) Each column of $\Psi(0)$ contains exactly two entries equal to 1, all other entries being equal to 0; hence $(1, 1, \ldots, 1)$ is a left eigenvector and $\lambda(0) := 2$ is an eigenvalue. The other properties of $\lambda(0)$ follow by Lemma 3.1(a). $(L_g)^*_{g\in\mathscr{A}}$ is seen to be a right eigenvector by letting $m\to\infty$ in

$$\left(\frac{\Psi(0)}{2}\right)^{m+1} = \frac{\Psi(0)}{2}\left(\frac{\Psi(0)}{2}\right)^m.$$

Now, we shall use Wielandt's theorem (Lemma 3.1(b)) to show that the eigenvalues of $\Psi(t)$, $t \neq 0$, are smaller in modulus than $\lambda(0)$. Obviously, $\Psi(0)$, $\Psi(t)$ are in the relation of A, B in Lemma 3.1. So we are left with showing that $\Psi(t)$ does not have the exceptional structure of Lemma 3.1(b) in relation to $\Psi(0)$. Lemma 3.3(a) will also be used in the following section to show that the maximum modulus eigenvalue of $\Psi(t)$ decreases quadratically as t moves away from 0 (see Lemma 4.3(b)).

LEMMA 3.3. (a) *For* $t\in[-\pi,\pi]^s$, $t\neq 0$, $\Psi(t)$ *does not have the form*

$$\Psi(t) = e^{i\eta}D\Psi(0)D^{-1},$$

with $\eta\in\mathbf{R}$ *and* D *a diagonal matrix all of whose diagonal elements have modulus* 1.

(b) *Consequently: Each eigenvalue of* $\Psi(t)$, $t\in[-\pi,\pi]^s$, $t\neq 0$, *is smaller in modulus than* $\lambda(0)$.

Proof. (a) Suppose that for some $t=(t_\nu)_{\nu=1}^s\in[-\pi,\pi]^s$

$$\Psi(t) = e^{i\eta}D\Psi(0)D^{-1}.$$

We want to show that then $t=0$.

(i) The relation between D and $\Psi(t)$: It is advantageous to use the $x\in[0,1)$ instead of the $h\in\mathscr{A}$ here. Therefore, we define $\varrho_\nu(x)$ and $\theta(x)$ by

$$\varrho_\nu(x) := p(h_\nu),\qquad \nu=1,2,\ldots,s,$$

and

$$e^{i\theta(x)} = (D)_{h,h},\qquad -\pi\leqslant\theta(x)<\pi,$$

where $h=([k_\nu x])_{\nu=1}^s$. (p like in Lemma 2.1.) Then, by the definition of $\Psi(t)$ and Lemma 2.3(a),

$$e^{i\sum_{\nu=1}^{s}t_\nu(\varrho_\nu(x)-\frac12)} = (\Psi(t))_{h',h} = e^{i\eta}e^{i\theta\left(\frac{x}{2}\right)}\cdot 1\cdot e^{-i\theta(x)},$$

and

$$e^{i\sum_{\nu=1}^{s}t_\nu(\frac12-\varrho_\nu(x))} = (\Psi(t))_{h'',h} = e^{i\eta}e^{i\theta\left(\frac{x+1}{2}\right)}\cdot 1\cdot e^{-i\theta(x)}.$$

Thus we can express $\theta(x)$ by $\theta(x/2)$ and $\theta((x+1)/2)$, respectively:

$$\theta(x)\equiv\eta+\theta\left(\frac{x}{2}\right)-\sum_{\nu=1}^{s}t_\nu\left(\varrho_\nu(x)-\tfrac12\right)$$

$$\equiv\eta+\theta\left(\frac{x+1}{2}\right)-\sum_{\nu=1}^{s}t_\nu\left(\tfrac12-\varrho_\nu(x)\right)\pmod{2\pi}.$$

Iterating each of these representations m-times yields

$$\theta(x)\equiv m\eta+\theta\left(\frac{x}{2^m}\right)-\sum_{\nu=1}^{s}t_\nu\sum_{\mu=0}^{m-1}\left(\varrho_\nu\left(\frac{x}{2^\mu}\right)-\frac12\right)$$

$$\equiv m\eta+\theta\left(\frac{x+(2^m-1)}{2^m}\right)-\sum_{\nu=1}^{s}t_\nu\sum_{\mu=0}^{m-1}\left(\frac12-\varrho_\nu\left(\frac{x+(2^\mu-1)}{2^\mu}\right)\right)\pmod{2\pi}.$$

For sufficiently large m (independent of x)

$$\theta\left(\frac{x}{2^m}\right)=\theta(0)\quad\text{and}\quad\theta\left(\frac{x+(2^m-1)}{2^m}\right)=\lim_{x\to1-}\theta(x);$$

and hence, for any $x,x'\in[0,1)$

$$\theta(x)-\theta(x')\equiv\sum_{\nu=1}^{s}t_\nu\sum_{\mu=0}^{m-1}\left(\varrho_\nu\left(\frac{x'}{2^\mu}\right)-\varrho_\nu\left(\frac{x}{2^\mu}\right)\right)$$

$$\equiv\sum_{\nu=1}^{s}t_\nu\sum_{\mu=0}^{m-1}\left(\varrho_\nu\left(\frac{x+(2^\mu-1)}{2^\mu}\right)-\varrho_\nu\left(\frac{x'+(2^\mu-1)}{2^\mu}\right)\right)\pmod{2\pi}.$$

(ii) Proof that these representations cannot both hold except for $t=0$: W.l.o.g., let

$$k_1<k_2<\ldots<k_s.$$

Let $\varkappa\in\{1,2,\ldots,s\}$ such that $t_\nu=0$ $\forall\nu>\varkappa$. We show that then even $t_\varkappa=0$.

