The Annals of Applied Probability
1997, Vol. 7, No. 3, 772–801

ON THE CONVERGENCE OF MULTITYPE BRANCHING PROCESSES WITH VARYING ENVIRONMENTS[1]

BY OWEN DAFYDD JONES

University of Sheffield

Using the ergodic theory of nonnegative matrices, conditions are obtained for the $\mathscr{L}^2$ and almost sure convergence of a supercritical multitype branching process with varying environment, normed by its mean. We also give conditions for the extinction probability of the limit to equal that of the process.

The theory developed allows for different types to grow at different rates, and an example of this is given, taken from the construction of a spatially inhomogeneous diffusion on the Sierpinski gasket.

1. Introduction and statement of results. A multitype branching process with varying environment (MTBPVE) generalizes the classical multitype branching or Galton–Watson process. For a finite number d of types, we allow the number of type j offspring of a type i parent at time n to depend on i, j and n. In what follows, we give second moment conditions under which a MTBPVE normed by its mean, whose mean matrices are weakly ergodic, converges a.s. and in $\mathscr{L}^2$ to a nontrivial limit. These conditions generalize those of Harris (1963) for multitype fixed environment processes *and* those of Fearn (1971) and Jagers (1974) for single-type varying environment processes. Notably, if the mean matrices are well behaved in some sense, then our $\mathscr{L}^2$ convergence condition is best possible.

Our results give conditions under which a MTBPVE grows like its mean, which in this case is given by a forward product of nonnegative matrices. Nonnegative matrix products can exhibit more than one rate of growth, in the sense that, as additional factors are added to the product, different elements of the product can grow at different rates. This opens up the possibility of MTBPVE with more than one rate of growth. Indeed, in Section 4 we give an example of a MTBPVE with two distinct growth rates, arising from the construction of a spatially inhomogeneous diffusion on the Sierpinski gasket (a simple fractal). In order to analyze growth rates better, the discussion of ergodic theory given in Section 2 goes beyond that strictly required for our convergence results, in particular looking at strong ergodicity and some

Received July 1995; revised December 1996, March 1997.

[1]Research partially supported by U.K. Engineering and Physical Sciences Research Council.

AMS 1991 *subject classifications*. Primary 60J80; secondary 15A48.

Key words and phrases. Branching process, multitype, varying environment, ergodic matrix products.

related ideas. However, this extra analysis will be needed for the example in Section 4.

In addition to results on the convergence of the normed process, we also derive conditions for the extinction probability of the limit to equal that of the process. This result will also be applied in Section 4.

We will adopt the following notation for the remainder of the paper. For a matrix $A \in \mathbb{R}^{d\times d}$, write $A(i,j)$ for its (i,j)th element, $A(i,\cdot)$ for the row vector given by its ith row and $A(\cdot,j)$ for the column vector given by its jth column. Similarly, for a vector $a \in \mathbb{R}^d$, write $a(i)$ for its ith component. The vector of 1s will be written 1 and the unit vector with a 1 in position i will be written e_i. A (nonnegative) matrix is called row/column allowable if each row/column has a nonzero component. A row *and* column allowable matrix is simply called allowable. Clearly, the product of allowable matrices is also allowable. Write $A \ge 0$ or $a \ge 0$ if every element of A or a is ≥ 0, and write $A > 0$ or $a > 0$ if every element is greater than 0. Unless stated otherwise, we will assume that all matrices and vectors dealt with are nonnegative.

A nonnegative matrix $A \in \mathbb{R}^{d\times d}$ is called primitive if there exists an n such that $A^n > 0$. For such A we write $\mathrm{PF}(A)$ for its (unique, real) largest eigenvalue, that is, its spectral radius, and $\mathrm{LPF}(A)$ and $\mathrm{RPF}(A)$ for the corresponding (unique, strictly positive) left and right eigenvectors respectively, normed to be probability vectors. Here, PF stands for Perron–Frobenius.

Suppose that the offspring distributions of the process are given by a sequence of $\mathbb{Z}_+^{d\times d}$ valued r.v.s. $\{X_n\}_{n=0}^{\infty}$. That is, the distribution of the number of type j children born to a single type i parent at time n is the same as that of $X_n(i,j)$. Define $M_n = \mathrm{E}X_n$, $V_n[i] = \mathrm{Cov}\,X_n(i,\cdot)$ and $\sigma_n^2(i,j) = \mathrm{Var}\,X_n(i,j) = V_n[i](j,j)$. We will assume that the $\{M_n\}_{n=0}^{\infty}$ are finite in all that follows. Unless otherwise stated, we will also assume that they are allowable. For fixed $m \ge 0$, let $Z_m = \{Z_{m,n}\}_{n=m}^{\infty}$ be the branching process defined in the usual way [see, e.g., Asmussen and Hering (1983) or Athreya and Ney (1972)], letting $Z_{m,n}(i,j)$ be the number of type j descendants at time n of a single type i parent at time m. Note that, as defined, Z_m takes on values in $\mathbb{Z}_+^{d\times d}$, where the rows of Z_m are independent processes.

For a sequence of matrices $\{A_n\}_{n=0}^{\infty}$, we will write $A_{m,n}$ for the forward product from m to $n-1$. That is, $A_{m,n} = A_m A_{m+1} \cdots A_{n-1}$. It follows from the branching property of Z_m that for any $m \le n \le p$,

$$\mathrm{E}\big(Z_{m,p}|Z_{m,n}\big) = Z_{m,n} M_{n,p}.$$

Our tool for dealing with the matrix product $M_{m,n}$ is the ergodic theory of nonnegative matrices.

Ergodic theory for nonnegative matrices can be viewed as a generalization of the Perron–Frobenius theory, as it describes the growth and limiting behavior of matrix products. We say that the matrices $\{M_n\}$ are *weakly ergodic* if for all $m \ge 0$ the forward product $M_{m,n}$ is strictly positive and of rank 1 in the limit as $n \to \infty$. A precise definition is given in Section 2.

A fundamental tool used in the development of ergodic theory for matrix products is Birkhoff's contraction coefficient. For $x, y \in \mathbb{R}^d$, $x, y > 0$, put

$$\rho(x^T, y^T) = \log \frac{\max_i x(i)/y(i)}{\min_i x(i)/y(i)} = \max_{i,j} \log \frac{x(i)y(j)}{x(j)y(i)}.$$

The character ρ is often called a *projective distance* since $\rho(x^T, y^T) = 0$ if and only if $x = \lambda y$ for some $\lambda > 0$ and $\rho(\alpha x^T, \beta y^T) = \rho(x^T, y^T)$ for all scalars $\alpha, \beta > 0$. For a nonnegative column allowable matrix $A \in \mathbb{R}^{d \times d}$, Birkhoff's contraction coefficient is defined as

$$\tau(A) = \sup_{x, y > 0,\, x \neq \lambda y} \frac{\rho(x^T A, y^T A)}{\rho(x^T, y^T)}.$$

It is easily shown that $0 \leq \tau(A) \leq 1$ and that for any other nonnegative column allowable matrix $B \in \mathbb{R}^{d \times d}$, $\tau(AB) \leq \tau(A)\tau(B)$. If A is allowable then

$$\tau(A) = \frac{1 - \sqrt{\phi(A)}}{1 + \sqrt{\phi(A)}} \quad \text{where } \phi(A) = \begin{cases} \min_{i,j,k,l} \dfrac{A(i,k)A(j,l)}{A(j,k)A(i,l)}, & A > 0, \\ 0, & A \not> 0. \end{cases}$$

[This result is due originally to Birkhoff (1957), but see also Seneta (1981), Section 3.4.] It follows that for allowable A, $\tau(A) < 1$ if and only if $A > 0$ and $\tau(A) = 0$ if and only if $A = wv^T$ for some strictly positive $w, v \in \mathbb{R}^d$. Given this and our definition of weak ergodicity, it should come as no surprise that

$$\{M_n\} \textit{ are weakly ergodic} \quad \Leftrightarrow \quad \forall\, m \geq 0,\ \tau(M_{m,n}) \to 0 \textit{ as } n \to \infty.$$

[See Seneta (1981), Lemma 3.3.] Define diagonal matrices

$${}^m R_n = \operatorname{diag}({}^m R_n(1), \ldots, {}^m R_n(d))$$

for $0 \leq m \leq n$ by

$${}^m R_n(j) = 1^T M_{m,n}(\cdot, j).$$

The main results of the paper follow.

THEOREM 1 ($\mathscr{L}^2$ convergence theorem). *If the $\{M_n\}$ are allowable and weakly ergodic with column limit vectors $\{w_m\}$ and if for some $m \geq 0$,*

$$\sum_{n=m}^{\infty} \sum_{i,j} \frac{{}^m R_n(i) \sigma_n^2(i,j)}{{}^m R_{n+1}^2(j)} < \infty, \tag{1}$$

then there exists a r.v. $L_m \geq 0$ such that $\mathrm{E} L_m = w_m$ and

$$Z_{m,n} {}^m R_n^{-1} \to_{\mathscr{L}^2} L_m 1^T \quad \textit{as } n \to \infty.$$

THEOREM 2 (Almost sure convergence theorem). *If in addition to the conditions of Theorem 1 we have*

$$\sum_{n=m}^{\infty} \sum_{i,j} \frac{(n+1-m)\, {}^m R_n(i) \sigma_n^2(i,j)}{{}^m R_{n+1}^2(j)} < \infty \tag{2}$$

and there exists a $C < \infty$ *such that for all* $n \geq m$,

$$\sum_{p=n}^{\infty} \tau(M_{n,p})^2 \leq (n+1-m)C, \tag{3}$$

then in addition to $\mathscr{L}^2$ *convergence we have*

$$Z_{m,n}{}^{m}R_n^{-1} \to L_m 1^T \quad a.s. \text{ as } n \to \infty.$$

Here condition (2) is a strengthening of the variance condition (1), while condition (3) constrains the speed at which the mean matrix $M_{m,n}$ tends to a rank 1 matrix.

The proofs are given in Section 3. If the $\{M_n\}$ are well behaved, then not only is condition (1) necessary for $\mathscr{L}^2$ convergence, but we can also dispense with condition (3) for a.s. convergence. The following corollary details what we mean by "well behaved," and is the form of these results used in Section 4. Its proof can also be found in Section 3.

COROLLARY 3 (Necessary and sufficient variance condition). *Suppose we are given rescaling matrices* $D_n = \operatorname{diag}(D_n(1), \dots, D_n(d))$ *for all* $n \geq 0$, *such that* $Q_n := D_n M_n D_{n+1}^{-1}$ *converges elementwise to a primitive matrix* Q. *Then the condition*

$$\sum_{n=m}^{\infty} \sum_{i,j} \frac{D_n(i)\sigma_n^2(i,j)}{1^T Q_{m,n} 1 D_{n+1}^2(j)} < \infty \tag{4}$$

is necessary and sufficient for the $\mathscr{L}^2$ *convergence of* $Z_{m,n}D_n^{-1}/1^T Q_{m,n}1$ *as* $n \to \infty$. *When it exists, this limit is of the form* $\overline{L}_m \bar{v}^T$, *for some r.v.* $\overline{L}_m$ *with* $\mathrm{E}\,\overline{L}_m = D_m^{-1}\overline{w}_m$. *Here* $\bar{v}$ *is the left Perron–Frobenius eigenvector of* Q (*normed as a probability vector*) *and the* $\{\overline{w}_m\}_{m=0}^{\infty}$ *are strictly positive probability vectors, which converge as* $m \to \infty$ *to* $\overline{w}$, *the right Perron–Frobenius eigenvector of* Q (*normed as a probability vector*). *Also, if*

$$\sum_{n=m}^{\infty} \sum_{i,j} \frac{(n+1-m)D_n(i)\sigma_n^2(i,j)}{1^T Q_{m,n} 1 D_{n+1}^2(j)} < \infty \tag{5}$$

then we get a.s. convergence.

Finally, we have a result that shows that (under certain conditions) $Z_{m,n}(i,\cdot)$ either dies out, or else the number of type j individuals grows like ${}^{m}R_n(j)$, for all j, with probability 1.

PROPOSITION 4 (Extinction probabilities). *Suppose that the conditions of Theorem* 1 *hold, and for all* $m \geq 0$ *and* $1 \leq i \leq d$, *let* $q_m(i)$ *be the extinction probability of the process* $Z_m(i) := \{Z_{m,n}(i,\cdot)\}_{n=m}^{\infty}$. *Then, if there exists some constant* K *and vectors* $h_n \in \mathbb{R}_+^d$, $n = 0, 1, \dots,$ *such that for all* $m \geq 0$,

$1 \le i \le d$ and $x \in \mathbb{R}_+$,

$$\mathrm{P}(Z_{m,n}(i,\cdot)h_n \le x) \to q_m(i) \quad \text{as } n \to \infty \tag{6}$$

and

$$\sum_{n=m}^{\infty} \sum_{j,k} \frac{M_{m,n}(i,j)\,\sigma_n^2(j,k)}{w_m(i)^{2\,m}R_{n+1}^2(k)} \le K/h_m(i) - 1 \tag{7}$$

then $q_m(i) = \mathrm{P}(L_m(i) = 0)$.

