Locally Restricted Compositions II. General Restrictions and Infinite Matrices

Edward A. Bender
Department of Mathematics
University of California, San Diego
La Jolla, CA 92093-0112
ebender@ucsd.edu

E. Rodney Canfield*
Department of Computer Science
University of Georgia
Athens, GA 30602
erc@cs.uga.edu

Submitted: Feb 11, 2009; Accepted: Aug 14, 2009; Published: Aug 21, 2009
AMS Subject Classification: 05A15, 05A16

Abstract

We study compositions $\vec{\mathbf{c}} = (c_1, \ldots, c_k)$ of the integer n in which the value c_i of the ith part is constrained based on previous parts within a fixed distance of c_i. The constraints may depend on i modulo some fixed integer m. Periodic constraints arise naturally when m-rowed compositions are written in a single row. We show that the number of compositions of n is asymptotic to Ar^{-n} for some A and r and that many counts can be expected to have a joint normal distribution with means vector and covariance matrix asymptotically proportional to n. Our method of proof relies on infinite matrices and does not readily lead to methods for accurate estimation of the various parameters. We obtain information about the longest run. In many cases, we obtain almost sure asymptotic estimates for the maximum part and number of distinct parts.

1 Introduction

Carlitz compositions are compositions in which adjacent parts are distinct. We were led to this work by proposing a generalization of ordinary and Carlitz compositions which we call *regular, locally restricted compositions.* Roughly speaking, locally restricted compositions are defined by looking at pairs of parts in a moving window and regularity deals with the recurrence of patterns in a composition. Precise definitions are given in the next two sections.

Example 1 (Carlitz-type compositions) In [2] we studied compositions in which the difference between adjacent parts must lie in a set D. Such compositions are locally

*Research supported by NSA Mathematical Sciences Program.

restricted but may not be regular. Additional conditions were imposed on D. For example, D must contain both positive and negative integers. (If D were the nonnegative integers we would be studying partitions which, as we shall see later, are not regular.) When D consists of all nonzero integers, the result is *Carlitz compositions* [13].

The methods in [2] are inadequate for dealing with more general locally restricted compositions. One such example are what we call *two-rowed Carlitz compositions.* A two-rowed Carlitz composition of n is an array

$$\begin{matrix} c_{1,1} & c_{1,2} & \cdots & c_{1,k} \\ c_{2,1} & c_{2,2} & \cdots & c_{2,k} \end{matrix}$$

of positive integers whose parts sum to n, such that vertically and horizontally adjacent parts are distinct. The parts can be written out in one dimension in column order, $c_{1,1}, c_{2,1}, c_{1,2}, \ldots$. Those parts in an odd position are required to be different from their neighbor two positions earlier, and those in an even position are required to be different from both their two previous neighbors; also, the number of parts is required to be even. These compositions are regular and locally restricted and so, as a consequence of our Theorem 3, the counting sequence a_n satisfies

$$a_n \sim Ar^{-n},\ n \to \infty. \tag{1}$$

The proof of Theorem 3 does not seem to provide an efficient method for estimating A or r. Computations of a_n through $n = 100$ suggest the values $A \doteq 0.284\cdots$ and $r \doteq 0.590\cdots$.

When two-rowed Carlitz compositions counted by a_n are sampled randomly, our results show that many counts have a joint distribution which is asymptotically normal, having means vector and covariance matrix asymptotically proportional to n. However, we have no reasonable method for estimating the limits. Examples of such counts are the number of columns, the number of fives, the number of odd parts, and the number of times columns three apart are identical.

These concepts and results extend to m-rowed Carlitz compositions, which may require either adjacent elements in a column to be distinct or all elements in a column to be distinct.

Another generalization of Carlitz compositions, studied by Munarini, Poneti and Rinaldi[15], require that adjacent columns differ (rather than adjacent parts). Again, our results apply when the parts are strictly positive; however, we do not obtain explicit generating functions as Munarini *et al.* do. They also allow parts to be zero, but require that no column be zero. Again, our results apply since the transfer matrix $T(x)$, which is defined later, still satisfies Theorem 1. □

Example 2 (Palindromes) Palindromes are compositions that read the same in both directions. For two-rowed compositions, the palindromes with k columns may be either those with $c_{i,j} = c_{i,k+1-j}$ for $i = 1, 2$ and $1 \leqslant j \leqslant k$ or those with $c_{1,j} = c_{2,k+1-j}$ for $1 \leqslant j \leqslant k$. Palindromes are not locally restricted since parts arbitrarily far apart must be equal. Nevertheless, we can apply our methods to the study of the set of palindromes in a collection of regular, locally restricted compositions. □

Example 3 (Avoiding or forcing patterns) The notion of a "pattern" in a composition may be limited to adjacent parts or may allow arbitrarily many intervening parts. The former is a local condition whereas the latter is not. Our results usually apply to the local form, both for avoiding, requiring and counting patterns. We have said "usually" because there is a recurrence condition, which roughly says that any pattern that has been seen in the interior of a composition can be seen again. For example, partitions can be described by pattern avoidance, but violate the recurrence condition. The paper [12] by Kitaev, McAllister and Petersen contains some explicit generating functions for local patterns. Savage and Wilf [19] study the non-local situation. Heubach, Kitaev and Mansour [10] count compositions which avoid certain patterns using recursion formulas for the generating functions. When the patterns to be avoided each consist of a sequence of specific, adjacent parts, Myers [16] and Heubach and Kitaev [9] obtain explicit generating functions containing $k \times k$ determinants when there are k sequences to be avoided. $\square$

Example 4 (Periodic local constraints) Suppose we have a periodic local constraint; e.g., $a_{3k+1} < a_{3k+2} < a_{3k+3} > a_{3k+4}$. In such a situation, one may want to restrict the number of parts to be some value modulo the period or may not wish to do so. If Theorem 3 applies (as it does in this example), then the value of r in the theorem will be the same in all cases, but the value of A may change. This is also true if we shift the period; e.g., $a_{3k} < a_{3k+1} < a_{3k+2} > a_{3k+3}$.

This follows because the transfer matrix T in Section 2.3 is unchanged but the vectors $\mathbf{s}$ and/or $\mathbf{f}$ are changed. The dominant eigenvalue of T determines r. $\square$

Lest the reader assume that our results apply only to "reasonable" restrictions, we hasten to point out they are more general. For example, we might require that every three adjacent parts sum to a prime unless at least one of the parts is a sum of two cubes, though why one would want to do this is unclear.

These examples did not discuss the counting of local events. See Theorem 4 at the end of this section.

We conclude this section with a statement of the main results. As noted earlier, some of the terminology will not be defined until later sections. Nevertheless, we believe stating the results now will give the reader the flavor of the paper without the need to plow through later sections.

Consideration of infinite matrices, essentially infinite-state machines for constructing all compositions of a prescribed class, led to general conditions implying (1). The main tool is isolated in Theorem 1, which concerns only infinite matrices, and nothing of a combinatorial nature. These matrices are associated with generating functions having the most basic analytic behavior: a single simple pole on the real axis. Theorem 2 asserts that Theorem 1 is applicable to the enumeration of regular, locally restricted compositions. We anticipate Theorem 1 will be applicable to counting sequences of other combinatorial objects, leading to their asymptotic form and asymptotic normality.

The complex, infinite matrices T and vectors $\mathbf{v}$ used in this paper are absolutely square summable:

$$||T||_2 \;=\; \Big(\sum_{i,j} |T_{i,j}|^2\Big)^{1/2} \;<\; \infty, \qquad \sum_i |(\mathbf{v})_i|^2 \;<\; \infty.$$

Define $\mathcal{M}$ to be the class of complex, infinite matrices T whose entries are absolutely square summable, and as usual let ℓ^2 denote the class of square summable, complex vectors. The classes $\mathcal{M}(\Omega)$ and $\ell^2(\Omega)$, where $\Omega \subseteq \mathbb{C}$ is a domain, will be defined shortly. These are natural extensions of $\mathcal{M}$ and ℓ^2 to matrices and vectors whose components are functions holomorphic in Ω.

Absolute value and weak inequality of matrices and vectors are componentwise:

$$|T|_{i,j} = |T_{i,j}| \text{ for all } i, j \quad \text{and} \quad T \leqslant S \text{ means } T_{i,j} \leqslant T_{i,j} \text{ for all } i, j.$$

Strong inequality $T < S$ means $T \leqslant S$ and $T \neq S$.

Definition 1 (Recurrent matrix) *The matrix T is* recurrent *provided that*

(1) for each i, j there exists k such that $(T^k)_{ij} \neq 0$ and

(2) for each j_1, j_2 there exist k, i such that $(T^k)_{ij_1} \neq 0$ and $(T^k)_{ij_2} \neq 0$.

Theorem 1 (Dominant eigenvalues) *Let $\rho > 0$ and let T_n be a sequence of infinite matrices. Suppose that the power series*

$$T(x) = xT_1 + x^2T_2 + \cdots$$

satisfies:

(a) $\sum_n |x|^n \, ||T_n||_2$ is convergent for $|x| < \rho$,

(b) $T_n \geqslant 0$,

(c) $T(x_0)$ is recurrent for all $x_0 \in (0, \rho)$.

Then for each $x_0 \in (0, \rho)$ the matrix $T(x_0)$ has an eigenvalue $\lambda(x_0) > 0$ which is simple and strictly larger in absolute value than the other eigenvalues of $T(x_0)$. On the interval $(0, \rho)$ the function $\lambda(x)$ is analytic and $\lambda'(x) > 0$.

Assume further that we have $r \in (0, \rho)$, an integer k_0, and functions $\mathbf{s}(x), \mathbf{f}(x) \in \ell^2(|x| < \rho)$ such that:

(d) $\lambda(r) = 1$,

(e) $\mathbf{s}(r), \mathbf{f}(r) > \mathbf{0}$,

(f) $|T(x)^{k_0}| < T(|x|)^{k_0}$ for $x \neq \pm|x|$, $0 < |x| < \rho$.

Then the function

$$\phi(x) = \mathbf{s}(x)^t \Big(\sum_{k=0}^{\infty} T(x)^k\Big)\, \mathbf{f}(x)$$

is analytic for $|x| < r$, has a simple pole at $x = r$, and has at most one additional singularity on the circle of convergence at $x = -r$.

Precise definitions of *locally restricted* and *regular* are in Definitions 4 and 9.

Theorem 2 (Compositions and matrices) *Let $\mathcal{C}$ be a regular, locally restricted class of compositions, and let $F(x)$ be the ordinary generating function (ogf) for $\mathcal{C}$. There is a power series $T(x) = xT_1 + x^2T_2 + \cdots$ satisfying hypotheses (a)–(c) of Theorem 1 with $\rho = 1$, as well as $k_0, r, \mathbf{s}(x), \mathbf{f}(x)$ satisfying (d)–(f), such that*

$$F(x^2) = \phi(x) + F_{NR}(x^2),$$

where

$$\phi(x) = \mathbf{s}(x)^t \left(\sum_{k=0}^{\infty} T(x)^k\right) \mathbf{f}(x)$$

and $F_{NR}(x)$ has radius of convergence at least 1. ($F_{NR}(x)$ is the ogf for a subclass of $\mathcal{C}$.)

Theorem 3 (Asymptotic number of compositions) *Let $\mathcal{C}$ be a regular, locally restricted class of compositions, and let a_n be the number of compositions of n in the class $\mathcal{C}$. Then $a_n \sim Ar^{-n}$ for some $A > 0$ and $r < 1$. Furthermore $a_n = Ar^{-n}(1 + O(\delta^n))$ for some $0 < \delta < 1$.*

Roughly speaking *recurrent events* are events that can occur arbitrarily often in $\mathcal{C}$. Recurrent events are *related* if a linear combination of their counts is always nearly a (possibly zero) multiple of the sum of parts. Precise definitions are given in Definitions 14 and 15 of Section 8.

Theorem 4 (Asymptotic normality) *Let $\mathcal{C}_n$ be the compositions of n in $\mathcal{C}$ made into a probability space with the uniform distribution. Let the random variables $Y_i(n)$, $1 \leqslant i \leqslant \kappa$ count occurrences of recurrent local events. Then $\mathrm{E}(Y_i(n)) = nm_i + o(n)$ where $m_i > 0$. Let $\vec{Z}(n) = n^{-1/2}\Big(\vec{Y}(n) - \mathrm{E}(\vec{Y}(n))\Big)$. If the $Y_i(n)$ are unrelated, then $\vec{Z}(n)$ converges in distribution to a k-dimensional normal.*

With further restrictions on the $Y_i(n)$, it would be possible to extend this central limit theorem to a local limit theorem, but we have not worked out the details.

Let the random variable M_n (resp. D_n) be the largest part (resp. number of distinct parts) in a locally restricted composition of n selected uniformly at random. We show that $M_n \leqslant (1+o(1))\log_{1/r}(n)$ almost surely and that often $M_n \sim D_n \sim \log_{1/r}(n)$ almost surely. See Section 9 for details and further results. That section can be read after Section 2.

Suppose k copies of $\vec{\mathbf{p}}$ are adjacent in a composition. This is a *run* of $\vec{\mathbf{p}}$. If it does not have $\vec{\mathbf{p}}$ on either side, it is a maximal run and its length is k.

Theorem 5 (Run lengths) *If a locally restricted composition can have arbitrarily long runs of* $\vec{\mathbf{p}}$*, then the length of the longest run of* $\vec{\mathbf{p}}$ *is almost surely asymptotic to*

$$\frac{\log_{1/r}(n)}{\Sigma(\vec{\mathbf{p}})} \text{ where } \Sigma(\vec{\mathbf{p}}) \text{ is the sum of the parts in } \vec{\mathbf{p}}. \tag{2}$$

Let $\mathcal{R}$ *be a set of* $\vec{\mathbf{p}}$ *such that the restricted compositions can have arbitrarily long runs of* $\vec{\mathbf{p}}$*. Then,*

(a) *The longest run in a random composition will almost surely be due to a composition* $\vec{\mathbf{p}} \in \mathcal{R}$ *for which* $\Sigma(\vec{\mathbf{p}})$ *is a minimum.*

(b) *If* $\mathcal{R}$ *is finite, the run with the greatest number of parts will almost surely be due to a composition* $\vec{\mathbf{p}} \in \mathcal{R}$ *for which the average part size,* $\Sigma(\vec{\mathbf{p}})/\mathrm{len}(\vec{\mathbf{p}})$*, is a minimum.*

The last part of the theorem implies that for many local restrictions the longest run is almost surely repetitions of the part 1 and the length of that run is almost surely asymptotic to $\log_{1/r}(n)$. To see that some restriction on $\mathcal{R}$ is needed in (b), consider unrestricted compositions and let $\mathcal{R}$ be those compositions where the number of parts is a power of 2, $p_k = 2$ if k is a power of 2 and $p_k = 1$ otherwise. It follows from (2) that as $n \to \infty$ the longest run in the sense of (b) will involve longer and longer compositions in $\mathcal{R}$. The theorem is proved in Section 10, which can be read after Sections 2 and 9.

