Multitype branching processes in random environments with not strictly positive expectation matrices

VILMA ORGOVÁNYI[1,a] and KÁROLY SIMON[2,b]

[1] *Department of Stochastics, Institute of Mathematics, Budapest University of Technology and Economics, Műegyetem rkp. 3., H-1111 Budapest, Hungary,* [a]*orgovanyi.vilma@gmail.com*
[2] *HUN-REN-BME Stochastics Research Group, Műegyetem rkp. 3., H-1111 Budapest, Hungary,* [b]*karoly.simon51@gmail.com*

It is well known that under some conditions the almost sure survival probability of a multitype branching process in random environment is positive if the Lyapunov exponent corresponding to the expectation matrices is positive, and zero if the Lyapunov exponent is negative. The goal of this note is to establish similar results when certain positivity conditions on the expectation matrices are not met. One application of such a result is to classify the positivity of Lebesgue measure of certain overlapping random self-similar sets.

MSC2020 subject classifications: Primary 60J80; 60J85; secondary 28A80
Keywords: Multitype branching process in random environment; random fractals

1. Introduction

In this paper we consider the extinction probability of multitype branching processes in random environments $\{\mathbf{Z}_n\}_{n=1}^{\infty}$, formally defined starting in Section 1.5. Our main theorem (Theorem 4.4) states that, under mild conditions, the positivity of the Lyapunov exponent corresponding to the expectation matrices determines the positivity of the survival probability. Similar results had already been established (see the next subsection), we extend the scope of these results and show an application in fractal geometry, where the extension we give is necessary.

Informally, a branching process is:

- *multitype* if each individual has a type and the type determines the distribution according to which it gives birth to different types of individuals;
- *temporally non-homogeneous* or *in varying environment* if we allow the offspring distribution to change over time in a predefined deterministic manner; and
- *in a random environment* if the temporal non-homogeneity is non-deterministic.

More precisely, let $N \geq 2$ (the number of types) and assume we are given a stationary ergodic sequence $\{\theta_n\}_{n\geq 1}$ of random variables called the environmental sequence. For the definition and properties of a stationary ergodic sequence see (Karlin and Taylor, 1975, Chapter 9.5.). For almost every realization, we consider the associated sequence of N-dimensional vectors of N-variate probability generating functions (pgfs)

$$\left\{\mathbf{f}_{\theta_n}(\mathbf{s}) = \left(f_{\theta_n}^{(0)}(\mathbf{s}), \ldots, f_{\theta_n}^{(N-1)}(\mathbf{s})\right)\right\}_{n\geq 1}.$$

The i-th component $f_{\theta_n}^{(i)}(\mathbf{s})$ of $\mathbf{f}_{\theta_n}(\mathbf{s})$ is

$$f_{\theta_n}^{(i)}(\mathbf{s}) = \sum_{\mathbf{j}\in\mathbb{N}_0^N} f_{\theta_n}^{(i)}[\mathbf{j}] \cdot \prod_{k=0}^{N-1} s_k^{j_k} \text{ for } \mathbf{s} = (s_0, \ldots, s_{N-1}) \in [0,1]^N,\, \mathbf{j} = (j_0, \ldots, j_{N-1})$$

where $f_{\theta_n}^{(i)}[\mathbf{j}]$ is the probability that a level $n-1$ individual of type i gives birth to j_k individuals of type k in environment θ_n.

If we condition on a realization $\overline{\boldsymbol{\theta}} = (\theta_n)_{n\geq 1}$ of the environmental process then

$$\left\{\mathbf{Z}_n(\overline{\boldsymbol{\theta}}) = \left(Z_n^{(0)}(\overline{\boldsymbol{\theta}}), \ldots, Z_n^{(N-1)}(\overline{\boldsymbol{\theta}})\right)\right\}_{n=1}^{\infty}$$

behaves like an N-dimensional temporally non-homogeneous branching process, where $Z_n^{(i)}(\overline{\boldsymbol{\theta}})$ is the number of type i individuals in the n-th generation in environment $\overline{\boldsymbol{\theta}}$. In this way the offspring distribution of a type i individual in the n-th generation is given by the pgf $f_{\theta_{n+1}}^{(i)}$. For a θ_n we consider the $(N \times N)$ expectation matrix $\mathbf{M}_{\theta_n}$,

$$\mathbf{M}_{\theta_n}(i,j) := \frac{\partial f_{\theta_n}^{(i)}}{\partial s_j}(\mathbf{1}),$$

where $\mathbf{1} \in \mathbb{R}^N$ is the vector with all components equal to 1. For a non-negative matrix $\mathbf{M}$ let $\|\mathbf{M}\|$ denote the sum of all of its elements. The Lyapunov exponent corresponding to the matrices and the environmental sequence is

$$\lambda := \lim_{n\to\infty} \frac{1}{n} \log \|\mathbf{M}_{\theta_1} \cdots \mathbf{M}_{\theta_n}\|,$$

where the limit exists and is the same constant for almost all $(\theta_n)_{n\geq 1}$, see Appendix C for more details.

1.1. Our result in a special case

In the special case where $\overline{\boldsymbol{\theta}} = (\theta_n)_{n\geq 1}$ is a stationary ergodic process over a finite alphabet, our result implies the following:

Theorem 1.1. *Assume that*

(a) $\overline{\boldsymbol{\theta}} = (\theta_n)_{n\geq 1}$ *is a stationary ergodic process over a finite alphabet* $[L] := \{0, \ldots, L-1\}$.
(b) For every $\theta \in [L]$ *the expectation matrix* $\mathbf{M}_\theta$ *is allowable, i.e. it contains a strictly positive number in each row and column.*
(c) There exists an n *and a* $(\theta_1, \ldots, \theta_n) \in [L]^n$ *such that all elements of the product matrix* $\mathbf{M}_{\theta_1} \cdots \mathbf{M}_{\theta_n}$ *are strictly positive.*
(d) There exists a $K < \infty$ *such that for every* $\theta \in [L]$ *and* $i, j, k \in [N]$, $\frac{\partial^2 f_\theta^{(k)}}{\partial s_j \partial s_i}(\mathbf{1}) < K$.

Under these conditions we have:

1. *If* $\lambda > 0$ *then in almost every environment* $\overline{\boldsymbol{\theta}}$, *starting with a single individual of arbitrary type, the process* $\mathbf{Z}_n(\overline{\boldsymbol{\theta}})$ *does not die out with positive probability. Moreover,* $\lim_{n\to\infty} n^{-1}\log(\|\mathbf{Z}_n\|) = \lambda$, *conditioned on the process does not die out.*

2. *If $\lambda < 0$ then in almost every environment $\overline{\boldsymbol{\theta}}$, starting with a single individual of arbitrary type, the process $\mathbf{Z}_n(\overline{\boldsymbol{\theta}})$ dies out almost surely.*
3. *If $\lambda = 0$ and with positive probability we can find θ such that for every type i the probability that a type i individual gives birth to only 0 or 1 child is less than 1 (i.e. for all i $f_\theta^{(i)}[\mathbf{0}] + \sum_{j\in[N]} f_\theta^{(i)}[\mathbf{e}_j] < 1$) then starting with a single individual of arbitrary type, the process $\mathbf{Z}_n(\overline{\boldsymbol{\theta}})$ dies out almost surely.*

In the text, the proof of part 1 is given in Corollary 4.6 and 4.10, for the proof of part 2 and part 3 see Corollary 4.9 and Lemma 4.8.

1.2. Corresponding literature

For a recent and detailed survey about MBPREs see Vatutin and Dyakonova (2021). For the convenience of the reader we give a short description of the literature here as well. The extinction problem for MBPREs was investigated in (Athreya and Karlin, 1971, Theorem 8). They proved that under some conditions in almost every environment:

$$\lambda < 0 \implies \text{almost sure extinction, and } \lambda > 0 \implies \text{survival with positive probability.} \tag{1.1}$$

The conditions of (Athreya and Karlin, 1971, Theorem 8) include the assumption that for every θ and i, j we have $\mathbf{M}_\theta(i,j) > 0$. In the same year in Weissner (1971) and later in Tanny (1981) this assumption was weakend—they required that there exists a k such that for all $\theta_1, \dots, \theta_k$ of positive probability the corresponding product of the expectation matrices is strictly positive, i.e.

$$\exists k,\ \forall \theta_1, \dots, \theta_k \text{ with } \nu(\theta_1, \dots, \theta_k) > 0,\ \forall i, j:\ (\mathbf{M}_{\theta_1} \cdots \mathbf{M}_{\theta_k})(i,j) > 0, \tag{1.2}$$

where ν is the distribution of the environmental sequence (see Definition 2.5). This condition always fails in the case that at least one of the matrices is triangular and the environment is an i.i.d. sequence (assuming that every letter has positive probability). In particular this happens in our motivating example described in Section 1.3. In this note we weaken the positivity assumptions. More precisely, we assume that

- each of the expectation matrices is "uniformly" allowable (see Definition 4.1, which always holds when we assume that the underlying alphabet is finite, and all of our matrices have a strictly positive element in every row and every column) and
- we require (Cf. (1.2))

$$\exists k,\ \exists \theta_1, \dots, \theta_k \text{ with } \nu(\theta_1, \dots, \theta_k) > 0,\ \forall i, j:\ (\mathbf{M}_{\theta_1} \cdots \mathbf{M}_{\theta_k})(i,j) > 0. \tag{1.3}$$

1.3. Application to random self-similar sets

Our motivation to consider multitype branching processes in random environments (MBPREs) comes from the theory of random self-similar sets. These sets are generalizations of the Mandelbrot percolation (or fractal percolation) set in a sense that we run a similar random process on the cylinder sets of a d-simensional self-similar IFS, the IFS is usually denoted by $\mathcal{S}$ and the probability parameter is denoted by p. See subsection Random Sierpiński carpet below for a more detailed explanation. In the application of MBPREs to random self-similar sets in most natural cases the finite set of induced expectation matrices contain triangular matrices. Hence, the positivity conditions of the already existing results are not satisfied. Namely, no power of a triangular matrix is strictly positive. We present this application on a concrete example: the 45-degree projection of the random Sierpiński carpet. Many other examples of random self-similar sets for which the theory has been applied can be found in Orgoványi and Simon (2024).

The positivity of Lebesgue measure of this projection depends on the extinction probability of a corresponding multitype branching processes in finitely generated random environment (where at each step we choose the driving distribution from the same finite family of distributions uniformly and independently). The survival of these type of multitype branching processes is addressed in Theorem 1.1 1. The explanations corresponding to this connection between the Lebesgue measure of the projection of the random Sierpiński carpet and an MBPRE given in this section are heuristic, their purpose is to enlighten the underlying ideas. Eventually the connection between MBPREs and random self-similar sets and the numerical estimation of the Lyapunov exponent for the expectation matrices done by Pollicott and Vytnova leads to the following proposition, which in more detail is repeated at the end of this section. Here p denotes the probability parameter of the random system.

Proposition 1.2. *There exists a $p_* \in [0.3745\ldots, 0.386\ldots]$ such that the 45-degree projection of the random Sierpiński carpet has positive Lebesgue measure almost surely conditioned on non-extinction if and only if $p > p_*$.*

The random Sierpiński carpet was introduced in (Dekking and Meester, 1990, Example 1.1), projections of similar random constructions have been studied, for example in Falconer (1989) and Falconer and Grimmett (1992) and in particular positivity of Lebesgue measure of projections (only onto the coordinate axes, which require a different treatment) of such sets has been considered in Dekking and Grimmet (1988). The size (existence of interior points and positivity of Lebesgue measure) of algebraic differences of 1-dimensional random Cantor sets has been investigated in for example Dekking and Simon (2008), Móra, Simon and Solomyak (2009).

Deterministic Sierpiński carpet

The deterministic Sierpiński carpet is the attractor of the iterated function system (IFS)

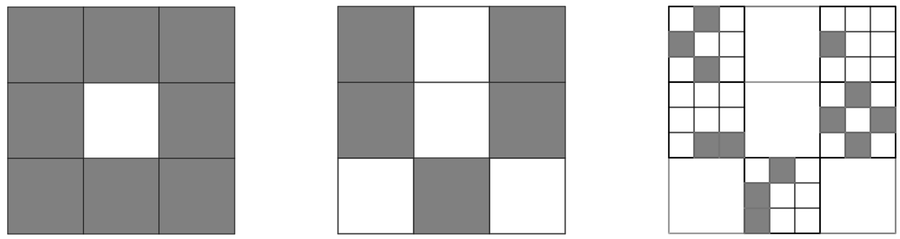
$$\mathcal{S} = \left\{ S_i(\mathbf{x}) = \frac{1}{3}\mathbf{x} + \mathbf{t}_i \right\}_{i=0}^{7},$$

of finitely many contracting maps in $\mathbb{R}^2$, where $\{t_i\}_{i=0}^7$ is an enumeration of the set $\{0, 1/3, 2/3\}^2 \setminus \{1/3, 1/3\}$. For the first level approximation, see Figure 1a. For more details on iterated function systems see Bárány, Simon and Solomyak (2023).

Random Sierpiński carpet

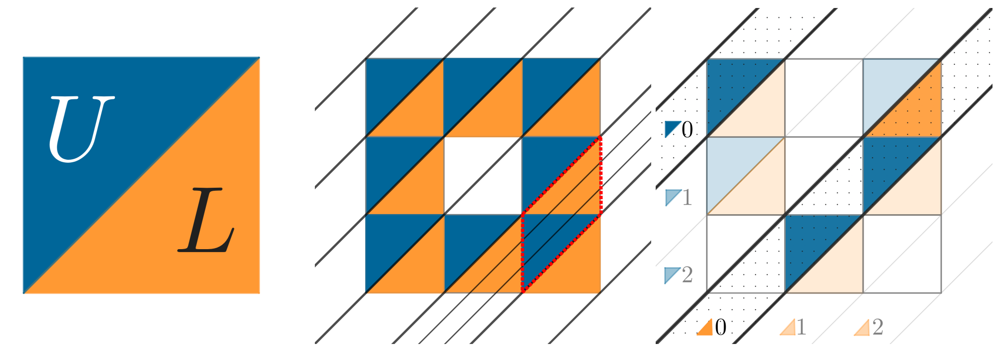
Now we give an informal description of how to obtain the random Sierpiński carpet. First, we fix a probability parameter $p \in [0, 1]$. In a given square we repeat the following two steps:

- We subdivide the square into 9 congruent subsquares and discard the middle one to achieve the first level approximation of the deterministic Sierpiński carpet. (For the first level approximation of the deterministic Sierpiński carpet see Figure 1a.)
- Each of the remaining 8 congruent squares is retained with probability p and discarded with probability $1 - p$ independently of each other. (For a possible realization of this step see Figure 1b).



(a) Sierpiński carpet. (b) Realization (level 1). (c) Realization (level 2).

Figure 1



(a) Upper and lower triangle. (b) Types and columns. (c) Zeroth columns.