Case 1: $k_\varkappa>1$. Note that ϱ_ν is continuous except at most at the end-points of the I_h (see Lemma 2.3(b)). We shall choose x such that for all $\nu<\varkappa$ ϱ_ν is continuous at the $x/2^\mu$ and $(x+(2^\mu-1))/2^\mu$, and for $\nu=\varkappa$ ϱ_ν is continuous at these points except at x and $x/2$. x' will be chosen in relation to x such that the respective ϱ_ν-values are equal in the continuity cases, and x' ($x'/2$) lies in the interval I_h immediately preceding the interval with left endpoint x ($x/2$). Namely, let

$$x:=\frac{2}{k_\varkappa},\qquad x':=\frac{2}{k_\varkappa}-\varepsilon\quad\text{with } 0<\varepsilon<\min_{h\in\mathscr{A}}L_h.$$

For $\nu<\varkappa$ $k_\nu x=k_\nu 2/k_\varkappa\notin\mathbf{Z}$ (because $k_\nu<k_\varkappa$ and $k_\varkappa$ odd), hence

$$[k_\nu x]=[k_\nu x'],$$

which implies by (2.7(a) & (b))

$$\left[k_\nu\frac{x}{2^\mu}\right]=\left[k_\nu\frac{x'}{2^\mu}\right]\quad\text{and}\quad\left[k_\nu\frac{x+(2^\mu-1)}{2^\mu}\right]=\left[k_\nu\frac{x'+(2^\mu-1)}{2^\mu}\right]\ \forall\mu\in\mathbf{N}_0.$$

For $\varkappa$ itself $k_\varkappa x=2$, hence

$$[k_\varkappa x]=2,\qquad[k_\varkappa x']=1,$$

$$\left[k_\varkappa\frac{x}{2}\right]=1,\qquad\left[k_\varkappa\frac{x'}{2}\right]=0,$$

$$\left[k_\varkappa\frac{x}{2^\mu}\right]=\left[k_\varkappa\frac{x'}{2^\mu}\right]=0\qquad\forall\mu>1.$$

Moreover, $k_\varkappa \frac{x+1}{2} = \frac{k_\varkappa}{2}+1 \notin \boldsymbol{Z}$, hence

$$\left[k_\varkappa \frac{x+1}{2}\right] = \left[k_\varkappa \frac{x'+1}{2}\right],$$

$$\left[k_\varkappa \frac{x+(2^\mu-1)}{2^\mu}\right] = \left[k_\varkappa \frac{x'+(2^\mu-1)}{2^\mu}\right] \quad \forall \mu \in \boldsymbol{N}.$$

This means that for $\nu \leqslant \varkappa$ all ϱ_ν-differences in $\theta(x)-\theta(x')$ vanish except

$$\varrho_\varkappa(x)-\varrho_\varkappa(x') = 0-1 = -1, \qquad \varrho_\varkappa(x/2)-\varrho_\varkappa(x'/2) = 1-0 = 1.$$

Then, because $t_\nu = 0$ $\forall \nu > \varkappa$,

$$\theta(x)-\theta(x') \equiv t_\varkappa\big(1+(-1)\big) \equiv t_\varkappa(-1) \pmod{2\pi},$$

consequently $t_\varkappa = 0$.

Case 2: $k_\varkappa = 1$ (whence $\varkappa = 1$). Then $\varrho_1(x) = 0$ for $0 \leqslant x < 1$, hence

$$\theta(x) \equiv \eta+\theta(x/2)+t_1/2$$

and

$$\theta(x) \equiv \eta+\theta\left(\frac{x+1}{2}\right)-\frac{t_1}{2} \pmod{2\pi}.$$

Letting $x \to 0+$ in the first, $x \to 1-$ in the second case yields

$$0 \equiv \eta+\frac{t_1}{2} \equiv \eta-\frac{t_1}{2} \pmod{2\pi},$$

consequently $t_1 = 0$. We thus successively have $t_s = 0$, $t_{s-1} = 0, \ldots$ $\ldots, t_1 = 0$. ∎

(b) Immediate by part (a) and Lemma 3.1(b).

Concludingly, we want to prove a symmetry relation for $\Psi(t)$, which will be used in the following section to prove that the maximum modulus eigenvalue of $\Psi(t)$ is an even function in t.

LEMMA 3.4. *Let* $\hat{k} := (k_\nu-1)_{\nu=1}^s$. *Then, for every* $h \in \mathscr{A}$ *also* $\hat{k}-h \in \mathscr{A}$ *and*

$$\big(\Psi(t)\big)_{g,h} = \big(\Psi(-t)\big)_{\hat{k}-g,\hat{k}-h} \quad \forall g, h \in \mathscr{A}.$$

Proof. Let

$$h = ([k_\nu x])_{\nu=1}^s,$$

where w.l.o.g. x is none of the endpoints of the intervals I_f, $f \in \mathscr{A}$, that is, no $k_\nu x$ is an integer. Then $[k_\nu(1-x)] = k_\nu-1-[k_\nu x]$, hence

$$\hat{k}-h = \big([k_\nu(1-x)]\big)_{\nu=1}^s \in \mathscr{A}. \tag{3.1}$$

By definition

$$\big(\Psi(t)\big)_{g,h} = \begin{cases} e^{i\sum_\nu t_\nu(\varrho_\nu(x)-\frac12)} & \text{if } g = h', \\ e^{i\sum_\nu t_\nu(\frac12-\varrho_\nu(x))} & \text{if } g = h'', \\ 0 & \text{otherwise.} \end{cases}$$