We thank the referee for suggesting that we consider runs.

2 Basic Concepts: Compositions

Let $\mathbb{N}$ denote the natural numbers $\{1, 2, \ldots\}$, $\mathbb{N}_0$ denote $\mathbb{N} \cup \{0\}$, and $\mathbb{Z}$ denote the integers $\{\cdots, -1, 0, 1, \cdots\}$.

Definition 2 (Composition notation) *A composition of the integer* n *into* k *parts is a* k*-tuple of* strictly positive *integers summing to* n*; that is,* $(c_1, \ldots, c_k)$*,* $c_i \in \mathbb{N}$*, such that* $\sum_{i=1}^{k} c_i = n$*. We write compositions in vector notation,* $\vec{\mathbf{c}}$*. The* sum n *and* number of parts k *are denoted* $\Sigma(\vec{\mathbf{c}})$ *and* $\mathrm{len}(\vec{\mathbf{c}})$*, respectively. We adopt the convention that* $c_i = 0$ *when* $i \leqslant 0$ *or* $i > k$*. The* empty composition*,* $\vec{\mathbf{e}}$*, is the only composition of 0, and has no parts. Thus* $\Sigma(\vec{\mathbf{e}}) = 0$*,* $\mathrm{len}(\vec{\mathbf{e}}) = 0$ *and* $c_i = 0$ *for all* i*.*

2.1 Local Restrictions

We impose additional constraints on compositions that can be tested by looking in a moving window at parts of the composition. The desired constraints are encoded in a "local restriction function" as described in the next two definitions.

Definition 3 (Local restriction function) *Let* $m, p \in \mathbb{N}$*. A* local restriction function of type (m, p) *is a function*

$$\Phi : \{0, 1, \ldots, m-1\} \times (\mathbb{N}_0)^{p+1} \to \{0, 1\}$$

with $\Phi(i;0,\ldots,0)=1$ *for all* i*. The integers* m *and* p *are called, respectively, the* modulus *and* span *of* Φ*.*

Definition 4 (Class of compositions determined by Φ) *Let* Φ *be a local restriction function. The* class of compositions determined by Φ *is*

$$\mathcal{C}_\Phi = \{\vec{\mathbf{c}} : \vec{\mathbf{c}} \text{ is a composition, and } \Phi(i \bmod m; c_i, c_{i-1}, \ldots, c_{i-p}) = 1 \text{ for } i \in \mathbb{Z}\}.$$

A class $\mathcal{C}$ *of compositions is* locally restricted *if* $\mathcal{C} = \mathcal{C}_\Phi$ *for some local restriction function* Φ*.*

Example 5 (Encoding properties) It might appear that the condition $c_i > 0$ could be encoded in Φ, but this is not the case—separate two compositions by a string of zeroes whose length exceeds the span: $\vec{\mathbf{c}}, 0, \ldots, 0, \vec{\mathbf{d}}$.

On the other hand divisibility conditions on the number of parts can be encoded because of the zeroes at the ends of a composition: If $\Phi(i;0,a_1,\ldots,a_p)=0$ whenever $a_1>0$ and $i \notin S$, then $\operatorname{len}(\vec{\mathbf{c}})+1$ modulo m is in S for all nonempty $\vec{\mathbf{c}} \in \mathcal{C}_\Phi$.

The zeroes also allow the encoding of special conditions at the beginning and end. For example, $\Phi(i;0,a_1,\ldots,a_p)=0$ whenever $a_1 \neq k$ ensures that the last part in a composition is k. $\square$

Example 6 (Adjacent differences) Let $\mathcal{D} \subseteq \mathbb{Z}$, and consider the class $\mathcal{C}$ of compositions $\vec{\mathbf{c}} = (c_1, c_2, \ldots, c_k)$ such that $c_i - c_{i-1} \in \mathcal{D}$ for $1 < i \leqslant k$. We may take $m=1$ and $\Phi(0;j,k)=1$ if and only if $jk=0$ or $j-k \in \mathcal{D}$. $\square$

Example 7 (m-rowed Carlitz compositions) Suppose our composition consists of m rows, say $b_{i,j}$ where $1 \leqslant i \leqslant m$ and $1 \leqslant j \leqslant \ell$ (m is fixed but ℓ is not). Adjacent parts are required to be different: $b_{i,j} \neq b_{i-1,j}$ for $i>1$ and $b_{i,j} \neq b_{i,j-1}$ for $j>1$. We convert it to a standard composition by writing the parts in column order:

$$b_{1,1}, b_{2,1}, \ldots, b_{m,1}, b_{1,2}, \ldots, b_{m,\ell} \;=\; c_1, \ldots, c_{m\ell}.$$

The modulus and span of Φ are m and the local restrictions are of three types. First there are those to force the number of parts to be a multiple of m. It suffices to set $\Phi(i;0,a_1,\ldots,a_m)=0$ when $i=1$ and $a_1 \neq 0$. Second there are those to force adjacent parts in row to be different: for all i, $\Phi(i;a_0,\ldots,a_m)=0$ when $a_0=a_m \neq 0$. Finally there are those to force adjacent parts in the same column to be different. It suffices to set $\Phi(i;a_0,\ldots,a_m)=0$ when $i \neq 0$ and $a_0=a_1 \neq 0$. $\square$

Example 8 (Distance d compositions) These are compositions having the property that within every window of width d or less there is no repeated part. (Of course, this means no repeated positive part; the imaginary leading and trailing zeros must be exempt from the no-repeat rule.) The case $d=1$ are traditional compositions; the case $d=2$ Carlitz compositions. It is straightforward to construct a Φ with modulus 1 and span $d-1$. Our theorems apply, but we are unable to obtain generating functions or effectively estimate constants when $d>2$. It would be interesting to have a direct combinatorial approach to the generating function in the case $d=3$. $\square$

Given a local restriction function Φ with span p and modulus m, it is clear that there is an equivalent (meaning defining the same class of compositions) local restriction function with any larger span desired. Likewise, there are equivalent local restriction functions whose modulus is any multiple of m. Consequently, for any class $\mathcal{C}_\Phi$, one may assume that Φ has equal span and modulus.

We will generally assume that modulus = span

and denote the common value by m.

2.2 The Digraph D_Φ and Recurrent Subcompositions

We shall define a digraph D_Φ, naturally associated with Φ, with the property that certain directed paths in D_Φ correspond bijectively with the compositions in $\mathcal{C}_\Phi$.

Let Φ be a local restriction function with modulus and span m. Define a *word* to be an m-tuple of integers. We distinguish compositions and words notationally by the use of bold: $\vec{\mathbf{c}}$ denotes a composition, and $\vec{c}$ denotes a word. We say that a word $\vec{\nu}$ *appears in* the composition $\vec{\mathbf{c}}$ if for some $i \equiv 0 \bmod m$ we have $c_{i+j} = \nu_j$ for $1 \leqslant j \leqslant m$. (In this definition it may be necessary to observe the convention about the meaning of c_i when $i > \text{len}(\vec{\mathbf{c}})$.) For example, when $m = 2$ the words in $c_1c_2c_3c_4c_5$ are 00, c_1c_2, c_3c_4 and c_50. Note that if zero appears in a word in $\vec{\mathbf{c}}$, then numbers to its right are also zero.

Define the vertex set $V(D_\Phi)$ to be all words which appear in some $\vec{\mathbf{c}} \in \mathcal{C}_\Phi$. Define the edge set $E(D_\Phi)$ to be all ordered pairs $(\vec{\nu}, \vec{\tau})$ of words which can be adjacent in some composition. The precise definition is the following.

Definition 5 (The digraph D_Φ) *Let Φ be a local restriction function whose span and modulus equal m. The vertex set $V(D_\Phi)$ of D_Φ consists of all words $\vec{\nu}$ of length m which appear in some composition of $\mathcal{C}_\Phi$. The edge set $E(D_\Phi)$ is all pairs $(\vec{\nu}, \vec{\tau})$ such that*

$$\Phi(i, \tau_i, \tau_{i-1}, \dots, \tau_1, \nu_m, \dots, \nu_{m-i+1}) = 1 \text{ for } 1 \leqslant i \leqslant m; \tag{3}$$

in other words, $\vec{\nu}\vec{\tau}$ can appear in a composition in $\mathcal{C}_\Phi$. We allow loops in D_Φ. (In fact, D_Φ always contains the edge $(\vec{0}, \vec{0})$.)

Notice that (3) is the same as saying that if the $2m$ long sequence $\nu_1, \dots \nu_m, \tau_1, \dots, \tau_m$ were part of a composition, starting at a position which is congruent mod m to 1, then the local restriction function Φ is satisfied when only $\vec{\nu}$ and $\vec{\tau}$ are within the span.

Definition 6 (Path in D_Φ) *A $(\vec{\nu}, \vec{\tau})$-*path *is a path π in the digraph D_Φ such that*

- *the initial and final vertices of π are $\vec{\nu}$ and $\vec{\tau}$, respectively;*
- *π includes at least one edge;*
- *the vertex $\vec{0}$ is not an interior vertex of π.*

The set of all $(\vec{\nu}, \vec{\tau})$-paths is denoted $\mathrm{Path}_\Phi(\vec{\nu}, \vec{\tau})$. *We* allow repeated vertices and edges *in paths. (In graph theory what we are calling a path here is often referred to as a walk.)*

It is easily seen that *there is a bijection between* $\mathrm{Path}_\Phi(\vec{0}, \vec{0})$ *and* $\mathcal{C}_\Phi$: The path $\vec{0}, \vec{\nu}_1, \ldots, \vec{\nu}_k, \vec{0}$ corresponds to the composition obtained by concatenating the $\vec{\nu}_i$. In particular, the path $\vec{0}, \vec{0}$ corresponds to the empty composition. We may think of $\vec{\nu}_1, \ldots, \vec{\nu}_k$ as a kind of "super" composition with parts in $\mathbb{N} \times \mathbb{N}_0^{m-1}$. The local restrictions of $\mathcal{C}_\Phi$ become adjacent restrictions for the parts of these super compositions.

Definition 7 (Recurrent vertex in D_Φ) *A vertex $\vec{\nu} \in V(D_\Phi)$ is* recurrent *if $\vec{\nu} \neq \vec{0}$ and* $\mathrm{Path}_\Phi(\vec{\nu}, \vec{\nu}) \neq \emptyset$. *Since vertices are words, we also speak of recurrent words.*

It can be checked that a vertex $\vec{\nu}$ is recurrent if and only if there is a composition $\vec{\mathbf{c}} \in \mathcal{C}_\Phi$ in which the word $\vec{\nu}$ appears at least twice. If a vertex $\vec{\nu}$ is recurrent, then there are obviously paths in $\mathrm{Path}_\Phi(\vec{\nu}, \vec{\nu})$ which contain $\vec{\nu}$ arbitrarily often and so there are compositions $\vec{\mathbf{c}} \in \mathcal{C}_\Phi$ containing the word $\vec{\nu}$ arbitrarily often.

2.3 The Transfer Matrix and Generating Function

We assume that $V(D_\Phi)$ contains recurrent vertices. Recall that $\vec{0}$ is not considered a recurrent vertex.

Let a_n be the number of compositions of n belonging to $\mathcal{C}_\Phi$, and let $F(x)$ be the ogf (ordinary generating function) of the numbers a_n:

$$F(x) = \sum_{n \geqslant 0} a_n x^n = \sum_{\vec{\mathbf{c}} \in \mathcal{C}_\Phi} x^{\Sigma(\vec{\mathbf{c}})}.$$

Let $F_{NR}(x)$ be the ogf for those compositions containing no recurrent words, and let $F_R(x)$ be the ogf for those compositions containing at least one recurrent word. Thus, $F(x) = F_R(x) + F_{NR}(x)$. In the compositions counted by $F_R(x)$ one may speak unambiguously of the first recurrent word and the last recurrent word in the composition. (These might be the same.)

Definition 8 (Transfer matrix associated with Φ) *Let Φ be a local restriction function, and let $\vec{\nu}_1, \vec{\nu}_2, \ldots$ be an ordered listing of all recurrent vertices in $V(D_\Phi)$, fixed once and for all. Define the* transfer matrix $T(x)$ associated with Φ *by*

$$(T(x))_{ij} = \begin{cases} x^{\Sigma(\vec{\nu}_i)+\Sigma(\vec{\nu}_j)} & \textit{if } (\vec{\nu}_i, \vec{\nu}_j) \in E(D_\Phi), \\ 0 & \textit{otherwise.} \end{cases} \tag{4}$$

By induction we find that for all $k \geqslant 1$

$$(T(x)^k)_{ij} = x^{\Sigma(\vec{\nu}_i)+\Sigma(\vec{\nu}_j)} \sum_{\pi} x^{2\Sigma(\vec{c}_1)+\cdots+2\Sigma(\vec{c}_{k-1})}, \tag{5}$$

in which the sum is over all paths

$$\pi = (\vec{\nu}_i, \vec{c}_1, \ldots, \vec{c}_{k-1}, \vec{\nu}_j)$$

belonging to $\mathrm{Path}_\Phi(\vec{\nu}_i, \vec{\nu}_j)$ and containing k edges.

Define the *start vector* $\mathbf{s}(x)$ by:

$$(\mathbf{s}(x))_i = x^{\Sigma(\vec{\nu}_i)} \sum_\pi x^{2\Sigma(\vec{c}_1)+\cdots+2\Sigma(\vec{c}_\ell)}, \tag{6}$$

where the sum is over all paths

$$\pi = (\vec{0}, \vec{c}_1, \ldots, \vec{c}_\ell, \vec{\nu}_i)$$

belonging to $\mathrm{Path}_\Phi(\vec{0}, \vec{\nu}_i)$ that contain only one recurrent vertex, the endpoint $\vec{\nu}_i$. In like manner the *finish vector* $\mathbf{f}(x)$ is defined by

$$(\mathbf{f}(x))_j = x^{\Sigma(\vec{\nu}_j)} \sum_\pi x^{2\Sigma(\vec{c}_1)+\cdots+2\Sigma(\vec{c}_\ell)}, \tag{7}$$

where the sum is over all paths

$$\pi = (\vec{\nu}_j, \vec{c}_1, \ldots, \vec{c}_\ell, \vec{0})$$

belonging to $\mathrm{Path}_\Phi(\vec{\nu}_j, \vec{0})$ that contain only one recurrent vertex, the initial vertex $\vec{\nu}_j$.