Figure 2

We start with the $[0,1]^2$ unit square, and repeat the above-described process in the retained cubes independently of each other ad infinitum, or until there are no cubes left. For a realization of a level 2 approximation see Figure 1c. Formally the construction is defined for example in (Orgoványi and Simon, 2023, Definition 1.1).

Projection and multitype branching processes in random environments

Here we give an informal description of how to obtain the MBPRE corresponding to the 45-degree projection of the random Sierpiński carpet. To investigate the 45-degree projection of the random Sierpiński carpet instead of analyzing the birth and death processes in squares, it is more practical to subdivide the square into two triangles as shown in Figure 2a for the level zeroth and 2b for the level 1 triangles (in the deterministic Sierpiński case, before the randomization) and analyze the process in terms of triangles. This strategy can be familiar from for example Dekking and Simon (2008) and Móra, Simon and Solomyak (2009). We begin by identifying the types and the environments.

Consider the two level 0 triangles, U (as upper) and L (as lower) shown in Figure 2a. These are the prototypes of the two *type* $\mathfrak{U}$ and $\mathfrak{L}$. Every triangle that will occur in the process is of type $\mathfrak{U}$ or type $\mathfrak{L}$. Upper facing triangles are of type $\mathfrak{U}$ and lower facing triangles are of $\mathfrak{L}$. (More precisely, upper and lower facing triangles are the rescaled and translated version of the level 0 triangles, U and L respectively.)

The 6 level 1 columns are the stripes which are bounded by 2 consecutive of the 7 45-degree lines going through the vertices (the lines are the thicker lines of Figure 2b). We number these columns according to Figure 2c.

The columns are the building blocks of the environments. In level 1, both level 0 triangles (U and L) are subdivided into 3 columns, containing $1/3$-rd smaller upper and lower facing triangles (Figure 2b), numbered as shown in Figure 2c. For each level 1 triangle we consider the corresponding level 2 sub-columns in an analogous way—the lines defining the level 2 columns (for the level 1 upper and lower triangles in the framed area) are drawn with thinner lines on Figure 2b. In levels deeper than the zeroth the upper and lower triangles in a given column are always positioned in a way, that their sub-columns pairwise coincide, as shown in Figure 2b; for example here on Figure 2b inside the red dotted area a level 1 upper and a level 1 lower triangle is visible, considering their next-level columns, it is clear that these column-boundaries coincide. For this reason hence we pair the 3 sub-columns of the upper and the 3 sub-columns of the lower triangle already in level 0, in the way that the corresponding column boundaries coincide. The result is 3 pairs of columns which we identify with the numbers $0, 1, 2$ as depicted in Figure 2c. Inside a level n column-pair we again have 3 column-pairs. The *environment* is an infinite sequence of column-pairs, eventually determining exactly two sections: one in the zeroth level U and one in the zeroth level L triangle. This particular numbering comes handy, because if we rescale the projection to $[0, 2]$ then the projection of the level 0 U and L triangles are the intervals $[0, 1]$ and $[1, 2]$ respectively, inside which the environmental sequence gives the triadic code of the point.

For example, in the randomized case in Figure 2c the zeroth column in a particular realization is highlighted. The particular realization is the first level of a random Sierpiński carpet from Figure 1b which we further inspect now using Figure 2c. We focus on the number 0 column-pair. The level-1 type $\mathfrak{U}$ triangle contains one level-2 type $\mathfrak{U}$ with probability p (this is the case in this realization), and nothing with probability $1 - p$. The number 0 column of the level 0 type $\mathfrak{L}$ triangle, however, contains two level 1 type $\mathfrak{U}$ and one level 1 type $\mathfrak{L}$ triangle in this realization. Here the number of type $\mathfrak{U}$ and $\mathfrak{L}$ triangles both follows a binomial distribution with parameters 2 and p.

Wherever we are in the process in a number 0 column the *expected number* of level $n + 1$ type $\mathfrak{U}$ triangles in one retained level n type $\mathfrak{U}$ triangle is p. The expected number of level $n + 1$ type $\mathfrak{L}$ triangles and of level $n + 1$ type $\mathfrak{U}$ triangles in a retained level n type $\mathfrak{L}$ triangle are both $2p$. We use the terminology that a level $n + 1$ triangle is the child of the level n triangle containing it (and the containing triangle is the parent). We summarize the above information in an *expectation matrix*, indexed with zero to denote that we are in the number 0 column-pair,

$$\mathbf{M}_0 = p \begin{bmatrix} 1 & 0 \\ 2 & 2 \end{bmatrix}$$

in which the first row corresponds to the parent type $\mathfrak{U}$, while the second to the parent type $\mathfrak{L}$ triangle, and the columns of the matrix correspond to the number of children of type $\mathfrak{U}$ and $\mathfrak{L}$. For example the first element of the second row gives the expected number of child type $\mathfrak{U}$ triangles given birth by a type $\mathfrak{L}$ triangle.

The expectation matrices corresponding to the other two columns are

$$\mathbf{M}_1 = p \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, \quad \mathbf{M}_2 = p \begin{bmatrix} 2 & 2 \\ 0 & 1 \end{bmatrix}.$$

When the survival probability of the MBPRE, informally described above (with two types: $\mathfrak{U}$ and $\mathfrak{L}$, and environments described by infinite sequences of the three column-pairs numbered by $0, 1, 2$), is positive then the Lebesgue measure of the 45-degree projection of the Sierpiński carpet is also positive (almost surely conditioned on the set being non-empty). The assumptions of Theorem 1.1 are satisfied: the environmental process (uniform and i.i.d. over the alphabet $[2]$) is stationary and ergodic, the expectation matrices are allowable, $\mathbf{M}_1$ is strictly positive, and assumption (d) is trivial. It is also clear, that for all $n \in \mathbb{N}$, we have $\mathbf{M}_0^n(1,2) = 0$. Therefore, condition (1.2) does not hold and the results preceding this paper are not applicable. The Lyapunov exponent (λ) in this case is estimated by Pollicott and Vytnova (based on personal communication in 2023, and can be found in Pollicott (2023)) to be $0.953 + \log(p) < \lambda < 0.982 + \log(p)$. We present the following proposition, which summarizes the usage of these theorems to the random Sierpiński carpet, without proof. A precise exposition of the connection between the geometry and the branching process, as well as the proof of an analogous statement for a broader class of random self similar sets is given in (Orgoványi and Simon, 2024, Theorem 3.4.).

Proposition 1.3 (Application of the result to the random Sierpiński carpet). *Let $p \in [0,1]$ and let $\mathbf{M}_0, \mathbf{M}_1, \mathbf{M}_2$ be the expectation matrices defined above. Then the 45-degree projection of the random Sierpiński carpet has positive Lebesgue measure almost surely conditioned on non-extinction for a given parameter p if and only if the corresponding MBPRE in a uniform environment does not die out with positive probability if and only if the Lyapunov exponent λ, corresponding to the uniform $\{1/3, 1/3, 1/3\}^{\mathbb{N}}$ measure on $[2]^{\mathbb{N}}$ is positive.*

Hence, by the above estimation of Pollicott and Vytnova there exists a $p_* \in [0.3745\ldots, 0.386\ldots]$ such that the random Sierpiński carpet has positive Lebesgue measure almost surely conditioned on non-extinction if and only if $p > p_*$.

Note that it is well known that the Hausdorff dimension of the random Sierpiński carpet is $\log(8p)/\log(3)$ almost surely conditioned on the set being non-empty. Combining this with the above we get that there exists an interval of p-s so that the dimension of the set is already exceeds 1 (i.e. $p > 0.375$); however, the projection has 0 Lebesgue measure.

1.4. Structure of the paper

For the remainder of this section, we introduce the notation that we will use throughout this paper. Then, in the second section, we define the multitype branching process in varying and in random environments, and we state our main results precisely. The remainder of the paper is devoted to the proof of the main theorem.

Our introduction of multitype branching processes in a random environment in Sections 1.5-2.2 follows closely that of (Kersting and Vatutin, 2017, Chapter 10) with slight modifications in the notation.

1.5. Notation

For $k > 0$, define $[k] := \{0, \ldots, k-1\}$. We denote the vectors and matrices by boldface letters; in particular

$$\mathbf{0} := (0, \ldots, 0) \quad \mathbf{1} := (1, \ldots, 1) \text{ and } \quad \mathbf{e}_i := (\underbrace{0, \ldots, 0}_{i}, 1, \underbrace{0, \ldots, 0}_{N-i-1}).$$

For two N-dimensional vectors $\mathbf{u} = (u_0, \dots, u_{N-1})$ and $\mathbf{v} = (v_0, \dots, v_{N-1})$ and the $N \times N$ matrices $\mathbf{U} = (u_{i,j})_{i,j\in[N]}$ and $\mathbf{V} = (u_{i,j})_{i,j\in[N]}$ let

$$\mathbf{u} \cdot \mathbf{v} := u_0 v_0 + \cdots + u_{N-1} v_{N-1}, \quad \mathbf{u}^{\mathbf{v}} = \prod_{i=0}^{N-1} u_i^{v_i} \text{ and} \tag{1.4}$$

$$\begin{matrix} \mathbf{u} \le \mathbf{v} \\ \mathbf{U} \le \mathbf{V} \end{matrix} \quad \text{if and only if} \quad \begin{matrix} u_i \le v_i \\ u_{i,j} \le v_{i,j} \end{matrix} \quad \text{for all } i, j \in [N]. \tag{1.5}$$

We further use the strict equality version of (1.5), when all $\le$ is replaced with $<$. Given the functions $\mathbf{f} = (f^{(0)}, \dots, f^{(N-1)}), \{\mathbf{f}_i = (f_i^{(0)}, \dots, f_i^{(N-1)})\}_{i\in\mathcal{I}}$ and $\mathbf{f}, \mathbf{f}_i : \mathbb{R}^N \to \mathbb{R}^N$ and $g : \mathbb{R}^N \to \mathbb{R}$ for some $N \in \mathbb{N}$ we write

- for $K \in \mathbb{N}, K \ge 0$ let g^K be the product of g with itself K times and
- for $0 \le \mathbf{K} = (K_0, \dots, K_{N-1}) \in \mathbb{N}^N$ let $\mathbf{f}^K = (f^{(0)})^{K_0} \dots (f^{(N-1)})^{K_{N-1}}$;
- for $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n) \in \mathcal{I}^n$ we write $\mathbf{f}_{\boldsymbol{\theta}} := \mathbf{f}_{\theta_1} \circ \cdots \circ \mathbf{f}_{\theta_n}$.

The convention throughout the paper is to denote vectors and matrices by boldface letters, infinite vectors are also overlined to be distinguishable from finite vectors. Some further notation we use throughout the paper, and the place of the first occurrence:

symbol	explanation	link
$\mathscr{A}$	$N \times N$ non-negative, allowable matrices	Sec. 3
λ	Lyapunov exponent corresponding to the expectation matrices and the ergodic measure ν	Def. 3.2
λ_*	column-sum exponent corresponding to the expectation matrices and the ergodic measure ν	Def. 3.3
α	the uniform allowability constant	(4.1)
$\mathbf{M}_\theta$	expectation matrix in the environment θ	(2.7)
$\mathbf{A}_\theta, \rho$	$\mathbf{A}_\theta$ is the ρ-decreased expectation matrix	(5.1)
$\mathfrak{W}$, $\mathfrak{W}^{\theta,k}$	$\{(k,i,\theta) \in [N]^2 \times \mathcal{I} : \mathbf{M}_\theta(k,i) > 0\}$ and $\{i \in [N] : (k,i,\theta) \in \mathfrak{W}\}$	
$t_\theta^{(k)}(\mathbf{s})$	$1 - \mathbf{r}_k(\mathbf{M}_\theta) \cdot (\mathbf{1} - \mathbf{s})$	(5.6)
$\mathbf{r}_k(\mathbf{B}), \mathbf{c}_k(\mathbf{B})$	the k-th row and column vector of a matrix $\mathbf{B}$, respectively	
$g_\theta^{(k)}(\mathbf{s})$, $\mathbf{g}_\theta(\mathbf{s})$	$1 - \mathbf{r}_k(\mathbf{A}_\theta) \cdot (\mathbf{1} - \mathbf{s})$ and $(g_\theta^{(0)}(\mathbf{s}), \dots, g_\theta^{(N-1)}(\mathbf{s}))=\mathbf{1} - \mathbf{A}_\theta(\mathbf{1} - \mathbf{s})$ resp.	(5.4)
B_δ	$\{\mathbf{s} \in [0,1]^N : \Vert\mathbf{1} - \mathbf{s}\Vert_\infty \le \delta\}$	(5.7)
$\psi(\mathbf{s})$		(5.9)
$R(t)$, $R^C(t)$	$\{\mathbf{s} \in [0,1]^N : \Vert\mathbf{s}\Vert < t\}$ and $\{\mathbf{s} \in [0,1]^N : \Vert\mathbf{s}\Vert \ge t\}$ resp.	(5.11)
$\varphi_w(t)$	$N - Nw + wt$	(5.15)
$q^{(k)}(\overline{\boldsymbol{\theta}})$, $\mathbf{q}(\overline{\boldsymbol{\theta}})$	the probability that the process starting with one individual of type-k becomes extinct, and the vector of these, $(q^{(k)}(\overline{\boldsymbol{\theta}}))_{k\in[N]}$ resp.	(2.9)

2. Introduction to multitype branching processes

We now introduce some preliminary notation regarding the N-type branching process for $N \geq 2$. Denote by $\mathcal{P}(\mathbb{N}_0^N)$ the probability distributions on $\mathbb{N}_0^N$. Furthermore, we identify the distributions in $\mathcal{P}(\mathbb{N}_0^N)$ with their probability generating functions (pgfs) f. That is, the pgf f corresponding to the distribution $f \in \mathcal{P}\left(\mathbb{N}_0^N\right)$ is denoted by:

$$f(\mathbf{s}) := \sum_{\mathbf{z} \in \mathbb{N}_0^N} f[\mathbf{z}]\mathbf{s}^{\mathbf{z}}, \text{ for } \mathbf{s} \in [0,1]^N, \tag{2.1}$$

where $f[\mathbf{z}]$ is the probability of $\{\mathbf{z}\}$, for $\mathbf{z} = (z_0, \dots, z_{N-1}) \in \mathbb{N}_0^N$. Throughout the paper we will only use this notation to pgfs, we will not consider the density or the cumulative distribution function.

We consider the N-dimensional vectors of probability measures which are identified by the vectors of their pgf

$$\mathbf{f} = \left(f^{(0)}, \dots, f^{(N-1)}\right) \in \mathcal{P}(\mathbb{N}_0^N) \times \cdots \times \mathcal{P}(\mathbb{N}_0^N) = \mathcal{P}^N(\mathbb{N}_0^N). \tag{2.2}$$

2.1. Multitype branching processes in varying environments

On the underlying probability space $(\Omega, \mathcal{F}, \mathbf{P})$ we define the N-type branching process in varying environment for an $N \geq 2$.

Definition 2.1. A sequence $\overline{\mathbf{v}} = (\mathbf{f}_1, \mathbf{f}_2, \dots)$, of N-dimensional probability measures $\mathbf{f}_n = (f_n^{(0)}, \dots, f_n^{(N-1)})$ on $\mathbb{N}_0^N$ is called a *varying environment*.