Furthermore, if there exist rescaling matrices $D_n = \mathrm{diag}(D_n(1), \dots, D_n(d))$ for all $n \ge 0$, such that $Q_n := D_n M_n D_{n+1}^{-1}$ converges elementwise to a primitive matrix Q, then (7) is equivalent to

$$\sum_{n=m}^{\infty} \sum_{j,k} \frac{D_n(j)\,\sigma_n^2(j,k)}{1^T Q_{m,n} 1 D_{n+1}^2(k)} \le (K/h_m(i) - 1)\big(D_m^{-1}\overline{w}_m\big)(i) \tag{8}$$

where the $\overline{w}_m$ are the same as those of Corollary 3.

In practice, condition (6) can be difficult to check. However, we can give some more practical conditions which imply it. If $\mathrm{P}(X_n > 0) = 1$ for all n, then Z can never die out. Let $\mathscr{M}_{m,n}$ denote the minimum family sizes of $Z_{m,n}$. Note that in many situations of interest—such as the example considered in Section 4—$\mathscr{M}_{m,n}$ can be explicitly determined. If we choose the $\{h_n\}$ so that

$$\mathscr{M}_{0,n} h_n \to \infty \quad \text{as } n \to \infty \tag{9}$$

then (6) follows. In practice we try and take the $\{h_n\}$ as small as possible, so that condition (7) can also be satisfied.

If the $\{h_n\}$ are constant, then (6) reduces to requiring that the only recurrent state of the branching process is 0. In the fixed environment case, Harris (1963) showed that "nonsingularity" of the offspring distribution is sufficient for the nonzero states to be transient. This result can be partially extended to MTBPVE. Suppose that a r.v. X, taking values in $\mathbb{Z}_+^{d\times d}$, describes the offspring distribution of a fixed environment multitype branching process. We say X is singular if

$$\mathrm{P}(X(i,\cdot)1 = 1) = 1 \quad \text{for all } i.$$

Now distinguish two cases.

1. The process Z can never die out. In this case, if we can find a nonsingular X such that $X \le_{\mathscr{D}} X_n$ for all $n \ge 0$, then the only recurrent state of $\{Z_{m,n}(i,\cdot)\}_{n=m}^{\infty}$ is 0, for all i.
2. Extinction is possible from any initial state. In this case, if we can find a nonsingular X such that $X \le_{\mathscr{D}} X_n$ for all $n \ge 0$, then the only recurrent state of $\{Z_{m,n}(i,\cdot)\}_{n=m}^{\infty}$ is 0, for all i.

An appropriate X can always be found if, for example, $X_n \to_{\mathscr{D}} X$ where X is nonsingular and $M = \mathrm{E}\,X$ is primitive. A proof of these results is given in Jones (1995). Note that the advantage of (9) over case (1) is that if the minimum population size grows to infinity, then we can take $h_n \to 0$, making (7) easier to satisfy. This is precisely the situation encountered in Section 4.

Background. Although the results of Harris (1963) and those of Fearn (1971) and Jagers (1974), which Theorems 1 and 2 generalize, were first proved more than twenty years ago, there has been little interest in MTBPVE until very recently. One reason for this new interest is the potential application of MTBPVE to the study of diffusion on fractals, as is pursued (with some success) in the work of Hattori and Watanabe (1993), Hattori, Hattori and Watanabe (1994) and Hattori (1994). The first of these uses an analytical approach to prove the weak convergence of the normed process to some limit, though it makes a number of rather restrictive conditions on the mean matrices $\{M_n\}$. The report by Hattori (1994) goes somewhat further, giving conditions for the $\mathscr{L}^2$ convergence of $Z_{m,n}(i,j)/M_{m,n}(i,j)$ and some results on the continuity of the limit. The conditions given make implicit use of weak ergodicity, though are somewhat more technical than those of Theorem 1. They also explicitly require supercriticality. The method used adapts some of the ideas of Cohn (1989) to the varying environment case.

Cohn himself has taken the results of Cohn (1989) further, in joint work with Jagers (1994) and also with Nerman and Biggins. The work with Jagers claims $\mathscr{L}^1$ convergence given weak ergodicity of the mean matrices and (essentially) uniform integrability in n of $\{Z_{m,n}{}^m R_n^{-1}\}_{n=m}^{\infty}$. Cohn also gives (without proof) some conditions for the $\mathscr{L}^2$ convergence of $Z_{m,n}(i,j)/M_{m,n}(i,j)$ as $n \to \infty$, in a recent research report [Cohn (1993)]. These again assume weak convergence of the mean matrices together with a uniform integrability condition and a variance condition, similar to but not the same as condition (1). Cohn also makes the observation that it is possible to move from an $\mathscr{L}^2$ result to an a.s. result using a Borel–Cantelli argument. At the time of writing, the work of Cohn, Nerman and Biggins referred to is still in preparation. It seems this work will treat MTBPVE more generally than the concept of weak ergodicity allows, using instead the concept of *space–time harmonic functions* to gain the required control over the matrix products $\{M_{m,n}\}_{n=m}^{\infty}$ [see Cohn and Nerman (1990) for a definition of space–time harmonic functions and a detailed analysis of how they relate to weak ergodicity].

2. Ergodicity of nonnegative matrix products. The use of "coefficients of ergodicity" such as τ in the study of products of nonnegative matrices, owes much of its modern development to the work of Hajnal (1976) and Cohen (1979). Their ideas in turn owe a lot to the study of products of positive stochastic matrices and inhomogeneous Markov chains. It is from this connection with Markov chains that we get the term ergodic. Hajnal (1976) suggested the more appropriate term "contractive" as an alternative, but this has yet to be widely adopted. Most of the standard results and ideas we will be using can be found in Seneta (1981), which provides a good summary of the work in the area and has an extensive bibliography. For more recent work in the area the reader is referred to Cohn and Nerman (1990).

In this section, we introduce the standard notions of weak and strong ergodicity and describe how they relate to each other. Although we do not use strong ergodicity explicitly in Theorems 1 and 2, it is of practical use when applying them, as can be seen in Section 4. It is also used in demonstrating that the variance condition (1) is best possible in certain situations: see Corollary 16.

In what follows, we will mean by a sequence of rescaling matrices a sequence of nonnegative diagonal matrices of full rank. For matrices $M \in \mathbb{R}^{d \times d}$ and $D = \operatorname{diag}(D(1), \dots, D(d))$, premultiplying M by D is equivalent to scaling each row i of M by $D(i)$, while postmultiplying M by D is equivalent to scaling each column j of M by $D(j)$. Define rescaling matrices ${}^m R_n = \operatorname{diag}({}^m R_n(1), \dots, {}^m R_n(d))$ by putting ${}^m R_m = I$ and requiring ${}^m R_n M_n {}^m R_{n+1}^{-1}$ to be column stochastic for all $n \geq m$. The term ${}^m R_n$ is allowable and thus also invertible, provided $M_{m,n}$ is column allowable. As products of column stochastic matrices are column stochastic, it is clear that, as defined in Section 1,

$$ {}^m R_n(j) = 1^T M_{m,n}(\cdot, j). $$

Definition 5 (Weak ergodicity). The matrices $\{M_n\}_{n=0}^{\infty}$ are said to be weakly ergodic if there exist strictly positive $w_{m,n}, v_{m,n} \in \mathbb{R}^d$ such that for all $m \geq 0$

$$ \frac{M_{m,n}(i,j)}{w_{m,n}(i) v_{m,n}(j)} \to 1 \quad \text{for all } i \text{ and } j \text{ as } n \to \circ. $$

For allowable M_n this is equivalent to requiring for all m the existence of an n such that $M_{m,n} > 0$ and a probability vector $w_m > 0$ such that

$$ \frac{M_{m,n}(i,k)}{M_{m,n}(j,k)} \to \frac{w_m(i)}{w_m(j)} \quad \text{for all } i, j \text{ and } k \text{ as } n \to \infty. $$

[See Hajnal (1976), Theorem 1 or Seneta (1981), Lemma 3.4 and Exercise 3.5.] Since $w_m > 0$, this is equivalent to requiring

$$ \frac{M_{m,n}(i,k)}{\sum_j M_{m,n}(j,k)} \to w_m(i) \quad \text{for all } i \text{ and } k \text{ as } n \to \infty. $$

We will generally write this in matrix form as $M_{m,n} {}^m R_n^{-1} \to w_m 1^T$ as $n \to \infty$.

So, weak ergodicity requires that as $n \to \infty$, the elements of any one column of the forward product $M_{m,n}$ all grow at the same rate and that within each column, the rows tend to fixed proportions.

The contraction coefficient τ is used to give more tractable conditions for weak ergodicity. We have already noted in Section 1 that the $\{M_n\}$ are weakly ergodic if and only if $\tau(M_{m,n}) \to 0$ as $n \to \infty$, for all m. Using the submultiplicity of τ, this is often enough to give a practical check for weak ergodicity. The form of τ for allowable matrices gives the following refinement for allowable M_n. The $\{M_n\}$ are weakly ergodic if and only if there exist $n(k) \uparrow \infty$,

$n(k) \neq n(k+1)$, such that

$$\sum_{k=0}^{\infty} \sqrt{\phi(M_{n(k),\, n(k+1)})} = \infty.$$

[See Hajnal (1976), Theorem 4 or Seneta (1981) Theorem 3.2.] A sufficient condition for allowable M_n [which also gives geometric decay of $\tau(M_{m,n})$ as $n \to \infty$, thereby satisfying condition (3) of Theorem 2] is that there exist some $n_0 \geq 1$ and $\gamma > 0$ such that for all n, $M_{n,\, n+n_0} > 0$ and

$$\frac{\min^{+}_{i,j} M_n(i,j)}{\max_{i,j} M_n(i,j)} \geq \gamma,$$

where $\min^{+}$ denotes the minimum over positive elements. [This is the well known Coale–Lopez theorem. See Seneta (1981) Theorem 3.3 for the form given here.] This condition also has consequences for strong ergodicity, as we will see below.

Say column-allowable matrices $\{M_n\}_{n=0}^{\infty}$ have the *RGR property* (for relative growth rates) if for all m, i and j there exists $r_m(i,j) \in [0, \infty]$ such that

$$\frac{{}^{m}R_n(i)}{{}^{m}R_n(j)} \to r_m(i,j) \text{ as } n \to \infty.$$

If the RGR property holds and $r_m(i,j) \in (0, \infty)$ for all m, i and j, then we say the $\{M_n\}$ have a *single growth rate*.

DEFINITION 6 (Strong ergodicity). The matrices $\{M_n\}$ are said to be strongly ergodic if for all m there exists a probability vector v_m such that

$$\frac{M_{m,n}(i,j)}{e_i^T M_{m,n} 1} \to v_m(j) \quad \text{as } n \to \infty \text{ independently of } i.$$

It follows immediately that if such v_m exist then they are in fact independent of m, that is, $v_m = v$ for all m.

Weak ergodicity can be thought of as requiring the columns of $M_{m,n}$ to tend to fixed proportions as $n \to \infty$, given by w_m. In an analogous manner, strong ergodicity is often thought of as requiring the rows of $M_{m,n}$ to tend to fixed proportions as $n \to \infty$, given by v. However, strong ergodicity only tells you about proportions with respect to the largest growth rate of $M_{m,n}$. Smaller growth rates, represented by the zeros in v, cannot be compared without extra information, such as that supplied by the RGR property.

The next lemma should give a better idea of how the concepts of weak and strong ergodicity are related.

PROPOSITION 7 (Relating weak and strong ergodicity).

(i) *If the* $\{M_n\}$ *are row allowable, then strong ergodicity with* $v > 0$ *implies weak ergodicity.*

(ii) *If the* $\{M_n\}$ *are allowable, then weak and strong ergodicity together imply* $M_{m,n}/1^T M_{m,n}1 \to w_m v^T$ *as* $n \to \infty$ *for all* m.

(iii) *For allowable* M_n, *if there exist probability vectors* $w_m > 0$ *and a probability vector* v *such that* $M_{m,n}/1^T M_{m,n}1 \to w_m v^T$ *as* $n \to \infty$ *for all* m, *then the* $\{M_n\}$ *are strongly ergodic.*

(iv) *If the* $\{M_n\}$ *are allowable, then strong ergodicity with* $v > 0$ *implies the RGR property, with* $r_m(i,j) = v(i)/v(j)$ (*thus giving a single growth rate*).

(v) *Weak ergodicity and the RGR property imply that for all* m, i *and* j, $r_m(i,j) = r(i,j)$ *is independent of* m *and they imply strong ergodicity, with* $v(j) = 1/\sum_i r(i,j)$.

PROOF. (i) It suffices to put $w_{m,n} = M_{m,n}1$ and $v_{m,n} = v$ in the first of our definitions of weak ergodicity. The $w_{m,n}$ are strictly positive provided the $\{M_n\}$ are row allowable.