Every path in D_Φ that contains a recurrent vertex may be uniquely parsed into the list of vertices (μ_1, ν, μ_2), where the (possibly empty) parts μ_1 and μ_2 contain no recurrent vertices and ν begins and ends with a recurrent vertex and may consist of just a single vertex. Keeping in mind the definition of $F_R(x)$ and the combinatorial interpretation (5) of $(T(x)^k)_{ij}$, we conclude that

$$F_R(x^2) = \mathbf{s}(x)^{\mathrm{t}} \sum_{k=0}^{\infty} T(x)^k \, \mathbf{f}(x). \tag{8}$$

Note that $F_R(x^2)$ counts each part of a composition twice, so to speak. It is not hard to arrange for a similar formula in which we obtain $F_R(x)$ on the left side, but the definition of $T(x)$ arising in that construction fails to have the desirable feature (from the functional analysis viewpoint) that its entries are square summable. We rely on this latter property to assure that the operator $T(x)$ is compact.

Definition 9 (Regularity) *Let $\mathcal{C} = \mathcal{C}_\Phi$ be a locally restricted class of compositions, for which Φ has both span and modulus equal to m. We say that Φ is* regular *provided:*

(R1) *The directed graph within D_Φ spanned by the recurrent vertices contains at least two vertices and is strongly connected. (Recall that $\vec{0}$ is* not *recurrent.)*

(R2) *Given any two recurrent vertices $\vec{\nu}_1, \vec{\nu}_2 \in V(D_\Phi)$ there is always a third recurrent vertex $\vec{\nu}_3$ and an integer k such that both* $\mathrm{Path}_\Phi(\vec{\nu}_3, \vec{\nu}_1)$ *and* $\mathrm{Path}_\Phi(\vec{\nu}_3, \vec{\nu}_2)$ *contain a path of length k.*

(R3) *There is an integer $k > 0$ and (possibly equal) recurrent vertices $\vec{\nu}_1$ and $\vec{\nu}_2$ such that*

$$\gcd\{m - n : m, n \in S\} \;=\; 1,$$

where

$$S \;=\; \Big\{n : n = \Sigma(\vec{c}_1) + \cdots + \Sigma(\vec{c}_{k-1}) \text{ for some } \vec{\nu}_1, \vec{c}_1, \ldots, \vec{c}_{k-1}, \vec{\nu}_2 \in \mathrm{Path}_\Phi(\vec{\nu}_1, \vec{\nu}_2)\Big\}.$$

(Thus k is the length of the path.)

(R4) *There exists a constant K such that any path π in D_Φ of length more than K contains at least one recurrent vertex.*

If Φ is regular, we say that $\mathcal{C}_\Phi$ is regular.

These are the compositions discussed in the theorems of Section 1. The locally restricted compositions discussed in earlier examples are regular, provided the differences D in Example 1 are appropriately restricted.

Example 9 (Comments on Definition 9) Partitions fail to satisfy (R1) because there are no recurrent vertices. Consider the composition where the first part is arbitrary and other parts must equal 1. The local restrictions allow $(0, a)$, $(a, 1)$ and $(1, 0)$ where a is arbitrary. Again (R1) fails because there is only one recurrent vertex, namely 1. The number of compositions in this case is linear in n. Conditions in the definition guarantee that the number of compositions grows exponentially.

More than one Φ may give rise to the same $\mathcal{C}$. It is natural to assume that regularity depends only on $\mathcal{C}$ and not on the particular choice of Φ. This is not the case. Let $\mathcal{C}$ be compositions such that $c_i = i$ modulo 2 for all i. Setting $\Phi(i, a_0, 0, \ldots, 0)$ to the parity of a_0 forces the first part of a composition to be odd. Setting $\Phi(i, a_0, a_1, \ldots)$ to be the parity of $a_0 + a_1$ when both are nonzero forces the parts to alternate in parity. If m is even, (R2) is satisfied. If m is odd, (R2) is not satisfied.

Condition (R2) is used in the proof of Lemma 2(e), which says that all eigenvalues of the infinite matrix $T(x)$ are strictly smaller in absolute value than the largest, which equals the spectral radius of $T(x)$, for each positive x. This is essential to insuring that $F(x)$ have only one singularity on its circle of convergence. $\square$

Lemma 1 (Regularity) *Suppose Φ shows that $\mathcal{C}_\Phi$ is a regular locally restricted class and Φ has span and modulus m.*

(i) (R2) is equivalent to the gcd *of the cycle lengths being 1.*

(ii) If (R3) is true for k and $k' > k$, then it is true for k'.

(iii) For $t > 0$, define Φ^t of span and modulus tm by $\Phi^t(i \bmod tm; c_i, \ldots, c_{i-tm}) = 1$ if and only if $\Phi(j \bmod m; c_j, \ldots, c_{j-m}) = 1$ for all $i - (t-1)m \leqslant j \leqslant i$. If Φ is regular, then Φ^t is regular and $\mathcal{C}_{\Phi^t} = \mathcal{C}_\Phi$.

Proof: **Proof of (i)**. Suppose the gcd of the cycle lengths is $k > 1$. Choose $0 < i < k$ and $\vec{\nu}_1$ and $\vec{\nu}_2$ on the same cycle a distance i apart. If ℓ_i is the length of a path in $\mathrm{Path}_\Phi(\vec{\nu}_3, \vec{\nu}_i)$, it is easily seen that ℓ_1 and ℓ_2 differ by i modulo k. Thus (R2) does not hold.

Suppose the gcd of the cycle lengths is 1. Pick any $\vec{\nu}_3$. Since the gcd of the cycle lengths is 1, $\mathrm{Path}_\Phi(\vec{\nu}_3, \vec{\nu}_i)$ contains a path of length k for all sufficiently large k.

Proof of (ii). We can find $\vec{\nu}_0$ such that there is a path $\vec{\nu}_0 \pi \vec{\nu}_1$ from $\vec{\nu}_0$ to $\vec{\nu}_1$ of length $k' - k$. If we prepend $\vec{\nu}_0 \pi$ to every path in $\mathrm{Path}_\Phi(\vec{\nu}_1, \vec{\nu}_2)$ and compute a set S' of part sums, the elements will differ from elements in S by a constant. Since S' is a subset of the set of part sums that would be produced from $\mathrm{Path}_\Phi(\vec{\nu}_0, \vec{\nu}_2)$, the gcd is still 1.

Proof of (iii). We limit our attention to the recurrent vertices of D_Φ. Each recurrent vertex in D_{Φ^t} is the sequence of recurrent vertices on a path of length in $t-1$ in D_Φ. The converse is true because (i) guarantees that, for any sufficiently large multiple M of m (and hence some multiple of tm), such a sequence will appear at offset M in some $\vec{\mathbf{c}} \in \mathcal{C}_\Phi$. Hence the recurrent vertices of D_{Φ^t} are precisely those produced from paths of length $t-1$ through recurrent vertices of D_Φ. Thus (R4) and (R1) hold for D_{Φ^t}. Furthermore, by (i) for D_Φ, the gcd of the cycle lengths in D_{Φ^t} is 1, and so by (i) for D_{Φ^t}, (R2) holds for D_{Φ^t}. To prove (R3), we need to adjust k so that the sequence $\vec{c}_1, \ldots, \vec{c}_{k-1}$ can be thought of as a sequence of vertices in D_{Φ^t}. This will be the case if $k-1$ is divisible by t. By (ii) we can increase k so that this is true. $\square$

Proposition 1 *Let $\mathcal{C}_\Phi$ be a regular locally restricted class, $T(x)$ the associated transfer matrix. Let $F(x) = \sum a_n x^n$ and $F_{NR}(x)$ the ogfs for all compositions in $\mathcal{C}_\Phi$ and for the nonrecurrent ones, respectively. Then:*

(a) The radius of convergence of $F(x)$ lies in the interval $[1/2, 1)$.

(b) $T(x)$ is recurrent for $0 < x < 1$. (See Definition 1 for "recurrent matrix".)

(c) There exists k such that $|T(x)^k| < T(|x|)^k$ for all $0 < x < 1$ and $x \neq \pm|x|$.

(d) $F_{NR}(x)$ has radius of convergence at least 1.

Proof of Proposition 1: **Proof of (a)**. Since the number of unrestricted compositions of n is 2^{n-1}, it follows that $a_n \leqslant 2^{n-1}$ and so $F(x)$ has radius of convergence at least $1/2$.

Let ν_1, ν_2, k, and S be as in (R3). Note that $\mathcal{C}_\Phi$ must contain a composition of the form

$$\vec{\mathbf{c}} \quad = \quad A, \vec{\nu}_1, B, \vec{\nu}_2, C, \vec{\nu}_1, D,$$

where A, B, C and D are (possibly empty) concatenations of words. The set S must contain two distinct integers n_1 and n_2. (Else, the difference set contains only 0, and does not have gcd 1.) Thus we have paths $\vec{\nu}_1, E_i, \vec{\nu}_2$ with $\Sigma(E_i) = n_i$ for $i = 1, 2$. Let

$$W_i \quad = \quad \vec{\nu}_1, E_i, \vec{\nu}_2, C$$

and note that $\Sigma(W_i) = n_i + f$ where $f = \Sigma(\nu_1) + \Sigma(\nu_2) + \Sigma(C)$. All concatenations $A, W_{\epsilon_1}, W_{\epsilon_2}, \ldots, W_{\epsilon_a}, \nu_1, D$ with each $\epsilon_h \in \{1,2\}$ are valid compositions. Any such concatenation containing μ W_1's and μ W_2's will be a composition of $K_1 + \mu K_2$ where

$$K_1 = \Sigma(A) + \Sigma(\vec{\nu}_1) + \Sigma(D) \quad \text{and} \quad K_2 = n_1 + n_2 + 2f.$$

Thus

$$a_{K_1+\mu K_2} \geqslant \binom{2\mu}{\mu}$$

for all $\mu \in \mathbb{N}$. For any fixed $1 < c < 2$ the binomial coefficient on the right is larger than c^μ for large μ. It follows that the radius of convergence of $F(x)$ is at most $(1/c)^{1/K_2}$.

Proof of (b). Let $0 < x < 1$. The (i,j) entry of $T(x)^k$ is nonzero if and only if there is a path of length k from $\vec{\nu}_i$ to $\vec{\nu}_j$. Referring to Definition 1, in Section 1, condition (1) for $T(x)$ to be recurrent is exactly property (R1), and condition (2) is exactly (R3).

Proof of (c). For a power-series $g(z) = \sum_n g_n z^n$ with nonnegative coefficients it is well known that $|g(z)| < g(|z|)$ if and only if the ratio of two nonzero terms is not a positive real. In other words, for some m and n, $g_m g_n \neq 0$ and $z^{m-n} \notin (0,\infty)$. It follows that $|g(z)| < g(|z|)$ for all $z \notin [0,\infty)$ if and only if

$$\gcd\{n_1 - n_2 : n_1, n_2 \in \{n : g_n > 0\}\} = 1.$$

The matrix inequality $|T(x)^k| < T(|x|)^k$ says that we must have the latter condition for $g(x) = ((T(x))^k)_{ij}$, some i,j. Applying (R3) to (5) we obtain (c).

Proof of (d). Since no path among nonrecurrent words can contain more than K words, and since each word contains m integers, the number of nonrecurrent compositions of n is $O(n^{Km})$, and so the radius of convergence of $F_{NR}(x)$ is at least 1. □

3 Basic Concepts: Infinite Matrices

To avoid having concepts spread out, we repeat some of the notation introduced in Section 1.

We work in the complex Hilbert space ℓ^2, the set of countably infinite, complex column vectors whose entries are absolutely square summable. Elements of ℓ^2 are denoted $\mathbf{v}$, $\mathbf{w}$, etc., and $||\mathbf{v}||$ denotes the ℓ^2 norm of the vector $\mathbf{v}$:

$$||\mathbf{v}||^2 = \sum_{k=1}^{\infty} |(\mathbf{v})_k|^2.$$

For a vector $\mathbf{v}$, we use $(\mathbf{v})_k$ for the k-th component, and $\mathbf{v}_k$ denotes a sequence of vectors. For infinite matrices T, T_{ij} denotes the (i,j) entry of the matrix. Absolute value is componentwise, namely $\mathbf{v} = |\mathbf{w}|$ means $(\mathbf{v})_k = |(\mathbf{w})_k|$. Notations $\mathbf{v}^t$ and T^t indicate the transpose of a vector and a matrix; $\mathbf{v}^*, T^*$ the adjoints (i.e. conjugate transposes). There are three types of inequalities:

- $\mathbf{v} \geqslant \mathbf{w}$ means that $(\mathbf{v})_k \geqslant (\mathbf{w})_k$ for all k;
- $\mathbf{v} > \mathbf{w}$ means that $\mathbf{v} \geqslant \mathbf{w}$ and $\mathbf{v} \neq \mathbf{w}$;
- $\mathbf{v} \gg \mathbf{w}$ means that $(\mathbf{v})_k > (\mathbf{w})_k$ for all k.

The same inequality notation is used for matrices.

An *operator* A is a continuous (equivalently, bounded) linear transformation on ℓ^2. We use the standard operator norm

$$||A|| = \sup_{||\mathbf{v}||=1} ||A\mathbf{v}||.$$

With respect to this norm, the operators on ℓ^2 form a Banach space. If an operator A is bijective, then the linear transformation A^{-1} is bounded and hence also an operator. Such an operator A is called *invertible*. The set of all $z \in \mathbb{C}$ such that $z - A$ is invertible is called the *resolvent set* of A, and is denoted $\rho(A)$. (In an expression like $z - A$ that requires the scalar z to be interpreted as an operator, the intended meaning is the product zI.) A complex number z such that $z - A$ is not invertible is called a *spectral value* for A. The set of all spectral values is called the *spectrum* of A, and is denoted $\sigma(A)$. The spectrum of A includes all eigenvalues of A, and (generally) other values, too. The *spectral radius* of A, denoted $\mathrm{spr}(A)$, is defined by

$$\mathrm{spr}(A) = \lim_{n\to\infty} \|A^n\|^{1/n} = \inf \|A^n\|^{1/n}.$$

(The limit exists, and equals the infimum as indicated.) We always have

$$\emptyset \neq \sigma(A) \subseteq \{z : |z| \leqslant \mathrm{spr}(A)\},$$

and

$$\mathrm{spr}(A) = \max\{|\lambda| : \lambda \in \sigma(A)\}.$$

An operator A is called *compact* if the image of every bounded sequence $\mathbf{v}_k$ contains a convergent subsequence. An infinite, bounded sequence in a finite dimensional space always contains a convergent (i.e., Cauchy) subsequence, and so, if the image of operator A is finite dimensional, then A is compact. The compact operators retain in the infinite dimensional setting many of the nice properties of finite dimensional operators. In particular, the nice spectrum properties are a natural extension to compact infinite matrices of the well-known Perron-Frobenius theorem for finite matrices. When A is compact and λ is a nonzero element of $\sigma(A)$, then λ is an eigenvalue, and the associated eigenspace is finite dimensional. The spectrum $\sigma(A)$ is countable, and the only possible accumulation point is 0.