2.1.1. Alternative description of the process

Assume that we are given a varying environment $\overline{\mathbf{v}} = (\mathbf{f}_n)_{n \geq 1}$ of N-dimensional probability measures. For each $i \in [N]$ and $n \geq 1$ there is an offspring vector random variable

$$\mathbf{Y}_n^{(i)} = \left(Y_n^{(i)}(0), \dots, Y_n^{(i)}(N-1)\right)$$

such that

$$\mathbf{P}\left(\mathbf{Y}_n^{(i)} = \mathbf{y}\right) = f_n^{(i)}[\mathbf{y}], \text{ for every } \mathbf{y} \in \mathbb{N}_0^N.$$

Now we define $\{\mathbf{Z}_n\}_{n \geq 0}$ the *N-type branching process in the varying environment* $\overline{\mathbf{v}}$. We start at level 0, where the number of different types of individuals is deterministic and is given by $\mathbf{z}_0 := (z_0^{(0)}, \dots, z_0^{(N-1)})$, that is $\mathbf{Z}_0 := \mathbf{z}_0$. The n-th element of the process is the vector random variable $\mathbf{Z}_n = (Z_n^{(0)}, \dots, Z_n^{(N-1)})$, where $Z_n^{(i)}$ is the number of level n individuals of type i. Given $\mathbf{Z}_0, \dots, \mathbf{Z}_{n-1}$ we define $\mathbf{Z}_n$ as follows.

We consider the sequence of vector random variables

$$\left\{\mathbf{Y}_{j,n}^{(i)} = \left(Y_{j,n}^{(i)}(0), \dots Y_{j,n}^{(i)}(N-1)\right) : i \in [N],\ j \in \{1, \dots, \mathbf{Z}_{n-1}^{(i)}\}\right\},$$

(a) $\left\{\mathbf{Y}_{j,n}^{(i)}\right\}_{i,j}$ are independent of each other and $\mathbf{Z}_{n-1}$, and

(b) $\mathbf{Y}_{j,n}^{(i)} \stackrel{d}{=} \mathbf{Y}_n^{(i)}$.

Informally the meaning of the ℓ-th component, $Y_{j,n}^{(i)}(\ell)$ of $\mathbf{Y}_{j,n}^{(i)}$ is the number of type ℓ children of the j-th individual of type i in the $n-1$-th generation. Then the vector of the numbers of various type level-n individuals is

$$\mathbf{Z}_n = \left(Z_n^{(0)}, \ldots, Z_n^{(N-1)}\right) := \sum_{i=0}^{N-1} \sum_{j=1}^{Z_{n-1}^{(i)}} \mathbf{Y}_{j,n}^{(i)}, \tag{2.3}$$

where recall that $Z_n^{(i)}$ stands for the number of type i individuals in the n-th generation.

Definition 2.2. Formally, the stochastic process $\mathcal{Z} = \{\mathbf{Z}_n\}_{n\geq 0}$ is called N*-type branching process with varying environment* $\overline{\mathbf{v}}$ if for any $\mathbf{z} \in \mathbb{N}_0^N$

$$\mathbf{P}_{\mathbf{z}_0,\overline{\mathbf{v}}}\left(\mathbf{Z}_n = \mathbf{z} \middle| \mathbf{Z}_1, \ldots, \mathbf{Z}_{n-1}\right) = \left(\mathbf{f}_n^{\mathbf{Z}_{n-1}}\right)[\mathbf{z}],$$

where we write $\mathbf{P}_{\mathbf{z}_0,\overline{\mathbf{v}}}$ instead of $\mathbf{P}$ to emphasize the initial population size $\mathbf{z}_0$ and the fixed (deterministic) environment $\overline{\mathbf{v}}$.

The inductive description above and the description using pgfs are connected by Fact B.1 in Appendix B.

2.2. Multitype branching processes in a random environment

In this section we describe the generalization of the above process, by instead of considering a fixed deterministic environment we consider random environments. Conditioning on the environment, the process behaves as a multitype branching process in a varying environment. We endow $\mathcal{P}^N(\mathbb{N}_0^N)$ with the metric of total variation (see (Kersting and Vatutin, 2017, p. 260)) and with the respective Borel σ-algebra. Hence, we can speak about random N-dimensional probability measures. These are random variables taking values in $\mathcal{P}^N(\mathbb{N}_0^N)$ of the form

$$\mathbf{F} = \left(F^{(0)}, \ldots, F^{(N-1)}\right),$$

where the components are the pgfs

$$F^{(i)}(\mathbf{s}) := \sum_{\mathbf{z}\in\mathbb{N}_0^N} F^{(i)}[\mathbf{z}]\mathbf{s}^{\mathbf{z}},\ i \in [N]. \tag{2.4}$$

Definition 2.3. A *random environment* is a sequence $\overline{\mathcal{V}} = (\mathbf{F}_n)_{n\geq 1}$ of N-dimensional random probability measures taking values in $\mathcal{P}^N(\mathbb{N}_0^N)$.

Now we introduce multitype branching processes $\mathcal{Z}$ in random environments as follows: first we consider a random environment $\overline{\mathcal{V}} = (\mathbf{F}_n)_{n\geq 1}$. For a realization $\overline{\mathbf{v}} = (\mathbf{f}_n)_{n\geq 1} = \left((f_n^{(0)}, \ldots, f_n^{(N-1)})\right)_{n\geq 1}$ of $\overline{\mathcal{V}}$, $\mathcal{Z}$ evolves as an N-dimensional temporally non-homogeneous branching process, where the offspring distribution of a type i individual on the $n-1$-th generation is governed by $f_n^{(i)}(\cdot)$.

Definition 2.4 (MBPRE). We say that the process $\mathcal{Z} = \left(\mathbf{Z}_n = (Z_n^{(0)}, \ldots, Z_n^{(N-1)})\right)_{n\in\mathbb{N}_0}$ taking values in $\mathbb{N}_0^N$ is a *multitype (N-type) branching process in the random environment* $\overline{\boldsymbol{\mathcal{V}}} = (\mathbf{F}_n)_{n\geq 1}$ (MBPRE) if for each realization $\overline{\mathbf{v}} = (\mathbf{f}_n)_{n\geq 1}$ of $\overline{\boldsymbol{\mathcal{V}}}$ and for each $\mathbf{z}_0, \mathbf{z}_1, \ldots, \mathbf{z}_k \in \mathbb{N}_0^N$

$$\mathbb{P}\left(\mathbf{Z}_1 = \mathbf{z}_1, \ldots, \mathbf{Z}_k = \mathbf{z}_k \,|\, \mathbf{Z}_0 = \mathbf{z}_0, \overline{\boldsymbol{\mathcal{V}}} = \overline{\mathbf{v}}\right) = \mathbf{P}_{\mathbf{z}_0,\overline{\mathbf{v}}}\left(\mathbf{Z}_1 = \mathbf{z}_1, \ldots, \mathbf{Z}_k = \mathbf{z}_k\right) \quad \text{a.s.},$$

where $\mathbf{P}_{\mathbf{z},\overline{\mathbf{v}}}$ denotes the probability measure corresponding to the N-type branching process in varying environment $\overline{\mathbf{v}}$ with initial distribution $\mathbf{Z}_0 = \mathbf{z}$. We write $\mathbb{P}(\cdot)$ and $\mathbb{E}(\cdot)$ for the probabilities and expectations in random environments.

From this it follows that for each realization $\overline{\mathbf{v}} = (\mathbf{f}_n)_{n\geq 1}$ and $\mathbf{z}, \mathbf{z}_0 \in \mathbb{N}_0^N$

$$\mathbb{P}\left(\mathbf{Z}_n = \mathbf{z} \,|\, \mathbf{Z}_0 = \mathbf{z}_0, \mathbf{Z}_1 = \mathbf{z}_1, \ldots, \mathbf{Z}_{n-1} = \mathbf{z}_{n-1}, \overline{\boldsymbol{\mathcal{V}}} = \overline{\mathbf{v}}\right) = \left(\mathbf{f}_n^{\mathbf{Z}_{n-1}}\right)[\mathbf{z}] \quad \text{a.s.} \tag{2.5}$$

from which we can conclude, that

$$\mathbb{E}\left(\mathbf{s}^{\mathbf{Z}_n} \,|\, \mathbf{Z}_0 = \mathbf{z}_0, \overline{\boldsymbol{\mathcal{V}}} = \overline{\mathbf{v}}\right) = \mathbf{f}_1(\mathbf{f}_2(\cdots(\mathbf{f}_n(\mathbf{s}))\cdots))^{\mathbf{z}_0}. \tag{2.6}$$

2.3. Our principal assumptions

From now on we restrict ourselves to the case when the environment is coming from the infinite product of a *countable* set of distributions from $\mathcal{P}^N(\mathbb{N}_0^N)$.

Namely, fix a countable set $\{\mathbf{f}_n\}_{n\in\mathcal{I}}$, $\mathbf{f}_n \in \mathcal{P}^N(\mathbb{N}_0^N)$ indexed by the set $\mathcal{I}$. This is the set of possible values of $\mathbf{F}_n$. In this way, the random environment $\overline{\boldsymbol{\mathcal{V}}}$ is a random variable which takes values in $\{\mathbf{f}_i,\, i \in \mathcal{I}\}^{\mathbb{N}}$.

It is more convenient to identify the environments with their "code" from $\mathcal{I}^{\mathbb{N}}$, and refer to the code instead. Namely, we define the map

$$\Phi : \{\mathbf{f}_n,\, n \in \mathcal{I}\}^{\mathbb{N}} \to \mathcal{I}^{\mathbb{N}} = \Sigma, \quad \Phi\left(\mathbf{f}_{\theta_1}, \mathbf{f}_{\theta_2}, \ldots\right) := (\theta_1, \theta_2, \ldots).$$

Definition 2.5. The probability space $(\Sigma, \mathcal{A}, \nu)$ with the shift map σ is defined as

- *(a)* $\Sigma = \mathcal{I}^{\mathbb{N}}$,
- *(b)* $\mathcal{A}$ is the usual σ-algebra on Σ,
- *(c)* $\nu := \Phi_* \mathfrak{m}$, where $\mathfrak{m}$ is the distribution of the environmental variable, $\overline{\boldsymbol{\mathcal{V}}}$. That is $\nu(H) = \mathfrak{m}(\Phi^{-1}H)$, for any Borel set $H \subset \Sigma$;
- *(d)* and for $\overline{\boldsymbol{\theta}} = (\theta_1, \theta_2, \ldots) \in \Sigma$, $\sigma(\overline{\boldsymbol{\theta}}) := (\theta_2, \theta_3, \ldots)$.

We will refer to an *environment* as $\overline{\boldsymbol{\theta}} = (\theta_1, \theta_2, \ldots) \in \Sigma$ instead of $(\mathbf{f}_{\theta_1}, \mathbf{f}_{\theta_2}, \ldots)$, and we write $\mathbf{Z}_n(\overline{\boldsymbol{\theta}})$ for $\mathbf{Z}_n$ in the environment $\overline{\boldsymbol{\theta}}$.

In our most important application we usually consider the following special case.

Example 2.6. When $\Sigma = [L]^{\mathbb{N}}$ for some $2 \leq L \in \mathbb{N}$ and the environmental sequence $\overline{\boldsymbol{\mathcal{V}}}$ is i.i.d. then we have a probability vector $\mathbf{p} = (p_1, \ldots, p_{L-1})$ such that $\mathbb{P}(\mathbf{F}_n = \mathbf{f}_k) = p_k$ for all n and $k \in [L]$. In this case the infinite product measure $\nu = (p_0, \ldots, p_{L-1})^{\mathbb{N}}$ on Σ corresponds to the distribution of $\overline{\boldsymbol{\mathcal{V}}}$ via the identification Φ.

From now on we always assume the following:

Principal Assumption I. The system $(\Sigma, \mathcal{A}, \sigma, \nu)$ defined in Definition 2.5 is ergodic.

2.4. Expectation matrices and survival probabilities

2.4.1. *Expectation matrices*

We define the $N \times N$ *expectation matrix* corresponding to a fixed $\theta \in \mathcal{I}$ as

$$\mathbf{M}_\theta(i,k) = \frac{\partial f_\theta^{(i)}}{\partial s_k}(\mathbf{1}). \tag{2.7}$$

In case the environment $\overline{\boldsymbol{\theta}} = (\theta_1, \dots, \theta_n, \dots)$ is fixed using the notations of Section 2.1.1,

$$\mathbf{M}_{\theta_n}(i,k) = \mathbb{E}(Y_n^{(i)}(k)).$$

From this, using induction it follows that for any $\overline{\boldsymbol{\theta}} = (\theta_1, \theta_2, \dots) \in \Sigma$ and $n \in \mathbb{N}$

$$\mathbb{E}\left[\mathbf{Z}_n | \overline{\boldsymbol{\mathcal{V}}} = \overline{\boldsymbol{\theta}}, \mathbf{Z}_0 = \mathbf{z}_0\right] = \mathbf{z}_0^T \mathbf{M}_{\theta_1} \cdots \mathbf{M}_{\theta_n}.$$

2.4.2. *Survival probabilities*

Fix $\overline{\boldsymbol{\theta}} = (\theta_1, \theta_2, \dots) \in \Sigma$. For every $\ell \in \mathbb{N}$ we consider the pgf vector

$$\mathbf{f}_{\theta_\ell}(\mathbf{s}) = \left(f_{\theta_\ell}^{(0)}(\mathbf{s}), \dots, f_{\theta_\ell}^{(N-1)}(\mathbf{s})\right), \text{ where } f_{\theta_\ell}^{(i)}(\mathbf{s}) = \sum_{\mathbf{j} \in \mathbb{N}_0^N} f_{\theta_\ell}^{(i)}[\mathbf{j}]\mathbf{s}^{\mathbf{j}}.$$

Recall from the introduction that, for a $\mathbf{j} = (j_0, \dots, j_{N-1}) \in \mathbb{N}_0^N$, $f_{\theta_\ell}^{(i)}[\mathbf{j}]$ is the probability that a level $\ell - 1$ individual of type i gives birth to j_1 individuals of type 1, ..., and j_N individuals of type N simultaneously.