(ii) Since the $\{M_n\}$ are weakly ergodic and allowable, we have that $\phi(M_{m,n}) \to 1$ as $n \to \infty$. Thus

$$\begin{aligned}\frac{M_{m,n}(i,j)\cdot 1^T M_{m,n}1}{e_i^T M_{m,n}1\cdot 1^T M_{m,n}e_j} &= \frac{M_{m,n}(i,j)\sum_{k,l}M_{m,n}(k,l)}{\sum_{k,l}M_{m,n}(i,k)M_{m,n}(l,j)}\\ &= \frac{M_{m,n}(i,j)\sum_{k,l}M_{m,n}(k,l)}{\displaystyle\sum_{k,l}\frac{M_{m,n}(i,k)M_{m,n}(l,j)}{M_{m,n}(l,k)M_{m,n}(i,j)}M_{m,n}(l,k)M_{m,n}(i,j)}\\ &\to 1 \quad \text{as } n \to \infty.\end{aligned}$$

Thus, dividing top and bottom of the left-hand side by $(1^T M_{m,n}1)^2$, we have

$$\frac{M_{m,n}(i,j)/1^T M_{m,n}1}{\left(e_i^T M_{m,n}1/1^T M_{m,n}1\right)\left(1^T M_{m,n}e_j/1^T M_{m,n}1\right)} \to 1 \quad \text{as } n \to \infty.$$

But $e_i^T M_{m,n}1/1^T A M_{m,n}1 \to w_m(i)$ and $1^T M_{m,n}e_j/1^T M_{m,n}1 \to v(j)$ and so $M_{m,n}(i,j)/1^T M_{m,n}1 \to w_m(i)v(j)$ as $n \to \infty$. Note that we do not in general need full weak ergodicity for this result to hold. All we need is that $M_{m,n}e_j/{}^m R_n(j) \to w_m$ for those j for which $v(j) > 0$. This is why we only get a partial converse to this result (see the next item).

(iii) It suffices to divide top and bottom of $M_{m,n}(i,j)/e_i^T M_{m,n}1$ by $1^T M_{m,n}1$ and send $n \to \infty$. Note that it also follows that $M_{m,n}e_j/{}^m R_n(j) \to w_m$ for all j for which $v(j) > 0$.

(iv) This follows from items (i) and (ii) on dividing top and bottom of ${}^m R_n(i)/{}^m R_n(j)$ by $1^T M_{m,n}1$. This argument fails if v has two or more zero elements, i_0 and i_1 say, as we do not know how quickly ${}^m R_n(i_0)/1^T M_{m,n}1$ and ${}^m R_n(i_1)/1^T M_{m,n}1$ go to zero. In particular we do not know if one of them tends to zero faster than the other or not.

(v) We observe to begin with that for all i

$$\frac{M_{m,n}(i,j)}{M_{m,n}(i,k)} = \frac{M_{m,n}(i,j)/{}^mR_n(j)}{M_{m,n}(i,k)/{}^mR_n(k)} \frac{{}^mR_n(j)}{{}^mR_n(k)}$$

$$\to \frac{w_m(i)}{w_m(i)} r_m(j,k) = r_m(j,k) \quad \text{as } n \to \infty.$$

Thus

$$r_m(i,j) = \lim_{n\to\infty} \frac{{}^mR_n(i)}{{}^mR_n(j)}$$

$$= \lim_{n\to\infty} \frac{\sum_l (\sum_k M_m(k,l)) M_{m+1,n}(l,i)}{\sum_l (\sum_k M_m(k,l)) M_{m+1,n}(l,j)}$$

$$\to r_{m+1}(i,j) \quad \text{as } n \to \infty$$

noting that for positive a_n, b_n, c_n and d_n, if $a_n/b_n \to x$ and $c_n/d_n \to x$ then $(a_n + c_n)/(b_n + d_n) \to x$. So $r_m(i,j)$ is independent of m. To show strong ergodicity, consider

$$\frac{M_{m,n}(i,j)}{\sum_k M_{m,n}(i,k)} = \frac{M_{m,n}(i,j)}{{}^mR_n(j)} \Bigg/ \sum_k \frac{M_{m,n}(i,k)}{{}^mR_n(k)} \frac{{}^mR_n(k)}{{}^mR_n(j)}$$

$$\to \frac{1}{\sum_k r(k,j)} \quad \text{as } n \to \infty.$$

Note that $1/\sum_k r(k,j) \le 1$ since $r(k,k) = 1$, and that $\sum_j (1/\sum_k r(k,j)) = 1$. □

The concept of strong ergodicity is a generalization of that used when dealing with (row) stochastic matrices. A sequence $\{A_n\}_{n=0}^{\infty}$ of stochastic matrices is strongly ergodic if for all m, i and j, $A_{m,n}(i,j) \to v(j)$ as $n \to \infty$ for some probability vector v. It is also usual when dealing with stochastic matrices to use *stochastic ergodicity* rather than weak ergodicity. The $\{A_n\}$ are stochastically ergodic if $A_{m,n}(i,k) - A_{m,n}(j,k) \to 0$ for all m, i, j and k as $n \to \infty$. Clearly strong ergodicity implies stochastic ergodicity (though it still does not imply weak ergodicity even in this setting). These definitions are generally sufficient in the stochastic setting, as the forward products $\{M_{m,n}\}_{n=m}^{\infty}$ are bounded and the question of growth rates is not particularly important. The extra concept of the RGR property is useful when you are interested in multiple growth rates, in particular, not just the largest growth rate.

A straightforward condition for strong ergodicity with $v > 0$ is the following. For allowable M_n, if there exists an $n_0 \ge 1$, a $\delta < 1$ and a probability vector $v > 0$ such that $\tau(M_{n,n+n_0}) \le \delta$ for all n and $x_n \to v$ for some sequence of M_n left eigenvectors $\{x_n\}$, then the $\{M_n\}$ are strongly ergodic with row limit vector v. [See Seneta and Sheridan (1981) Theorem 4.2.] This happens, for example, when the $\{M_n\}$ converge elementwise to some primitive matrix M, in which case v is the left Perron–Frobenius eigenvector of M. It is possible to say a little more in this case.

LEMMA 8 (Asymptotically primitive mean matrices). *If $M_n \to M$ elementwise where M is primitive, then the $\{M_n\}$ are strongly ergodic. Moreover, if we let $\lambda = PF(M)$ be the spectral radius of M, $v = LPF(M)$ be the left Perron–Frobenius eigenvector of M and $w = RPF(M)$ be the right Perron–Frobenius eigenvector of M (normed as probability vectors), then v is the row limit vector for the $\{M_n\}$, the column limit vectors $\{w_m\}$ converge elementwise to w as $m \to \infty$ and*

$$\lambda = \lim_{m\to\infty} \lim_{n\to\infty} \frac{1^T M_{m,n} 1}{1^T M_{m+1,n} 1} = \lim_{n\to\infty} \frac{1^T M_{m,n+1} 1}{1^T M_{m,n} 1} \quad \textit{for all } m.$$

PROOF. That the $\{M_n\}$ are strongly ergodic with row limit vector v follows from Theorem 4.2 of Seneta and Sheridan (1981). For the remainder consider the following:

$$\begin{aligned} w_m &= \lim_{n\to\infty} \frac{M_{m,n} 1}{1^T M_{m,n} 1} \\ &= M_m \lim_{n\to\infty} \frac{M_{m+1,n} 1}{1^T M_{m+1,n} 1} \frac{1^T M_{m+1,n} 1}{1^T M_{m,n} 1} \\ &= M_m w_{m+1} \lim_{n\to\infty} \frac{1^T M_{m+1,n} 1}{1^T M_{m,n} 1} \end{aligned}$$

which implies the existence of $\alpha_m^{m+1} := \lim_{n\to\infty} 1^T M_{m+1,n} 1 / 1^T M_{m,n} 1$ and shows that if $\lim_{m\to\infty} w_m$ exists then it must equal w. Moreover, if $\lim_{m\to\infty} w_m$ exists, then so does $\lim_{m\to\infty} \alpha_m^{m+1}$, which must equal λ^{-1}.

In fact, we can show that the limit of any convergent subsequence of the $\{w_m\}$ must be w. Let $\{w_{n(k)}\}_{k=0}^{\infty}$ be a convergent subsequence of the $\{w_n\}$ with limit x_0. Such a subsequence always exists, as the space of probability vectors is compact in $\mathbb{R}^d$. We have that $w_{n(k)} = M_{n(k)} w_{n(k)+1} \alpha_{n(k)}^{n(k)+1}$, whence sending $k \to \infty$,

$$x_0 = M x_1 \beta_0,$$

where the existence of $x_1 := \lim_{k\to\infty} w_{n(k)+1}$ and $\beta_0 := \lim_{k\to\infty} \alpha_{n(k)}^{n(k)+1}$ is implied by the existence of x_0 and M. That each limit exists separately follows from the fact that the $\{w_n\}$ are all probability vectors.

We can repeat this procedure with the $\{w_{n(k)+1}\}_{k=0}^{\infty}$ to show that $w_{n(k)+2}$ converges to some limit x_2. Repeating this ad infinitum gives us a sequence of probability vectors $x_0, x_1, x_2, \dots$ and a sequence of scalars $\beta_0, \beta_1, \beta_2, \dots$ such that $x_k = M x_{k+1} \beta_k$ for all k. That is, $x_0 = M^k x_k \prod_{l=0}^{k-1} \beta_l$ for all k. Let $\{x_{k(p)}\}_{p=0}^{\infty}$ be a convergent subsequence of the x_k with limit y. Then $x_0 = \lim_{p\to\infty} M^{k(p)} y \prod_{l=0}^{k(p)-1} \beta_l$. It follows immediately that $x_0 = w$, since the w coefficient of the eigenvalue expansion of y cannot be zero because $y \geq 0$.

Given that all convergent subsequences of the $\{w_m\}$ converge to w, it follows immediately that the $\{w_m\}$ themselves must converge to w, since $\{w_m\}_{m=0}^{\infty}$ is contained in the compact set of probability vectors.

The final part of the lemma follows directly from the following observation

$$\lim_{n\to\infty} \frac{M_{m,n}M_n}{1^T M_{m,n}1} = w_m v^T M = \lim_{n\to\infty} \frac{1^T M_{m+1,n}1}{1^T M_{m,n}1} w_m v^T. \qquad \square$$

By way of illustrating what is required to get more than one growth rate, we give the following. If the $\{M_n\}$ are strongly ergodic and irreducible, then the existence of some $\gamma > 0$ such that

$$\frac{\min^+_{i,j} M_n(i,j)}{\max_{i,j} M_n(i,j)} \ge \gamma \quad \text{for all } n$$

is sufficient to ensure that $v > 0$, that is, that there is a single growth rate. [See Seneta (1981), Theorem 3.4.] It is possible to adapt existing results on strong ergodicity with $v > 0$ to the multiple growth rate case by using rescaling arguments, as we will see in Section 2.1 below.

2.1. *Rescaling*. We will now take a closer look at rescaling in general and the rescaling matrices $\{{}^m R_n\}_{0\le m\le n}$ in particular.

LEMMA 9 (Rescaling lemma). *Let $D = \operatorname{diag}(D(1),\dots,D(d))$ be nonnegative and of full rank. Then for any nonnegative column allowable matrix $A \in \mathbb{R}^{d\times d}$,*

$$\tau(AD) = \tau(DA) = \tau(A)$$

and for allowable A

$$\phi(AD) = \phi(DA) = \phi(A).$$

PROOF. The result follows directly from the definitions of τ and ϕ. $\square$

Observe that for column-allowable column-stochastic matrices $\{P_n\}_{n=0}^{\infty}$, weak ergodicity is equivalent to strong ergodicity with $v = (1/d)1$. Now, define ${}^m P_n = {}^m R_n M_n {}^m R_{n+1}^{-1}$ for all $n \ge m \ge 0$ and suppose that the $\{M_n\}$ are column-allowable and weakly ergodic with column limit vectors $\{w_m\}$. Thus for any $m \le n \le p$, $\tau(M_{n,p}) \to 0$ as $p \to \infty$, so from Lemma 9, $\tau({}^m P_{n,p}) \to 0$ as $n \to \infty$. Thus as the ${}^m P_n$ are column-stochastic, they are strongly ergodic with $v = (1/d)1$. That is, there exist strictly positive probability vectors $\{{}^m w_n\}_{n=m}^{\infty}$ such that

$${}^m P_{n,p} \to {}^m w_n 1^T \quad \text{as } p \to \infty \text{ for all } n \ge m.$$

It is clear that ${}^m w_m = w_m$. More generally we have

$$\begin{aligned} {}^m w_n 1^T &= \lim_{p\to\infty} {}^m P_{n,p} \\ &= \lim_{p\to\infty} {}^m R_n M_{n,p} {}^n R_p^{-1} {}^n R_p {}^m R_p^{-1} \\ &= {}^m R_n w_n 1^T \lim_{p\to\infty} {}^n R_p {}^m R_p^{-1}. \end{aligned}$$

It follows that $\lim_{p\to\infty} {}^nR_p {}^mR_p^{-1}$ exists and equals $\alpha_m^n I$ for some constant α_m^n. Thus for $n \ge m$,

$$ {}^m w_n = \alpha_m^n {}^m R_n w_n. $$

It is easily checked that this definition of α_m^n is consistent with the definition given in Lemma 8, namely that $\alpha_m^n = \lim_{p\to\infty} 1^T M_{n,p} 1 / 1^T M_{m,p} 1$.