Now, in analogy to (3.1) (w.l.o.g., $x/2$ and $(x+1)/2$ are no endpoints, too),

$$\hat{k}-h' = \left(\left[k_\nu\left(1-\frac{x}{2}\right)\right]\right)_{\nu=1}^s = \left(\left[k_\nu \frac{(1-x)+1}{2}\right]\right)_{\nu=1}^s = (\hat{k}-h)''$$

and similarly

$$\hat{k}-h'' = (\hat{k}-h)'.$$

Therefore

$$\big(\Psi(t)\big)_{\hat{k}-g,\hat{k}-h} = \begin{cases} e^{i\sum_\nu t_\nu(\frac12-\varrho_\nu(1-x))} & \text{if } g = h', \\ e^{i\sum_\nu t_\nu(\varrho_\nu(1-x)-\frac12)} & \text{if } g = h'', \\ 0 & \text{otherwise.} \end{cases}$$

Since

$$\varrho_\nu(1-x) = p(k_\nu-1-h_\nu) = p(h_\nu) = \varrho_\nu(x),$$

the proposition follows in all three cases.

We have shown that $\Psi(0)$ has a simple, positive eigenvalue $\lambda(0)$ $(=2)$ exceeding all other eigenvalues of $\Psi(0)$ and all eigenvalues of $\Psi(t)$, $t \neq 0$, in modulus. The entries of $\Psi(t)$ are analytic functions of $t = (t_1, t_2, \ldots, t_s)$.

So the implicit function theorem yields an analytic function $\lambda(t)$ on a neighbourhood of $t = 0$ that extends $\lambda(0)$ as a simple zero of $\det\big(\Psi(t)-z\big)$, that is as a simple eigenvalue of $\Psi(t)$. (Moreover, for small $|t|$ $\lambda(t)$ exceeds all other eigenvalues of $\Psi(t)$ in modulus.)

Write

$$\ker(\Psi-\lambda) =: U_0 =: \langle\hat{\zeta}\rangle \qquad (\text{"kernel space"}),$$

$$\operatorname{im}(\Psi-\lambda) =: U_1 \qquad (\text{"image space"}).$$

Then there is a decomposition

$$\Psi = \lambda P+\Psi_1, \tag{3.2}$$

where P is the projection onto U_0 with kernel U_1 and

$$\Psi_1(\zeta) = \Psi(\zeta) \text{ on } U_1, \quad \Psi_1(\hat{\zeta}) = 0.$$

Therefore

(3.3a) $$P^2 = P, \quad P\Psi_1 = \Psi_1 P = 0;$$

alternatively

(3.3b) $$P(\Psi-\lambda) = (\Psi-\lambda)P = 0.$$

(For the proof one can assume that Ψ is in Jordan normal form.)

Also $P = P(t)$ and $\Psi_1 = \Psi_1(t)$ are analytic functions. This can be seen from the representation

$$P(t) = \frac{1}{2\pi i}\int_{\Gamma}\big(z-\Psi(t)\big)^{-1}dz,$$

Γ a small circle around $\lambda(t)$ (see, e.g., Kato [5], page 38 and following, especially (5.22)).

(3.2) and (3.3a) imply

(3.4) $$\Psi^m = \lambda^m P + \Psi_1^m.$$

Furthermore, all eigenvectors of Ψ other than multiples of $\hat{\zeta}$ are in U_1. The spectrum of Ψ_1, hence, is the spectrum of Ψ minus $\{\lambda\}$.

4. The maximum modulus eigenvalue $\lambda(t)$ and the back transform. From this position we can do our first step in exploiting (2.6):

LEMMA 4.1. *For sufficiently small $\varepsilon > 0$ there is an η: $0 < \eta < 1$ such that*

(4.1) $$S_l(h, a; 2^m) = \sum_{g\in\mathscr{A}}\frac{1}{(2\pi)^s}\int\cdots\int_{|t|\leqslant\varepsilon} e^{-i\left(a-\frac{m}{2}-b_l(g)\right)t^*}\lambda^m(t)\big(P(t)\big)_{g,h}dt + O\big(\lambda^m(0)\eta^m\big),$$

where the O-constant does not depend on h, a, and l.

Proof. We split the domain of integration in (2.6) into two parts:

$$[-\pi, \pi]^s = \{t;\ |t|\leqslant\varepsilon\}\cup([-\pi,\pi]^s\cap\{t;\ |t|\geqslant\varepsilon\}) =: K_1\cup K_2.$$

Using now (3.4) on K_1, we see that the integral in (2.6) equals the integral in (4.1) apart from a remainder term

$$O\Big(\int_{K_1}\cdots\int\|\Psi_1^m(t)\|dt\Big) + O\Big(\int_{K_2}\cdots\int\|\Psi^m(t)\|dt\Big)$$

($\|\ldots\|$ denoting any matrix norm).

By Lemma 3.2 and Lemma 3.3 the spectral radii of $\Psi_1(t)$ and $\Psi(t)$ are less than $\lambda(0)$ and, because of the continuity of the spectra and the compactness of K_1, K_2, this holds even uniformly on K_1 and K_2, respectively. Namely, for some $\delta > 0$

$$\operatorname{spr}\big(\Psi_1(t)\big) < (1-\delta)\lambda(0) \qquad \forall\, t\in K_1$$

and

$$\operatorname{spr}_i'\big(\Psi(t)\big) < (1-\delta)\lambda(0) \qquad \forall\, t\in K_2.$$

However, as can be inferred via the Jordan normal form,

$$\|\Psi_1^m(t)\| = O\Big(m^{\#\mathscr{A}}\operatorname{spr}^m\big(\Psi_1(t)\big)\Big)$$

($\Psi(t)$ analogously). This completes the proof.

We now want to obtain information on the analytic expansion of $\lambda(t)$ and $P(t)$.