The symbol T always denotes an infinite, complex matrix. All such matrices in this paper belong to the class $\mathcal{M}$:

Definition 10 (The class $\mathcal{M}$) *The class $\mathcal{M}$ is the set of all infinite, complex matrices T such that*

$$\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}|T_{ij}|^2 \;<\; \infty.$$

This class appears in the literature as Hilbert-Schmidt operators. Some of the basic facts about the class $\mathcal{M}$ are collected in the next proposition.

Proposition 2 (Properties of $\mathcal{M}$) *(i) If $\mathbf{v} \in \ell^2$ and $T \in \mathcal{M}$, then $T\mathbf{v} \in \ell^2$, and*

$$||T\mathbf{v}|| \;\leqslant\; \left(\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}|T_{ij}|^2\right)^{1/2} ||\mathbf{v}|| \;\stackrel{\text{def}}{=}\; ||T||_2\,||\mathbf{v}||.$$

Thus, acting from the left, T is an operator on the Hilbert space ℓ^2, and

$$||T|| \;\leqslant\; ||T||_2.$$

(ii) If $S, T \in \mathcal{M}$ then $S+T, ST \in \mathcal{M}$.

(iii) If $\mathbf{v}_n$ is an infinite, bounded sequence in ℓ^2 and $T \in \mathcal{M}$, then $T\mathbf{v}_n$ contains a convergent subsequence; thus, T is a compact linear operator on the Hilbert space ℓ^2.

Proof: The proofs of (i) and (ii) are straightforward using the Cauchy-Schwartz inequality. For (iii), let T_n agree with T in the upper left $n\times n$ corner, and be zero elsewhere. Clearly, $(T-T_n) \in \mathcal{M}$; and, since the image of T_n is finite dimensional, T_n is compact. Above we have seen that for any $S \in \mathcal{M}$, $||S|| \leqslant ||S||_2$. From

$$||T-T_n||^2 \;\leqslant\; (||T-T_n||_2)^2 \;=\sum_{\max(i,j)>n}|T_{ij}|^2,$$

we have $||T-T_n|| \to 0$. Since the compact operators are a closed subspace of the bounded operators ([11], page 158), it follows that T is compact. $\square$

The results and definitions that we need from functional analysis are stated in terms of operators, not infinite matrices. Without further comment, we interpret an infinite matrix $T \in \mathcal{M}$ as an operator in the obvious manner. Thus, for an infinite matrix $T \in \mathcal{M}$ all the notations $||T||$, $\sigma(T)$, $\rho(T)$, and $\text{spr}(T)$, as well as the notion of being compact, as defined earlier for operators, are meaningful. Under this identification of infinite matrices $T \in \mathcal{M}$ with operators, the sum, scalar multiple, and product of two matrices correspond as expected with the usual sum, scalar multiple, and composition of the operators.

We use the same conventions regarding absolute value as were stated earlier for vectors: $T = |S|$ means $T_{ij} = |S_{ij}|$. Similarly, the relations $\geqslant, >, \gg$ can be applied to infinite matrices. The boldface zero, $\mathbf{0}$, is used for the zero vector; the standard 0 is used for both the scalar zero and the zero matrix. We use the following version of the Krein-Rutman theorem. See, for example, [5] (Theorem 19.2) or [14].

Proposition 3 (The Krein-Rutman Theorem) *Let T be compact and $T > 0$. If* $\mathrm{spr}(T) > 0$, *then* $\mathrm{spr}(T)$ *is an eigenvalue for T and there is a corresponding eigenvector* $\mathbf{v}$ *which satisfies* $\mathbf{v} > \mathbf{0}$.

Remark 1 For any operator T there is always a spectral value λ with $|\lambda| = \mathrm{spr}(T)$. If $T > 0$, then a generalization of Pringsheim's argument, applied to the function $(1-zT)^{-1}$, shows that $\mathrm{spr}(T)$ itself is a spectral value. Because T is compact and $\mathrm{spr}(T) > 0$, the spectral value $\mathrm{spr}(T)$ is in fact an eigenvalue. The essential additional information provided by the Krein-Rutman theorem is that there is an eigenvector $\mathbf{v}$ corresponding to this eigenvalue which satisfies $\mathbf{v} > \mathbf{0}$. When we later attribute an assertion to the Krein-Rutman theorem, Proposition 3 is the result we have in mind. □

Let $\Omega \subset \mathbb{C}$ be a domain, (an open, connected set). The functions analytic (or holomorphic) on Ω are a subset of the functions $F : \Omega \to \mathbb{C}$. If we replace the range $\mathbb{C}$ with a Banach space X, then most of the definitions and theorems of complex analysis carry over, leading to the notion of vector-valued and operator-valued holomorphic functions.

Definition 11 (Holomorphic function with range a Banach space) *Let X be a Banach space. A function $F : \Omega \to X$ is* holomorphic *provided that the limit*

$$\lim_{\Delta x \to 0} \frac{F(x + \Delta x) - F(x)}{\Delta x}$$

exists. We shall say that $T(x)$ is holomorphic on Ω, or is a holomorphic family *for $x \in \Omega$.*

As pointed out by Kato ([11], p.10), such familiar results as the Cauchy integral formula, Taylor and Laurent expansions, and Liouville's Theorem, all hold in this more general setting, and can be proven in the same manner.

We define $\mathcal{M}(\Omega)$ to be a class of infinite matrices whose entries are holomorphic functions, and which satisfy the condition for membership in $\mathcal{M}$ uniformly on compact subsets. It will be seen in the next Proposition that such matrices are holomorphic families.

Definition 12 ($\mathcal{M}(\Omega)$) *Let $\Omega \subseteq \mathbb{C}$ be a domain. Define $\mathcal{M}(\Omega)$ to be the set of infinite matrices $T(x)$ such that each entry $T_{ij}(x)$ is holomorphic in Ω, and such that for every compact $K \subseteq \Omega$ there exists C with*

$$\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} |T_{ij}(x)|^2 \leqslant C \quad \textit{for all } x \in K.$$

The set $\ell^2(\Omega)$ is defined similarly.

Proposition 4 *(a) If $T \in \mathcal{M}(\Omega)$, then $T(x)$ is a holomorphic family of operators for $x \in \Omega$, and the derivative $T'(x)$ may be taken component-wise.*

(b) If $\mathbf{v}, \mathbf{w} \in \ell^2(\Omega)$ and $T \in \mathcal{M}(\Omega)$, then $\mathbf{w}^t\mathbf{v}$ is holomorphic on Ω, and $\mathbf{w}^t T, T\mathbf{v} \in \ell^2(\Omega)$.

(c) If $S, T \in \mathcal{M}(\Omega)$ then $S+T, ST \in \mathcal{M}(\Omega)$.

(d) Let $T \in \mathcal{M}(\Omega)$ and $\mathbf{v} \in \ell^2(\Omega)$. If $z_0 \in \Omega$, then $T(z_0), |T(z_0)| \in \mathcal{M}$ and $\mathbf{v}(z_0) \in \ell^2$.

Proof: Again, most of the proof is standard and omitted. A key ingredient in the proof of (b) and (c) is that if a sequence of holomorphic functions converges uniformly on compact subsets of Ω, then the limit is holomorphic. We prove part (a). Let $z_0 \in \Omega$. Choose $\rho > 0$ sufficiently small that

$$K = \{|z - z_0| \leqslant \rho\} \subseteq \Omega,$$

and let Γ be the boundary of K oriented in the counterclockwise direction. Let B be the infinite matrix of derivatives, $B_{ij}(x) = T'_{ij}(x)$. For $|h| \leqslant \rho/2$,

$$(T(z_0+h) - T(z_0) - hB(z_0))_{ij} \quad = \quad \frac{h^2}{2\pi i} \oint_\Gamma \frac{T_{ij}(w)}{(w - z_0 - h)(w - z_0)^2}\, dw.$$

By Cauchy-Schwartz

$$\left| \oint_\Gamma \frac{T_{ij}(w)}{(w - z - h)(w - z)^2}\, dw \right|^2 \quad \leqslant \quad \oint_\Gamma |T_{ij}(w)|^2\, |dw| \, \oint_\Gamma \frac{1}{|w - z_0 - h|^2 |w - z_0|^4}\, |dw|.$$

Hence,

$$\left| (T(z_0+h) - T(z_0) - hB(z_0))_{ij} \right|^2 \quad \leqslant \quad \frac{|h|^4}{4\pi^2} \, \frac{2\pi\rho}{(\rho/2)^2(\rho)^4} \oint_\Gamma |T_{ij}(w)|^2\, |dw|.$$

Since K is compact, we have C such that

$$\sum_i \sum_j |T_{ij}(w)|^2 \leqslant C \quad \text{for } w \in K.$$

By (i) of Proposition 2,

$$||T(z_0+h) - T(z_0) - hB(z_0)||^2 \leqslant \frac{|h|^4}{4\pi^2} \, (8\pi\rho^{-5})\, (2\pi\rho) C = 4C\rho^{-4}|h|^4.$$

Thus,

$$\lim_{h \to 0} \frac{T(z_0+h) - T(z_0)}{h} = B(z),$$

and (a) is proven. □

For further background information the reader may consult a text such as [4], [5], [6], [8], or [11].

4 Two Lemmas from Functional Analysis

In this section we prove the two lemmas which are used in the proof of Theorem 1. The following proposition will be used in the proof of the first.

Proposition 5 *Let $\rho > 0$, $T \in \mathcal{M}$, $T \geqslant 0$, and $\mathbf{z}$ be a real vector. Define the set of integers N by*

$$N = \{i : (\mathbf{z})_i < 0\}.$$

If $T\mathbf{z} \leqslant \rho\mathbf{z}$ and $N \neq \emptyset$, then

$$\rho^2 \leqslant \sum_{i \in N} \sum_{j \in N} (T_{ij})^2.$$

Proof: Because $T_{ij} \geqslant 0$, for $i \in N$ we have

$$0 > (\rho\mathbf{z})_i \geqslant (T\mathbf{z})_i = \sum_{j=1}^{\infty} T_{ij}(\mathbf{z})_j \geqslant \sum_{j \in N} T_{ij}(\mathbf{z})_j.$$

So,

$$\begin{aligned}
\rho^2 \sum_{i \in N} (\mathbf{z})_i^2 &\leqslant \sum_{i \in N} \Big(\sum_{j \in N} T_{ij}(\mathbf{z})_j \Big)^2 \\
&\leqslant \sum_{i \in N} \Big((\sum_{j \in N} T_{ij}^2)\,(\sum_{j \in N} (\mathbf{z})_j^2) \Big) \\
&= (\sum_{i \in N} \sum_{j \in N} T_{ij}^2)\,(\sum_{j \in N} (\mathbf{z})_j^2).
\end{aligned}$$

Since $N \neq \emptyset$, we may divide both sides by the positive quantity $\sum_{i \in N} (\mathbf{z})_i^2$ to obtain the desired conclusion. □

For any scalar Λ and operator T we have the inclusions

$$\{\mathbf{0}\} \subseteq \ker(\Lambda I - T) \subseteq \ker((\Lambda I - T)^2) \subseteq \cdots .$$

If $\Lambda \neq 0$ and T is compact, then there is a smallest $n = n_\Lambda$ such that $\ker((\Lambda I - T)^i) = \ker((\Lambda I - T)^{i+1})$ for $i \geqslant n$.

Definition 13 (Simple eigenvalue) *We say that $\Lambda \neq 0$ is a simple eigenvalue provided $\dim(\ker(\Lambda I - T)) = 1$ and $n_\Lambda = 1$.*

To verify that an eigenvalue is simple, one must show that the space of eigenvectors for Λ is spanned by a single vector $\mathbf{v} \neq \mathbf{0}$, and that no vector $\mathbf{w}$ satisfies the relation $(\Lambda I - T)\mathbf{w} = \mathbf{v}$.

Lemma 2 (Eigenvalues of recurrent matrices) *Let $T \in \mathcal{M}$ be recurrent and satisfy $T > 0$. Then* *(a) T is compact;*

(b) $\lambda = \operatorname{spr}(T) > 0$ is an eigenvalue of T;

(c) the space of eigenvectors associated with λ is spanned by a vector $\mathbf{v} \gg \mathbf{0}$;

(d) the eigenvalue λ is simple;

(e) if $\mu \neq \lambda$ is an eigenvalue, then $|\mu| < \lambda$;

(f) if $T > |B|$, then $\operatorname{spr}(T) > \operatorname{spr}(B)$.

Proof: **Part (a)**. This has been established earlier in Proposition 2 (iii).

Part (b). Let ℓ be such that $(T^\ell)_{i,i} \neq 0$ for some i. Then

$$\begin{aligned}\lim_{n\to\infty} \|T^n\|^{1/n} &= \lim_{k\to\infty} \|T^{\ell k}\|^{1/k\ell} \geqslant \lim_{k\to\infty} \left(\left(T^{\ell k}\right)_{i,i}\right)^{1/k\ell} \\ &\geqslant \lim_{k\to\infty} \left(\left((T^\ell)_{i,i}\right)^k\right)^{1/k\ell} = \left((T^\ell)_{i,i}\right)^{1/\ell} > 0.\end{aligned}$$

By the Krein-Rutman Theorem (Proposition 3), $\lambda = \operatorname{spr}(T)$ is an eigenvalue with an eigenvector $\mathbf{v}$ that satisfies $\mathbf{v} > \mathbf{0}$.