We can in fact bound the speed at which ${}^m P_{n,p}$ converges to ${}^m w_n 1^T$ as $p \to \infty$. To do this we make use of a second coefficient of ergodicity (or contraction coefficient), κ. It is normally used with row-stochastic matrices, but has been adapted here for use with column-stochastic matrices by the simple expedient of transposing everything. For a column-stochastic matrix $P \in \mathbb{R}^{d\times d}$ define

$$ \kappa(P) = \sup_{x\in\mathbb{R}^d,\, x\neq 0,\, 1^T x = 0} \frac{\|Px\|_1}{\|x\|_1}. $$

It can be shown that $\kappa(P) \le \tau(P)$ [Seneta (1981), Theorem 3.13]. Thus for a column stochastic P, if $x \in \mathbb{R}^d$ and $1^T x = 0$ then $\|Px\|_1 \le \|x\|_1 \tau(P)$. That is, if $\tau(P) < 1$ then P is a contraction mapping on the set $\{x \in \mathbb{R}^d\colon 1^T x = 0\}$. In particular we have here that $1^T(e_j - {}^m w_p) = 0$ and so

$$ \begin{aligned} \left\| {}^m P_{n,p}(\cdot, j) - {}^m w_n \right\|_1 &= \left\| {}^m P_{n,p}\left(e_j - {}^m w_p\right) \right\|_1 \\ &\le 2\tau\left({}^m P_{n,p}\right) \\ &= 2\tau\left(M_{n,p}\right). \end{aligned} $$

In practice we may be able to find natural rescaling matrices different from the $\{{}^m R_n\}$. Suppose that $\{D_n = \operatorname{diag}(D_n(1), \dots, D_n(d))\}_{n=0}^{\infty}$ is a sequence of rescaling matrices and put $Q_n = D_n M_n D_{n+1}^{-1}$ for all $n \ge 0$. It follows from the rescaling lemma that the $\{Q_n\}$ are weakly ergodic if and only if the $\{M_n\}$ are weakly ergodic. Suppose this is the case, and let $\{w_m\}_{m=0}^{\infty}$ and $\{\overline{w}_m\}_{m=0}^{\infty}$ be the column limit vectors for the $\{M_n\}$ and $\{Q_n\}$, respectively, then, noting that

$$ \overline{w}_m = \lim_{n\to\infty} \frac{Q_{m,n} e_j}{1^T Q_{m,n} e_j} = \lim_{n\to\infty} \frac{D_m M_{m,n} e_j}{1^T D_m M_{m,n} e_j} $$

and that

$$ \frac{1^T M_{m,n} e_j}{1^T D_m M_{m,n} e_j} = \frac{\sum_i D_m^{-1}(i) Q_{m,n}(i,j)}{1^T Q_{m,n} e_j} \to \sum_i D_m^{-1}(i)\overline{w}_m(i) \quad \text{as } n \to \infty, $$

it follows that

$$ w_m = \frac{D_m^{-1}\overline{w}_m}{1^T D_m^{-1}\overline{w}_m}. \tag{10} $$

For strong ergodicity we have the following.

PROPOSITION 10 (Sufficient condition for strong ergodicity). *If the $\{Q_n\}$ are strongly ergodic with row limit vector $\bar{v} > 0$ and $\lim_{n\to\infty} D_n(i)/D_n(j)$ exists*

$\in [0,\infty]$ *for all i and j, then the $\{M_n\}$ are strongly ergodic and have the RGR property, with*

$$r(i,j) = \frac{\bar{v}(i)}{\bar{v}(j)} \lim_{n\to\infty} \frac{D_n(i)}{D_n(j)}.$$

PROOF. As $\bar{v} > 0$, the $\{Q_n\}$ and thus the $\{M_n\}$ are weakly ergodic. Thus from Proposition 7(v), it suffices to establish the RGR property for $m = 0$. Consider

$$\begin{aligned}\lim_{n\to\infty} \frac{{}^0R_n(j)}{{}^0R_n(k)} &= \lim_{n\to\infty} \frac{\sum_i D_0^{-1}(i) Q_{0,n}(i,j)}{\sum_i D_0^{-1}(i) Q_{0,n}(i,k)} \frac{D_n(j)}{D_n(k)} \\ &= \frac{\bar{v}(j)}{\bar{v}(k)} \lim_{n\to\infty} \frac{D_n(j)}{D_n(k)}\end{aligned}$$

since $\lim_{n\to\infty} Q_{0,n}(i,j)/Q_{0,n}(i,k) = \bar{v}(j)/\bar{v}(k)$ independently of i. $\square$

This proposition provides a practical way of applying our existing conditions for strong ergodicity with $v > 0$ to situations where we have multiple growth rates. We also have the following (for use in Corollary 16 below).

PROPOSITION 11 (Rescaled limit matrices). *If the $\{Q_n\}$ converge elementwise to a primitive matrix Q then for all m, ${}^mP_n$ converges elementwise to a primitive matrix P given by*

$$\bar{\lambda} P(i,j) = \bar{v}(i) Q(i,j)/\bar{v}(j)$$

where $\bar{\lambda} = PF(Q)$ is the spectral radius of Q and $\bar{v} = LPF(Q)$ is the left Perron–Frobenius eigenvector of Q (normed as a probability vector).

PROOF. To begin with put ${}^mD_n = D_m^{-1} D_n$ and ${}^mQ_n = {}^mD_n M_n {}^mD_{n+1}^{-1}$ for all $n \ge m$. Then ${}^mQ_n \to {}^mQ := D_m^{-1} Q D_m$ as $n \to \infty$ where $\mathrm{PF}({}^mQ) = \bar{\lambda}$ and ${}^m\bar{v} := \mathrm{LPF}({}^mQ) = D_m \bar{v} / \sum_i D_m(i) \bar{v}(i)$. Now, define a further set of rescaling matrices $\{{}^mE_n\}_{0\le m\le n}$ by putting ${}^mE_m = I$ and requiring ${}^mE_n {}^mQ_n {}^mE_{n+1}^{-1}$ to be column stochastic for all $n \ge m$ (so the $\{{}^mE_n\}$ play the same role for the $\{{}^mQ_n\}$ that the $\{{}^mR_n\}$ play for the $\{M_n\}$). It should be clear that ${}^mE_n {}^mQ_n {}^mE_{n+1}^{-1} = {}^mP_n$, that is, that ${}^mE_n = {}^mR_n {}^mD_n^{-1}$.

Now, from Lemma 8, we have that

$$\frac{1^T\, {}^mE_n}{1^T\, {}^mQ_{m,n} 1} = \frac{1^T\, {}^mQ_{m,n}}{1^T\, {}^mQ_{m,n} 1} \to {}^m\bar{v} \quad \text{as } n \to \infty$$

and that

$$\frac{1^T\, {}^mQ_{m,n} 1}{1^T\, {}^mQ_{m,n+1} 1} \to \frac{1}{\bar{\lambda}} \quad \text{as } n \to \infty,$$

whence

$$\begin{aligned}\bar{\lambda}^m P_n(i,j) &= \bar{\lambda}\frac{1^{T\,m}Q_{m,n}1}{1^{T\,m}Q_{m,n+1}1}\frac{{}^mE_n(i)}{1^{T\,m}Q_{m,n}1}{}^mQ_n(i,j)\frac{1^{T\,m}Q_{m,n+1}1}{{}^mE_{n+1}(j)}\\ &\to \frac{{}^m\bar{v}(i){}^mQ(i,j)}{{}^m\bar{v}(j)} \quad \text{as } n\to\infty\\ &= \frac{\bar{v}(i)Q(i,j)}{\bar{v}(j)}. \qquad \square\end{aligned}$$

2.2. *Growth rates.* In this subsection we compare and bound various growth rates obtained from the matrix product $M_{m,n}$ as $n \to \infty$, with the aim of simplifying the application of Theorems 1 and 2 and Proposition 4. The results obtained will be put to practical use in Section 4.

For two sequences $\{x_n\}_{n=0}^{\infty}$ and $\{y_n\}_{n=0}^{\infty}$, write $\{x_n\} \equiv \{y_n\}$ if $\lim_{n\to\infty} x_n/y_n$ exists $\in (0,\infty)$. We say two such sequences have the same growth rate.

LEMMA 12. *Suppose we are given rescaling matrices $D_n = \operatorname{diag}(D_n(1), \dots, D_n(d))$ for $n \geq 0$ such that the matrices $\{Q_n := D_n M_n D_{n+1}^{-1}\}_{n=0}^{\infty}$ are strongly ergodic with row limit vector $\bar{v} > 0$. Then for any $m \geq 0$ and $1 \leq j \leq d$*

$$\{{}^mR_n(j)\}_{n=m}^{\infty} \equiv \{1^TQ_{m,n}1\cdot D_n(j)\}_{n=m}^{\infty}.$$

PROOF. We have ${}^mR_n(j) = 1^TD_m^{-1}Q_{m,n}e_j\cdot D_n(j)$. From Proposition 7 we know that $Q_{m,n}(i,j)/1^TQ_{m,n}1$ converges as $n\to\infty$ to $\overline{w}_m(i)\bar{v}(j)$, where the $\{\overline{w}_m\}$ are the column limit vectors for the $\{Q_n\}$. Thus

$$\frac{1^TD_m^{-1}Q_{m,n}e_j}{1^TQ_{m,n}1} = \frac{\sum_i D_m^{-1}(i)Q_{m,n}(i,j)}{1^TQ_{m,n}1} \to \sum_i D_m^{-1}(i)\overline{w}_m(i)\cdot\bar{v}(j) \quad \text{as } n\to\infty,$$

whence we get the result. $\square$

Under the conditions of Lemma 8 we can give bounds on the growth of $1^TM_{m,n}1$ or (more commonly) $1^TQ_{m,n}1$, as $n\to\infty$.

LEMMA 13 (Growth bounds). *If $M_n \to M$ elementwise where M is primitive, then for any $\varepsilon > 0$ we can find a constant $c_0 \in (0,\infty)$ such that for all $n \geq m \geq 0$,*

$$c_0^{-1}(\lambda-\varepsilon)^{n-m} \leq 1^TM_{m,n}1 \leq c_0(\lambda+\varepsilon)^{n-m}$$

where $\lambda = PF(M)$ is the spectral radius of M.

PROOF. From Lemma 8 we have for any $k \geq 0$ that

$$\lim_{m\to\infty}\alpha_{m+k}^m := \lim_{m\to\infty}\lim_{n\to\infty} 1^TM_{m,n}1/1^TM_{m+k,n}1 = \lambda^k.$$

Thus for any $\varepsilon > 0$ we can find some constant $c_1 \in (0, \infty)$ such that for all $m, k \geq 0$,

$$c_1^{-1}(\lambda - \varepsilon)^k \leq \alpha_{m+k}^m \leq c_1(\lambda + \varepsilon)^k. \tag{11}$$

Now since M is primitive and $M_n \to M$, there exists an m_0 and a $c_2 \in (0, \infty)$ such that for all $m \geq m_0$, $c_2^{-1}11^T \leq M_{m, m+n_0} \leq c_2 11^T$, where n_0 is such that $M^{n_0} > 0$. Thus for all $n \geq m \geq m_0$ and $p \geq n + n_0$,

$$c_2^{-1}1^T M_{m,p}1 \leq 1^T M_{m,n}11^T M_{n+n_0,p}1 \leq c_2 1^T M_{m,p}1.$$

Dividing through by $1^T M_{n+n_0,p}1$ and sending $p \to \infty$ gives us

$$c_2^{-1}\alpha_{n+n_0}^m \leq 1^T M_{m,n}1 \leq c_2\alpha_{n+n_0}^m.$$

The result now follows on applying inequalities (11) to these. $\square$

The next result gives a condition for the growth rate of $1^T M_{m,n}1$ to equal λ exactly.

LEMMA 14 (Asymptotically geometric growth function). *If $M_n \to M$ elementwise where M is primitive and $\sum_{n=0}^\infty \|M_n - M\|_1 < \infty$, then for any $m \geq 0$,*

$$\{1^T M_{m,n}1\}_{n=m}^\infty \equiv \{\lambda^{n-m}\}_{n=m}^\infty \equiv \{\textstyle\prod_{k=m}^{n-1}\lambda_k\}_{n=m}^\infty$$

where $\lambda = PF(M)$ is the spectral radius of M and $\lambda_k = PF(M_k)$ is the spectral radius of M_k, for all $k \geq m$.