$\lambda(t)$ is an even function in t: By Lemma 3.4,

$$\big(\Psi(t)\big)_{g,h} = \big(\Psi(-t)\big)_{\hat{k}-g,\hat{k}-h},$$

and hence, because a mere consistent relabelling of the rows and columns does not change the eigenvalues, $\Psi(t)$ and $\Psi(-t)$ have the same spectrum; especially $\lambda(t) = \lambda(-t)$. Thus:

For small $|t|$

(4.2) $$\lambda(t) = \lambda(0)e^{-\frac{1}{2}tVt^*+O(|t|^4)},$$

where V is a symmetric $s\times s$-matrix.

The exponential way of writing facilitates to form $\lambda^m(t)$.

Before showing that V is actually the matrix V of Section 1, we must prove some properties of $P(t)$:

LEMMA 4.2. (a) $\forall g, h\in\mathscr{A}$:

$$\big(P(0)\big)_{g,h} = L_g.$$

(b) *The entries of* $\dfrac{\partial P}{\partial t_\nu}(0)$ *depend only on the row index, too, and the column sums of* $\dfrac{\partial P}{\partial t_\nu}(0)$ *vanish, i.e. with*

$$Q_g := \left(\left(\frac{\partial P}{\partial t_\nu}(0)\right)_{g,h}\right)_{\nu=1}^{s}$$

we have

$$\sum_{g\in\mathscr{A}} Q_g = 0.$$

Proof. (a) By formula (3.4),

$$P(0) = \frac{\Psi^m(0)}{\lambda^m(0)} - \frac{\Psi_1^m(0)}{\lambda^m(0)}.$$

Herein, the first right side term tends to L_g as $m\to\infty$ by Lemma 3.2(a) and the second tends to 0, because all eigenvalues of $\Psi_1(0)$ are less than $\lambda(0)$.

(b) We need from part (a) that each row of $P(0)$ is proportional to $(1,1,\ldots,1)$ and each column sum equals 1. Now, each row of $\frac{\partial P}{\partial t_\nu}(0)$ is—unless 0—a left eigenvector of $\Psi(0)$ to $\lambda(0)$, i.e. proportional to $(1,1,\ldots,1)$, too: Differentiation of formula (3.3b) yields

$$P(0)\frac{\partial\Psi}{\partial t_\nu}(0)-P(0)\frac{\partial\lambda}{\partial t_\nu}(0)+\frac{\partial P}{\partial t_\nu}(0)\big(\Psi(0)-\lambda(0)\big)=0,$$

$$\frac{\partial P}{\partial t_\nu}(0)\big(\Psi(0)-\lambda(0)\big)=0,$$

by the definition of $\Psi(t)$ and (4.2).

Hence, $\frac{\partial P}{\partial t_\nu}(0)P(0)=\frac{\partial P}{\partial t_\nu}(0)$ and consequently we get by differentiating $P(t)=P^2(t)$:

$$\frac{\partial P}{\partial t_\nu}(0)=\frac{\partial P}{\partial t_\nu}(0)P(0)+P(0)\frac{\partial P}{\partial t_\nu}(0),$$

$$P(0)\frac{\partial P}{\partial t_\nu}(0)=0.$$

Now, back to $\lambda(t)$:

LEMMA 4.3. *For small* $|t|$

$$\lambda(t)=\lambda(0)e^{-\frac{1}{2}tVt^*+O(|t|^4)},\tag{4.2}$$

where:

(a) *V is the real, symmetric $s\times s$-matrix with entries*

$$v_{\mu,\nu}=\frac{1}{4}\frac{\gcd^2(k_\mu,k_\nu)}{k_\mu k_\nu}\quad(\mu,\nu=1,2,\ldots,s).$$

(b) *V is positive-definite.*

Proof. (a) Obviously,

$$v_{\mu,\nu}=-\frac{1}{2}\frac{\partial^2\lambda}{\partial t_\mu\partial t_\nu}(0)=:-\tfrac{1}{2}\lambda_{\mu,\nu}.$$

Now, differentiating formula (3.3b) yields

$$\frac{\partial^2 P}{\partial t_\mu\partial t_\nu}(0)\big(\Psi(0)-\lambda(0)\big)+\frac{\partial P}{\partial t_\mu}(0)\frac{\partial(\Psi-\lambda)}{\partial t_\nu}(0)+\frac{\partial P}{\partial t_\nu}(0)\frac{\partial(\Psi-\lambda)}{\partial t_\mu}(0)+$$
$$+P(0)\frac{\partial^2(\Psi-\lambda)}{\partial t_\mu\partial t_\nu}(0)=0.$$

Multiplying from the left by $P(0)$ suppresses the second and third term (see above), multiplying from the right by $P(0)$ the first term. Hence,

$$\lambda_{\mu,\nu}P(0)=P(0)\frac{\partial^2\Psi}{\partial t_\mu\partial t_\nu}(0)P(0),$$

$$\lambda_{\mu,\nu}=\frac{1}{\big(P(0)\big)_{f,j}}\sum_{g,h\in\mathscr{A}}\big(P(0)\big)_{f,g}\Big(\frac{\partial^2\Psi}{\partial t_\mu\partial t_\nu}(0)\Big)_{g,h}\big(P(0)\big)_{h,j}\qquad(f,j\in\mathscr{A}),$$

$$=\sum_{h\in\mathscr{A}}L_h\sum_{g\in\mathscr{A}}\Big(\frac{\partial^2\Psi}{\partial t_\mu\partial t_\nu}(0)\Big)_{g,h}$$

$$=i^2\sum_{h\in\mathscr{A}}L_h\Big(\big(p(h_\mu)-\tfrac{1}{2}\big)\big(p(h_\nu)-\tfrac{1}{2}\big)+\big(\tfrac{1}{2}-p(h_\mu)\big)\big(\tfrac{1}{2}-p(h_\nu)\big)\Big)$$