Applying (2.6) to $\mathbf{z}_0 = \mathbf{e}_i$ gives that

$$\mathbb{E}\left[\mathbf{s}^{\mathbf{Z}_n} \mid \overline{\boldsymbol{\mathcal{V}}} = \overline{\boldsymbol{\theta}}, \mathbf{Z}_0 = \mathbf{e}_i\right] = f_{\overline{\boldsymbol{\theta}}|_n}^{(i)}(\mathbf{s}),$$

where for $\overline{\boldsymbol{\theta}} = (\theta_1, \theta_2, \dots)$, $\overline{\boldsymbol{\theta}}|_n = (\theta_1, \dots, \theta_n)$: the vector containing the first n letters of $\overline{\boldsymbol{\theta}}$. This implies that

$$\mathbb{P}\left(\mathbf{Z}_n = \mathbf{0} \mid \overline{\boldsymbol{\mathcal{V}}} = \overline{\boldsymbol{\theta}}, \mathbf{Z}_0 = \mathbf{e}_i\right) = f_{\overline{\boldsymbol{\theta}}|_n}^{(i)}(\mathbf{0}) = f_{\theta_1}^{(i)}(\mathbf{f}_{\theta_2} \circ \cdots \circ \mathbf{f}_{\theta_n}(\mathbf{0})). \tag{2.8}$$

Let $q^{(k)}(\overline{\boldsymbol{\theta}})$ denote the probability that the process starting with one individual of type-k becomes extinct, and

$$\mathbf{q}(\overline{\boldsymbol{\theta}}) = (q^{(0)}(\overline{\boldsymbol{\theta}}), \dots, q^{(N-1)}(\overline{\boldsymbol{\theta}})). \tag{2.9}$$

Further, we denote the level-n extinction probability by

$$q_n^{(k)}(\overline{\boldsymbol{\theta}}) = \mathbb{P}(\mathbf{Z}_n(\overline{\boldsymbol{\theta}}) = \mathbf{0} \mid \mathbf{Z}_0 = \mathbf{e}_k), \text{ and } \mathbf{q}_n(\overline{\boldsymbol{\theta}}) = (q_n^{(0)}(\overline{\boldsymbol{\theta}}), \dots, q_n^{(N-1)}(\overline{\boldsymbol{\theta}})). \tag{2.10}$$

Using (2.8) we obtain

$$\mathbf{q}_n(\overline{\boldsymbol{\theta}}) = \mathbf{f}_{\overline{\boldsymbol{\theta}}|_n}(\mathbf{0}). \tag{2.11}$$

Hence,

$$\mathbf{q}(\overline{\boldsymbol{\theta}}) = \lim_{n\to\infty} \mathbf{q}_n(\overline{\boldsymbol{\theta}}) = \lim_{n\to\infty} \mathbf{f}_{\overline{\boldsymbol{\theta}}|_n}(\mathbf{0}). \tag{2.12}$$

3. Lyapunov and column-sum exponent

For this and the following subsections we fix $N \geq 2$. We call a matrix allowable if it has at least one strictly positive element in every row and in every column. Let $\mathscr{A}$ be the set of $N \times N$ allowable matrices having only non-negative elements. For a $\mathbf{B} \in \mathscr{A}$ we introduce the minimum and maximum column sums:

$$(\mathbf{B})_* = \min_{j\in[N]} \sum_{i\in[N]} \mathbf{B}_{i,j}, \quad \|\mathbf{B}\|_1 := \max_{j\in[N]} \sum_{i\in[N]} \mathbf{B}_{i,j}. \tag{3.1}$$

Finally, we define the norm we will mainly use throughout the paper

$$\|\mathbf{B}\| := \sum_{i\in[N]} \sum_{j\in[N]} \mathbf{B}_{i,j}. \tag{3.2}$$

For $\{\mathbf{B}_i\}_{i\in\mathcal{I}} \subset \mathscr{A}$ and $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n) \in \mathcal{I}^n$ we denote

$$\mathbf{B}_{\boldsymbol{\theta}} := \mathbf{B}_{\theta_1} \cdots \mathbf{B}_{\theta_n}.$$

Definition 3.1 (Good set of matrices). Let $\mathcal{B} = \{\mathbf{B}_i\}_{i\in\mathcal{I}} \subset \mathscr{A}$ and let ν be an ergodic invariant measure on $(\Sigma, \mathcal{A}, \sigma)$, where $\Sigma := \mathcal{I}^{\mathbb{N}}$. We say that $\mathcal{B}$ *is good* (with respect to ν) if

1. $m_1 := \int |\log \|\mathbf{B}_{\theta_1}\|_1|\, d\nu(\overline{\boldsymbol{\theta}}) + \int |\log(\mathbf{B}_{\theta_1})_*|\, d\nu(\overline{\boldsymbol{\theta}}) < \infty$;
2. There exists an $n \in \mathbb{N}$ and $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n) \in \mathcal{I}^n$ such that $\nu(\{\overline{\boldsymbol{\theta}} \in \Sigma : \overline{\boldsymbol{\theta}}|_n = \boldsymbol{\theta}\}) > 0$ and every element of $\mathbf{B}_{\boldsymbol{\theta}}$ is strictly positive.

Note that it follows from the fact that $(\mathbf{B})_* > 0$ if $\mathbf{B} \in \mathscr{A}$, that assumption (1) of Definition 3.1 always holds when $|\mathcal{I}| < \infty$.

Now we define the Lyapunov exponent of a random matrix product.

Definition 3.2 (Lyapunov exponent). We are given an ergodic measure ν on (Σ, σ) and a

$$\mathcal{B} = \{\mathbf{B}_i\}_{i\in\mathcal{I}} \subset \mathscr{A},$$

which is good with respect to ν. The *Lyapunov-exponent* corresponding to $\mathcal{B}$ and the ergodic measure ν is

$$\lambda := \lambda(\nu, \mathcal{B}) = \lim_{n\to\infty} \frac{1}{n} \log \|\mathbf{B}_{\overline{\boldsymbol{\theta}}|_n}\| \quad \text{for } \nu\text{-almost every } \overline{\boldsymbol{\theta}} \in \Sigma.$$

The explanation about the existence of λ as defined above (independent of $\overline{\boldsymbol{\theta}}$) can be found in Appendix C.

Using the super-multiplicativity of $(\cdot)_*$ for non-negative allowable matrices (namely, if $\mathbf{B}_1, \dots, \mathbf{B}_n \in \mathscr{A}$ then $(\mathbf{B}_1 \cdots \mathbf{B}_n)_* \geq (\mathbf{B}_1)_* \cdots (\mathbf{B}_n)_*$) it follows that we can define the analogue of the Lyapunov exponent for the minimal column sum, which we call the *column-sum exponent*. For more details see again Appendix C.

Definition 3.3. The *column-sum exponent* corresponding to a good set of matrices, $\mathcal{B} = \{\mathbf{B}_i\}_{i\in\mathcal{I}}$ and an ergodic measure ν is

$$\lambda_* := \lambda_*(\nu, \mathcal{B}) = \lim_{n\to\infty} \frac{1}{n} \log\left[(\mathbf{B}_{\overline{\boldsymbol{\theta}}|_n})_*\right] \quad \text{for } \nu\text{-almost every } \overline{\boldsymbol{\theta}} \in \Sigma. \tag{3.3}$$

Below we give conditions under which $\lambda = \lambda_*$.

Lemma 3.4. *Let ν be an ergodic measure on (Σ, σ). If $\mathcal{B} = \{\mathbf{B}_i\}_{i\in\mathcal{I}} \subset \mathscr{A}$ is good then*

$$\lambda(\nu, \mathcal{B}) = \lambda_*(\nu, \mathcal{B}). \tag{3.4}$$

Proof. The assertion follows from (Hennion, 1997, Theorem 2). For more details see Appendix A. $\square$

4. The main theorem: extinction probability for MBPRE

Before we state our theorem we define a condition.

Definition 4.1. We say that the MBPRE (see Definition 2.4) with N types and type space $\mathcal{I}$ is *uniformly allowable* if there exists an $\alpha > 0$ such that

$$\inf\left\{\sum_{\substack{w_i\neq 0\\ \mathbf{w}\in\mathbb{N}_0^N}} f_\theta^{(k)}[\mathbf{w}];\ \theta\in\mathcal{I}, \mathbf{M}_\theta(k,i) > 0, i,k\in[N]\right\} > \alpha. \tag{4.1}$$

Remark 4.2. Even though we call this property uniform allowability, this property is stronger than an actual uniform allowability condition would be. Namely, for each $\theta \in \mathcal{I}$; $k, i \in [N]$ such that $\mathbf{M}_\theta(k,i) > 0$ it holds that

$$\mathbf{M}_\theta(k,i) = \frac{\partial f_\theta^{(k)}}{\partial s_i}(\mathbf{1}) = \sum_{\mathbf{w}\in\mathbb{N}_0^N} f_\theta^{(k)}[\mathbf{w}] w_i \mathbf{1}^{\mathbf{w}} = \sum_{\substack{w_i\neq 0\\ \mathbf{w}\in\mathbb{N}_0^N}} f_\theta^{(k)}[\mathbf{w}] w_i \mathbf{1}^{\mathbf{w}} > \alpha. \tag{4.2}$$

Remark 4.3.
1. It is easy to see that if $\mathcal{I}$ is finite and the corresponding expectation matrices are good (in particular allowable), then the MBPRE is uniformly allowable.
2. An example with infinite alphabet is as follows: Choose the alphabet be $\mathcal{I} = \mathbb{N}$ and let us have two types, for simplicity call these type A and B.
   - For $\theta = 1$ the offspring distribution for A and B is the same: both give birth to exactly one type A individual with probability $1/2$ and to exactly one type B individual with probability $1/2$. The pgf is then $\mathbf{f}_\theta(\mathbf{s}) = (f_\theta^{(1)}(\mathbf{s}), f_\theta^{(2)}(\mathbf{s}))$, where

$$f_\theta^{(1)}(\mathbf{s}) = f_\theta^{(2)}(\mathbf{s}) = \frac{1}{2}s_1 + \frac{1}{2}s_2, \tag{4.3}$$

   for $\mathbf{s} = (s_1, s_2) \in [0,1]^2$.

- For $\theta = N \in \mathbb{N}$ A can only give birth to type B individuals and B can only give birth to type A individuals, both are according to the same rules, written down for a type A individual:
  - will have number 1 type B offspring with probability $1/2$;
  - will have number N offspring of type B with probability $1/N^2$;
  - will have number N offspring of type B with probability $1/2 - 1/N^2$;

  The pgf is then $\mathbf{f}_\theta(\mathbf{s}) = (f_\theta^{(1)}(\mathbf{s}), f_\theta^{(2)}(\mathbf{s}))$, where

$$f_M^{(1)}(\mathbf{s}) = \frac{1}{2}s_2 + \frac{1}{N^2}s_2^N + \frac{1}{2} - \frac{1}{N^2} \tag{4.4}$$

$$f_M^{(2)}(\mathbf{s}) = \frac{1}{2}s_1 + \frac{1}{N^2}s_1^N + \frac{1}{2} - \frac{1}{N^2}. \tag{4.5}$$

Hence, we have

$$\mathbf{M}_1 = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}, \quad \mathbf{M}_N = \begin{bmatrix} 0 & 1/2 + 1/N \\ 1/2 + 1/N & 0 \end{bmatrix}.$$

Given that we choose a nice ergodic measure on the space of environments this process satisfies all the requirements of the following theorem, in particular the uniform allowability condition, with $\alpha = 1/2$.

The main theorem of the paper is as follows (for a special case see Theorem 1.1 and an application of that can be found in Section 1.1).

Theorem 4.4. *Let ν be an ergodic measure on (Σ, σ). Consider the N-type MBPRE $\mathcal{Z} = \{\mathbf{Z}_n\}_{n\in\mathbb{N}}$ as it was defined in Definition 2.4. We assume that*

(a) $\mathcal{M} = \{\mathbf{M}_i\}_{i\in\mathcal{I}}$ (defined in (2.7)) is good (see Definition 3.1) with respect to ν.
(b) $\mathcal{Z}$ is uniformly allowable ((4.1)).
(c) For every $\theta \in \mathcal{I}$ and $i, j \in [N]$, $f_\theta^{(k)}$ is twice differentiable in $\mathbf{1}$ and there exists an $M < \infty$ such that for every $\theta \in \mathcal{I}$ and $i, j \in [N]$

$$\frac{\partial^2 f_\theta^{(k)}}{\partial s_j \partial s_i}(\mathbf{1}) < M. \tag{4.6}$$

Then we have that

1. *if $\lambda(\nu, \mathcal{M}) > 0$, then*

$$\mathbf{q}(\overline{\boldsymbol{\theta}}) = (q^{(0)}(\overline{\boldsymbol{\theta}}), \dots, q^{(N-1)}(\overline{\boldsymbol{\theta}})) \neq \mathbf{1} \text{ for } \nu\text{-almost every } \overline{\boldsymbol{\theta}} \in \Sigma.$$

2. *$\mathbf{q}(\overline{\boldsymbol{\theta}}) \neq \mathbf{1}$ for ν-almost every $\overline{\boldsymbol{\theta}} \in \Sigma$ implies that $\mathbf{q}(\overline{\boldsymbol{\theta}}) < \mathbf{1}$ for ν-almost every $\overline{\boldsymbol{\theta}} \in \Sigma$.*

Remark 4.5. From the twice differentiability assumption *(c)* above, it follows that the function is twice differentiable on $[0, 1]^N$. This is because the pgf is an infinite sum, and the first and second derivative of the sum converges uniformly.

The following corollary is an immediate consequence of the theorem.

Corollary 4.6. *Under the conditions of Theorem 4.4 if $\lambda(\nu,\mathcal{M}) > 0$, then $\mathbf{q}(\overline{\boldsymbol{\theta}}) < \mathbf{1}$ for ν-almost every $\overline{\boldsymbol{\theta}} \in \Sigma$.*

Definition 4.7. We say that an MBPRE $\{\mathbf{Z}_n\}_{n=0}^{\infty}$ is *strongly regular* if there exists an n such that $\mathbb{P}(\min_{i\in[N]} \mathbb{P}(\|\mathbf{Z}_n\| > 1 \,|\, \mathbf{Z}_0 = \mathbf{e}_i, \overline{\boldsymbol{\mathcal{V}}} = \overline{\boldsymbol{\theta}}) > 0) > 0$.

Lemma 4.8. *Assume that the pgfs satisfies that for a $\theta \in \mathcal{I}$ such that $\nu(\{\overline{\boldsymbol{\theta}} \in \Sigma : \overline{\boldsymbol{\theta}}_1 = \theta\}) > 0$, we have that for all $i \in [N]$*

$$f_{\theta}^{(i)}[\mathbf{0}] + \sum_{j\in[N]} f_{\theta}^{(i)}[\mathbf{e}_j] < 1, \tag{4.7}$$

then the MBPRE $\{\mathbf{Z}_n\}_{n=0}^{\infty}$ is strongly regular.