Part (c). To see that $\mathbf{v} \gg \mathbf{0}$, suppose to the contrary that $(\mathbf{v})_i = 0$. Let j be such that $(\mathbf{v})_j > 0$ and let ℓ be such that $(T^\ell)_{ij} > 0$. Then

$$\lambda^\ell (\mathbf{v})_i = (T^\ell \mathbf{v})_i \geqslant (T^\ell)_{ij} (\mathbf{v})_j > 0,$$

a contradiction. For the rest of the proof of the lemma, $\mathbf{v}$ refers to this vector. We want to show that $\ker(\lambda I - T) = \operatorname{Span}(\mathbf{v})$. Let $\mathbf{w} \in \ker(\lambda I - T)$. Since this implies that $\operatorname{Re}\mathbf{w}, \operatorname{Im}\mathbf{w} \in \ker(\lambda I - T)$, we may assume without loss that $\mathbf{w}$ is real. If $\mathbf{w} = \mathbf{0}$ we are done. So assume, again without loss, that $(\mathbf{w})_i > 0$ for some i. Then

$$t = \max\{x \in \mathbb{R}_{\geqslant 0} : x\mathbf{w} \leqslant \mathbf{v}\}$$

is well defined. If $\mathbf{v} = t\mathbf{w}$, we are done, so assume to the contrary that $\mathbf{v} > t\mathbf{w}$. Since $\mathbf{v} - t\mathbf{w} \in \ker(\lambda I - T)$, it follows as it did for $\mathbf{v}$ a moment ago that, in fact, $\mathbf{v} - t\mathbf{w} \gg \mathbf{0}$. For each $\epsilon > 0$ define $k(\epsilon)$ to be the smallest index i such that $(\mathbf{v} - (t+\epsilon)\mathbf{w})_i < 0$. It must be the case that $k(\epsilon) \to \infty$ as $\epsilon \to 0$, since otherwise we find $(\mathbf{v} - t\mathbf{w})_i = 0$ for some i, a contradiction. But Proposition 5 applies to $\rho = \lambda$ and $\mathbf{z} = \mathbf{v} - (t+\epsilon)\mathbf{w}$, showing

$$\lambda^2 \leqslant \sum_{i,j \in N_\epsilon} (T_{ij})^2 \leqslant \sum_{i,j \geqslant k(\epsilon)} (T_{ij})^2. \tag{9}$$

Since $\lambda > 0$ and $k(\epsilon) \to \infty$, this is a contradiction.

Part (d). We want to show that the eigenvector $\mathbf{v} \gg \mathbf{0}$ is simple. We have seen in Part (c) that the space of eigenvectors associated with λ is one dimensional, so for λ not to be simple we would have a vector $\mathbf{w}$ such that $(\lambda I - T)\mathbf{w} = \mathbf{v}$. Passing to the real part, we may assume $\mathbf{w}$ is real. By adding a sufficiently large multiple of $\mathbf{v}$, which belongs to $\ker(\lambda I - T)$, we may assume that $(\mathbf{w})_i > 0$ for some i. Define $t \geqslant 0$ and $k(\epsilon)$ exactly as in the previous part of the proof. Letting $\mathbf{z} = \mathbf{v} - t\mathbf{w}$ we have, since $\mathbf{0} \leqslant \mathbf{z}$,

$$\mathbf{0} \leqslant T\mathbf{z} = T\mathbf{v} - tT\mathbf{w} = \lambda\mathbf{v} - t(\mathbf{v} + \lambda\mathbf{w}) = \lambda\mathbf{z} - t\mathbf{v}.$$

If $t > 0$, it follows from $\mathbf{0} \ll \mathbf{v}$ that $\mathbf{0} \ll \mathbf{z}$. If $t = 0$, then $\mathbf{0} \ll \mathbf{v} = \mathbf{z}$. In either case, $\mathbf{0} \ll \mathbf{z}$, and again we must have $k(\epsilon) \to \infty$. For $\epsilon > 0$

$$T(\mathbf{z} - \epsilon\mathbf{w}) \;=\; \lambda(\mathbf{z} - \epsilon\mathbf{w}) - (t + \epsilon)\mathbf{v},$$

and so $T(\mathbf{z} - \epsilon\mathbf{w}) \ll \lambda(\mathbf{z} - \epsilon\mathbf{w})$. Hence, the pair $\rho = \lambda$ and $\mathbf{z} = \mathbf{z} - \epsilon\mathbf{w} = \mathbf{v} - (t + \epsilon)\mathbf{w}$ satisfy the hypotheses of Proposition 5, leading again to the contradiction (9).

Part (e). The adjoint T^* satisfies the same conditions as T: $T^* \in \mathcal{M}$, $T^* > 0$, and T^* recurrent. Also, $\mathrm{spr}(T^*) = \mathrm{spr}(T)$. It follows that $T^*\mathbf{z} = \lambda\mathbf{z}$ for some $\mathbf{z} \gg \mathbf{0}$. We use this $\mathbf{z}$ here and in the proof of (f). Let μ be a nonzero eigenvalue for T, different from λ.

Since $\lambda = \mathrm{spr}(T)$, $|\mu| \leqslant \lambda$ and we can assume μ is not a positive real number. Let $\mathbf{w}$ be an eigenvector for μ. Without loss of generality, the first nonzero component of $\mathbf{w}$ is a positive real. We cannot have $\mathbf{w} > \mathbf{0}$ for then $T\mathbf{w} = \mu\mathbf{w}$ would imply that $\mu > 0$. Suppose $(\mathbf{w})_{j_1} > 0$ and $(\mathbf{w})_{j_2} \neq 0$ is not a positive real. By property (R2) there exist ℓ, i such that $(T^\ell)_{i,j_1} > 0$ and $(T^\ell)_{i,j_2} > 0$. Thus $|T^\ell\mathbf{w}| < T^\ell|\mathbf{w}|$ by considering the ith component. We have

$$\lambda^\ell\mathbf{z}^*|\mathbf{w}| \;=\; \mathbf{z}^*T^\ell|\mathbf{w}| \;>\; \mathbf{z}^*|T^\ell\mathbf{w}| \;=\; \mathbf{z}^*|\mu^\ell\mathbf{w}| \;=\; |\mu|^\ell\mathbf{z}^*|\mathbf{w}|.$$

Since $\mathbf{z} \gg \mathbf{0}$, it follows that $\mathbf{z}^*|\mathbf{w}| > 0$ and so $\lambda^\ell > |\mu|^\ell$.

Part (f). We first consider $B \geqslant 0$. From $0 \leqslant B < T$ it follows that $B^n \leqslant T^n$ for all n, and then $||B^n|| \leqslant ||T^n||$. So, at least, $\mathrm{spr}(B) \leqslant \mathrm{spr}(T)$. Suppose that we have equality, so that λ is an eigenvalue of B as well as T. By Krein-Rutman, B has an eigenvector $\mathbf{w} > \mathbf{0}$ associated with λ. We have

$$\mathbf{z}^*(T - B)\mathbf{w} \;=\; (\mathbf{z}^*T)\mathbf{w} - \mathbf{z}^*(B\mathbf{w}) \;=\; (\lambda - \lambda)\mathbf{z}^*\mathbf{w} \;=\; 0.$$

Since $\mathbf{z} \gg \mathbf{0}$, it follows that $(T - B)\mathbf{w} = \mathbf{0}$. By this and the definition of $\mathbf{w}$, $T\mathbf{w} = B\mathbf{w} = \lambda\mathbf{w}$ and so $\mathbf{w}$ is an eigenvector of T. Hence $\mathbf{w} \gg \mathbf{0}$. Since $\mathbf{z}^*(T - B)\mathbf{w} = 0$, it follows that $T - B = 0$, contradicting $T > B$. Hence, the assumption of equality leads to a contradiction, and we have $\mathrm{spr}(B) < \mathrm{spr}(T)$. For general B, we simply observe that $\|B^n\|^{1/n} \leqslant \||B|^n\|^{1/n}$, and so $\mathrm{spr}(B) \leqslant \mathrm{spr}(|B|)$. □

Let T be a compact operator on a Hilbert space V and let $\Lambda \in \sigma(T) - \{0\}$. If $V = M \oplus N$, a direct sum of closed, T-invariant subspaces, then $\sigma(T) = \sigma(T|_M) \cup \sigma(T|_N)$. It is a theorem of F. Riesz [17], see also p.178 of [18], that there is a unique such decomposition with the additional property that $\sigma(T|_M) = \{\Lambda\}$ and $\sigma(T|_N) = \sigma(T) - \{\Lambda\}$. We say that the decomposition *separates* Λ *from the rest of the spectrum* $\sigma(T)$. Suppose that E is a bounded idempotent (i.e. projection) which commutes with T and satisfies

$$T = \Lambda E + B, \quad EB = BE = 0, \quad \mathrm{spr}(B) < |\Lambda| \quad \text{and} \quad \mathrm{rank}(E) = 1. \tag{10}$$

Then $M = \text{image}(E), N = \ker(E)$ is the unique Riesz decomposition which separates Λ from the rest of $\sigma(T)$; moreover, the spectral value Λ is a simple eigenvalue and strictly larger than all other spectral values of T in absolute value. The converse is also true: if $V = M \oplus N$ is the Riesz decomposition which separates a simple eigenvalue Λ from the rest of the spectrum, and if Λ is the unique spectral value of maximum absolute value, then the projection E determined from this decomposition by $\text{image}(E) = M, \ker(E) = N$ commutes with T and satisfies (10). The role of the next lemma is to say that if $T = T(x_0)$, with $T(x)$ holomorphic, has a decomposition with the properties (10), then the same is true of $T(x)$ locally. In this lemma, and henceforth, we say that a matrix (or operator) U is *real* if $\langle \mathbf{e}_i, U\mathbf{e}_j \rangle$ is real for all i, j; equivalently, $\mathbf{v} \geqslant \mathbf{0} \Longrightarrow U\mathbf{v} \geqslant \mathbf{0}$. For $\Omega \subseteq \mathbb{C}$, $\text{Re}(\Omega)$ denotes the real elements of the set Ω.

Lemma 3 *Let $\Omega \subseteq \mathbb{C}$ be a domain, and $T(x) \in \mathcal{M}(\Omega)$. Assume for a certain point $x_0 \in \Omega$ that $\Lambda \in \sigma(T(x_0)) - \{0\}$, and that all other spectral values of $T(x_0)$ are strictly smaller in absolute value than $|\Lambda|$. Then there exists $\delta > 0$ such that in the neighborhood $\mathcal{N} = \{|x - x_0| < \delta|\}$*

$$T(x) \;=\; \Lambda(x)E(x) + B(x);$$

and

- *(a) All four functions T, Λ, E and B are holomorphic in $\mathcal{N}$*
- *(b) $\Lambda = \Lambda(x_0)$*
- *(c) $E(x)^2 = E(x)$ and $\text{rank}(E(x)) = 1$*
- *(d) $B(x)E(x) = E(x)B(x) = 0$*
- *(e) $\text{spr}(B(x)) < |\Lambda(x)|$.*

Moreover, if $T(x)$ is real for $x \in \text{Re}(\Omega)$, then the same is true of $E(x)$.

Proof: Let $T = T(x_0)$, and let Λ, E, B satisfy (10). Because T is compact and $\Lambda \neq 0$, it is isolated from other spectral values. Let Γ be a positively oriented, circular arc in the complex z-plane which encloses Λ, encloses no other spectral values for T, and lies itself entirely in $\rho(T)$. By formula (13), p.418 of [18], or page 180 of [11], the operator E is given by the integral

$$E \;=\; \frac{1}{2\pi i} \oint_\Gamma \frac{dz}{z - T}.$$

By a standard compactness argument, there exists δ sufficiently small that

$$|x - x_0| \;<\; \delta \quad\Longrightarrow\quad \Gamma \;\subseteq\; \rho(T(x)).$$

Then

$$E(x) \;=\; \frac{1}{2\pi i} \oint_\Gamma \frac{dz}{z - T(x)} \tag{11}$$

is holomorphic on $\mathcal{N}$, $E(x)^2 = E(x)$, and $E(x)$ commutes with $T(x)$. (One way to see that $E(x)$ is holomorphic is to note that it is the uniform limit on compact subsets of $\mathcal{N}$ of the finite Riemann sums defining the integral.)

By taking δ smaller still, if necessary, we may assume that $||E(x)-E|| < 1$ throughout $\mathcal{N}$. By Lemma 4.10, p.34 [11], $\text{rank}(E(x)) = 1$ for all $x \in \mathcal{N}$. In fact, the similarity condition

$$E \;=\; U(x)^{-1}E(x)U(x)$$

holds with $U(x)$ an invertible linear operator on ℓ^2. By examining the proof of Kato's Lemma 4.10 [11], we see that $U(x)$ is holomorphic, and we have the explicit formula

$$U(x) \;=\; \hat{U}(x)\sum_{n=0}^{\infty}(-1)^n\binom{-1/2}{n}R(x)^n,$$

where

$$\begin{aligned}
R(x) &= 1-\hat{U}(x)\hat{V}(x),\\
\hat{U}(x) &= E(x)E+(1-E(x))(1-E),\\
\hat{V}(x) &= EE(x)+(1-E)(1-E(x)).
\end{aligned}$$

Since Λ is a simple eigenvalue, image(E) equals the span of a single eigenvector. Let $\mathbf{v}$ be such an eigenvector. We have $E\mathbf{v} = \mathbf{v}$, and by the above similarity relation

$$E(x)U(x)\mathbf{v} \;=\; U(x)E(x)\mathbf{v} \;=\; \mathbf{v}.$$

Thus $U(x)\mathbf{v} \in$ image$(E(x))$. If δ is sufficiently small, $U(x)\mathbf{v}$ is nonzero for $x \in \mathcal{N}$, and since the image of $E(x)$ is one-dimensional, the nonzero vector $U(x)\mathbf{v}$ spans it. Also, for any $z \in \rho(T(x))$, $T(x)$ commutes with $(z-T(x))^{-1}$, and so by the above integral formula, $E(x)$ commutes with $T(x)$. We have then

$$\begin{aligned}
T(x)U(x)\mathbf{v} &= T(x)E(x)U(x)\mathbf{v}\\
&= E(x)T(x)U(x)\mathbf{v}\\
&= \text{an element of the image of } E(x)\\
&= \Lambda(x)U(x)\mathbf{v}
\end{aligned}$$

for some well-defined scalar $\Lambda(x)$. This gives us the eigenvalue function $\Lambda(x)$, with $\Lambda(x_0) = \Lambda$. We wish to show that $\Lambda(x)$ is holomorphic in $\mathcal{N}$. We know that $U(x_0)$ is the identity operator, and that $U(x_0)\mathbf{v} = \mathbf{v}$ is nonzero. Thus, for some j, the inner product $\langle U(x_0)\mathbf{v}, \mathbf{e}_j\rangle$ is nonzero. With δ small, the same is true throughout $\mathcal{N}$, and the formula

$$\Lambda(x) \;=\; \frac{\langle T(x)U(x)\mathbf{v}, \mathbf{e}_j\rangle}{\langle U(x)\mathbf{v}, \mathbf{e}_j\rangle},$$

shows that $\Lambda(x)$ is indeed holomorphic in $\mathcal{N}$.