(by the definition of $\Psi(t)$),

$$=-\sum_{\substack{h\in\mathscr{A}\\ p(h_\mu)=p(h_\nu)}}L_h(\tfrac{1}{2})-\sum_{\substack{h\in\mathscr{A}\\ p(h_\mu)\neq p(h_\nu)}}L_h(-\tfrac{1}{2})$$

$= -\frac{1}{2}\Big(\big($measure of the set of all $x\in[0,1)$ with $p([k_\mu x])=p([k_\nu x])\big)-$
$-\big($measure of the set of all $x\in[0,1)$ with $p([k_\mu x])\neq p([k_\nu x])\big)\Big)$

$$=-\frac{1}{2}\frac{\gcd^2(k_\mu,k_\nu)}{k_\mu k_\nu}.$$

For a proof of the last identity the reader is referred to W. Schmidt's paper (see W. Schmidt [7], §6, proof of Lemma 6).

(b) Positive-semidefiniteness is immediate, because $|\lambda(t)|<\lambda(0)$ $\forall t\neq 0$.

Positive-definiteness was proved by W. Schmidt (see W. Schmidt [7], § 8, Lemma 9) from the formula in Lemma 4.3(a).

An alternative proof, due to the present author, shows that the positive-definiteness of V is already rooted in the structure of Ψ: For all irreducible matrices $\Psi=\Psi(t)$ with entries of the form e^{irt^*} which fulfill Lemma 3.3(a) the corresponding V of the maximum modulus eigenvalue is positive-definite (see J. Schmid [6]).

We now can complete the back transform of Lemma 2.2 and Lemma 4.1:

MAIN LEMMA 4.4. *For* $\mathrm{ld}(l+2) = O(\mathrm{ld}\, m)$

$$S_l(h, a; 2^m) \tag{4.3}$$

$$= \frac{2^m}{\sqrt{2\pi m}^s \sqrt{\det V}} e^{-\frac{1}{2m}\left(a-\frac{m}{2}\right)V^{-1}\left(a-\frac{m}{2}\right)^*} \left(1 + \frac{1}{m}\left(a - \frac{m}{2}\right) V^{-1} c_l^*\right) + $$
$$+ O\left(\frac{2^m \mathrm{ld}^2(l+2)}{\sqrt{m}^s m}\right),$$

where

$$c_l := \left(\frac{1}{k_\nu} \sum_{k_\nu l \leqslant q < k_\nu(l+1)} B(q)\right)_{\nu=1}^s.$$

Herein, the O-constant does not depend on h, a.

(Remark. $\mathrm{ld}(l+2) = O(\sqrt{m/\log m})$ would suffice for our proof, but we do not need this.)

Proof. (i) (So called saddle point method, cf. Höglund [3], Theorem 3.1, page 132).

The remainder term in Lemma 4.1 is already $O(2^m/m^q)$ for every $q > 0$ since $\lambda(0) = 2$ (Lemma 3.2(b)) and $0 < \eta < 1$.

Under the integral we use approximations for $\lambda^m(t)$ and $P(t)$: By Lemma 4.3, for small $|t|$ (w.l.o.g. for $|t| \leqslant \varepsilon$ with the ε of (4.1))

$$|\lambda^m(t) - 2^m e^{-\frac{m}{2}tVt^*}| = 2^m e^{-\frac{m}{2}tVt^*} |e^{mO(|t|^4)} - 1| = 2^m e^{-\frac{m}{2}tVt^*} O(m|t|^4 e^{\frac{m}{4}tVt^*})$$
$$= O(2^m m |t|^4 e^{-\frac{m}{4}tVt^*}),$$

where in the second step the positive-definiteness of V is vital, namely

$$|t|^4 \leqslant \tfrac{1}{4} tVt^* \quad (|t| \text{ small}).$$

On the other hand, by Lemma 4.2,

$$\left(P(t)\right)_{g,h} = L_g + Q_g t^* + O(|t|^2).$$

With these expansions,

$$\frac{1}{(2\pi)^s} \int_{|t|\leqslant\varepsilon} \cdots \int e^{-i\left(a-\frac{m}{2}-b_l(g)\right)t^*} \lambda^m(t) \left(P(t)\right)_{g,h} dt$$
$$= \frac{1}{(2\pi)^s} 2^m \int_{|t|\leqslant\varepsilon} \cdots \int e^{-i\left(a-\frac{m}{2}-b_l(g)\right)t^*} e^{-\frac{1}{2}mtVt^*} (L_g + Q_g t^*)\, dt +$$
$$+ O\left(2^m \int_{\mathbf{R}^s} \cdots \int |t|^2 e^{-\frac{1}{4}mtVt^*} dt\right) + O\left(2^m m \int_{\mathbf{R}^s} \cdots \int |t|^4 e^{-\frac{1}{4}mtVt^*} dt\right).$$

These integrals can be brought into a handier form by the substitution

$$u = tV^{1/2} m^{1/2},$$
$$du = dt \det(m^{1/2} V^{1/2}) = \sqrt{m}^s \sqrt{\det V}\, dt.$$

It is seen, then, that the domain of integration of the main integral can be changed to $\mathbf{R}^s$ without enlarging the remainder term.