Proof of Lemma 4.8. The assumption of strong regularity is satisfied for $n = 1$. Namely, recall that $f_{\theta_1}^{(i)}[\mathbf{0}] + \sum_{j\in[N]} f_{\theta_1}^{(i)}[\mathbf{e}_j]$ is the probability that the process starting with one individual of type i conditioned on the environments first letter being θ_1 has at most one individual at level 1. From the assumptions of this lemma it follows that for almost every $\overline{\boldsymbol{\theta}} = (\theta_1, \theta_2, \dots)$

$$\min_{i\in[N]} \mathbb{P}(\|\mathbf{Z}_1\| > 1 \,|\, \mathbf{Z}_0 = \mathbf{e}_i, \overline{\boldsymbol{\mathcal{V}}} = \overline{\boldsymbol{\theta}}) = 1 - \max_{i\in[N]} \left(f_{\theta_1}^{(i)}[\mathbf{0}] + \sum_{j\in[N]} f_{\theta_1}^{(i)}[\mathbf{e}_j] \right) > 0, \tag{4.8}$$

from which strong regularity follows. $\square$

That is $\{\mathbf{Z}_n\}_{n=0}^{\infty}$ is <u>not</u> strongly regular if and only if for every n, for ν-almost every $\overline{\boldsymbol{\theta}}$ there exists $i \in [N]$ such that

$$\mathbb{P}(\|\mathbf{Z}_n\| \leq 1 \,|\, \mathbf{Z}_0 = \mathbf{e}_i, \overline{\boldsymbol{\mathcal{V}}} = \overline{\boldsymbol{\theta}}) = 1.$$

Corollary 4.9. *Under the conditions of Theorem 4.4:*

1. *If $\lambda(\nu,\mathcal{M}) < 0$, then $\mathbf{q}(\overline{\boldsymbol{\theta}}) = \mathbf{1}$ for ν-almost every $\overline{\boldsymbol{\theta}} \in \Sigma$.*
2. *If $\lambda(\nu,\mathcal{M}) = 0$, then either $\mathcal{Z}$ is not strongly regular, or $\mathbf{q}(\overline{\boldsymbol{\theta}}) = \mathbf{1}$ for ν-almost every $\overline{\boldsymbol{\theta}} \in \Sigma$.*

Proof. Tanny proved (see (Tanny, 1981, Theorem 9.6)) that the assertions of this corollary hold if a certain condition (called Condition Q in Tanny (1981)) is satisfied. The fact that this Condition Q holds in our case follows from part (2) of Theorem 4.4. $\square$

Corollary 4.10. *If $\lambda(\nu,\mathcal{M}) > 0$ then*

$$\lim_{n\to\infty} \frac{1}{n} \log \|\mathbf{Z}_n\| = \lambda(\nu,\mathcal{M}) \tag{4.9}$$

almost surely conditioned on $\{\|\mathbf{Z}_n\| \not\to 0$ as $n \to \infty\}$.

Proof of Corollary 4.10. The assertion follows from a combination of Theorem 4.4, a result of Hennion and (Tanny, 1981, Theorem 9.6). Namely, Corollary 4.10 is essentially the conclusion of (Tanny, 1981, Theorem 9.6, (3)). We only need to verify that both of the assumptions of this theorem are satisfied. One of them is called Condition Q. This holds because this is exactly the assertion of Theorem 4.4 part (2). The other assumption is the stability of the MBPRE (see (Tanny, 1981, Definition 9.5)). This holds, because of a result of Hennion, see Remark A.1 in particular (3.2). $\square$

5. Proof of Theorem 4.4

Throughout this section we always use the notation of Theorem 4.4.

5.1. Preparation for the proof of Theorem 4.4 part (1)

Let

$$\lambda := \lim_{n\to\infty} \frac{1}{n} \log \|\mathbf{M}_{\overline{\boldsymbol{\theta}}|_n}\|, \text{ for } \nu \text{ almost every } \overline{\boldsymbol{\theta}} \in \mathcal{I}^{\mathbb{N}}.$$

Assume (the assumption of Theorem 4.4 part (1)) $\lambda > 0$. Then we can choose

$$0 < \rho < 1 \quad \text{ such that } \quad 1 < \rho e^{\lambda}. \tag{5.1}$$

Define the $N \times N$ matrices $\mathbf{A}_\theta$ for $\theta \in \mathcal{I}$ as

$$\mathbf{A}_\theta = \rho \mathbf{M}_\theta. \tag{5.2}$$

Lemma 5.1. *For ν almost every $\overline{\boldsymbol{\theta}} \in \Sigma$*

$$\lim_{n\to\infty} \frac{1}{n} \log \left((\mathbf{A}_{\overline{\boldsymbol{\theta}}|_n})_* \right) = \log(\rho) + \lambda > 0.$$

Proof. Let

$$H := \left\{ \overline{\boldsymbol{\theta}} \in \Sigma : \lim_{n\to\infty} \frac{1}{n} \log(\mathbf{M}_{\overline{\boldsymbol{\theta}}|_n})_* \text{ exists and equals to } \lambda \right\} \subset \Sigma. \tag{5.3}$$

It follows from Lemma 3.4 that $\nu(H) = 1$. Fix $\overline{\boldsymbol{\theta}} = (\theta_1, \theta_2, \dots) \in H$. Then from $(\mathbf{A}_{\overline{\boldsymbol{\theta}}|_n})_* = \rho^n (\mathbf{M}_{\overline{\boldsymbol{\theta}}|_n})_*$ and the definition of ρ the assertion follows. $\square$

Define the set

$$\mathfrak{W} := \left\{ (k, i, \theta) \in [N]^2 \times \mathcal{I} : \mathbf{M}_\theta(k, i) > 0 \right\} = \left\{ (k, i, \theta) \in [N]^2 \times \mathcal{I} : \mathbf{A}_\theta(k, i) > 0 \right\}.$$

Moreover, for $k \in [N]$, $\theta \in \mathcal{I}$ we define

$$\mathfrak{W}^{\theta,k} := \{ i \in [N] : (k, i, \theta) \in \mathfrak{W} \}.$$

For a matrix $\mathbf{B}$ let $\mathbf{r}_k(\mathbf{B})$ and $\mathbf{c}_k(\mathbf{B})$ denote the k-th row and column vector of $\mathbf{B}$ respectively. For $\theta \in \mathcal{I}$, $\mathbf{s} \in [0, 1]^N$ let

$$g_\theta^{(k)}(\mathbf{s}) = 1 - \mathbf{r}_k(\mathbf{A}_\theta) \cdot (\mathbf{1} - \mathbf{s}), \text{ and } \mathbf{g}_\theta(\mathbf{s}) = (g_\theta^{(0)}(\mathbf{s}), \dots, g_\theta^{(N-1)}(\mathbf{s})) = \mathbf{1} - \mathbf{A}_\theta(\mathbf{1} - \mathbf{s}). \tag{5.4}$$

This immediately implies (which we frequently use without mentioning) that for $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n) \in \mathcal{I}^n$ we have for $\mathbf{s} \in [0, 1]^N$

$$\mathbf{g}_{\boldsymbol{\theta}}(\mathbf{s}) := \mathbf{g}_{\theta_1} \circ \mathbf{g}_{\theta_2} \circ \cdots \circ \mathbf{g}_{\theta_n}(\mathbf{s}) = \mathbf{1} - \mathbf{A}_{\boldsymbol{\theta}}(\mathbf{1} - \mathbf{s}). \tag{5.5}$$

Remark 5.2 (The meaning of $\mathbf{g}_\theta$ and $\mathbf{g}_{\boldsymbol{\theta}}$). For $\theta \in \mathcal{I}$, $k \in [N]$ consider $\{(\mathbf{s}, f_\theta^{(k)}(\mathbf{s})) \in \mathbb{R}^{N+1} : \mathbf{s} \in [0,1]^N\}$ the graph of the function $f_\theta^{(k)}$. We denote the tangent plane of this graph at $\mathbf{1} \in \mathbb{R}^N$ by

$$t_\theta^{(k)}(\mathbf{s}) := f_\theta^{(k)}(\mathbf{1}) - (f_\theta^{(k)})'(\mathbf{1}) \cdot (\mathbf{1}-\mathbf{s}) = 1 - \mathbf{r}_k(\mathbf{M}_\theta) \cdot (\mathbf{1}-\mathbf{s}). \tag{5.6}$$

By Taylor's Theorem

$$f_\theta^{(k)}(\mathbf{s}) = t_\theta^{(k)}(\mathbf{s}) + \frac{1}{2}(\mathbf{1}-\mathbf{s})^T \left(f_\theta^{(k)}\right)''(\mathbf{w})(\mathbf{1}-\mathbf{s}),$$

for some $\mathbf{w} \in \{\mathbf{s} + t\,(\mathbf{1}-\mathbf{s})\,, t \in (0,1)\}$ the line segment connecting $\mathbf{s}$ and $\mathbf{1}$.

Hence, $g_\theta^{(k)}(\mathbf{s})$ the analogue of $t_\theta^{(k)}(\mathbf{s})$ using the matrices $\mathbf{A}_\theta$ instead of $\mathbf{M}_\theta$. From (5.4), it follows that an analogous description can be given for $g_{\boldsymbol{\theta}}^{(k)}(\mathbf{s})$, using $f_{\boldsymbol{\theta}}^{(k)}(\mathbf{s})$ and $\mathbf{A}_{\boldsymbol{\theta}}$.

For a visual depiction in the $N = 1$ case see Figure 3c.

Let

$$B_\delta := \{\mathbf{s} \in [0,1]^N : \|\mathbf{1}-\mathbf{s}\|_\infty \le \delta\}, \tag{5.7}$$

where for $\mathbf{z} = (z_0, \dots, z_{N-1}) \in \mathbb{R}^N$

$$\|\mathbf{z}\|_\infty = \max_{i \in [N]} |z_i|$$

Lemma 5.3. *There exists a $\delta > 0$ such that for all $\theta \in \mathcal{I}$ and $\mathbf{s} \in B_\delta$, $\mathbf{g}_\theta(\mathbf{s}) \ge \mathbf{f}_\theta(\mathbf{s})$.*

Proof of Lemma 5.3. Recall from Remark 5.2, that for $\mathbf{s} \in [0,1]^N$ $\theta \in \mathcal{I}$ and $k \in [N]$,

$$f_\theta^{(k)}(\mathbf{s}) = t_\theta^{(k)}(\mathbf{s}) + \frac{1}{2}(\mathbf{1}-\mathbf{s})^T \left(f_\theta^{(k)}\right)''(\mathbf{w})(\mathbf{1}-\mathbf{s}),$$
$$g_\theta^{(k)}(\mathbf{s}) = t_\theta^{(k)}(\mathbf{s}) + \mathbf{r}_k(\mathbf{M}_\theta - \mathbf{A}_\theta)(\mathbf{1}-\mathbf{s}),$$

where $t_\theta^{(k)}(\mathbf{s}) = 1 - \mathbf{r}_k(\mathbf{M}_\theta) \cdot (\mathbf{1}-\mathbf{s})$, as we stated in (5.6).

Hence, we only have to prove that there exists a $\delta > 0$ such that for all $\theta \in \mathcal{I}$ and $k \in [N]$ and $\mathbf{s} \in B_\delta$

$$(\mathbf{1}-\mathbf{s})^T \left(f_\theta^{(k)}\right)''(\mathbf{w})(\mathbf{1}-\mathbf{s}) \le \mathbf{r}_k(\mathbf{M}_\theta - \mathbf{A}_\theta)(\mathbf{1}-\mathbf{s}).$$

It follows from Lemma B.2 in Appendix B, that

$$(\mathbf{1}-\mathbf{s})^T \left(f_\theta^{(k)}\right)''(\mathbf{w})(\mathbf{1}-\mathbf{s}) = \sum_{i \in \mathfrak{W}^{\theta,k}} \left(\sum_{j \in \mathfrak{W}^{\theta,k}} (\mathbf{1}-\mathbf{s})_j \frac{\partial^2 f_\theta^{(k)}}{\partial s_j \partial s_i}(\mathbf{w}) \right) (\mathbf{1}-\mathbf{s})_i.$$

Clearly, $\mathbf{r}_k(\mathbf{M}_\theta - \mathbf{A}_\theta)(\mathbf{1}-\mathbf{s}) = \sum_{i \in \mathfrak{W}^{\theta,k}} \mathbf{r}_k(\mathbf{M}_\theta - \mathbf{A}_\theta)_i (\mathbf{1}-\mathbf{s})_i$. Choose any $0 < \delta < 1$ so that

$$\delta \le \frac{(1-\rho)\alpha}{2 \cdot N \cdot M}, \tag{5.8}$$

where recall that α is from the definition of uniform allowability, Definition 4.1 and M is from the last assumption of the theorem, see (*(c)*). Clearly, we can choose $\delta \in (0,1)$. Let $\mathbf{s} \in B_\delta$. With this choice for all $\theta \in \mathcal{I}, k \in [N]$ and $i, j \in \mathfrak{W}^{\theta,k}$ we have that $0 \le 1 - s_j \le \delta$. By definition, we have $\frac{\partial^2 f_\theta^{(k)}}{\partial s_j \partial s_i}(\mathbf{w}) \le \frac{\partial^2 f_\theta^{(k)}}{\partial s_j \partial s_i}(\mathbf{1})$ for all $\mathbf{w} \in [0,1]^N$, hence

$$\sum_{j \in \mathfrak{W}^{\theta,k}} (\mathbf{1} - \mathbf{s})_j \frac{\partial^2 f_\theta^{(k)}}{\partial s_j \partial s_i}(\mathbf{w}) \le \delta \cdot N \cdot \max_j \frac{\partial^2 f_\theta^{(k)}}{\partial s_j \partial s_i}(\mathbf{1}) < \mathbf{r}_k(\mathbf{M}_\theta - \mathbf{A}_\theta)_i .$$

The second inequality follows from $\frac{\partial^2 f_\theta^{(k)}}{\partial s_j \partial s_i}(\mathbf{1}) < M$, which is assumption *(c)* of Theorem 4.4. Whereas by the choice of the matrix $\mathbf{A}_\theta$ (see (5.2)), $\mathbf{r}_k(\mathbf{M}_\theta - \mathbf{A}_\theta)_i = (1-\rho)\mathbf{M}_\theta(k,i)$ then the uniform allowability assumption (see Remark 4.2) guarantees that $\mathbf{M}_\theta(k,i) > \alpha$, hence we get that $\mathbf{r}_k(\mathbf{M}_\theta - \mathbf{A}_\theta)_i > (1-\rho)\alpha$ for $i \in \mathfrak{W}^{\theta,k}$. □

Fix the value of δ such that the assertion of Lemma 5.3 holds.

Now we define $\psi : [0,1]^N \to [0,1]^N$ such that whenever $\mathbf{s} \in B_\delta$ then $\psi(\mathbf{s}) = \mathbf{s}$ but when $\mathbf{s} \notin B_\delta$ then $\psi(\mathbf{s}) \in B_\delta$. Namely,

$$(\psi(\mathbf{s}))_j := \begin{cases} s_j, & \text{if } s_j \ge 1 - \delta; \\ 1 - \delta, & \text{if } s_j < 1 - \delta \end{cases} \qquad \text{for} j \in [N]. \tag{5.9}$$

It immediately follows from the definition that the function ψ has the following monotonicity properties.

Fact 5.4. *For all $\theta \in \mathcal{I}$, $k \in [N]$ and $\mathbf{s}, \mathbf{t} \in [0,1]^N$ we have*

(a) $\mathbf{s} \le \psi(\mathbf{s})$*, and*
(b) for $\mathbf{s} \le \mathbf{t}$*,* $\psi(\mathbf{s}) \le \psi(\mathbf{t})$.