PROOF. Put $\lambda_{m,n} = \prod_{k=m}^{n-1}\lambda_k$. Also, recall that for any matrix $A \in \mathbb{R}^{d\times d}$, the $\mathscr{L}^1$ operator norm $\|A\|_1$ is equivalent to the matrix norm $\|A\| := \max_{i,j}|A(i,j)|$. The following result is taken from Markus and Minc (1964), Theorem 3.1.6. For $A, B \in \mathbb{R}^{d\times d}$, $B \geq A \geq 0$, A, B primitive with $\mathrm{PF}(A) = \alpha$, $\mathrm{PF}(B) = \beta$ we have

$$\frac{m}{\max_{i,j} C(i,j)/\sum_k C(k,j)} \leq \beta - \alpha \leq \frac{M}{\min_{i,j} C(i,j)/\sum_k C(k,j)}$$

where $m = \min_{i,j}(B(i,j) - A(i,j))$, $M = \max_{i,j}(B(i,j) - A(i,j))$ and $C \geq 0$ is column allowable and commutes with either A or B.

Suppose $M_n \uparrow M$. Let $C = M^{n_0}$ where n_0 is such that $M^{n_0} > 0$ and put $c_1^{-1} = \min_{i,j} C(i,j)/\sum_k C(k,j)$. Then from the above result, $\lambda - \lambda_n \leq c_1\|M_n - M\|$. Thus $\sum_n\|M_n - M\| < \infty$ implies $\sum_n(\lambda - \lambda_n) < \infty$ which implies that $\lambda_{m,n}/\lambda^{n-m}$ converges $\in (0,\infty)$ as $n \to \infty$, that is, that $\{\lambda_{m,n}\}_{n=m}^\infty \equiv \{\lambda^{n-m}\}_{n=m}^\infty$.

The same holds if $M_n \downarrow M$. If $M_n \to M$ but not monotonically, then consider the following elementwise minima and maxima:

$$M_n^- = M_n \wedge M \quad \text{and} \quad M_n^+ = M_n \vee M.$$

Clearly $\|M_n - M\| = \|M_n^- - M\| \vee \|M_n^+ - M\|$, so $\sum_n\|M_n^- - M\| < \infty$ and $\sum_n\|M_n^+ - M\| < \infty$. Putting $\lambda_n^- = \mathrm{PF}(M_n^-) \leq \lambda_n \leq \lambda_n^+ = \mathrm{PF}(M_n^+)$, we get $\{\lambda_{m,n}^-\}_{n=m}^\infty \equiv \{\lambda_{m,n}^+\}_{n=m}^\infty \equiv \{\lambda^{n-m}\}_{n=m}^\infty$, whence $\{\lambda_{m,n}\}_{n=m}^\infty \equiv \{\lambda^{n-m}\}_{n=m}^\infty$.

Now

$$\begin{aligned}\lambda_{m,n}^{-1} M_{m,n} &- \lambda^{m-n} M^{n-m} \\ &= \sum_{k=m}^{n-1} \lambda_{m,k}^{-1} M_{m,k} \left(\lambda_k^{-1} M_k - \lambda^{-1} M \right) \lambda^{(k+1)-n} M^{n-(k+1)}\end{aligned}$$

(a telescoping sum). Thus, since $\|M_n\|_1 \le \lambda_n$, $\sum_{n=0}^{\infty} \|M_n - M\|_1 < \infty$ and $\{\lambda_{m,n}\}_{n=m}^{\infty} \equiv \{\lambda^{n-m}\}_{n=m}^{\infty}$,

$$\begin{aligned}\|\lambda_{m,n}^{-1} M_{m,n} &- \lambda^{m-n} M^{n-m}\|_1 \\ &\le \sum_{k=m}^{n-1} \|\lambda_k^{-1} M_k - \lambda^{-1} M\|_1 \\ &\le \sum_{k=m}^{\infty} \left(\lambda^{-1} \|M_k - M\|_1 + |\lambda_k^{-1} - \lambda^{-1}| \|M_k\|_1 \right) \\ &\to 0 \quad \text{as } m \to \infty.\end{aligned}$$

Finally, if we let $v = \mathrm{LPF}(M)$ and $w = \mathrm{RPF}(M)$ then since

$$\lim_{n\to\infty} \lambda^{m-n} M^{n-m} = wv^T / w^T v,$$

we have that

$$\begin{aligned}\lim_{m\to\infty} \lim_{n\to\infty} M_{m,n} / 1^T M_{m,n} 1 &= \lim_{m\to\infty} w_m v^T \\ &= wv^T \\ &= w^T v \lim_{m\to\infty} \lim_{n\to\infty} \lambda_{m,n}^{-1} M_{m,n}\end{aligned}$$

and so $\{1^T M_{m,n} 1\}_{n=m}^{\infty} \equiv \{\lambda_{m,n}\}_{n=m}^{\infty} \equiv \{\lambda^{n-m}\}_{n=m}^{\infty}$. $\square$

Our final result for this section gives conditions for the uniform equivalence of some particular growth rates.

LEMMA 15. *Suppose we are given rescaling matrices* $D_n = \mathrm{diag}(D_n(1), \ldots, D_n(d))$ *for all* $n \ge 0$, *such that the matrices* $\{Q_n := D_n M_n D_{n+1}^{-1}\}_{n=0}^{\infty}$ *converge elementwise to a primitive matrix* Q. *If we let* $\{w_m\}_{m=0}^{\infty}$ *be the column limit vectors for the* $\{M_n\}$ *and let* $\{\overline{w}_m\}_{m=0}^{\infty}$ *be the column limit vectors and* $\overline{v}$ *the row limit vector for the* $\{Q_n\}$, *then for any* $1 \le i, j \le d$,

$$\frac{Q_{m,n}(i,j)}{1^T Q_{m,n} 1} \to \overline{w}_m(i) \overline{v}(j) \quad \text{as } n \to \infty \text{ uniformly in } m \ge 0, \tag{12}$$

whence

$$\frac{M_{m,n}(i,j)}{w_m(i)\, {}^m R_n(j)} \to 1 \quad \text{as } n \to \infty \text{ uniformly in } m \ge 0. \tag{13}$$

and

$$\frac{{}^m R_n(j)}{1^T D_m^{-1} \overline{w}_m \cdot 1^T Q_{m,n} 1 \cdot D_n(j) \overline{v}(j)} \to 1 \quad \text{as } n \to \infty \text{ uniformly in } m \ge 0. \tag{14}$$

PROOF. Note to begin with that from Section 2.1, we have that for any $0 \le m \le n$ and $1 \le j \le d$, $\|Q_{m,n}e_j/1^T Q_{m,n}e_j - \overline{w}_m\| \le 2\tau(Q_{m,n})$. Since $Q_n \to Q$, we can find constants $c_0 > 0$ and $\delta_0 < 1$ such that $\tau(Q_{m,n}) \le c_0 \delta_0^{n-m}$ for all $0 \le m \le n$. Thus, $Q_{m,n}e_j/1^T Q_{m,n}e_j \to \overline{w}_m$ as $n \to \infty$ uniformly in m (and geometrically fast). So, (12) will follow if we can show that the convergence $1^T Q_{m,n}e_j/1^T Q_{m,n}1 \to \overline{v}(j)$ as $n \to \infty$ is also uniform in m.

From Lemma 3.10 of Seneta (1981) (on uniform strong ergodicity) we can find constants $c_1 > 0$ and $\delta_1 < 1$ such that for any $0 \le m \le n$,

$$\rho\left(1^T Q_{m,n}/1^T Q_{m,n}1, 1^T Q_{0,n}/1^T Q_{0,n}1\right) \le c_1 \delta_1^{n-m}.$$

Thus since $\rho(1^T Q_{0,n}/1^T Q_{0,n}1, \overline{v}^T) \to 0$, we have that

$$\rho\left(1^T Q_{m,n}/1^T Q_{m,n}1, \overline{v}^T\right) \to 0 \quad \text{as } n \to \infty \text{ uniformly in } m \ge 0.$$

Moreover, as $\overline{v} > 0$ and $Q_n \to Q$, it is clear that $\Omega := \mathrm{cl}(\{1^T Q_{m,n}/1^T Q_{m,n}1\colon 0 \le m \le n\})$ is such that $\Omega \cap \delta\mathbb{R}_+^d = \varnothing$ (where δ indicates the set boundary). Thus, if Θ is the set of probability vectors in $\mathbb{R}^d$, $\rho(\cdot,\cdot)$ is equivalent to $\|\cdot - \cdot\|_1$ on $\Theta \cap \Omega$, and we have that $1^T Q_{m,n}/1^T Q_{m,n}1 \to \overline{v}$ as $n \to \infty$ uniformly in m, as required.

Now, from (10) we have that

$$\frac{M_{m,n}(j,k)}{w_m(j)^m R_n(k)} = \frac{\sum_i D_m^{-1}(i)\overline{w}_m(i)/\overline{w}_m(j)}{\sum_i D_m^{-1}(i) Q_{m,n}(i,k)/Q_{m,n}(j,k)}.$$

Thus (13) follows from (12) and the fact that $\overline{w}_m(i)/\overline{w}_m(j) \to \overline{w}(i)/\overline{w}(j) > 0$. Again from (10) we have that

$$\frac{{}^m R_n(j)}{1^T D_m^{-1}\overline{w}_m \cdot 1^T Q_{m,n}1 \cdot D_n(j)\overline{v}(j)} = \frac{\sum_i D_m^{-1}(i) Q_{m,n}(i,j)/1^T Q_{m,n}1}{\sum_i D_m^{-1}(i)\overline{w}_m(i)\overline{v}(j)}.$$

Thus (14) follows from (12) and the fact that $\overline{w}_m(i)\overline{v}(j) \to \overline{w}(i)\overline{v}(j) > 0$. □

3. Proofs of the theorems.

PROOF OF THEOREM 1. We are given that the $\{M_n\}$ are allowable and weakly ergodic with column limit vectors $\{w_m\}$ and that

$$\sum_{n=m}^{\infty} \sum_{i,j} \frac{{}^m R_n(i)\sigma_n^2(i,j)}{{}^m R_{n+1}^2(j)} < \infty$$

[condition (1)]. We will show that $Z_{m,n}{}^m R_n^{-1}$ converges in $\mathscr{L}^2$ to some $L_m 1^T$ where $\mathrm{E}\, L_m = w_m$. We use a straightforward Cauchy convergence condition.

Consider for any $0 \le m \le n \le p$,

$$\begin{aligned}&\mathrm{E}\left(e_i^T Z_{m,n}{}^m R_n^{-1} - e_i^T Z_{m,p}{}^m R_p^{-1}\right)^T \left(e_i^T Z_{m,n}{}^m R_n^{-1} - e_i^T Z_{m,p}{}^m R_p^{-1}\right)\\ &\quad = C_{m,n,n}[i] - C_{m,n,p}[i] - C_{m,p,n}[i] + C_{m,p,p}[i],\end{aligned}$$

where

$$
\begin{aligned}
C_{m,n,p}[i] &= \mathrm{E}\,{}^mR_n^{-1}Z_{m,n}^T e_i e_i^T Z_{m,p}\,{}^mR_p^{-1} \\
&= {}^mR_n^{-1}M_{m,n}^T e_i e_i^T M_{m,p}\,{}^mR_p^{-1} \\
&\quad + \sum_{k=m}^{(n\wedge p)-1} {}^mR_n^{-1}M_{k+1,n}^T\left(\sum_{j=1}^{d} V_k[j]\cdot M_{m,k}(i,j)\right)M_{k+1,p}\,{}^mR_p^{-1}.
\end{aligned}
$$

It follows (given the $\{M_n\}$ are allowable and weakly ergodic) that $e_i^T Z_{m,n}\,{}^mR_n^{-1}$ converges in $\mathscr{L}^2$ if and only if $\lim_{n,p\to\infty} C_{m,n,p}[i]$ is finite. From Section 2.1 we have that

$$
\begin{aligned}
M_{k+1,n}\,{}^mR_n^{-1} &= {}^mR_{k+1}^{-1}\,{}^mP_{k+1,n} \\
&\to {}^mR_{k+1}^{-1}\,{}^mw_{k+1}1^T \text{ as } n\to\infty.
\end{aligned}
$$

Thus our normed process converges in $\mathscr{L}^2$ if and only if

$$
\sum_{k=m}^{\infty} 1\,{}^mw_{k+1}^T\,{}^mR_{k+1}^{-1}\left(\sum_{j=1}^{d} V_k[j]\cdot M_{m,k}(i,j)\right){}^mR_{k+1}^{-1}\,{}^mw_{k+1}1^T < \infty \tag{15}
$$

which is certainly the case if

$$
\sum_{k=m}^{\infty} {}^mR_{k+1}^{-1}V_k[j]\,{}^mR_{k+1}^{-1}\cdot M_{m,k}(i,j) < \infty \quad \text{for all } j. \tag{16}
$$

This condition is also necessary if the $\{{}^mw_{k+1}\}_{k=m}^{\infty}$ are uniformly bounded away from 0. Since $M_{m,k}(i,j)/{}^mR_k(j) \to w_m(i) > 0$, (16) holds if and only if

$$
\sum_{k=m}^{\infty} {}^mR_{k+1}^{-1}V_k[j]\,{}^mR_{k+1}^{-1}\cdot{}^mR_k(j) < \infty \quad \text{for all } j \tag{17}
$$

which is independent of i. We can rewrite (17) as

$$
\sum_{k=m}^{\infty} \frac{{}^mR_k(x)V_k[x](y,z)}{{}^mR_{k=1}(y)\,{}^mR_{k+1}(z)} < \infty \quad \text{for all } x, y \text{ and } z
$$

which is equivalent to

$$
\sum_{k=m}^{\infty} \frac{{}^mR_k(x)V_k[x](y,y)}{{}^mR_{k+1}^2(y)} < \infty \quad \text{for all } x \text{ and } y. \tag{18}
$$

To see this note for positive sequences $\{a_n\}$ and $\{b_n\}$, $\sum_n a_n < \infty$ and $\sum_n b_n < \infty$ together imply $\sum_n \sqrt{a_n b_n} < \infty$ (Schwarz inequality) and that

$$
\frac{{}^mR_k(x)V_k[x](y,z)}{{}^mR_{k+1}(y)\,{}^mR_{k+1}(z)} \le \sqrt{\frac{{}^mR_k(x)V_k[x](y,y)}{{}^mR_{k+1}^2(y)}}\sqrt{\frac{{}^mR_k(x)V_k[x](z,z)}{{}^mR_{k+1}^2(z)}}.
$$

Since $V_k[x](y,y) = \sigma_k^2(x,y)$, condition (18) is just condition (1) and as this is independent of i, we have proved the $\mathscr{L}^2$ convergence of $Z_{m,n}\,{}^mR_n^{-1}$ as $n\to\infty$ to some limit, call it W_m.