We omit the details of computing this integral and just state the result:

$$S_l(h, a; 2^m) = H_m \sum_{g\in\mathscr{A}} E_m(\hat{a}-b)\left(L_g - \frac{i}{m}(\hat{a}-b)V^{-1}Q_g^*\right) + O\left(\frac{2^m}{\sqrt{m}^s m}\right),$$

with the abbreviations

$$\hat{a} := a - m/2, \qquad b := b_l^{\text{̃}}(g),$$

$$H_m := \frac{2^m}{\sqrt{2\pi m}^s \sqrt{\det V}}, \tag{4.4}$$

and

$$E_m(\xi) := e^{-\frac{1}{2m}\xi V^{-1}\xi^*} \qquad (\xi \in \mathbf{R}^s). \tag{4.5}$$

(ii) Now

$$E_m(\hat{a}-b) = E_m(\hat{a})\left(1 + \frac{\hat{a}V^{-1}b^*}{m} + O\left(\frac{|\hat{a}|^2|b|^2}{m^2}\right)\right)\cdot\left(1 + O\left(\frac{|b|^2}{m}\right)\right). \tag{4.6}$$

Herein we used $b = O\left(\mathrm{ld}(l+2)\right) = O(\mathrm{ld}\, m)$ and supposed without loss of generality that

$$|\hat{a}| \leqslant C\sqrt{m \log m} \quad \text{for some } C.$$

The latter is justified because in view of Remark 3 to Theorem 1.1 the Main Lemma holds trivially if $|\hat{a}| \geqslant C\sqrt{m\log m}$ and C is large. (The fact that we have to use $x = (l+1)2^m$ instead of $x = 2^m$ here does not matter, because l is of smaller order than 2^m).

Hence

$$S_l(h, a; 2^m)$$
$$= H_m E_m(\hat{a}) \sum_{g\in\mathscr{A}} \left(1 + \frac{\hat{a}V^{-1}b^*}{m} + O\text{-terms}\right)\left(L_g - \frac{i}{m}\hat{a}V^{-1}Q_g^* + \frac{i}{m} bV^{-1}Q_g^*\right) +$$
$$+ O\left(\frac{2^m}{\sqrt{m}^s m}\right).$$

The sum over the terms containing L_g is

$$\sum_{g\in\mathcal{A}} L_g + \frac{\hat{a}V^{-1}}{m}\sum_{g\in\mathcal{A}} b_l^*(g)L_g + O\text{-terms} = 1 + \frac{\hat{a}V^{-1}c_l^*}{m} + O\text{-terms},$$

the latter because

$$\sum_{g\in\mathcal{A}} b_l(g)L_g = \Big(\sum_g B(k_\nu l+g_\nu)L_g\Big)_{\nu=1}^s = \Big(\int_0^1 B([k_\nu l+k_\nu x])\,dx\Big)_{\nu=1}^s$$
$$= \Big(\frac{1}{k_\nu}\sum_{q=0}^{k_\nu-1} B(k_\nu l+q)\Big)_{\nu=1}^s.$$

All other terms contribute only to the remainder term of (4.3), because the terms of larger order add up to 0 (see Lemma 4.2(b)):

$$\sum_{g\in\mathcal{A}} \frac{-i}{m}\hat{a}V^{-1}Q_g^* = 0.$$

We are left with remainder terms of the form

$$O\left(\frac{2^m}{\sqrt{m}^s}E_m(\hat{a})\frac{|\hat{a}|^\beta|b|^\gamma}{m^\delta}\right), \qquad \beta,\gamma,\delta\in\mathbf{N}_0.$$

Because $E_m(\hat{a})\dfrac{|\hat{a}|^\beta}{\sqrt{m}^\beta} = O(1)$, these are

$$O\left(\frac{2^m|b|^\gamma}{\sqrt{m}^s m^{\delta'}}\right).$$

With $|b| = O\big(\mathrm{ld}(l+2)\big) = O(\mathrm{ld}\,m)$ these terms are seen to be

$$O\left(\frac{2^m}{\sqrt{m}^s}\,\frac{\mathrm{ld}^2(l+2)}{m}\right)$$

and the result follows.

5. Proof of the theorems. We now want to derive the theorems of Section 1 from the Main Lemma 4.4.

Proof of Theorem 1.1. (i) For

$$x = \sum_{\mu=1}^{q} 2^{m_\mu}+\delta, \qquad m_1>m_2>\ldots>m_q\geqslant 0,\ m_\mu\in\mathbf{N}_0,\ 0<\delta<1,$$

we can decompose $[0,[x])\cap\mathbf{N}_0$ thus:

$$[0,[x])\cap\mathbf{N}_0 = \sum_{\mu=1}^{q}{}' [l_\mu 2^{m_\mu}, (l_\mu+1)2^{m_\mu})\cap\mathbf{N}_0,$$

where

$$l_\mu := \Big(\sum_{\varkappa=1}^{\mu-1} 2^{m_\varkappa}\Big)/2^{m_\mu}.$$

Hence,

$$S(a;x) := \#\{0\leqslant n<x;\ \forall\nu=1,2,\ldots,s:\ B(k_\nu n)=a_\nu\}$$
$$= \sum_{\mu=1}^{q}{}' S_{l_\mu}(0,a;2^{m_\mu}) + O\left(\frac{x}{\sqrt{\mathrm{ld}\,x}^s\sqrt{\mathrm{ld}\,x}}\right),$$

where $\sum_{\mu=1}^{q}{}'$ is the restriction of $\sum_{\mu=1}^{q}$ to those terms with

$$m_\mu \geqslant \mathrm{ld}\,x - \frac{s+1}{2}\mathrm{ld}\,\mathrm{ld}\,x;$$

the rest being put into the remainder term.

The Main Lemma yields

$$S(a;x) = \sum_{\mu=1}^{q}{}' H_{m_\mu}E_{m_\mu}(\hat{a}-d_\mu) + O\left(\frac{x}{\sqrt{\mathrm{ld}\,x}^s\sqrt{\mathrm{ld}\,x}}\right), \tag{5.1}$$

where

$$\hat{a} := a - \frac{\mathrm{ld}\,x}{2}, \qquad d_\mu := \frac{m_\mu}{2} - \frac{\mathrm{ld}\,x}{2}.$$

(H_m and $E_m(\xi)$ as in (4.4) and (4.5).)