Now define $B(x)$ by

$$B(x) \;=\; T(x)-\Lambda(x)E(x).$$

It is immediate that $B(x)$ is holomorphic, since all three of $T(x), \Lambda(x), E(x)$ are. The previous part of the proof has shown that is possible to define $\Lambda(x)$ in such a way that

$$T(x)E(x) \;=\; \Lambda(x)E(x).$$

Because $E(x)$ and $T(x)$ commute, the same is true for $E(x)$ and $B(x)$. Since

$$E(x)B(x) \;=\; E(x)\left(T(x) - \Lambda(x)E(x)\right)E(x) = T(x)E(x) - \Lambda(x)E(x) = 0,$$

we have $B(x)E(x) = E(x)B(x) = 0$. This proves parts (a)–(d). We come now to (e). Since $\sigma(T(x_0)) = \{\Lambda\} \cup \sigma(B(x_0))$, the hypothesis that Λ is larger in absolute value than all other spectral values tells us

$$\operatorname{spr}(B(x_0)) \;<\; |\Lambda| \;=\; |\Lambda(x_0)|.$$

Since $x \mapsto |\Lambda(x)|$ is continuous on $\mathcal{N}$ and $x \mapsto \operatorname{spr}(B(x))$ is upper semicontinuous, assertion (e) follows for δ sufficiently small.

Assume now that $T(x)$ is real for $x \in \operatorname{Re}(\Omega)$. The last step of the proof is to show that the same is true of $E(x)$. Let $z_j = e^{2\pi i j/d}$, the d-th roots of unity and let $w_j = (z_{j+1}+z_j)/2$, the midpoints between the z_j's. Consider the integral formula (11). For x real, $T(x)$ is real, and

$$(\overline{w} - T(x))^{-1} \;=\; \overline{(w - T(x))^{-1}}.$$

It follows that for even integral d

$$\sum_{j=0}^{d-1}(z_{j+1} - z_j)(w_j - T(x))^{-1}, \;\; z_d \equiv z_0,$$

is a pure imaginary, because for each term in the sum its negative conjugate is also present. But $\int_\Gamma (z - T(x))^{-1}dz$ is the limit as $d \to \infty$ of the latter Riemann sums; thus the integral is a pure imaginary, and after multiplication by $(2\pi i)^{-1}$ we see that $E(x)$ is real, as claimed. □

5 Proof of Theorem 1

We begin with a lemma that extends the first part of Theorem 1 slightly by allowing constant terms.

Lemma 4 *Let $\rho > 0$ and let U_n be a sequence of infinite matrices. Suppose that the power series*

$$U(x) = U_0 + xU_1 + x^2U_2 + \cdots$$

satisfies:

(a) $\sum_n |x|^n \, ||U_n||_2$ is convergent for $|x| < \rho$,

(b) $U_n \geqslant 0$,

(c) $U(x_0)$ *is recurrent for all* $x_0 \in (0,\rho)$.

Then for each $x_0 \in (0,\rho)$ *the matrix* $U(x_0)$ *has an eigenvalue* $\lambda(x_0) > 0$ *which is simple and strictly larger in absolute value than the other eigenvalues of* $U(x_0)$. *On the interval* $(0,\rho)$ *the function* $\lambda(x)$ *is analytic and strictly increasing.*

Proof: For each x_0 in the real interval $(0,\rho)$ the matrix $U(x_0)$ satisfies the hypotheses of Lemma 2. Hence, $\mathrm{spr}(U(x_0))$ is a positive, simple eigenvalue which is strictly larger than the absolute value of all other eigenvalues of $U(x_0)$. Denote this eigenvalue by $\lambda(x_0)$, so that the function $\lambda(x)$ is defined on the interval $(0,\rho)$. Hypothesis (c) implies that not all U_n are zero, and so, in view of (b), $U(x_0) > 0$ for all $x_0 \in (0,\rho)$. Moreover, for $x_1 < x_2$, we have $U(x_1) < U(x_2)$. By Lemma 2(f), $\lambda(x_1) < \lambda(x_2)$, and so the function $\lambda(x)$ is increasing.

We wish to prove next that the function $\lambda(x)$ is continuous. Hypothesis (a) implies $U(x) \in \mathcal{M}(|x|<\rho)$, and so by Proposition 4(i) $U(x)$ is holomorphic in the domain $|x| < \rho$. The family $U(x)$ satisfies the hypotheses of Lemma 3 with $\Omega = \{|x| < \rho\}$ and x_0 any point in the real interval $(0,\rho)$. Hence, for each such x_0 we have a neighborhood $\mathcal{N}$ of x_0 and functions $\Lambda(x), E(x), B(x)$ analytic in $\mathcal{N}$ which satisfy the conclusions (a)–(e) of Lemma 3. We claim that $\Lambda(x) = \lambda(x)$ for x belonging to $\mathrm{Re}(\mathcal{N})$.

Indeed, since $\Lambda(x) \in \sigma(U(x))$, we have

$$|\Lambda(x)| \;\leqslant\; \mathrm{spr}(U(x)) \;=\; \lambda(x), \text{ when } x \in \mathrm{Re}(\mathcal{N}).$$

If for some $x \in \mathrm{Re}(\mathcal{N})$ it were the case that $\lambda(x) \neq \Lambda(x)$, then $\lambda(x)$ would be a spectral value for $U(x)$ which differs from $\Lambda(x)$ and whose absolute value is at least $|\Lambda(x)|$, contradicting conclusion (e) of Lemma 3. Thus, as claimed, $\Lambda(x) = \lambda(x)$ for $x \in \mathrm{Re}(\mathcal{N})$, and so not only is $\lambda(x)$ continuous at each $x_0 \in (0,\rho)$, but in fact analytic. □

We now turn our attention to the proof of Theorem1. In the proof, $\lambda(x)$ denotes the eigenvalue of $T(x)$.

By setting $U_0 = 0$ in Lemma 4, we obtain part of the theorem. Now apply the lemma to $U(x) = T(x)/x$ and let its largest eigenvalue be $\mu(x) > 0$, Then $\mu(x)$ is increasing, $\lambda(x) = x\mu(x)$, and $\lambda'(x) = \mu(x) + x\mu'(x) \geqslant \mu(x)$.

For the rest of the proof, assume that $r, k_0, \mathbf{s}(x), \mathbf{f}(x)$ satisfy the hypotheses (d)–(f) of Theorem 1. Define

$$S(x) \;=\; \sum_{k=0}^{\infty} T(x)^k.$$

The sum converges uniformly on compact subsets of $\mathrm{spr}(T(x)) < 1$ and defines an analytic function on the latter open set. In particular, $S(x)$ is analytic in $|x| < r$. For any x_0 satisfying $|T(x_0)|^{k_0} < T(r)^{k_0}$ we have

$$\mathrm{spr}(T(x_0)^{k_0}) \;\leqslant\; \mathrm{spr}(|T(x_0)^{k_0}|) \;\leqslant\; \mathrm{spr}(|T(x_0)|^{k_0}) \;<\; \mathrm{spr}(T(r)^{k_0}) \;=\; 1.$$

By hypothesis (f) of Theorem 1, then, $S(x)$ is holomorphic near each x on the circle $|x| = r$, with the possible exceptions of $x = r$ and $x = -r$. Thus $\phi(x) = \mathbf{s}(x)^t S(x)\mathbf{f}(x)$, too, is holomorphic for $|x| < r$, with $x = \pm r$ the only possible singularities.

By Lemma 3 we have a neighborhood $\mathcal{N}$ of $x = r$ in which

$$T(x) \;=\; \lambda(x)E(x) + B(x)$$

with all four functions holomorphic in $\mathcal{N}$, $E(x)^2 = E(x)$, $E(x)B(x) = B(x)E(x) = 0$, and $\lambda(x) = \mathrm{spr}(T(x))$ for $x \in \mathrm{Re}(\mathcal{N})$. It follows by induction that for $k \geqslant 1$,

$$T(x)^k \;=\; \lambda(x)^k E(x) + B(x)^k \tag{12}$$

in $\mathcal{N}$. Assume $\mathcal{N}$ sufficiently small that $\mathrm{spr}(B(x)) < 1$ for $x \in \mathcal{N}$. Then, in the intersection $\mathcal{N} \cap \{|x| < r\}$

$$S(x) \;=\; \frac{\lambda(x)}{1-\lambda(x)}\,E(x) \;+\; (I - B(x))^{-1}.$$

This gives

$$\phi(x) \;=\; \frac{g(x)}{1-\lambda(x)} \;+\; h(x) \tag{13}$$

with $g(x) = \lambda(x)\mathbf{s}(x)^t E(x)\mathbf{f}(x)$ and $h(x) = \mathbf{s}(x)^t(I - B(x))^{-1}\mathbf{f}(x)$. Initially we have proven (13) for $x \in \mathcal{N} \cap \{|x| < r\}$. However, both $g(x)$ and $f(x)$ are holomorphic throughout $\mathcal{N}$. Thus the right side of (10) is holomorphic in the punctured neighborhood $\mathcal{N} - \{r\}$. To complete the proof that $\phi(x)$ has a simple pole at $x = r$, we need to show that $g(r) \neq 0$ and $\lambda'(r) \neq 0$.

From equation (12) and $\lambda(r) = 1$, for all $k \geqslant 1$,

$$E(r) \;=\; T(r)^k - B(r)^k.$$

Now the matrices $B(r)^k$ are summable, so in particular $||B(r)^k|| \to 0$ for $k \to \infty$. Suppose we show there exist $i, j \geqslant 1$ and $\eta > 0$ such that for k sufficiently large

$$(T(r)^k)_{ij} \;\geqslant\; \eta. \tag{14}$$

Let i_1, j_1, ℓ, m satisfy

$$(\mathbf{s}(r))_{i_1} > 0, \quad (\mathbf{f}(r))_{j_1} > 0, \quad (T(r)^\ell)_{i_1 i} > 0, \quad (T(r)^m)_{jj_1} > 0,$$

and let k be so large that $(T(r)^{k-\ell-m})_{ij} \geqslant \eta$. Then

$$\mathbf{s}(r)^t T(r)^k \mathbf{f}(r) \;\geqslant\; (\mathbf{s}(r))_{i_1}\,(T(r)^\ell)_{i_1 i}\,(T(r)^{k-\ell-m})_{ij}\,(T(r)^m)_{jj_1}(\mathbf{f}(r))_{j_1} \;\geqslant\; \eta' \;>\; 0.$$

Since

$$\begin{aligned} g(r) &= \lambda(r)\mathbf{s}(r)^t E(r)\mathbf{f}(r) \\ &= \mathbf{s}(r)^t T(r)^k \mathbf{f}(r) - \mathbf{s}(r)^t B(r)^k \mathbf{f}(r) \\ &= \mathbf{s}(r)^t T(r)^k \mathbf{f}(r) + o(1) \end{aligned}$$

as $k \to \infty$, we may conclude that $g(r) \neq 0$, as needed.

So, let us prove assertion (14). Let $\mathbf{v} \gg \mathbf{0}$ be an eigenvector corresponding to the eigenvalue 1 and spanning image(E). Let j be such that in the decomposition

$$\mathbf{e}_j \;=\; \alpha\mathbf{v} + \mathbf{w}, \quad \text{with } \mathbf{w} \in \ker(E)$$

the scalar α is nonzero. By taking the inner product of both sides of the latter with $\mathbf{v}$, we see that $\alpha > 0$. For an arbitrary integer i we have

$$\begin{aligned}(T^k)_{ij} &= \langle \mathbf{e}_i, T^k\mathbf{e}_j\rangle \\ &= \alpha(\mathbf{v})_i + \langle \mathbf{e}_i, B^k\mathbf{w}\rangle \\ &= \alpha(\mathbf{v})_i + o(1),\end{aligned}$$

and the proof of (14) is complete. This completes the proof of Theorem 1.

6 Proof of Theorem 2

Let $\mathcal{C} = \mathcal{C}_\Phi$ and D_Φ be the digraph associated with Φ. Define $T(x)$ to be the transfer matrix for D_Φ as given in (4). Define $\mathbf{s}(x)$ and $\mathbf{f}(x)$ by (6) and (7). When $T(x)$ is written as a power series, it is seen that the coefficient T_n of x^n is a $(0,1)$-matrix, with $(T_n)_{ij} = 1$ if and only if $(\vec{\nu}_i, \vec{\nu}_j) \in E(D_\Phi)$ and $\Sigma(\vec{\nu}_i) + \Sigma(\vec{\nu}_j) = n$. Since the number of such edges is bounded by the number of compositions of n into $2m$ parts, (m is the common value of the span and modulus of Φ), we see $||T_n||_2 = O(n^m)$, and hypothesis (a) of Theorem 1 is confirmed for $\rho = 1$. Hypothesis (b) is clear, since, as just indicated, each T_n is a $(0,1)$-matrix. Hypothesis (c) is follows from Proposition 1(b).

Let $\mathbf{s}(x), \mathbf{f}(x)$ be defined by (6) and (7). The number of paths π in the sum defining $\mathbf{s}(x)$ is $O(n^{Km})$, constant K coming from property (R4) of regularity, and so $\mathbf{s}(x) \in \ell^2(|x| < 1)$ as claimed. Similarly for $\mathbf{f}(x)$. By the mere existence of recurrent states ν_i we know that $\mathbf{s}(x), \mathbf{f}(x) > 0$ for all $0 < x < 1$, hypothesis (e). The integer k_0 needed to verify hypothesis (f) is supplied from Proposition 1(c). The desired relationship, $F(x^2) = \phi(x) + F_{NR}(x^2)$, follows from (8), and the fact that $F_{NR}(x)$ has radius of convergence 1 is given in Proposition 1(d).

It remains only to show that there exists r such that $\lambda(r) = 1$. By Proposition 1(a) the radius of convergence for $F(x^2)$ is some r^2, $r^2 < 1$. Because $a_n \geqslant 0$, $x = r^2$ is itself a singularity for $F(x^2)$. Because the radius of convergence of $F_{NR}(x^2)$ is 1, it must be the case that $\phi(x)$ has radius of convergence r. It is clear that $\lambda(x) \to 0$ as $x \to 0$. Thus, if there is no root for the equation $\lambda(x) = 1$ in the range $0 < x < 1$, we must have $\lambda(r) < 1$. But then $\phi(x)$ is regular in a neighborhood of $x = r$, as shown in the proof of Theorem 1, a contradiction. This completes the proof of Theorem 2.

Remark 2 Suppose that for any two recurrent vertices ν_i, ν_j there exists k such that all four entries of the 2×2 (i,j) principle subminor of $T(x)^k$ are nonzero. This is a stronger form of recurrence, and implies properties (R1) and (R2). Moreover, the same subminor of $T(x)^{k\ell}$ will dominate $x^{\beta\ell}$ times the 2×2 matrix containing four entries 2^ℓ. Hence, if $(2x^\beta) > 1$, $T(x)$ has spectral radius greater than 1. Using all compositions (2^{n-1}) for a lower bound, this gives

$$\frac{1}{2} \leqslant r \leqslant \left(\frac{1}{2}\right)^{1/\beta}.$$

This stronger form of recurrence is available in all our examples. Since we can prove the existence of r without it (although without the above upper bound), we decided to use the weaker properties (R1) and (R2). □

7 Proof of Theorem 3

Theorem 3 is an immediate corollary of Theorems 1 and 2 when used with this simple observation:

Fact. Let $F(x)$ be a power series and $r > 0$. Suppose that $F(x^2)$ has radius of convergence $\sqrt{r}$, and that the only singularity of $F(x^2)$ on

$$\{|z| = \sqrt{r}\} \cap \{\mathrm{Re}(z) \geqslant 0\}$$

is a simple pole at $z = \sqrt{r}$. Then the radius of convergence of $F(x)$ is r, and the only singularity of $F(x)$ on its circle of convergence is a simple pole at $x = r$.