We now define for all $k \in [N], \theta \in \mathcal{I}$ and $\mathbf{s} \in [0,1]^N$

$$h_\theta^{(k)}(\mathbf{s}) := g_\theta^{(k)}(\psi(\mathbf{s})) \text{ and } \mathbf{h}_\theta(\mathbf{s}) := (h_\theta^{(0)}(\mathbf{s}), \dots, h_\theta^{(N-1)}(\mathbf{s})); \tag{5.10}$$

recall the definition of $g_\theta^{(k)}$ from (5.5).

We summarize the important properties of $\mathbf{h}_\theta(\mathbf{s})$, but before that we introduce some more notation. Let

$$u := \min\{1, \min\{\mathbf{A}_\theta(k,i) : (k,i,\theta) \in \mathfrak{W}\}\}. \tag{u}$$

It follows from the uniform allowability condition, together with the choice of the matrix $\mathbf{A}$ (see (5.1)), that $u > 0$. Further, let $R(t)$ denote the open ball with respect to the 1-norm centered at the origin with radius t in $\mathbb{R}^N$ and $R^C(t)$ its complement, namely

$$R(t) := \{\mathbf{s} \in [0,1]^N : \|\mathbf{s}\| < t\} \text{ and } R^C(t) := \{\mathbf{s} \in [0,1]^N : \|\mathbf{s}\| \ge t\}. \tag{5.11}$$

Lemma 5.5. *For any $\theta \in \mathcal{I}$, $k \in [N]$ and $\mathbf{s} \le \mathbf{t} \in [0,1]^N$, the following holds.*

(a) If $\mathbf{s} \in B_\delta$*, we have* $\mathbf{h}_\theta(\mathbf{s}) = \mathbf{g}_\theta(\mathbf{s})$*;*
(b) $\mathbf{h}_\theta(\mathbf{1}) = \mathbf{1}$*;*
(c) $\mathbf{h}_\theta(\mathbf{s}) \le \mathbf{h}_\theta(\mathbf{t})$*;*
(d) For any $n \in \mathbb{N}$*,* $\boldsymbol{\theta} \in \mathcal{I}^n$*, we have that* $\mathbf{h}_{\boldsymbol{\theta}}(\mathbf{s}) \ge \mathbf{f}_{\boldsymbol{\theta}}(\mathbf{s})$*;*
(e) $\mathbf{0} \le \mathbf{h}_\theta(\mathbf{s})$*;*

(f) *For any $w > N - u\delta$ (u was defined in (u)) if $\mathbf{h}_\theta(\mathbf{s}) \in R^C(w)$, then $\mathbf{s} \in B_\delta$ and in particular $\mathbf{h}_\theta(\mathbf{s}) = \mathbf{g}_\theta(\mathbf{s})$.*

Proof. *(a)* follows from the definition of $\mathbf{h}_\theta$, and *(b)* immediately follows from *(a)* combined with the fact that $\mathbf{1} \in B_\delta$. Part *(c)* is inherited from the monotonicity properties (see Lemma 5.4) of ψ and $\mathbf{g}_\theta$. *(e)* follows from *(d)*, since $\mathbf{f}_\theta(\mathbf{s}) \geq 0$.

(d) We use induction on n. First if $n = 1$, then for $\mathbf{s} \in B_\delta$, $\mathbf{h}_\theta(\mathbf{s}) = \mathbf{g}_\theta(\mathbf{s}) \geq \mathbf{f}_\theta(\mathbf{s})$ by the definition of $\mathbf{h}_\theta$ and the choice of δ (according to Lemma 5.3). For $\mathbf{s} \notin B_\delta$, $\mathbf{h}_\theta(\mathbf{s}) = \mathbf{g}_\theta(\psi(\mathbf{s}))$ on the one hand, $\mathbf{g}_\theta(\psi(\mathbf{s})) \geq \mathbf{f}_\theta(\psi(\mathbf{s})) \geq \mathbf{f}_\theta(\mathbf{s})$ and on the other hand, $\mathbf{g}_\theta(\psi(\mathbf{s})) \geq \mathbf{g}_\theta(\mathbf{s})$. Here we used that $\mathbf{g}_\theta$ and $\mathbf{f}_\theta$ are monotone increasing and that $\psi(\mathbf{s}) \geq \mathbf{s}$. Now assume that $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n) \in \mathcal{I}^n$ and we know that the assumption holds for $\boldsymbol{\theta}^- = (\theta_1, \ldots, \theta_{n-1})$, then from the hypothesis and the monotonicity of $\mathbf{f}_\theta$, $\mathbf{h}_{\boldsymbol{\theta}}(\mathbf{s}) = \mathbf{h}_{\boldsymbol{\theta}^-}(\mathbf{h}_{\theta_n}(\mathbf{s})) \geq \mathbf{f}_{\boldsymbol{\theta}^-}(\mathbf{h}_{\theta_n}(\mathbf{s})) \geq \mathbf{f}_{\boldsymbol{\theta}^-}(\mathbf{f}_{\theta_n}(\mathbf{s})) = \mathbf{f}_{\boldsymbol{\theta}}(\mathbf{s})$.

(f) Assume $\mathbf{s} \notin B_\delta$, then $\|\mathbf{1} - \mathbf{s}\|_\infty > \delta$, i.e. there exist a $j^* \in [N]$ such that $1 - s_{j^*} > \delta$, by the definition of ψ, $(\psi(\mathbf{s}))_{j^*} = 1 - \delta$. Since $\mathbf{A}_\theta$ is allowable, there exists a k^* such that $\mathbf{A}_\theta(k^*, j^*) > 0$, in particular $\mathbf{A}_\theta(k^*, j^*) \geq u$. Fix an arbitrary $w > N - u\delta$. It follows that

$$w \leq \|\mathbf{h}_\theta(\mathbf{s})\| = \sum_{k\in[N]} h_\theta^{(k)}(\mathbf{s}) = \sum_{k\in[N]} g_\theta^{(k)}(\psi(\mathbf{s})) = \sum_{k\in[N]} 1 - \sum_{j\in\mathfrak{W}^{\theta,k}} \mathbf{A}_\theta(k,j)(1 - \psi(\mathbf{s})_j)$$

$$\leq N - \mathbf{A}_\theta(k^*, j^*)(1 - (\psi(\mathbf{s}))_{j^*}) \leq N - u\delta < w,$$

which is a contradiction. □

5.2. Proof of Theorem 4.4, part (1)

Now we are ready to prove the first part of our main theorem.

Proof of Theorem 4.4, Part (1). From Lemma 5.1 it follows that there exists a set $H \subset \Sigma$, with $\nu(H) = 1$, such that for every $\overline{\boldsymbol{\theta}}$ there exists a $\gamma > 1$ and an $\widetilde{N} = \widetilde{N}(\boldsymbol{\theta})$ such that for $n > \widetilde{N}$

$$\left(\mathbf{A}_{\overline{\boldsymbol{\theta}}|_n}\right)_* \geq \gamma^n. \tag{5.12}$$

We fix such a $\gamma > 1$ and $\widetilde{N}$.

For a vector $\mathbf{v} \in \mathbb{R}^N$ analogously to the matrix case, we will use the 1-norm, that is we denote $\|\mathbf{v}\| := \sum_{i\in[N]} |v_i|$. In what follows we will show that

$$\text{for all } \overline{\boldsymbol{\theta}} \in H, \quad \|\mathbf{q}(\overline{\boldsymbol{\theta}})\| = \lim_{n\to\infty} \|\mathbf{f}_{\overline{\boldsymbol{\theta}}|_n}(\mathbf{0})\| < N, \tag{5.13}$$

proving that $\mathbf{q}(\overline{\boldsymbol{\theta}}) \neq \mathbf{1}$ for ν-almost every $\overline{\boldsymbol{\theta}}$.

Instead of inspecting the behavior of $\mathbf{f}_{\overline{\boldsymbol{\theta}}|_n}(\mathbf{0})$ directly, we consider $\mathbf{h}_{\overline{\boldsymbol{\theta}}|_n}(\mathbf{0}) \geq \mathbf{f}_{\overline{\boldsymbol{\theta}}|_n}(\mathbf{0})$ for $n > \widetilde{N}$. By the uniform allowability condition we can choose $0 < \mu < 1$ such that

$$\mu \leq \min_{k\in[N],\theta\in\mathcal{I}} \|\mathbf{c}_k(\mathbf{A}_\theta)\|. \tag{5.14}$$

For the visual explanation of the following part see Figure 3a, 3b. For $w, t \in \mathbb{R}$ define

$$\varphi_w(t) := N - Nw + wt, \tag{5.15}$$

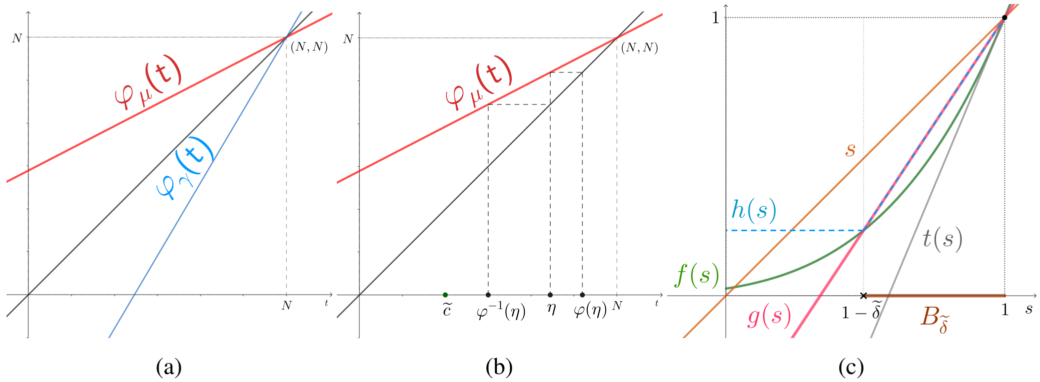


Figure 3

Choose $\eta < N$ such that

$$\varphi_\mu^{-1}(\eta) > \widetilde{c} := N - u\delta, \tag{5.16}$$

where the value of $\widetilde{c}$ occurred in Lemma 5.5 part *(f)* and u was defined in (u). Since $\mu < 1$ there exists an $\varepsilon > 0$, such that

$$\varphi_\mu(\eta) < \varphi_\mu^{\widetilde{N}}(\eta) =: N - \varepsilon, \tag{5.17}$$

where $\varphi_\mu^{\widetilde{N}}(\eta) = \varphi_\mu \circ \cdots \circ \varphi_\mu(\eta)$. Now we fix an $m > \widetilde{N}$. Then there are two cases,

1. either $\|\mathbf{h}_{\overline{\boldsymbol{\theta}}|_m}(\mathbf{0})\| \le \varphi_\mu(\eta)$, or
2. $\|\mathbf{h}_{\overline{\boldsymbol{\theta}}|_m}(\mathbf{0})\| > \varphi_\mu(\eta)$.

In case (1), since for all $k \in [N]$ we have $f_\theta^{(k)}(\mathbf{s}) \le h_\theta^{(k)}(\mathbf{s})$ for any $\mathbf{s} \in [0,1]^N$ (by part *(d)* of Lemma 5.5) and $\varphi_\mu(\eta) < N - \varepsilon$ (see (5.17)) we get that

$$\|\mathbf{q}_m(\overline{\boldsymbol{\theta}})\| = \|\mathbf{f}_{\overline{\boldsymbol{\theta}}|_m}(\mathbf{0})\| \le \varphi_\mu(\eta) < N - \varepsilon. \tag{5.18}$$

In the rest of this section, we consider case (2), namely when

$$\|\mathbf{h}_{\overline{\boldsymbol{\theta}}|_m}(\mathbf{0})\| > \varphi_\mu(\eta). \tag{5.19}$$

Set $\boldsymbol{\theta} := \overline{\boldsymbol{\theta}}|_m$. In the rest of this section, we always assume that $\mathbf{s} \in [0,1]^N$.

Lemma 5.6. (i) *For all $\theta \in \mathcal{I}$, $k \in [N]$, if $g_\theta^{(k)}(\mathbf{s}) \ge 0$, then*

$$\|\mathbf{g}_\theta(\mathbf{s})\| \le \varphi_\mu(\|\mathbf{s}\|).$$

(ii) *For $n > \widetilde{N}$ and $\widehat{\boldsymbol{\theta}} = \overline{\boldsymbol{\theta}}|_n$, if for all $k \in [N]$, $g_{\widehat{\boldsymbol{\theta}}}^{(k)}(\mathbf{s}) \ge 0$, then*

$$\|\mathbf{g}_{\widehat{\boldsymbol{\theta}}}(\mathbf{s})\| \le \varphi_\gamma(\|\mathbf{s}\|).$$

Proof. We only present the proof of the first statement since the second one can be proven using the same steps and the fact (see (5.12)) that $(\mathbf{A}_{\widehat{\boldsymbol{\theta}}})_* \geq \gamma^n > \gamma$.

$$
\begin{aligned}
\|\mathbf{g}_\theta(\mathbf{s})\| = \|\mathbf{1} - \mathbf{A}_\theta(1-\mathbf{s})\| &= \sum_{k=0}^{N-1}(1 - \sum_{\ell=0}^{N-1}\mathbf{A}_\theta(k,\ell)(1-s_\ell)) \\
&= N - \sum_{\ell=0}^{N-1}\sum_{k=0}^{N-1}\mathbf{A}_\theta(k,\ell)(1-s_\ell) = N - \sum_{\ell=0}^{N-1}\|\mathbf{c}_\ell(\mathbf{A}_\theta)\|(1-s_\ell) \\
&\leq N - (\mathbf{A}_\theta)_* N + (\mathbf{A}_\theta)_*\|\mathbf{s}\| = N - (\mathbf{A}_\theta)_*(N - \|\mathbf{s}\|) \\
&\leq N - \mu(N - \|\mathbf{s}\|) = \varphi_\mu(\|\mathbf{s}\|).
\end{aligned}
$$

□

Now we continue the proof of the first part of Theorem 4.4. Define

$$
T := \left\{p \leq m : \forall k \leq p,\ \mathbf{h}_{\boldsymbol{\theta}_k^m}(\mathbf{0}) \in R^C(\eta)\right\}, \tag{5.20}
$$

where $\boldsymbol{\theta}_k^m = (\theta_k, \dots, \theta_m)$ and m was fixed earlier in the proof. By our assumption (5.19) we get that $\mathbf{h}_{\boldsymbol{\theta}}(\mathbf{0}) \in R^C(\varphi_\mu(\eta)) \subset R^C(\eta)$, this implies that $1 \in T$. On the other hand, Lemma 5.5 *(f)* and (5.16) together imply that $m \notin T$. Namely, by (5.16) $\mathbf{h}_{\theta_m}(\mathbf{0}) \leq \widetilde{c} < \varphi_\mu^{-1}(\eta) < \eta$. That is $\mathbf{h}_{\theta_m}(\mathbf{0}) \in R(\eta)$. So, $m \notin T$. This does not contradict $1 \in T$ in case of $m = 1$, because if $m = 1$ then it is not possible that $\|\mathbf{h}_{\overline{\boldsymbol{\theta}}|_1}(\mathbf{0})\| > \varphi_\mu(\eta) > N - u\delta$, because in this case by Lemma 5.5 *(f)* we have that $\mathbf{0} \in B_\delta$, which is not possible since $\delta < 1$. Then $1 \leq Q := \max T \leq m - 1$.