Here, the second main term of Main Lemma 4.4 does only contribute to the remainder term; namely with

$$c_{l_\mu} = O\big(\mathrm{ld}(l_\mu+2)\big) = O(m_1-m_\mu+1)$$

it yields

(5.2)
$$O\left(\sum_{\mu=1}^{q}{}' \frac{2^{m_\mu}(m_1-m_\mu+1)}{\sqrt{m_\mu}^s\sqrt{m_\mu}}\right) = O\left(\sum_{\mu=1}^{m_1}\frac{2^\mu(m_1-\mu+1)}{\sqrt{\mu}^s\sqrt{\mu}}\right) = O\left(\frac{x}{\sqrt{\mathrm{ld}\,x}^s\sqrt{\mathrm{ld}\,x}}\right).$$

(ii) In the following we may assume w.l.o.g. that

$$|\hat{a}| \leqslant C\sqrt{\mathrm{ld}\,x\log\mathrm{ld}\,x} \quad \text{for some } C$$

in view of Remark 3 to Theorem 1.1.

Now we use (4.6) with d_μ instead of b taking into account that

$$d_\mu = O(m_1 - m_\mu + 1) = O(\mathrm{ld}\,\mathrm{ld}\,x)$$

and obtain

$$E_{m_\mu}(\hat{a} - d_\mu) = E_{m_\mu}(\hat{a}) + O(d_\mu/\sqrt{m_\mu}).$$

But

$$\sum_{\mu=1}^{q}{}' \frac{2^{m_\mu} d_\mu}{\sqrt{m_\mu}^s \sqrt{m_\mu}} = O\left(\frac{x}{\sqrt{\mathrm{ld}\,x}^s \sqrt{\mathrm{ld}\,x}}\right)$$

like in (5.2). Hence (5.1) holds with $\hat{a} - d_\mu$ replaced by $\hat{a}$.

As well we can eliminate the m_μ's in $E_{m_\mu}(\hat{a})$ and the divisor of H_{m_μ} by $\mathrm{ld}\,x$. Namely,

$$\frac{1}{\sqrt{m_\mu}^s} = \frac{1}{\sqrt{\mathrm{ld}\,x}^s} \sqrt{\frac{\mathrm{ld}\,x}{m_\mu}}^s = \frac{1}{\sqrt{\mathrm{ld}\,x}^s} + O\left(\frac{1}{\sqrt{\mathrm{ld}\,x}^s} \frac{\mathrm{ld}\,\mathrm{ld}\,x}{\mathrm{ld}\,x}\right),$$

and

$$\frac{1}{m_\mu} - \frac{1}{\mathrm{ld}\,x} = O\left(\frac{\mathrm{ld}\,\mathrm{ld}\,x}{\mathrm{ld}^2 x}\right),$$

whence

$$E_{m_\mu}(\hat{a}) = E_{\mathrm{ld}\,x}(\hat{a})\, e^{-\frac{1}{2}\hat{a}V^{-1}\hat{a}^*\left(\frac{1}{m_\mu} - \frac{1}{\mathrm{ld}\,x}\right)} = E_{\mathrm{ld}\,x}(\hat{a}) + O\left(\frac{\mathrm{ld}\,\mathrm{ld}\,x}{\mathrm{ld}\,x}\right)$$

$$\left(m_\mu \geqslant \mathrm{ld}\,x - \frac{s+1}{2}\mathrm{ld}\,\mathrm{ld}\,x\right).$$

The rest is immediate by $x = \sum_{\mu=1}^{q} 2^{m_\mu} + \delta$.

Proofs of Theorem 1.2 and Theorem 1.3 (sketched). The heuristic idea for a proof suggested in Section 1, namely summing up the terms of the Main Theorem, can indeed be utilized.

We shall demonstrate this generally applicable technique on Theorem 1.3, whose proof we shall therefore treat first. For Theorem 1.2 we shall sketch a different, more elegant way relating the Fourier-transforms of this problem to those of the corresponding two-dimensional problem.

Proof of Theorem 1.3. Let, for abbreviation,

$$\bar{a} := \frac{a - \frac{\mathrm{ld}\,x}{2}}{\sqrt{\mathrm{ld}\,x}}.$$

Then

$$S_{\leqslant}(\xi; x) := \#\left\{0 \leqslant n < x;\ \forall \nu = 1, 2, \ldots, s: \frac{B(k_\nu n) - \frac{\mathrm{ld}\,x}{2}}{\sqrt{\mathrm{ld}\,x}} \leqslant \xi_\nu\right\}$$

$$= \sum_{a:\bar{a} \leqslant \xi} S(a; x).$$

We cannot immediately insert the right side of the result in Theorem 1.1 for $S(a; x)$ because the remainder terms would add up desastrously. However, it suffices to sum over those a with $\bar{a}$ componentwise "small" in modulus:

$$\sum_{\substack{a: \\ \exists \nu: |\bar{a}_\nu| \geqslant C\sqrt{\log \mathrm{ld}\,x}}} S(a; x) = O\left(\frac{x}{\sqrt{\mathrm{ld}\,x}}\right)$$

for large C by Remark 3 to Theorem 1.1.