8 Definitions and Proofs for Theorem 4

Definition 14 (Recurrent local event) *A* local event *is a subset E of*

$$\{0, 1, \ldots, m-1\} \times \mathbb{N}_0^{m+1}$$

for some sufficiently large m. If $\{(i \bmod m), c_i, \ldots, c_{i+m}\} \in E$, we say that E occurs at position i in the composition $\vec{\mathbf{c}}$, with the usual convention that we pad the composition out with zeros. The event E is recurrent *if, for some recurrent $\vec{\mu}$ and $\vec{\nu}$, $(\vec{\mu}, \vec{\nu})$ is an edge of D_Φ such that E occurs in the composition $\vec{\mu}, \vec{\nu}$ with $1 \leqslant i \leqslant m$. (That is, the "window" in E lies entirely within $\vec{\mu}, \vec{\nu}$.)*

Example 10 (Recurrent local events) The requirement that a part belong to a specified subset of $\mathbb{N}$ is a local event for most $\mathcal{C}$. Equal columns at distance two in 2-rowed Carlitz compositions are a local event. A strict local maximum is local event. In this case we take $m \geqslant 3$ and

$$E \;=\; \Big\{\{0, 1, \ldots, m-1\}, \ell_1, \ell_2, \ldots, \ell_m\Big\},$$

where $\ell_i \in \mathbb{N}_0$ and $\ell_1 < \ell_2 > \ell_3$. Suppose the conditions on $\mathcal{C}$ allow arbitrarily long strings of equal parts and we look for local maxima without the strictness condition. In this case, no finite window suffices since we must know the relative size of the parts that border an arbitrarily long string of equal parts. □

Suppose we are keeping track of κ local events We introduce variables $\vec{y}$ into the formula $F_R(x) = \mathbf{s}(x)\sum T(x)^k\mathbf{f}(x)$ to keep track of the random variables as follows. $T(x,\vec{y})_{\mu,\nu}$ will have the same power of x as before. The power of y_i will be the number of times local event E_i occurs in μ,ν, with the first part occurring in μ. The vectors $\mathbf{s}$ and $\mathbf{f}$ are adjusted to keep track of the occurrences of events near the start and end of the compositions.

As with x, the resulting function is holomorphic in $\vec{y}$. Inverting $\lambda(x,\vec{y}) = 1$ gives $x = r(\vec{y})$ a function holomorphic in $\vec{y}$ and Theorem 3 becomes $a_n(\vec{y}) \sim A(\vec{y})r(\vec{y})^{-n}$ for holomorphic A.

Definition 15 (Unrelated events) *With a power series $f(\vec{y})$ in κ variables associate a set $\mathcal{P}(f) \subset \mathbb{R}^\kappa$ of vectors one for each nonzero term as follows. Associate a vector $\vec{m}$ with each nonzero term $C_{\vec{m}}\prod y_i^{m_i}$. Let $\mathcal{V}(f) \subseteq \mathbb{R}^\kappa$ be the vector space spanned over $\mathbb{R}$ by the differences of vectors in $\mathcal{P}(f)$. Given κ recurrent local events E_i, introduce variables $y_1,\ldots,y_\kappa$ in the transfer matrix T to count them. If there are i,j,k,n such that*

$$\mathcal{V}\Big(\sum_{p=1}^{k}[x^n](T^p)_{ij}\Big) = \mathbb{R}^\kappa, \tag{15}$$

we say that the events are unrelated.

Remark 3 This may not sound like the concept roughly stated before Theorem 4: What does $\mathcal{V}$ have to do with the counts being nearly linearly independent? If $\mathcal{V}$ were not the whole space, we would have some $\vec{w}$ such that $\vec{w}\cdot\vec{k}$ were nearly a constant multiple of n for all counts $\vec{\mathbf{k}}$ of compositions of n; a fact we will not prove since it is not needed. Rather, it explains why the precise definition of unrelated is close to the imprecise definition before Theorem 4. $\square$

Make $\mathcal{C}_n$, those $\vec{\mathbf{c}} \in \mathcal{C}$ with $\Sigma(\vec{\mathbf{c}}) = n$, into a probability space by using the uniform distribution. Introduce random variables $\vec{Y}(n)$ that count various recurrent local events. As we shall see,

$$\mathrm{E}(\vec{Y}(n)) \;=\; n\vec{m} + o(n) \quad \text{for some } \vec{m} \gg \vec{0} \qquad \text{and} \qquad \mathrm{cov}(\vec{Y}(n)) \;=\; nB + o(n),$$

where computing the means vector $\vec{m}$ and covariance matrix B seems impossible since they require partial derivatives of an eigenvalue we can only crudely estimate. A standard result tells us that $\vec{Y}(n)$ is asymptotically normal *provided B is nonsingular*. The need to prove nonsingularity of B without recourse to calculating its value motivates Definition 15.

8.1 Three Lemmas

Lemma 5 (Non-overlapping paths) *For this lemma, we call two paths σ_1,σ_2 in D_Φ* non-overlapping *if, whenever π is a path in D_Φ containing σ_1 and σ_2, σ_1 and σ_2 never partially overlap.*

Suppose $\mathcal{C}_\Phi$ is regular and locally restricted and that $\pi_1, \pi_2 \in \mathrm{Path}_\Phi(\vec{\nu}_1, \vec{\nu}_2)$ are not necessarily distinct. Then there are (possibly single vertex) paths $\alpha\vec{\nu}_1$ and $\vec{\nu}_2\omega$ (i.e. α and/or ω may be empty) such that the two paths $\alpha\pi_i\omega$ are non-overlapping.

Proof: Since there are at least two recurrent vertices by Definition 9(R1), it follows from (R2) that the recurrent vertices contain at least two cycles (including loops). Let σ be a cycle in D_Φ of shortest length (possibly just a loop). Let $\vec{\mu}$ be a vertex not on σ. Let α be a path that goes from $\vec{\mu}$ to σ, then traverses σ many more times than σ appears in the π_i, then goes to $\vec{\mu}$ and, finally, goes to $\vec{\nu}_1$, but does not include, $\vec{\nu}_1$. Let ω be a path that goes from $\vec{\nu}_2$ to $\vec{\mu}$, then to σ and finally traverses σ more than α does. $\square$

Lemma 6 (Distribution of a recurrent event) *Let the random variable $Y(n)$ be the number of occurrences of some recurrent local event E in $\mathcal{C}_n$. Then*

(i) $\mathrm{E}(Y(n)) \sim mn$ *for some* $m > 0$ *and*

(ii) the random variable $Z(n) = \frac{Y(n)-\mathrm{E}(Y(n))}{n^{1/2}}$ *converges in distribution to either a point distribution or a normal* $\mathcal{N}(0,\sigma)$.

Proof: The proof of the central limit theorem (Theorem 1) of [3] actually shows a bit more than claimed, namely, that $\mathrm{E}(\vec{Y}) = n\vec{m} + o(n)$ and $\mathrm{cov}(\vec{Y}) = nB + o(n)$ as $n \to \infty$, regardless of whether or not B is nonsingular. We apply it here to the case when $\vec{Y}$ has a single component.

Let $\lambda(x, y)$ be the eigenvalue of the transfer matrix, where x keeps track of sum of parts and y keeps track of the the occurrences of the recurrent event. To study the asymptotic distribution of $Y(n)$, we must solve $1 = \lambda(r(y), y)$ for y near 1. By Lemma 4, λ is holomorphic and is increasing for x and y positive reals. By Theorem 1, $\lambda'(x, 1) > 0$ and so $1 = \lambda(r(y), y)$ has a holomorphic solution near $y = 1$ and, by the previous sentence, is decreasing for positive reals. By Lemma 2(f), $\lambda(x, y)$ is not a constant function of y and so $r(y)$ is not constant. Recall that "mean shifting" [1] looks at the distribution of

$$\Pr(Y(n)\!=\!k) \;=\; \frac{a_{n,k}s^k}{\sum_k a_{n,k}s^k}.$$

One gets $\mathrm{E}(Y(n)) = m(s)n + o(n)$ where $m(s) = sr'(s)/r(s)$. By the definition of $Y(n)$, $m(s)$ is clearly an increasing function of s. Consider $s \in (1-\delta, 1)$. Since $r(y)$ is nonconstant, we must have $r'(s) \neq 0$ somewhere in $(1-\delta, 1)$ and so $m(s) > 0$ for this s. Claim (i) follows from monotonicity of $m(s)$. Claim (ii) follows from the opening paragraph of the proof. $\square$

Lemma 7 *Let $X, Y_1, Y_2, \ldots, Y_k$ be random variables. A value is assigned X as follows. First choose $1 \leqslant i \leqslant k$ with probability p_i. Then let X be the value of Y_i. It follows that* $\mathrm{var}(X) \geqslant \sum_{i=1}^k p_i \mathrm{var}(Y_i)$.

Proof: We have $\mathrm{var}(X) = \sum_{i=1}^k p_i \mathrm{E}((Y_i - \mathrm{E}(X))^2)$. Since the minimum value of $\mathrm{E}((Z-t)^2)$ occurs at $t = \mathrm{E}(Z)$, the result follows. $\square$

8.2 Proof of Theorem 4

As remarked in the proof of Lemma 6, the proof of the central limit theorem (Theorem 1) of [3] actually shows a bit more than claimed, namely, that $\mathrm{E}(\vec{Y}) = n\vec{m} + o(n)$ and $\mathrm{cov}(\vec{Y}) = nB + o(n)$ as $n \to \infty$, regardless of whether or not B is nonsingular.

By Lemma 6(i), $\vec{m} \gg \vec{0}$. It remains to prove that B is nonsingular.

Suppose B is singular and let $\vec{w} \neq \vec{0}$ be such that $B\vec{w} = \vec{0}$. Define the random variable $X = \vec{w} \cdot \vec{Y}$. Then

$$\mathrm{var}(X) \;=\; n\vec{w}^t B\vec{w} + o(n) \;=\; o(n).$$

By (15) there exists

$$\vec{p} \;=\; \vec{m}_1 - m_2\vec{b} \;\in\; \Big(\sum_{p=1}^{k}[x^n](T^p)_{ij}\Big)$$

such that $\vec{w}\cdot\vec{p} \neq 0$. Let $\vec{\nu}_1, \vec{\nu}_2$ correspond to the indices i, j of T. Let $\pi_i \in \mathrm{Path}_\Phi(\vec{\nu}_1, \vec{\nu}_2)$ give rise to the counts $\vec{m}_i$. Apply Lemma 5 to obtain two non-overlapping paths $\sigma_i = \alpha\pi_i\omega$. Note that σ_i gives rise to some count $\vec{m}_i + \vec{k}$, where $\vec{k}$ does not depend on i because π_1 and π_2 have been extended by using the same paths and events are local. Hence the difference of these counts is still $\vec{p}$.

Apply Lemma 6 to conclude that, for some δ and ϵ and all sufficiently large n, at least a fraction ϵ of the compositions of n contain at least δn copies of σ_1. To do this, one needs to increase the modulus m since a local event's span is limited by the modulus and we are looking for σ_1 which has a multiple of m parts. Lemma 1(iii) justifies this.

For a composition $\vec{\mathbf{c}}$, define $\chi(\vec{\mathbf{c}})$ to be the composition obtained by replacing all occurrences of σ_2 with σ_1. Partition the compositions of n into sets S_i as follows.

- If the total number of copies of σ_1 in $\chi(\vec{\mathbf{c}})$ is less than δn, place $\vec{\mathbf{c}}$ in S_1.
- Partition those compositions $\vec{\mathbf{c}}$ for which $\chi(\vec{\mathbf{c}})$ has at least δn copies of σ_1 into equivalence classes where $\chi(\vec{\mathbf{c}}) = \chi(\vec{\mathbf{d}})$ defines equivalence.

We now apply Lemma 7. When $i \neq 1$, the random variable Y_i has a binomial distribution on at least δn items and so has variance at least $\delta n/4$. Since the probability that $\vec{\mathbf{c}} \in S_1$ is at most $(1-\epsilon)n$, it follows that $\mathrm{var}(X) \geqslant \epsilon\delta n/4$, contradicting the earlier result that its variance was $o(n)$. This contradiction completes the proof of the nonsingularity of B.

9 The Largest Part and Number of Distinct Parts

Let Φ be a local restriction function with modulus and span m. (See Sections 2.1 and 2.2 for definitions and notation.) Let the random variable M_n (resp. D_n) be the largest part (resp. number of distinct parts) in a locally restricted composition of n selected uniformly at random. We begin with upper bounds on M_n and D_n. Since $D_n \leqslant M_n$, it suffices to bound M_n. Consider compositions of n that contain the part p. We now overcount such compositions. All of them can be constructed by choosing $0 \leqslant t \leqslant n - p$, forming compositions $\vec{\mathbf{a}}$ of t and $\vec{\mathbf{b}}$ of $n-t-p$, and then considering the composition $\vec{\mathbf{c}} = \vec{\mathbf{a}}, p, \vec{\mathbf{b}}$. We allow $\vec{\mathbf{a}}$ to end in any fashion whatsoever and $\vec{\mathbf{b}}$ to begin in any fashion whatsoever (the vectors $\mathbf{s}$ and $\mathbf{f}$ of (8)). According to Example 4, this alters the constant factor in the asymptotics but does not change the exponential growth rate. Many of the compositions $\vec{\mathbf{c}}$ produced in this manner will be illegal and compositions with k copies of p will be produced k times; however, none of this matters since we are interested in an overcount. It follows that the number of compositions of n with largest part exceeding L is bounded by

$$\sum_{p>L}\sum_{t=0}^{n-p}(C_1 r^{-t})(C_2 r^{-(n-t-p)}) \;<\; C_1 C_2 r^{-n} n \sum_{p>L} r^p \;<\; C_3 n r^{L-n}.$$

If $nr^L = o(1)$, it follows that $M_n \leqslant L$ almost surely. Hence, for any function $g(n) \to \infty$,

$$M_n \;\leqslant\; \log_{1/r}(n) + g(n) \;\text{ almost surely.}$$

Unlike the upper bounds, our proof of lower bounds requires the construction of valid compositions and hence requires more information about Φ. In fact, it was shown in [2] that, in some situations, one can have $M_n = \Theta((\log n)^{1/2})$ almost surely.