Since $Z_{m,n}{}^m R_n^{-1} \to_{\mathscr{L}^2} W_m$, we have that

$$\mathrm{E}\, W_m = \lim_{n\to\infty} \mathrm{E}\, Z_{m,n}{}^m R_n^{-1} = w_m 1^T$$

and

$$\begin{aligned}\operatorname{Cov} W_m(i,\cdot) &= \lim_{n\to\infty} \operatorname{Cov} Z_{m,n}(i,\cdot)^m R_n^{-1}\\ &= \sum_{k=m}^{\infty} 1^m w_{k+1}^T {}^m R_{k+1}^{-1}\left(\sum_{j=1}^{d} V_k[j]\cdot M_{m,k}(i,j)\right){}^m R_{k+1}^{-1}{}^m w_{k+1} 1^T.\end{aligned}$$

The fact that $\operatorname{Cov} W_m(i,\cdot) = c_0 1 1^T$ and $\mathrm{E}\, W_m(i,\cdot) = c_1 1^T$ for some constants $c_0 = c_0(m)$ and $c_1 = c_1(m)$ is enough to give us that $W_m(i,\cdot) = L_m(i)1^T$ for some real valued $L_m(i)$, as it implies that $\mathrm{E}(W_m(i,j) - W_m(i,k))^2 = 0$ for all j and k. □

COROLLARY 16 (Best possible variance condition). *If there exist diagonal scaling matrices $\{D_n\}_{n=0}^{\infty}$ such that $Q_n := D_n M_n D_{n+1}^{-1}$ converges to some primitive matrix Q, then the condition* (1) *is necessary and sufficient for the $\mathscr{L}^2$ convergence of $Z_{m,n}{}^m R_n^{-1}$ as $n \to \infty$.*

PROOF. From Proposition 11 we have that ${}^m P_n := {}^m R_n M_n {}^m R_{n+1}^{-1}$ converges elementwise as $n \to \infty$ to a primitive matrix P given by $\bar{\lambda} P(i,j) = \bar{v}(i)Q(i,j)/\bar{v}(j)$ where $\bar{\lambda} = \mathrm{PF}(Q)$ and $\bar{v} = \mathrm{LPF}(Q)$. It follows from Lemma 8 that as $n \to \infty$, ${}^m w_n$ converges to some $w := \mathrm{LPF}(P) > 0$. Thus the $\{{}^m w_n\}_{0\le m\le n}$ are uniformly bounded away from 0 and conditions (15) and (16) in the above proof of Theorem 1 are equivalent. □

Before proving the second of our main theorems (on a.s. convergence), it is worth noting that there is a natural martingale present. Recall from Section 2.1 that for any $m \le n$, $\lim_{p\to\infty} {}^n R_p {}^m R_p^{-1}$ exists and equals $\alpha_m^n I$. It follows that

$$w_m 1^T = \lim_{p\to\infty} M_{m,p} {}^m R_p^{-1} = M_{m,n} \lim_{p\to\infty} M_{n,p} {}^m R_p^{-1} = \alpha_m^n M_{m,n} w_n 1^T.$$

Let $\mathscr{F}_{m,n}$ be the σ-field generated by $\{Z_{m,k}\}_{k=m}^{n}$, then

$$\mathrm{E}\big(\alpha_m^{n+1} Z_{m,n+1} w_{n+1} | \mathscr{F}_{m,n}\big) = \alpha_m^n Z_{m,n} w_n.$$

That is, $\{\alpha_m^n Z_{m,n} w_n\}_{n=m}^{\infty}$ is a martingale with respect to the filtration $\{\mathscr{F}_{m,n}\}_{n=m}^{\infty}$. In the terminology of Cohn and Nerman (1990), $\{\alpha_m^n w_n\}_{n=m}^{\infty}$ is a harmonic sequence for the matrices $\{M_n\}_{n=m}^{\infty}$. Unfortunately, even given $\mathscr{L}^2$ convergence, the a.s. convergence of a particular linear combination of the $Z_{m,n}(i,\cdot)$ is not sufficient to give the a.s. convergence of the random vector itself.

PROOF OF THEOREM 2. We are given that the $\{M_n\}$ are weakly ergodic, that

$$\sum_{n=m}^{\infty} \sum_{i,j} \frac{(n+1-m)^m R_n(i)\sigma_n^2(i,j)}{{}^m R_{n+1}^2(j)} < \infty$$

[condition (2)] and that there exists a $C < \infty$ such that for all $n \geq m$,

$$\sum_{p=n}^{\infty} \tau(M_{n,p})^2 \leq (n+1-m)C$$

[condition (3)]. Note that weak ergodicity is in fact implied by (3), since it implies that $\tau(M_{m,n}) \to 0$ as $n \to \infty$. We will show that $Z_{m,n}{}^m R_n^{-1}$ converges almost surely as $n \to \infty$ to $L_m 1^T$, using a Borel–Cantelli type argument.

The first Borel–Cantelli lemma gives the a.s. convergence of

$$Z_{m,n}(i,j)/{}^m R_n(j)$$

to $L_m(i)$ if $\sum_{n=m}^{\infty} \mathrm{P}(|Z_{m,n}(i,j)/{}^m R_n(j) - L_m(i)| > \varepsilon) < \infty$ for all $\varepsilon > 0$. Applying Chebyshev's inequality, it is sufficient for this that

$$\sum_{n=m}^{\infty} \mathrm{E}\big(Z_{m,n}(i,j)/{}^m R_n(j) - L_m(i)\big)^2 < \infty.$$

It is easily checked that

$$\begin{aligned}
&\mathrm{E}\big(Z_{m,n}(i,j)/{}^m R_n(j) - L_m(i)\big)^2 \\
&\quad = \left(\frac{M_{m,n}(i,j)}{{}^m R_n(j)} - w_m(i)\right)^2 + \sum_{k=m}^{n-1} \left(\frac{M_{k+1,n}(\cdot,j)}{{}^m R_n(j)} - {}^m R_{k+1}^{-1}\,{}^m w_{k+1}\right)^T \\
&\qquad\qquad \times \left(\sum_{l=1}^{d} V_k[l] \cdot M_{m,k}(i,l)\right)\left(\frac{M_{k+1,n}(\cdot,j)}{{}^m R_n(j)} - {}^m R_{k+1}^{-1}\,{}^m w_{k+1}\right) \\
&\qquad + \sum_{k=n}^{\infty} {}^m w_{k+1}^T\,{}^m R_{k+1}^{-1}\left(\sum_{l=1}^{d} V_k[l] \cdot M_{m,k}(i,l)\right){}^m R_{k+1}^{-1}\,{}^m w_{k+1}.
\end{aligned}$$

To show the sum of these converges we need to know first, how fast ${}^m R_{k+1} M_{k+1,n}(\cdot,j)/{}^m R_n(j)$ converges to ${}^m w_{k+1}$ as $n \to \infty$, and second, how fast ${}^m R_{k+1}^{-1}(\sum_{l=1}^{d} V_k[l] \cdot M_{m,k}(i,l)){}^m R_{k+1}^{-1}$ converges to 0 as $k \to \infty$.

From Section 2.1, we have that for $m \leq n \leq p$ and $1 \leq j \leq d$,

$$\left\|{}^m R_n M_{n,p}(\cdot,j)/{}^m R_p(j) - {}^m w_n\right\| = \left\|{}^m P_{n,p}\big(e_j - {}^m w_p\big)\right\| \leq 2\tau(M_{n,p}).$$

Applying this above we get

$$\begin{aligned}
&\mathrm{E}\big(Z_{m,n}(i,j)/{}^m R_n(j) - L_m(i)\big)^2 \\
&\quad \leq 4\tau(M_{m,n})^2 \\
&\qquad + \sum_{k=m}^{n-1} 4\tau(M_{k+1,n})^2 1^T\,{}^m R_{k+1}^{-1}\left(\sum_{l=1}^{d} V_k[l] \cdot M_{m,k}(i,l)\right){}^m R_{k+1}^{-1} 1 \\
&\qquad + \sum_{k=n}^{\infty} {}^m w_{k+1}^T\,{}^m R_{k+1}^{-1}\left(\sum_{l=1}^{d} V_k[l] \cdot M_{m,k}(i,l)\right){}^m R_{k+1}^{-1}\,{}^m w_{k+1}.
\end{aligned}$$

We deal with each term separately.

Putting $n = m$, condition (3) implies that $\sum_{n=m}^{\infty} \tau(M_{m,n})^2 < \infty$. This takes care of the first term.

Summing the third term over $n \geq m$, we get

$$\sum_{k=m}^{\infty} (k + 1 - m)^m w_{k+1}^T {}^m R_{k+1}^{-1} \left(\sum_{l=1}^{d} V_k[l] \cdot M_{m,k}(i,l) \right) {}^m R_{k+1}^{-1} {}^m w_{k+1}.$$

Using the same argument that was used in the proof of Theorem 1, this is finite if condition (2) holds.

Summing the second term over $n \geq m$ we get

$$\sum_{k=m}^{\infty} 1^T {}^m R_{k+1}^{-1} \left(\sum_{l=1}^{d} V_k[l] \cdot M_{m,k}(i,l) \right) {}^m R_{k+1}^{-1} 1 \sum_{n=k+1}^{\infty} \tau(M_{k+1,n})^2.$$

From condition (3), $\sum_{n=k+1}^{\infty} \tau(M_{k+1,n})^2 \leq (k + 2 - m)C$. Now apply condition (2) as was done for the third term, to show that the second term summed over $n \geq m$ is also finite. □

PROOF OF COROLLARY 3. We have rescaling matrices $D_n = \operatorname{diag}(D_n(1), \dots, D_n(d))$ for all $n \geq 0$, such that the matrices $\{Q_n := D_n M_n D_{n+1}^{-1}\}_{n=0}^{\infty}$ converge elementwise to some primitive matrix Q. Let $\bar{v} = \operatorname{LPF}(Q)$ be the row limit vector and $\{\bar{w}_m\}_{m=0}^{\infty}$ the column limit vectors for the $\{Q_n\}$, so that $\bar{w}_m \to \bar{w} = \operatorname{RPF}(Q)$.

We show to begin with that

$$\sum_{n=m}^{\infty} \sum_{i,j} \frac{D_n(i) \sigma_n^2(i,j)}{1^T Q_{m,n} 1 D_{n+1}^2(j)} < \infty$$

[condition (4)] is necessary and sufficient for the $\mathscr{L}^2$ convergence of $Z_{m,n} D_n^{-1} / 1^T Q_{m,n} 1$ as $n \to \infty$ for all fixed $m \geq 0$.

From Lemmas 8 and 12 it is clear that (4) is in fact equivalent to (1), whence we have from Corollary 16 that (4) is necessary and sufficient for the $\mathscr{L}^2$ convergence of $Z_{m,n} {}^m R_n^{-1}$ as $n \to \infty$, to some $L_m 1^T$ with $\mathrm{E}\, L_m = w_m$. Thus, as ${}^m R_n(j) / 1^T Q_{m,n} 1 D_n(j) \to 1^T D_m^{-1} \bar{w}_m \cdot \bar{v}(j)$ as $n \to \infty$ (from the proof of Lemma 12), putting $\bar{L}_m = L_m \cdot 1^T D_m^{-1} \bar{w}_m$, we get from (10) that as $n \to \infty$,

$$Z_{m,n} D_n^{-1} / 1^T Q_{m,n} 1 \to_{\mathscr{L}^2} \bar{L}_m \bar{v}^T,$$

where $\mathrm{E}\, \bar{L}_m = D_m^{-1} \bar{w}_m$.