Summing up $O\left(\frac{x}{\sqrt{\mathrm{ld}\,x}^s \sqrt{\mathrm{ld}\,x}}\right)$ over all $a \in \mathbf{Z}^s$ with $|\bar{a}_\nu| \leqslant C\sqrt{\log \mathrm{ld}\,x}$ $\forall \nu$ would still yield an estimate too large by the factor $\sqrt{\log \mathrm{ld}\,x}^s$. But a review of the proof of Theorem 1.1 from the Main Lemma shows that all O-terms larger than $O\left(\frac{x}{\sqrt{\mathrm{ld}\,x}^s} \frac{\mathrm{ld}^2 \mathrm{ld}\,x}{\mathrm{ld}\,x}\right)$ originated from terms of the form $\frac{H_{m_\mu}}{\sqrt{m_\mu}} E_{m_\mu}(\hat{a}) \frac{\hat{a}}{\sqrt{m_\mu}}(\ldots)$, namely when estimating the second main term of the Main Lemma and $E_{m_\mu}(\hat{a} - d_\mu)$.

Therefore we may sum up these terms even over all $a \in \mathbf{Z}^s$ without getting a remainder term worse than $O(x/\sqrt{\mathrm{ld}\,x})$; and $O\left(\frac{x}{\sqrt{\mathrm{ld}\,x}^s} \frac{\mathrm{ld}^2 \mathrm{ld}\,x}{\mathrm{ld}\,x}\right)$ summed up over all a with $|\bar{a}_\nu| \leqslant C\sqrt{\log \mathrm{ld}\,x}$ $\forall \nu$ yields $O\left(\frac{x}{\sqrt{\mathrm{ld}\,x}}\right)$, too.

Thus we obtain

$$S_{\leqslant}(\xi; x) = \frac{x}{\sqrt{2\pi}^s \sqrt{\det V}} \sum_{a:\bar{a} \leqslant \xi} \frac{1}{\sqrt{\mathrm{ld}\,x}^s} E_{\mathrm{ld}\,x}(\hat{a}) + O\left(\frac{x}{\sqrt{\mathrm{ld}\,x}}\right).$$

Going over from this Riemann-sum to the respective integral yields the desired result.

Proof of Theorem 1.2. Given $s = 2$ and k_1, k_2, for $h \in \mathscr{A}$ and $\tilde{a} \in \mathbf{Z}$

$$\tilde{S}_l(h, \tilde{a}; 2^m) := \#\{0 \leqslant n < 2^m; B(k_1 n + h_1) - B(k_2 n + h_2) = \tilde{a}\} = \sum_{\tilde{b} \in \mathbf{Z}} S_l\big(h, (\tilde{a} + \tilde{b}, \tilde{b}); 2^m\big).$$

We observe that

$$\big(\tilde{\varphi}_{l,m}(\tilde{t})\big)_h := \sum_{\tilde{a} \in \mathbf{Z}} \tilde{S}_l(h, \tilde{a}; 2^m) e^{i\tilde{a}\tilde{t}} = \big(\varphi_{l,m}\big((\tilde{t}, -\tilde{t})\big)\big)_h;$$

i.e. the Fourier-transform of $\tilde{S}_l(h, \tilde{a}; 2^m)$ at $\tilde{t} \in \mathbf{R}$ equals the Fourier-transform of the respective two-dimensional problem at $t = (\tilde{t}, -\tilde{t})$.

Moreover, let $\tilde{\Phi}(\tilde{t})$, $\tilde{\lambda}(\tilde{t})$, and $\tilde{P}(\tilde{t})$ be defined for $\tilde{S}_l(h, \tilde{a}; 2^m)$ as $\Phi(t)$, $\lambda(t)$, and $P(t)$ were defined for $S_l(h, a; 2^m)$. Then

$$\tilde{\Phi}(\tilde{t}) = \Phi\big((\tilde{t}, -\tilde{t})\big),$$
$$\lambda(\tilde{t}) = \lambda\big((\tilde{t}, -\tilde{t})\big),$$
$$\tilde{P}(\tilde{t}) = P\big((\tilde{t}, -\tilde{t})\big);$$

and the whole proof of Theorem 1.1 is translated straightforward into a proof of Theorem 1.2.

Acknowledgement. I am greatly indebted to Professor E. Wirsing of Ulm University. An earlier version of this paper was written as a doctoral thesis on his suggestion and advice.

References

[1] Kai Lai Chung, *Markov Chains with Stationary Transition Probabilities*, Springer-Verlag, Berlin–Heidelberg 1967.

[2] F. R. Gantmacher, *The Theory of Matrices*, Chelsea Publishing Company, New York, N. Y. 1960.

[3] Thomas Höglund, *Central limit theorems and statistical inference for finite Markov chains*, Z. Wahrscheinlichkeitstheorie verw. Gebiete 29 (1974), pp. 123–151.

[4] I. Kátai, *Change of the sum of digits by multiplication*, Acta Sci. Math. 39 (1977), pp. 319–328.

[5] Tosio Kato, *Perturbation Theory for Linear Operators*, Springer-Verlag, New York 1966.

[6] Johannes Schmid, *The maximum modulus eigenvalues of certain non-negative matrices under perturbation*, Analysis 2 (1982), pp. 163–175.

[7] Wolfgang M. Schmidt, *The joint distribution of the digits of certain integer s-tuples* (to appear).

[8] K. B. Stolarsky, *Integers whose multiples have anomalous digital frequencies*, Acta Arith. 38 (1980), pp. 117–128.

[9] K. B. Stolarsky and J. B. Muskat, *The number of binary digits in multiples of n* (to appear).

[10] Helmut Wielandt, *Unzerlegbare, nicht negative Matrizen*, Math. Zeitschr. 52 (1950), pp. 642–648.

UNIVERSITÄT ULM
ABTEILUNG MATHEMATIK II
Oberer Eselsberg, Postfach 4066
D-7900 Ulm (West Germany)

Received on 19.11.1982 (1328)