Let $\mathbf{v} := \mathbf{h}_{\boldsymbol{\theta}_{Q+1}^m}(\mathbf{0})$. By the definition of Q,

$$
\mathbf{v} \in R(\eta). \tag{5.21}
$$

Also, for any $k \leq Q$, $\mathbf{h}_{\theta_k}(\mathbf{h}_{\boldsymbol{\theta}_{k+1}^m}(\mathbf{0})) \in R^C(\eta)$, hence by $\eta > \widetilde{c}$ and Lemma 5.5 *(f)* it follows that for any $k \leq Q$

$$
\mathbf{h}_{\theta_k}(\mathbf{h}_{\boldsymbol{\theta}_{k+1}^m}(\mathbf{0})) = \mathbf{g}_{\theta_k}(\mathbf{h}_{\boldsymbol{\theta}_{k+1}^m}(\mathbf{0})).
$$

By repeated applications of this we get that for any $k \leq Q$,

$$
\mathbf{h}_{\boldsymbol{\theta}_k^Q}(\mathbf{v}) = \mathbf{g}_{\boldsymbol{\theta}_k^Q}(\mathbf{v}). \tag{5.22}
$$

Now we show that Q can not be too big. Namely, to get a contradiction assume that $Q > \widetilde{N}$. Using (5.19), (5.22), the second part of Lemma 5.6 (we can use this, since for $k \leq Q$ it is true that $\mathbf{g}_{\theta_k} = \mathbf{h}_{\theta_k} \geq 0$, by Lemma 5.5 *(e)*) together with the assumption $Q > \widetilde{N}$, the fact that $\gamma > 1$, and (5.21), in this order, we obtain that

$$
\varphi_\mu(\eta) < \|\mathbf{h}_{\boldsymbol{\theta}_1^m}(\mathbf{0})\| = \|\mathbf{g}_{\boldsymbol{\theta}_1^Q}(\mathbf{v})\| \leq \varphi_\gamma(\|\mathbf{v}\|) \leq \|\mathbf{v}\| < \eta,
$$

which is a contradiction since $\eta < \varphi_\mu(\eta)$.

Hence, $Q \leq \widetilde{N}$. However, in this case using the first part of Lemma 5.6 $\widetilde{N}$-times (again for $k \leq Q$ it is true that $\mathbf{g}_{\theta_k} = \mathbf{h}_{\theta_k} \geq 0$, by Lemma 5.5 *(e)*), we get that

$$
\|\mathbf{h}_{\boldsymbol{\theta}}(\mathbf{0})\| = \|\mathbf{g}_{\boldsymbol{\theta}_1^Q}(\mathbf{v})\| \leq \varphi_\mu(\|\mathbf{g}_{\boldsymbol{\theta}_2^Q}(\mathbf{v})\|) \leq \dots \leq \varphi_\mu^{Q-1}(\|g_{\boldsymbol{\theta}_Q^Q}\mathbf{v}\|) \leq \varphi_\mu^Q(\|\mathbf{v}\|) \leq \varphi_\mu^{\widetilde{N}}(\eta) = N - \varepsilon,
$$

where $\varepsilon > 0$ was defined in (5.17). This means that (similarly to (5.18)),

$$\|\mathbf{q}_m(\overline{\boldsymbol{\theta}})\| = \|\mathbf{f}_{\boldsymbol{\theta}|_m}(\mathbf{0})\| \le \|\mathbf{h}_{\boldsymbol{\theta}}(\mathbf{0})\| \le N - \varepsilon.$$

This finishes the treatment of case 2. It follows that for $m > \widetilde{N}$ we have $\|\mathbf{q}_m(\overline{\boldsymbol{\theta}})\| \le N - \varepsilon$. In this way we have verified that (5.13) holds, which completes the proof of the first part of Theorem 4.4. □

5.3. The proof of Theorem 4.4, part (2)

It follows from the assumption *(a)* of Theorem 4.4 that there exists a finite word $\widetilde{\boldsymbol{\theta}} = (\widetilde{\theta}_1, \dots, \widetilde{\theta}_p) \in \mathcal{I}^p$ ($p \in \mathbb{N}$) such that for $[\widetilde{\boldsymbol{\theta}}] = \{\overline{\boldsymbol{\theta}} = (\theta_1, \theta_2, \dots) \in \Sigma : \ \theta_i = \widetilde{\theta}_i,\ i \le p\}$ we have $\nu([\widetilde{\boldsymbol{\theta}}]) > 0$ and further all elements of $\mathbf{M}_{\widetilde{\boldsymbol{\theta}}} := \mathbf{M}_{\widetilde{\theta}_1} \cdots \mathbf{M}_{\widetilde{\theta}_p}$ are strictly positive. From the ergodicity of ν (Principal Assumption I), it follows that there exists $\widetilde{\Sigma} \subset \Sigma$ with $\nu(\widetilde{\Sigma}) = 1$, such that all $\overline{\boldsymbol{\theta}} \in \widetilde{\Sigma}$ contain $\widetilde{\boldsymbol{\theta}}$ as a subword (in any position).

For a $k_1 \in [N]$ we define

$$\mathbf{Bad}_{k_1} := \left\{ \overline{\boldsymbol{\theta}} \in \widetilde{\Sigma} : \lim_{n \to \infty} f^{(k_1)}_{\overline{\boldsymbol{\theta}}|_n}(\mathbf{0}) = 1 \right\}. \tag{5.23}$$

Lemma 5.7. *Under the conditions of Theorem 4.4, for any $k_1 \in [N]$, $\nu(\mathbf{Bad}_{k_1}) = 0$.*

Proof. Fix an $n \ge p$. Let

$$A_n := \left\{ \overline{\boldsymbol{\theta}} \in \Sigma : \overline{\boldsymbol{\theta}}|_n \text{ does } \underline{\text{not}} \text{ contain the word } \widetilde{\boldsymbol{\theta}} \right\}, \text{ then } \bigcup_{n=p}^{\infty} A_n^C = \widetilde{\Sigma}.$$

Hence, it is enough to prove that for any $n \ge p$ and any ε

$$\nu\left(\mathbf{Bad}_{k_1} \cap A_n^C\right) < \varepsilon. \tag{5.24}$$

Let n and $\varepsilon > 0$ be arbitrary. Recall that $R^C(t) := \{\mathbf{s} \in [0,1]^N : \|\mathbf{s}\| \ge t\}$. By the assumption $\lim_{\ell \to \infty} \mathbf{f}_{\overline{\boldsymbol{\theta}}|_\ell}(\mathbf{0}) = \mathbf{q}(\overline{\boldsymbol{\theta}}) \ne \mathbf{1}$ for ν-almost every $\overline{\boldsymbol{\theta}}$, we can choose a $\delta > 0$ which is so small that for

$$t := N(1 - \delta p_*^{-n}), \text{ and } X_t := \left\{ \overline{\boldsymbol{\theta}} \in \widetilde{\Sigma} : \lim_{\ell \to \infty} \mathbf{f}_{\overline{\boldsymbol{\theta}}|_\ell}(\mathbf{0}) \in R^C(t) \right\}, \tag{5.25}$$

we have $\nu(X_t) < \varepsilon$, where $p_* = \frac{\alpha}{2}$ (recall that α was defined in (4.1)).

Lemma 5.8. *Let $\theta \in \mathcal{I}$, $k \in [N]$ and $0 < \widetilde{\delta} < 1$. Then for any $\mathbf{s} \in [0,1]^N$*

$$\textit{if there exists an } i \in \mathfrak{W}^{\theta,k} \textit{ such that } s_i < 1 - \widetilde{\delta} \Longrightarrow f^{(k)}_\theta(\mathbf{s}) < 1 - p_* \widetilde{\delta}.$$

Proof. Fix an $\mathbf{s} \in [0,1]^N$ such that $s_i < 1 - \widetilde{\delta}$ for some $i \in \mathfrak{W}^{\theta,k}$.

$$f^{(k)}_\theta(\mathbf{s}) \le f^{(k)}_\theta(\mathbf{1} - \mathbf{e}_i \widetilde{\delta}) = \sum_{\mathbf{z} \in \mathbb{N}_0^N} f^{(k)}_\theta[\mathbf{z}](1 - \widetilde{\delta})^{z_i} \le \sum_{\substack{\mathbf{z} \in \mathbb{N}_0^N \\ z_i = 0}} f^{(k)}_\theta[\mathbf{z}]$$

$$+\sum_{\substack{\mathbf{z}\in\mathbb{N}_0^N\\ z_i\neq 0}} f_\theta^{(k)}[\mathbf{z}](1-\widetilde{\delta}) \leq 1-\widetilde{\delta}\sum_{\substack{\mathbf{z}\in\mathbb{N}_0^N\\ z_i\neq 0}} f_\theta^{(k)}[\mathbf{z}] < 1-p_*\widetilde{\delta}.$$

□

Now we fix a $\overline{\boldsymbol{\theta}} \in A_n^C \bigcap \mathbf{Bad}_{k_1}$. For an $r\in\mathbb{N}$, $r\geq 2$ let

$$C := \left\{k\in[N] : \exists(k_2,\dots,k_n)\in[N]^{n-1} \text{ such that } k_i \in \mathfrak{W}^{\theta_{i-1},k_{i-1}}, i\in\{2,\dots,n\}, k\in\mathfrak{W}^{\theta_n,k_n}\right\}$$
$$= \left\{k\in[N] : \mathbf{M}_{\overline{\boldsymbol{\theta}}|_n}(k_1,k)>0\right\}$$

Lemma 5.9. $C=[N]$.

Proof of Lemma 5.9. This follows from the fact that all the expectation matrices are allowable and that $\overline{\boldsymbol{\theta}}|_n$ contains the word $\widetilde{\boldsymbol{\theta}}$, which implies that $\mathbf{M}_{\overline{\boldsymbol{\theta}}|_n}$ is a strictly positive matrix. □

Since we assumed that $\overline{\boldsymbol{\theta}}\in\mathbf{Bad}_{k_1}$ we can find a $\mathfrak{K}>n$ such that

$$f_{\overline{\boldsymbol{\theta}}|_{\mathfrak{K}}}^{(k_1)}(\mathbf{0}) > 1-\delta. \tag{5.26}$$

As earlier, we write $\overline{\boldsymbol{\theta}}_j^{\ell} = (\theta_j,\theta_{j+1},\dots,\theta_\ell)$. For every $h<\mathfrak{K}$ let

$$\mathbf{s}_h = (\mathbf{s}_h(0),\dots,\mathbf{s}_h(N-1)) := \mathbf{f}_{\overline{\boldsymbol{\theta}}_{h+1}^{\mathfrak{K}}}(\mathbf{0}). \tag{5.27}$$

Then again for $h<\mathfrak{K}$

$$f_{\overline{\boldsymbol{\theta}}|_{\mathfrak{K}}}^{(k_1)}(\mathbf{0}) = f_{\theta_1}^{(k_1)}(\mathbf{s}_1), \qquad \mathbf{s}_h(\ell) = f_{\theta_{h+1}}^{(\ell)}(\mathbf{s}_{h+1}) = f_{\theta_{h+1}}^{(\ell)}(f_{\overline{\boldsymbol{\theta}}_{h+2}^{\mathfrak{K}}}(\mathbf{0})).$$

It follows from Lemma 5.8, and formulae (5.26) and (5.27) that for all $k_2\in\mathfrak{W}^{\theta_1,k_1}$ we have $\mathbf{s}_1(k_2) > 1-\delta p_*^{-1}$. By repeated application of Lemma 5.8 we get that for the n fixed in the beginning of the proof we have

$$\mathbf{s}_n(k_{n+1}) > 1-\delta p_*^{-n}, \quad \forall(k_2,\dots,k_{n+1})\in C^n.$$

Using this, by Lemma 5.9 we get that $\mathbf{s}_n(k) > 1-\delta p_*^{-n}$ for all $k\in[N]$. This means that

$$\mathbf{f}_{\overline{\boldsymbol{\theta}}_{n+1}^{\mathfrak{K}}}(\mathbf{0}) = \mathbf{s}_n \in B_{\delta p_*^{-n}}. \tag{5.28}$$

Using that $\mathbf{f}_\theta$ is componentwise monotone and $\mathbf{f}_\theta : [0,1]^N \to [0,1]^N$, we get from (5.28) that

$$\lim_{M\to\infty} \mathbf{f}_{\overline{\boldsymbol{\theta}}_{n+1}^{M}}(\mathbf{0}) \in B_{\delta p_*^{-n}} \subset R^C(t),$$

where t was defined in (5.25). That is we have proved that $\overline{\boldsymbol{\theta}}\in A_n^C\bigcap\mathbf{Bad}_{k_1} \Longrightarrow \sigma^n\overline{\boldsymbol{\theta}}\in X_t$. In other words we have verified that $A_n^C\cap\mathbf{Bad}_{k_1}\subset\sigma^{-n}X_t$, where for $(\theta_1,\theta_2,\dots)=\overline{\boldsymbol{\theta}}\in\Sigma$, $\sigma^{-1}(\overline{\boldsymbol{\theta}}) = \{(\theta,\theta_1,\theta_2);\ \theta\in\mathcal{I}\}$. In particular $\sigma^{-n}\overline{\boldsymbol{\theta}}$ contains the infinite words that have $\overline{\boldsymbol{\theta}}$ as a subword starting at the $n+1$-st position. Using that ν is measure preserving, we get that $\nu(A_n^C\cap\mathbf{Bad}_{k_1}) \leq \nu(\sigma^{-n}X_t) = \nu(X_t) < \varepsilon$. □

Proof of Theorem 4.4, Part (2). Using that $k_1\in[N]$ was arbitrary in Lemma 5.7, we get that the second assertion of Theorem 4.4 holds. □

Appendix A: Theorem of Hennion

Let $\mathbf{B}$ be an $N \times N$ non-negative and allowable ($\mathbf{B} \in \mathscr{A}$) matrix. It is easy to check that the norms defined in Section 3 agree with the ones used in Hennion (1997), i.e.

$$\|\mathbf{B}\|_1 = \max \left\{ \sum_{i\in[N]} \sum_{j\in[N]} \mathbf{B}_{i,j} x_j : x_j \geq 0, \sum_{j\in[N]} x_j = 1 \right\}, \text{ and}$$

$$(\mathbf{B})_* = \min \left\{ \sum_{i\in[N]} \sum_{j\in[N]} \mathbf{B}_{i,j} x_j : x_j \geq 0, \sum_{j\in[N]} x_j = 1 \right\}.$$

Let $\mathcal{B} = \{\mathbf{B}_i\}_{i\in\mathcal{I}} \subset \mathscr{A}$. For $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n) \in \mathcal{I}^n$ as before we denote $\mathbf{B}_{\boldsymbol{\theta}} := \mathbf{B}_{\theta_1} \cdots \mathbf{B}_{\theta_n}$ and $\mathbf{B}_{\overleftarrow{\boldsymbol{\theta}}} := \mathbf{B}_{\theta_n} \cdots \mathbf{B}_{\theta_1}$.