For a.s. convergence, we have from Lemmas 8 and 12 that condition (2) is equivalent to

$$\sum_{n=m}^{\infty} \sum_{i,j} \frac{(n + 1 - m) D_n(i) \sigma_n^2(i,j)}{1^T Q_{m,n} 1 D_{n+1}^2(j)} < \infty$$

[condition (5)]. Also, because we have geometric decay of $\tau(M_{m,n})$, condition (3) is satisfied automatically here. To see this, let n_0 be such that $Q^{n_0} > 0$. Then as $n \to \infty$ we get $\tau(M_{n,n+n_0}) = \tau(Q_{n,n+n_0}) \to \tau(Q^{n_0}) < 1$. Thus, from Theorem 2, under the given conditions on the $\{M_n\}$, if (5) holds then

$$Z_{m,n} D_n^{-1} / 1^T Q_{m,n} 1 \to_{\text{a.s.}\ \mathscr{L}^2} \bar{L}_m \bar{v}^T \quad \text{as } n \to \infty. \qquad \square$$

Before proving Proposition 4 we need some extra definitions. We will suppose that the conditions of Theorem 1 hold in what follows.

Recall that $q_m(i)$ is the extinction probability of the Z_+^d valued process $Z_m(i,\cdot) = \{Z_{m,n}(i,\cdot)\}_{n=m}^\infty$. Put $q_{m,n}(i) = \mathrm{P}(Z_{m,n}(i,\cdot) = 0)$. Then as $n \to \infty$, $q_{m,n}(i) \uparrow q_m(i)$ for all $m \geq 0$ and $1 \leq i \leq d$. Put $l_m(i) = \mathrm{P}(L_m(i) = 0)$. Then in general $q_m := (q_m(1), \dots, q_m(d)) \leq l_m := (l_m(1), \dots, l_m(d))$. We seek conditions under which $q_m = l_m$.

Let f_n^i be the joint p.g.f. of $X_n(i,\cdot)$ and for $x \in [0,1]^d$ put $f_n(x) = (f_n^1(x), \dots, f_n^d(x))$. Similarly, let $f_{m,n}^i$ be the joint p.g.f. of $Z_{m,n}(i,\cdot)$ and for $x \in [0,1]^d$ put $f_{m,n}(x) = (f_{m,n}^1(x), \dots, f_{m,n}^d(x))$. Clearly

$$f_{m,n}(x) = f_m\big(f_{m+1}\big(\cdots\big(f_{n-1}(x)\big)\cdots\big)\big).$$

Let ϕ_m^i be the Laplace transform of $L_m(i)$ and for $s \in \mathbb{R}_+$ put $\phi_m(s) = (\phi_m^1(s), \dots, \phi_m^d(s))$. Similarly, let $\phi_{m,n}^i$ be the Laplace transform of $Z_{m,n}(i,1)/{}^m R_n(1)$ and for $s \in \mathbb{R}_+$ put $\phi_{m,n}(s) = (\phi_{m,n}^1(s), \dots, \phi_{m,n}^d(s))$. Since $Z_{m,n}(i,1)/{}^m R_n(1) \to_{\mathscr{L}^2} L_m(i)$, $\phi_{m,n}(s) \to \phi_m(s)$ as $n \to \infty$ for all $m \geq 0$ and $s \in \mathbb{R}_+$.

Conditioning on $Z_{m,n}$, we have for all $m \leq n \leq p$,

$$\phi_{m,p}(s) = f_{m,n}\Big(\phi_{n,p}\big(s^n R_p(1)/{}^m R_p(1)\big)\Big). \tag{19}$$

Sending $p \to \infty$ then $s \to \infty$ in (19) gives

$$l_m = f_{m,n}(l_n) \quad \text{for all } 0 \leq m \leq n. \tag{20}$$

Conversely, sending $s \to \infty$ then $p \to \infty$ gives

$$q_m = f_{m,n}(q_n) \quad \text{for all } 0 \leq m \leq n.$$

So $\vec{l} := (l_0, l_1, \dots)$, $\vec{q} := (q_0, q_1, \dots)$ and (trivially) $\vec{1} := (1, 1, \dots)$ are all solutions to equations (20). Clearly, taking the inequalities elementwise, $\vec{q} \leq \vec{l} \leq \vec{1}$. In fact $\vec{q}$ is the minimal nonnegative solution to (20). To see this, suppose that $\vec{p} = (p_0, p_1, \dots)$ is some other nonnegative solution. Then for all m,

$$p_m = f_{m,n}(p_n) \geq f_{m,n}(0) = q_{m,n} \uparrow q_m \quad \text{as } n \to \infty.$$

LEMMA 17. *Suppose there exist vectors* $h_n \in \mathbb{R}_+^d$ *such that* $\mathrm{P}(Z_{m,n}(i,\cdot)\, h_n \leq x) \to q_m(i)$ *as* $n \to \infty$ *for all* $m \geq 0$, $1 \leq i \leq d$ *and* $x \in \mathbb{R}_+$. *Then for any sequence* $\vec{p} = (p_0, p_1, \dots)$ *satisfying* $\log p_n(i) \leq -c_0 h_n(i)$ *for some* $c_0 > 0$ *and all* $n \geq 0$ *and* $1 \leq i \leq d$,

$$\lim_{n \to \infty} f_{m,n}(p_n) = q_m \quad \textit{for all } m \geq 0.$$

PROOF. For all $m \geq 0$ and $1 \leq i \leq d$ we have, for any $x \in \mathbb{R}_+$,

$$\begin{aligned} f_{m,n}^i(p_n) &\leq \sum_{0 \leq z^T h_n \leq x} \mathrm{P}\big(Z_{m,n}(i,\cdot) = z^T\big) \\ &\quad + \sum_{x \leq z^T h_n} \mathrm{P}\big(Z_{m,n}(i,\cdot) = z^T\big) \prod_i p_n(i)^{z(i)} \\ &\leq \sum_{0 \leq z^T h_n \leq x} \mathrm{P}\big(Z_{m,n}(i,\cdot) = z^T\big) + e^{-c_0 x}. \end{aligned}$$

Send $n \to \infty$, then $x \to \infty$ to complete the proof. □

PROOF OF PROPOSITION 4. We are given that the conditions of Theorem 1 hold, and that, for some vectors $h_n \in \mathbb{R}_+^d$ and for all $m \geq 0$, $1 \leq i \leq d$ and $x \in \mathbb{R}_+$,

$$\mathrm{P}\big(Z_{m,n}(i,\cdot)h_n \leq x\big) \to q_m(i) \quad \text{as } n \to \infty$$

[condition (6)] and

$$\sum_{n=m}^{\infty} \sum_{j,k=1}^{d} \frac{M_{m,n}(i,j)\,\sigma_n^2(j,k)}{w_m(i)^2\, {}^mR_{n+1}^2(k)} \leq K/h_m(i) - 1$$

[condition (7)]. We will show (to begin with) that $q_m(i) = \mathrm{P}(L_m(i) = 0)$.

From (20) and Lemma 17 we see that the result will hold if we can find a constant $c_0 > 0$ such that $\log l_m(i) \leq -c_0 h_m(i)$ for all $m \geq 0$ and $1 \leq i \leq d$. Applying the Cauchy–Schwarz inequality to $\mathrm{E}\, L_m(i) I(L_m(i) > 0)$ gives (noting $\mathrm{E}\, L_m = w_m$)

$$\mathrm{P}\big(L_m(i) = 0\big) \leq 1 - w_m(i)^2 / \big(w_m(i)^2 + \operatorname{Var} L_m(i)\big).$$

Thus the result will hold if, putting $K = c_0^{-1}$,

$$\operatorname{Var} L_m(i)/w_m(i)^2 \leq K/h_m(i) - 1 \tag{21}$$

for all $m \geq 0$ and $1 \leq i \leq d$. But from Theorem 1 we have

$$\begin{aligned} \operatorname{Var} L_m(i) &= \sum_{k=m}^{\infty} {}^m w_{k+1}^T\, {}^m R_{k+1}^{-1} \left(\sum_{j=1}^{d} V_k[j] \cdot M_{m,k}(i,j) \right) {}^m R_{k+1}^{-1}\, {}^m w_{k+1}^T \\ &\leq \sum_{k=m}^{\infty} \sum_{j,k=1}^{d} \frac{M_{m,n}(i,j)\,\sigma_n^2(j,k)}{{}^mR_{n+1}^2(k)} \end{aligned}$$

noting that

$$\frac{V_n[j](x,y)}{{}^mR_{n+1}(x)\,{}^mR_{n+1}(y)} \leq \frac{1}{2}\left(\frac{V_n[j](x,x)}{{}^mR_{n+1}^2(x)} + \frac{V_n[j](y,y)}{{}^mR_{n+1}^2(y)} \right).$$

So, (21) follows from (7), and we have proved the first part of the proposition.

The second part of the proposition is a simplification of (7), in the case where there exist diagonal rescaling matrices $D_n = \operatorname{diag}(D_n(1), \dots, D_n(d))$ for all $n \geq 0$, such that $Q_n := D_n M_n D_{n+1}^{-1}$ converges elementwise to a primitive matrix Q. It follows immediately from (10) and Lemma 15. □

4. An example from diffusions on fractals. The following example is a reworking of one originally given by Hattori, Hattori and Watanabe (1994), applying the results above.

The Sierpinski gasket is a simple fractal, defined as follows. Let $V_0 = \{(0,0), (1,0), (1/2, \sqrt{3}/2)\}$ and $E_0 = \{((0,0),(1,0)), ((1,0),(0,0)), ((0,0),(1/2,\sqrt{3}/2)), ((1/2,\sqrt{3}/2),(0,0)), ((1,0),(1/2,\sqrt{3}/2)), ((1/2,\sqrt{3}/2),(1,0))\}$ and recursively define $(V_1, E_1), (V_2, E_2), (V_3, E_3), \dots$ by

$$V_{n+1} = V_n \cup \left[(2^n, 0) + V_n\right] \cup \left[(2^{n-1}, 2^{n-1}\sqrt{3}) + V_n\right]$$

and

$$E_{n+1} = E_n \cup \left[(2^n, 0) + E_n\right] \cup \left[(2^{n-1}, 2^{n-1}\sqrt{3}) + E_n\right]$$

taking the sums elementwise over the given sets. Let $V = V_\infty \cup [-V_\infty]$ and $E = E_\infty \cup [-E_\infty]$ and write G_0 for the graph (V, E) and G_n for $2^{-n}G_0$. The Sierpinski gasket G is the closure of the set $\bigcup_{n=0}^{\infty} 2^{-n}V$.

The direct approach to the construction of a diffusion on G is to consider for each n some random walk Y_n on G_n and then look at the limit as $n \to \infty$. See for example Barlow and Perkins (1988) or Kumagai (1993). The essential requirement of such a construction is that for any $n \geq m$, the random walk obtained by observing Y_n on G_m (the G_m *decimation* of Y_n) is exactly Y_m. Note that we observe Y_n only when it moves from one G_m vertex to a different G_m vertex. A sequence $\{Y_n\}_{n=0}^{\infty}$ of random walks satisfying this requirement is called *nested*. We will be looking at a nested sequence of symmetric, spatially inhomogeneous random walks.

Observe that topologically the graphs G_n are identical. We distinguish three distinct types of vertex depending on the relative positions of their nearest neighbors. In G_0 define a type I vertex (x, y) as one with neighbors $(x-1, y)$, $(x+1, y)$, $(x - 1/2, y + \sqrt{3}/2)$ and $(x + 1/2, y + \sqrt{3}/2)$. A type II vertex has neighbors $(x + 1/2, y + \sqrt{3}/2)$, $(x+1, y)$, $(x - 1/2, y - \sqrt{3}/2)$ and $(x + 1/2, y - \sqrt{3}/2)$. A type III vertex has neighbors $(x + 1/2, y - \sqrt{3}/2)$, $(x-1, y)$, $(x - 1/2, y - \sqrt{3}/2)$ and $(x + 1/2, y - \sqrt{3}/2)$ (the y-axis reflection of those of a type II vertex). The analogous definitions in G_n should be clear.

We also distinguish twelve types of directed edge, depending on the orientation of the edge and its initial vertex. There are six possible orientations: 0; $\pi/3$; $2\pi/3$; π; $4\pi/3$ and $5\pi/3$ radians. The possible types of directed edge are I-0; I-π; I-$\pi/3$; I-$2\pi/3$; II-$\pi/3$; II-0; II-$4\pi/3$; II-$5\pi/3$; III-$2\pi/3$; III-π; III-$4\pi/3$ and III-$5\pi/3$.

Now, weight the horizontal edges of G_n (those with direction 0 or π) with weight 1 and the diagonal edges with weight a_n. We define a random walk Y_n on G_n in the usual manner, setting the probability of moving along a given

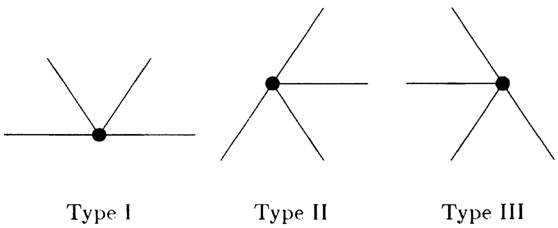

FIG. 1. *Three types of vertex.*

edge proportional to the weight of that edge. Requiring the Y_n to be nested leads to the following relation between the a_n:

$$a_n = \frac{a_{n+1}(4 + 6a_{n+1})}{3 + 6a_{n+1} + a_{n+1}^2}.$$

See, for example, Barlow (1993) for a discussion of how to calculate weights (or conductivities) for a decimated graph (or network). Note that this requires separate verification for each type of vertex.