Let $\vec{\nu}$ be a recurrent vertex in D_Φ. By Theorem 4 there is a $\delta > 0$ such that a composition of n almost surely contains more than δn copies of $\vec{\nu}$. For each n, let this set of compositions be $\mathcal{S}_n$ and let $\mathcal{S}_n(p)$ be the subset containing no part of size p.

For each $p \in \mathbb{N}$, let $\pi = \mathrm{Path}_\Phi(\vec{\nu}, \vec{\nu})$ be a composition containing p for which $\Sigma(\pi)$ is a minimum. Let $w(p) = \Sigma(\pi) - \Sigma(\vec{\nu})$. If no such π exists, set $w(p) = \infty$. We can construct compositions containing p by choosing a composition in $\mathcal{S}_n(p)$ and replacing one occurrence of $\vec{\nu}$ with π. This gives us at least δn compositions of $n + w(p)$. Using Theorem 3 we obtain

$$\delta n |\mathcal{S}_n(p)| < (A+\delta) r^{-(n+w(p))}$$

for all sufficiently large n. Thus

$$|\mathcal{S}_n(p)| < \frac{A+\delta}{\delta n r^{w(p)}}.$$

For $\mathcal{P} \subset \mathbb{N}$, the probability that a composition in $\mathcal{S}_n$ contains no parts in $\mathcal{P}$ is at most

$$\mathrm{miss}(\mathcal{P}) \;=\; \frac{A+\delta}{n}\sum_{p\in\mathcal{P}} r^{-w(p)}. \tag{16}$$

The plan is to choose $\mathcal{P}$ depending on n so that (16) goes to zero as $n \to \infty$. To get a lower bound on M_n, we choose $\mathcal{P}$ to contain one large part since almost surely M_n is at least as large as this part. To get a lower bound on D_n, we choose $|\mathcal{P}|$ as large as possible since almost surely $D_n \geqslant |\mathcal{P}|$. Because of the variety of possibilities, we simply describe some results.

(a) Suppose $w(p) = p + O(1)$ for all p. Then (16) and the upper bounds obtained earlier give $M_n \sim D_n \sim \log_{1/r}(n)$ almost surely.

(b) This is easily extended to the case when $w(p_i) = p_i + o(1)$ for an infinite sequence of increasing parts satisfying $p_{i+1}/p_i \to 1$.

(c) If one still has $w(p_i) = p_i + o(1)$ but has only that p_{i+1}/p_i is bounded as $i \to \infty$, then the lower bounds on M_n and D_n become $\theta(\log n)$.

(d) If limitations on differences between nearby parts lead to $w(p) = O(p^\alpha)$, one obtains lower bounds on M_n and D_n of $O((\log n)^{1/\alpha})$. This arose with $\alpha = 2$ in [2] when the set of allowed adjacent differences was finite.

(e) If some condition on p_{i+1}/p_i such as that in (c) is added to (d), we get $\Theta((\log n)^{1/\alpha})$.

The ideas in this section are adaptations of the material in Sections 7 and 8 of [2].

10 Longest Runs

Considerable information has been obtain by Grabner et al. [7] about runs in unrestricted compositions. Their method uses explicit generating functions, which we lack in the general case. Consequently, we obtain the cruder results contained in Theorem 5.

Let $\vec{\mathbf{p}}$ be a sequence of parts that can form arbitrarily long runs. Given a locally restricted composition of n selected uniformly at random, let $R_n(\vec{\mathbf{p}})$ be the maximum run length of $\vec{\mathbf{p}}$ in n. Let $p = \Sigma(\vec{\mathbf{p}})$.

We begin with an upper bound for $R_n(\vec{\mathbf{p}})$ following the method in the previous section, again by overcounting. All compositions containing a run of length at least k can be constructed by choosing $0 \leqslant t \leqslant n - kp$, forming compositions $\vec{\mathbf{a}}$ of t and $\vec{\mathbf{b}}$ of $n - kp - t$, and then considering the composition $\vec{\mathbf{c}} = \vec{\mathbf{a}}, (\vec{\mathbf{p}},)^k \vec{\mathbf{b}}$. We allow $\vec{\mathbf{a}}$ to end in any fashion whatsoever and $\vec{\mathbf{b}}$ to begin in any fashion whatsoever. The total number of compositions constructed is

$$\sum_{t=0}^{n-kp-t} (C_1 r^{-t})(C_2 r^{-(n-kp-t)}) \ < \ C_3 r^{-n}(n r^{kp}). \tag{17}$$

If $nr^{kp} = o(1)$, it follows that $R_n(\vec{\mathbf{p}}) < k$ almost surely. Hence, for any function $g(n) \to \infty$,

$$R_n(\vec{\mathbf{p}}) \ \leqslant \ \frac{\log_{1/r}(n)}{\Sigma(\vec{\mathbf{p}})} + g(n) \ \text{ almost surely.} \tag{18}$$

We now obtain lower bounds, again following the method of the previous section. Let $\vec{\nu}$, δ and $\mathcal{S}_n$ be as there: $\vec{\nu}$ is recurrent, a composition of n almost surely has δn copies of $\vec{\nu}$, and $\mathcal{S}_n$ is that set of compositions of n. Let $\mathcal{S}_n(k)$ be the subset that does not contain a run of $\vec{\mathbf{p}}$ of length at least k.

Let $\pi_k = \text{Path}_\Phi(\vec{\nu}, \vec{\nu})$ contain a run of $\vec{\mathbf{p}}$ of length k. Choose π_k so that $\Sigma(\pi_k)$ is minimal. When k is sufficiently large, we can insert m additional copies of $\vec{\mathbf{p}}$ into π to obtain an m longer run of $\vec{\mathbf{p}}$. (Recall that m is the modulus and span of the local restriction.) Thus $s_k = \Sigma(\pi_k) - \Sigma(\vec{\nu})$ grows linearly with growth rate $\Sigma(\vec{\mathbf{p}})$. Thus $s_k < a + k\Sigma(\vec{\mathbf{p}})$.

For each composition in $\mathcal{S}_n(k)$ replace each $\vec{\nu}$ in turn with π_k, obtaining at least $\delta n|\mathcal{S}_n(k)|$ distinct compositions of $n+s_k$ containing a run of length at least k. Since this cannot exceed the total number of compositions, it is bounded by $(A+\delta)r^{-(n+s_k)}$ and so

$$|\mathcal{S}_n(k)| \;<\; \frac{A+\delta}{\delta n r^{n+s_k}} \;<\; \frac{Br^{-n}}{nr^{k\Sigma(\vec{\mathbf{p}})}}$$

for some B. Thus $\mathcal{S}_n(k)$ will be a negligibly small fraction of all compositions provided

$$k \;<\; \frac{\log_{1/r}(n)}{\Sigma(\vec{\mathbf{p}})} - g(n)$$

for some $g(n) \to \infty$. Combining this with (18), we obtain

$$R_n(\vec{\mathbf{p}}) \;\sim\; \frac{\log_{1/r}(n)}{\Sigma(\vec{\mathbf{p}})} \text{ almost surely,}$$

which is the first part of Theorem 5.

We now turn to (a) in the theorem. Let L be the minimum of $\Sigma(\vec{\mathbf{p}})$ over $\mathcal{R}$. We sum (17) over all $\vec{\mathbf{p}}$ with $\Sigma(\vec{\mathbf{p}}) > L$, not just those in $\mathcal{R}$. Group terms by the value p of $\Sigma(\vec{\mathbf{p}})$. Since there are less than 2^p compositions of p, an upper bound on the number of compositions of n containing a run of length at least k of anything with $\Sigma > L$ is

$$\sum_{p>L} 2^p C_3 r^{-n}(nr^{kp}) \;=\; C_3 r^{-n}\, n \sum_{p>L} (2r^k)^p \;<\; C_4 r^{-n}\, n(2r^k)^{L+1}. \tag{19}$$

Fix $\delta > 0$ so that $(1-2\delta)(L+1) > L$. Let $k = (1-\delta)\Big((\log_{1/r}(n))/L\Big)$ and let n be large. The right side of (19) is bounded above by

$$C_4 r^{-n}\, n \left(\frac{2}{n^{(1-\delta)/L}}\right)^{L+1} < \; C_4 r^{-n} \frac{n}{n^{(1-2\delta)(1+L)/L}} \;=\; o(a_n),$$

the last by Theorem 5. We know that the smallest case almost certainly has a run of length asymptotic to $(\log_{1/r}(n))/L$. This completes the proof of (a).

The proof of (b) is easier than that for (a). Let $\mu(\vec{\mathbf{p}}) = \Sigma(\vec{\mathbf{p}})/\mathrm{len}(\vec{\mathbf{p}})$, the average part size in $\vec{\mathbf{p}}$. By the first part of the theorem,

$$\mathrm{len}(\vec{\mathbf{p}})\, R_n(\vec{\mathbf{p}}) \;\sim\; \frac{\log_{1/r}(n)}{\mu(\vec{\mathbf{p}})} \text{ almost surely,} \tag{20}$$

Let μ_0 be the minimum of μ over $\mathcal{R}$ and let μ_1 be the smallest value of μ in $\mathcal{R}$ that is not equal to μ_0. The bound in (17) can be written $C_3 r^{-n}(nr^{k\mu})$, where k is the number of parts in the run. Thus we get a bound of $C_3|\mathcal{R}|r^{-n}(nr^{k\mu_1})$ for a run using any $\vec{\mathbf{p}}$ with $\mu(\vec{\mathbf{p}}) > \mu_0$. When this is combined with (20) for $\mu(\vec{\mathbf{p}}) = \mu_0$, part (b) follows easily.

11 Palindromes

For locally restricted palindromes, we assume that Φ is invariant under reversal, obtaining results as for regular locally restricted compositions. In this case,

$$\mathbf{a}(x) \;=\; \mathbf{s}(x)^{\mathrm{t}} \sum_{k=0}^{\infty} T(x)^k \tag{21}$$

Is essentially the generating function for the first half of the palindrome evaluated at x^2. Each composition $\vec{\mathbf{c}}$ counted in $\mathbf{a}(x)_{\vec{n}u}$ followed by a symmetric connecting composition $\vec{\mathbf{d}}$ of at most $2m-1$ parts and then by $\vec{\mathbf{c}}^r$, the reversal of $\vec{\mathbf{c}}$. Thus the components of $\mathbf{a}(x)$ are multiplied by generating functions for the various $\vec{\mathbf{d}}$'s and summed to produce the generating function for palindromes. If the value of Φ varies with its first argument, this effect must be taken into account when $\vec{\mathbf{d}}$ is chosen so that $\vec{\mathbf{c}}^r$ is acceptable at its position. This could be annoying. Regardless, it is the sum in (21) that determines the asymptotics and so one obtains the same result as for general locally restricted compositions.

With m-rowed compositions, there are at least two possible definitions of palindrome: The reversal of

$$\begin{matrix} a_{1,1} & a_{1,2} & \cdots & a_{1,k} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,k} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,k} \end{matrix} \quad \text{may be} \quad \begin{matrix} a_{1,k} & \cdots & a_{1,2} & a_{1,1} \\ a_{2,k} & \cdots & a_{2,2} & a_{2,1} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m,k} & \cdots & a_{m,2} & a_{m,1} \end{matrix} \quad \text{or} \quad \begin{matrix} a_{m,k} & \cdots & a_{m,2} & a_{m,1} \\ \cdots & \cdots & \cdots & \cdots \\ a_{2,k} & \cdots & a_{2,2} & a_{2,1} \\ a_{1,k} & \cdots & a_{1,2} & a_{1,1} \end{matrix}.$$

Both can be dealt with in the manner described in the previous paragraph.

Acknowledgment

We are grateful to the referee for suggesting that we consider runs; this led to Section 10.

References

[1] E.A. Bender, Central and local limit theorems applied to asymptotic enumeration, *J. Combinatorial Theory* **15** (1973) 91– 111.

[2] E.A. Bender and E.R. Canfield, Locally restricted compositions I. Restricted adjacent differences, *Elec. J. Combin.* **12**(1) (2005) R57, 27pp.

[3] E.A. Bender and L.B. Richmond, Central and local limit theorems applied to asymptotic enumeration II: Multivariate generating functions, *J. Combin. Theory, Series A* **34** (1983) 255–265.

[4] J.B. Conway, *A Course in Functional Analysis*, 2nd ed., Springer (1990).

[5] K. Deimling, *Nonlinear Functional Analysis*, Springer (1985).

[6] N. Dunford and J.T. Schwartz, *Linear Operators Part I: General Theory*, (1957); reprinted by J. Wiley & Sons (1988).

[7] P. Grabner, A. Knopfmacher and H. Prodinger, Combinatorics of geometrically distributed random variables: Run statistics, *Theoret. Computer Sci.* **297** (2003) 261–270.

[8] F. Hirsch and G. Lacombe, *Elements of Functional Analysis*, Springer (1999).

[9] S. Heubach and S. Kitaev, Avoiding substrings in a composition, `arXiv:0903.5135v2`.

[10] S. Heubach, S. Kitaev, and T. Mansour, Avoidance of partially ordered patterns in compositions, *Pure Math. Appl.* **17** (2006) 123–134.

[11] T. Kato, *Perturbation Theory for Linear Operators*, Springer (1980).

[12] S. Kitaev, T.B. McAllister and T.K. Petersen, Enumerating segmented patterns in compositions and encoding by restricted permutations, *Integers* **6** (2006), A34, 16 pp. (electronic).

[13] A. Knopfmacher and H. Prodinger, On Carlitz compositions, *European J. Combin.* **19** (1998) 579–589.

[14] M.G. Krein and M.A. Rutman, Linear operators leaving invariant a cone in a Banach space, translation by M. M. Day of the 1948 paper in Russian, appears in *Functional Analysis and Measure Space*, AMS Translations, Series One, Volume 10 (1962).

[15] E. Munarini, M. Poneti and S. Rinaldi, Matrix compositions, in *Formal Power Series and Algebraic Combinatorics*, San Diego, California, 2006. Preprint at `http://garsia.math.yorku.ca/fpsac06/papers/49.pdf`

[16] Amy N. Myers, Forbidden substrings on weighted alphabets, *Australasian J. Math.*, to appear.

[17] F. Riesz, Über lineare Funktionalgleichungen, *Acta Mathematica* **41** (1918) 71–98.

[18] F. Riesz and B. Sz.-Nagy, *Functional Analysis*, Frederick Ungar Publishing Company, 1955.

[19] C. Savage and H.S. Wilf, Pattern avoidance in compositions and multiset permutations, *Adv. Appl. Math.* **36** (2006) 194–201.