A.1. A corollary of a theorem of Hennion

In Hennion (1997), X_n, a sequence of random variables taking values in the set of non-negative allow able matrices is defined. The random matrix $X^{(n)}$, which appears in (Hennion, 1997, Theorem 2), corresponds to $\mathbf{B}^T_{\overleftarrow{\overline{\boldsymbol{\theta}}|_n}}$ where $\overline{\boldsymbol{\theta}}$ is chosen randomly according to the probability measure ν. (Hennion, 1997, Theorem 2) has two conditions: The first one is that $m_1 < \infty$ and the second one is Condition $\mathscr{C}$ which corresponds to the first and second part of our assumption that $\mathcal{B}$ is good (see Definition 3.1). Hence, the conditions of (Hennion, 1997, Theorem 2) always holds whenever $\mathcal{B} \subset \mathscr{A}$ is good with respect to the ergodic measure ν. The conclusion of (Hennion, 1997, Theorem 2) immediately implies that

$$\lim_{n\to\infty} \sup_{i\in[N]} \left| \frac{1}{n} \log\left(\mathbf{1}^T \mathbf{B}_{\overline{\boldsymbol{\theta}}|_n} \mathbf{e}_i\right) - \lambda \right| = 0, \text{ for } \nu\text{-a.e. } \overline{\boldsymbol{\theta}} \in \Sigma. \tag{A.1}$$

Observe that $\mathbf{1}^T \mathbf{B}_{\overline{\boldsymbol{\theta}}|_n} \mathbf{e}_i$ is the i-th column sum of the matrix $\mathbf{B}_{\overline{\boldsymbol{\theta}}|_n}$. Hence, for ν-almost every $\overline{\boldsymbol{\theta}} \in \Sigma$

$$\lim_{n\to\infty} \left| \frac{1}{n} \log(\mathbf{B}_{\overline{\boldsymbol{\theta}}|_n})_* - \lambda \right| = 0. \tag{A.2}$$

Remark A.1. In this paper we change the order of the matrix product from $\mathbf{B}_{\overleftarrow{\overline{\boldsymbol{\theta}}|_n}}$ to $\mathbf{B}_{\overline{\boldsymbol{\theta}}|_n}$, now we explain why we can do this. Observe that in (Hennion, 1997, Theorem 2) the role of $\mathbf{1}^T$ and $\mathbf{e}_i$ is interchangeable, meaning that we can consider row, instead of column sums of the matrices. Consequently, we get an analogous result to Lemma 3.4 but for the "row-sum exponent", namely that

$$\lim_{n\to\infty} \frac{1}{n} \log(\min_{i\in[N]} \mathbf{e}_i^T \mathbf{B}_{\overline{\boldsymbol{\theta}}|_n} \mathbf{1}) = \lambda \quad \text{ for } \nu\text{-a.e. } \overline{\boldsymbol{\theta}} \in \Sigma. \tag{A.3}$$

Remark A.2. In Hennion (1997) the notation X_n sometimes stands for the whole process starting at the n-th position and sometimes as the n-th element of the process; from the context there, it is clear what is the correct interpretation, in this paper we only use the second interpretation.

Appendix B: Basic properties of multivariate pgfs

The following is a well-known fact.

Fact B.1. *Let X, Y be independent random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ taking values in $\mathbb{N}_0^N$ with pgfs f_X and f_Y respectively. Then the pgf f_{X+Y} of the random variable $X+Y$ satisfies $f_{X+Y}(\mathbf{s}) = f_X(\mathbf{s}) \cdot f_Y(\mathbf{s})$.*

Lemma B.2. *If $(k, i, \theta) \notin \mathfrak{W}$ (or equivalently $\partial f_\theta^{(k)}/\partial s_i(\mathbf{1}) = 0$) then for all $\mathbf{w} \in (0,1]^N$*

1. *$\partial f_\theta^{(k)}/\partial s_i(\mathbf{w}) = 0$ and*
2. *$\mathbf{r}_i((f_\theta^{(k)})''(\mathbf{w}))$=$\mathbf{c}_i((f_\theta^{(k)})''(\mathbf{w})) = \mathbf{0}$.*

Proof. Since 2 immediately follows from 1, we only provide details for the first part. By definition

$$f_\theta^{(k)}(\mathbf{w}) := \sum_{\mathbf{z} \in \mathbb{N}_0^N} f_\theta^{(k)}[\mathbf{z}] \mathbf{w}^{\mathbf{z}}, \ \mathbf{w} \in [0,1]^N, \ k \in [N], \ \theta \in \mathcal{I},$$

hence

$$\frac{\partial f_\theta^{(k)}}{\partial s_i}(\mathbf{w}) = \sum_{\substack{\mathbf{z} \in \mathbb{N}_0^N \\ z_i \neq 0}} f_\theta^{(k)}[\mathbf{z}] z_i w_i^{z_i - 1} \prod_{\substack{\ell \in [N] \\ \ell \neq i}} w_\ell^{z_\ell}. \tag{B.1}$$

From

$$0 = \partial f_\theta^{(k)}/\partial s_i(\mathbf{1}) = \sum_{\mathbf{z} \in \mathbb{N}_0^N} f_\theta^{(k)}[\mathbf{z}] z_i,$$

since all the summands are non-negative, we conclude that $f_\theta^{(k)}[\mathbf{z}] z_i = 0$ for all $\mathbf{z} \in \mathbb{N}_0^N$, hence the assertion follows. □

Appendix C: Lyapunov and column-sum exponents

We recall the Subadditive Ergodic Theorem.

Theorem C.1 (Partially Theorem 10.1 Walters (2000)). *Let $(X, \mathcal{B}, \nu)$ be a probability space and let $T : X \to X$ be a measure preserving transformation. Let $\{f_n\}_{n=1}^\infty$ be a sequence of measurable functions $f_n : X \to \mathbb{R} \cup \{-\infty\}$ satisfying*

(1) *$f_1^+ \in L^1(\nu)$, for $f_i^+(x) = \max\{0, f_1(x)\}$;*
(2) *for each $k, n \geq 1$, $f_{n+k} \leq f_n + f_k$ a.e.*

Then there exists a measurable function $f : X \to \mathbb{R} \cup \{-\infty\}$ such that the following holds:

1. *$f^+ \in L^1(\nu)$*
2. *$f \circ T = f$ a.e. (hence if ν is ergodic, then f is constant a.e.)*
3. *$\lim_{n \to \infty} \frac{1}{n} f_n = f$ a.e.*

Let the probability space be given as in Definition 2.5, together with the left shift σ. Assume that this system is not just measure preserving but also ergodic. Assume further that the $N \times N$ matrices $\{\mathbf{M}_i\}_{i\in\mathcal{I}}$ for some countable index set $\mathcal{I}$ are good w.r.t. the system (see Definition 3.1 for the definition of good set of matrices).

In what follows we discuss the Lyapunov and the column sum exponent, starting with the Lyapunov exponent. For $\overline{\boldsymbol{\theta}} = (\theta_1, \theta_2, \dots)$ let $f_n(\overline{\boldsymbol{\theta}}) := \log \|\mathbf{M}_{\theta_1} \cdots \mathbf{M}_{\theta_n}\|$. Then assumption (1) of Theorem C.1 is satisfied because of the norm requirements (the first numbered point) in the definition of being good. By the sub-multiplicativity of the matrix norm (2) is satisfied as well, namely:

$$
\begin{aligned}
f_{n+k}(\overline{\boldsymbol{\theta}}) &:= \log\left(\|\mathbf{M}_{\theta_1} \cdots \mathbf{M}_{\theta_n} \cdot \mathbf{M}_{\theta_{n+1}} \cdots \mathbf{M}_{\theta_{n+k}}\|\right) \le \log\left(\|\mathbf{M}_{\theta_1} \cdots \mathbf{M}_{\theta_n}\| \cdot \|\mathbf{M}_{\theta_{n+1}} \cdots \mathbf{M}_{\theta_{n+k}}\|\right) \\
&\le \log\left(\|\mathbf{M}_{\theta_1} \cdots \mathbf{M}_{\theta_n}\|\right) + \log\left(\|\mathbf{M}_{\theta_{n+1}} \cdots \mathbf{M}_{\theta_{n+k}}\|\right) = f_n(\overline{\boldsymbol{\theta}}) + f_k(\sigma^n(\overline{\boldsymbol{\theta}})).
\end{aligned}
$$

Hence, the following corollary of the combination of Theorem C.1 parts 2 and 3 holds:

Corollary C.2. *In the above given setup there exists $\lambda \in \mathbb{R}$ (called (the maximal) Lyapunov exponent) such that the following holds:*

$$
\lambda := \lambda(\nu, \mathcal{B}) = \lim_{n\to\infty} \frac{1}{n} \log \|\mathbf{B}_{\overline{\boldsymbol{\theta}}|_n}\| \quad \text{for } \nu\text{-almost every } \overline{\boldsymbol{\theta}} \in \Sigma. \tag{C.1}
$$

Remark C.3. For the conclusion of Corollary C.2 it is enough that the above system is ergodic and $\int_\Sigma \max\{0, \log \|\mathbf{M}_{\theta_1}\|\} d\nu(\overline{\boldsymbol{\theta}}) < \infty$ and the matrix norm can be any sub-multiplicative norm. The remaining requirements of being good, namely, having non-negative elements, being allowable and having a positive product is important for the existence of the column-sum exponent and for the conclusion of our main theorem.

We now consider the columns sum exponent. It can be shown that for a good set of matrices the minimum column sum $(\dot{})_*$ defined in (3.1) is super-multiplicative in the sense that for the good matrices $\mathbf{M}_1, \mathbf{M}_2$ we have

$$
(\mathbf{M}_1\mathbf{M}_2)_* \ge (\mathbf{M}_1)_*(\mathbf{M}_2)_*. \tag{C.2}
$$

Since the direction of the inequality is the opposite of that in the case of the norm, we use the sub-additive ergodic theorem for the function:

$$
f_n(\overline{\boldsymbol{\theta}}) = -\log\left((\mathbf{M}_{\theta_1} \dots \mathbf{M}_{\theta_n})_*\right), \tag{C.3}
$$

for some $\overline{\boldsymbol{\theta}} \in \Sigma$. Similarly to the previous part assumption (1) of Theorem C.1 is satisfied because of the norm requirements (this time the second part of the first numbered point) in the definition of being good. The second assumption of the theorem is satisfied for similar reasons as in the previous case. Hence, by multiplying the resulting limit f of the conclusion of Theorem C.1 with -1 we can conclude that

Corollary C.4. *There exists a $\lambda_* \in \mathbb{R}$ such that*

$$
\lambda_* := \lambda_*(\nu, \mathcal{B}) = \lim_{n\to\infty} \frac{1}{n} \log(\mathbf{B}_{\overline{\boldsymbol{\theta}}|_n})_* \quad \text{for } \nu\text{-almost every } \overline{\boldsymbol{\theta}} \in \Sigma.
$$

Acknowledgments

We would like to say thanks to De-Jun Feng, who called our attention to the result in Hennion (1997) which played a crucial role in proving Theorem 4.4. We also would like to say thanks to the anonymous referees, Michel Dekking and Alex Rutar for their helpful comments and suggestions about the manuscript.

Funding

VO is supported by National Research, Development and Innovation Office - NKFIH, Project FK134251. KS is supported by National Research, Development and Innovation Office - NKFIH, Project K142169. Both authors have received funding from the HUN-REN BME Stochastics Research group and grant NKFI KKP144059 “Fractal geometry and applications”.

References

ATHREYA, K. B. and KARLIN, S. (1971). On branching processes with random environments: I: Extinction probabilities. *Ann. Math. Statist.* **42** 1499–1520. https://doi.org/10.1214/aoms/1177693150

BÁRÁNY, B., SIMON, K. and SOLOMYAK, B. (2023). *Self-Similar and Self-Affine Sets and Measures*. *Math. Surv. Monogr.* Providence, RI: American Mathematical Society (AMS).

DEKKING, F. M. and GRIMMET, G. R. (1988). Superbranching processes and projections of random Cantor sets. *Probab. Theory Related Fields* **78** 335–355. https://doi.org/10.1007/BF00334199

DEKKING, F. M. and MEESTER, R. W. J. (1990). On the structure of Mandelbrot's percolation process and other random Cantor sets. *J. Stat. Phys.* **58** 1109–1126. https://doi.org/10.1007/BF01026566

DEKKING, F. M. and SIMON, K. (2008). On the size of the algebraic difference of two random Cantor sets. *Random Structures Algorithms* **32** 205–222. https://doi.org/10.1002/rsa.20178

FALCONER, K. J. (1989). Projections of random Cantor sets. *J. Theoret. Probab.* **2** 65–70. https://doi.org/10.1007/BF01048269

FALCONER, K. J. and GRIMMETT, G. R. (1992). On the geometry of random Cantor sets and fractal percolation. *J. Theoret. Probab.* **5** 465–485. https://doi.org/10.1007/BF01060430

HENNION, H. (1997). Limit theorems for products of positive random matrices. *Ann. Probab.* **25** 1545 – 1587. https://doi.org/10.1214/aop/1023481103

KARLIN, S. and TAYLOR, H. M. (1975). *A First Course in Stochastic Processes*.

KERSTING, G. and VATUTIN, V. (2017). Discrete Time Branching Processes in Random Environment. *Math. Stat. Ser.* Hoboken, NJ: John Wiley & Sons; London: ISTE. https://doi.org/10.1002/9781119452898

MÓRA, P., SIMON, K. and SOLOMYAK, B. (2009). The Lebesgue measure of the algebraic difference of two random Cantor sets. *Indag. Math., New Ser.* **20** 131–149. https://doi.org/10.1016/S0019-3577(09)80007-4

ORGOVÁNYI, V. and SIMON, K. (2023). Projections of the random Menger sponge. *Asian J. Math.* **27** 893–936. https://doi.org/10.4310/AJM.2023.v27.n6.a4

ORGOVÁNYI, V. and SIMON, K. (2024). Interior points and Lebesgue measure of overlapping Mandelbrot percolation sets. Preprint.

POLLICOTT, M. (2023). Lectures on Equilibrium states, mixing and dimension. Lecture Notes from Banach Center, Warsaw.

TANNY, D. (1981). On multitype branching processes in a random environment. *Adv. in Appl. Probab.* **13** 464–497. https://doi.org/10.2307/1426781

VATUTIN, V. A. and DYAKONOVA, E. E. (2021). Multitype branching processes in random environment. *Russ. Math. Surv.* **76** 1019–1063. https://doi.org/10.1070/RM10012

WALTERS, P. (2000). *An Introduction to Ergodic Theory*. *Graduate Texts in Mathematics*. Springer New York.

WEISSNER, E. W. (1971). Multitype branching processes in random environments. *J. Appl. Probab.* **8** 17–31. https://doi.org/10.2307/3211834