An MTBPVE is obtained by considering the frequencies of the steps each Y_n makes along edges of each type (I-0, I-π, etc.). Take as our original ancestor at time 0 the first step made by Y_0. The children of this step are the steps made by Y_1 in going from $Y_0(0)$ to $Y_0(1)$. Continuing in this manner, the children of a step $Y_n(k)$ to $Y_n(k+1)$ are the corresponding group of steps made by Y_{n+1}. As the weights a_n, and thus the transition probabilities of Y_n, vary with n, we have a varying environment. As the frequencies of each type of step have different distributions, we have a multitype process. It is clear from the geometry of G_n and our choice of weights, that we only need to distinguish five different types of transition.

Type 1: type I-0 or I-π.
Type 2: type I-$\pi/3$ or I-$2\pi/3$.
Type 3: type II-0 or III-π.
Type 4: type II-$4\pi/3$, II-$5\pi/3$, III-$4\pi/3$ or III-$5\pi/3$.
Type 5: type II-$\pi/3$ or type III-$2\pi/3$.

Types 1 and 3 are the horizontal transitions, while types 2, 4 and 5 are the diagonal transitions.

The joint probability generating functions of the branching process are given in the Appendix. We use these to calculate the mean matrices M_n and the variance matrices σ_n^2 of the process. To the first order of magnitude (in terms of powers of a_n) these are, for $n \geq 1$,

$$M_{n-1} \simeq \begin{pmatrix} 4 & 8a_n/3 & 8/9 & 8a_n/3 & 0 \\ 5/2 & 1 & 1/3a_n & 5a_n/3 & 1 \\ 2 & 4a_n/3 & 10/3 & 4a_n & 8a_n/3 \\ 3/2 & a_n & 1/3a_n & 2 & 2a_n \\ 1 & 1/2 & 1/3a_n & 3a_n & 3/2 \end{pmatrix}$$

and

$$\sigma_{n-1}^2 \simeq \begin{pmatrix} 8 & 8a_n/3 & 16/27a_n & 8a_n/3 & 0 \\ 33/4 & 5a_n/3 & 1/9a_n^2 & 5a_n/3 & 0 \\ 2 & 4a_n/3 & 8/9a_n & 4a_n & 8a_n/3 \\ 9/2 & 2a_n & 1/9a_n^2 & 3a_n & 2a_n \\ 2 & 1/4 & 1/9a_n^2 & 3a_n & 1/4 \end{pmatrix}.$$

Putting $D_n = \operatorname{diag}(1, a_n, 1, a_n, a_n)$ we get $Q_n := D_n M_n D_{n+1}^{-1} \to Q$ where

$$Q = \begin{pmatrix} 4 & 8/3 & 8/9 & 8/3 & 0 \\ 0 & 4/3 & 4/9 & 0 & 4/3 \\ 2 & 4/3 & 10/3 & 4 & 8/3 \\ 0 & 0 & 4/9 & 8/3 & 0 \\ 0 & 2/3 & 4/9 & 0 & 2 \end{pmatrix};$$

Q is primitive with $\mathrm{PF}(Q) = 6$. Moreover $a_n \to 0$ geometrically fast and, as $a_n/a_{n+1} - 4/3 = O(a_{n+1})$, $a_n/a_{n+1} \to 4/3$ geometrically fast as well. Thus $Q_n \to Q$ geometrically fast and $1^T Q_{m,n} 1$ grows like 6^{n-m}. So the forward matrix product $M_{m,n}$ has two distinct growth rates, 6^{n-m} and $6^{n-m} a_n$, corresponding to horizontal and diagonal type transitions respectively.

Next, note that $D_n \sigma_n^2 D_{n+1}^{-2} = A_n/a_{n+1} + O(1)$ where $A_n \to A$, given by

$$A = \begin{pmatrix} 0 & 8/3 & 16/27 & 8/3 & 0 \\ 0 & 0 & 4/27 & 0 & 0 \\ 0 & 4/3 & 8/9 & 4 & 8/3 \\ 0 & 0 & 4/27 & 0 & 0 \\ 0 & 1/3 & 4/27 & 0 & 1/3 \end{pmatrix}.$$

It is easily shown that $\{a_n\}_{n=0}^{\infty} \equiv \{(3/4)^n\}_{n=0}^{\infty}$, whence for any $0 \le m \le n$,

$$\frac{D_n \sigma_n^2 D_{n+1}^{-2}}{1^T Q_{m,n} 1} = O\left(\left(\frac{2}{9}\right)^{n-m}\left(\frac{4}{3}\right)^{m}\right).$$

Applying Corollary 3, we see that $6^{m-n} Z_{m,n} D_n^{-1}$ converges a.s. and in $\mathscr{L}^2$ as $n \to \infty$ to a nontrivial limit, call it $W_m 1^T$.

We can now apply Proposition 4 to show that $\mathrm{P}(W_m(i) = 0) = 0$ for all m and i. For some $\varepsilon > 0$ small, put $h_n = (2 - \varepsilon)^{-n} 1$ for all $n \ge 0$. Then these h_n satisfy (6). Moreover

$$\sum_{n=m}^{\infty} \sum_{j,k} \frac{D_n(j)\sigma_n^2(j,k)}{1^T Q_{m,n} 1^T D_{n+1}^2(k)} = O\left(\left(\frac{4}{3}\right)^m\right)$$

and

$$\left(D_n^{-1}\overline{w}_m\right)(i)/h_m(i) \ge \begin{cases} K_0(2-\varepsilon)^m, & i = 1, 3, \\ K_0(2-\varepsilon)^m\left(\frac{4}{3}\right)^m, & i = 2, 4, 5, \end{cases}$$

for some constant $K_0 > 0$, so that (8) is satisfied and we have $\mathrm{P}(W_m(i) = 0) = q_m(i) = 0$.

Applying these results to the nested random walks $\{Y_n\}_{n=0}^{\infty}$, it is possible prove the existence of a scaled limit Y of the $\{Y_n\}$, which will be a symmetric, spatially inhomogeneous diffusion on the Sierpinski gasket G. The (time) scaling used is that of the largest growth rate of the branching process, namely $\{6^n\}_{n=0}^{\infty}$, and the normed limits $\{W_m\}$ give (on multiplying by a suitable scale factor) the G_m crossing times of Y for each of the five transition types and $m \geq 0$. Note that the existence of a nontrivial limit Y follows from the existence of the $\{W_m\}$, but to show Y is a diffusion requires that the $\{W_m\}$ have no mass at 0. The actual details of the construction are standard: see Barlow and Perkins (1988); Hattori, Hattori and Watanabe (1994) or Kumagai (1993).

APPENDIX

We give here the joint p.g.f.'s of the branching process described in Section 4. Write $f_n^i(x_1, \dots, x_5)$ for the joint p.g.f. of $X_n(i, \cdot)$. Then we have:

$$\begin{aligned}
&f_{n-1}^1(x_1, x_2, x_3, x_4, x_5)\\
&\quad = \big((1+a_n)(3+7a_n)\big(x_1^2 + 6a_n x_1^2 + 9a_n^2 x_1^2 - x_1^2 x_3^2 + 2a_n^2 x_1 x_2 x_4\\
&\qquad + 6a_n^3 x_1 x_2 x_4 + 2a_n^2 x_2 x_3 x_4 + 2a_n^3 x_2 x_3 x_4 + 2a_n^2 x_1 x_2 x_3 x_4 + a_n^4 x_2^2 x_4^2\big)\big)\\
&\qquad \times \big((-3 - 6a_n - a_n^2)\big(-2 - 16a_n - 44a_n^2 - 48a_n^3 - 18a_n^4 + x_1^2\\
&\qquad\quad + 6a_n x_1^2 + 9a_n^2 x_1^2 + 2x_3^2 + 4a_n x_3^2 + 2a_n^2 x_3^2 - x_1^2 x_3^2 + 4a_n^2 x_2 x_4\\
&\qquad\quad + 16a_n^3 x_2 x_4 + 12a_n^4 x_2 x_4 + 2a_n^2 x_1 x_2 x_4 + 6a_n^3 x_1 x_2 x_4 + 2a_n^2 x_2 x_3 x_4\\
&\qquad\quad + 2a_n^3 x_2 x_3 x_4 + 2a_n^2 x_1 x_2 x_3 x_4 - a_n^4 x_2^2 x_4^2\big)\big)^{-1},\\
&f_{n-1}^2(x_1, x_2, x_3, x_4, x_5)\\
&\quad = \big(a_n(1+a_n)(3+7a_n)(1+a_n+x_1)x_2(1+3a_n+x_3)x_5\big)\\
&\qquad \times \big((-2-3a_n)\big(-2 - 16a_n - 44a_n^2 - 48a_n^3 - 18a_n^4 + x_1^2 + 6a_n x_1^2\\
&\qquad\quad + 9a_n^2 x_1^2 + 2x_3^2 + 4a_n x_3^2 + 2a_n^2 x_3^2 - x_1^2 x_3^2 + 4a_n^2 x_2 x_4\\
&\qquad\quad + 16a_n^3 x_2 x_4 + 12a_n^4 x_2 x_4 + 2a_n^2 x_1 x_2 x_4 + 6a_n^3 x_1 x_2 x_4 + 2a_n^2 x_2 x_3 x_4\\
&\qquad\quad + 2a_n^3 x_2 x_3 x_4 + 2a_n^2 x_1 x_2 x_3 x_4 - a_n^4 x_2^2 x_4^2\big)\big)^{-1},\\
&f_{n-1}^3(x_1, x_2, x_3, x_4, x_5)\\
&\quad = \big((3 + 18a_n + 19a_n^2)\big(x_1 x_3 + 6a_n x_1 x_3 + 9a_n^2 x_1 x_3 - x_1 x_3^3 + a_n^2 x_2 x_3 x_4\\
&\qquad + 3a_n^3 x_2 x_3 x_4 + a_n^2 x_2 x_3^2 x_4 + a_n^2 x_1 x_4 x_5 + 3a_n^3 x_1 x_4 x_5 + 2a_n^2 x_3 x_4 x_5\\
&\qquad + 2a_n^3 x_3 x_4 x_5 + a_n^2 x_1 x_3 x_4 x_5 + a_n^4 x_2 x_4^2 x_5\big)\big)\\
&\qquad \times \big((-3 - 6a_n - a_n^2)\big(-2 - 20a_n - 72a_n^2 - 108a_n^3 - 54a_n^4 + x_1 x_3\\
&\qquad\quad + 6a_n x_1 x_3 + 9a_n^2 x_1 x_3 + 2x_3^2 + 8a_n x_3^2 + 6a_n^2 x_3^2 - x_1 x_3^3 + 2a_n^2 x_2 x_4\\
&\qquad\quad + 12a_n^3 x_2 x_4 + 18a_n^4 x_2 x_4 + 3a_n^2 x_2 x_3 x_4 + 9a_n^3 x_2 x_3 x_4 + a_n^2 x_2 x_3^2 x_4\\
&\qquad\quad + 6a_n^2 x_4 x_5 + 24a_n^3 x_4 x_5 + 18a_n^4 x_4 x_5 + a_n^2 x_1 x_4 x_5 + 3a_n^3 x_1 x_4 x_5\\
&\qquad\quad + 4a_n^2 x_3 x_4 x_5 + 4a_n^3 x_3 x_4 x_5 + a_n^2 x_1 x_3 x_4 x_5 - a_n^4 x_2 x_4^2 x_5\big)\big)^{-1},
\end{aligned}$$

$$
\begin{aligned}
&f^4_{n-1}(x_1, x_2, x_3, x_4, x_5)\\
&\quad = \big(a_n\big(3 + 18a_n + 19a_n^2\big)(1 + a_n + x_1)(1 + 3a_n + x_3)x_4^2\big)\\
&\qquad \times\big((-2 - 3a_n)\big(-2 - 20a_n - 72a_n^2 - 108a_n^3 - 54a_n^4 + x_1x_3\\
&\qquad\qquad + 6a_nx_1x_3 + 9a_n^2x_1x_3 + 2x_3^2 + 8a_nx_3^2 + 6a_n^2x_3^2\\
&\qquad\qquad - x_1x_3^3 + 2a_n^2x_2x_4 + 12a_n^3x_2x_4 + 18a_n^4x_2x_4\\
&\qquad\qquad + 3a_n^2x_2x_3x_4 + 9a_n^3x_2x_3x_4 + a_n^2x_2x_3^2x_4 + 6a_n^2x_4x_5\\
&\qquad\qquad + 24a_n^3x_4x_5 + 18a_n^4x_4x_5 + a_n^2x_1x_4x_5 + 3a_n^3x_1x_4x_5\\
&\qquad\qquad + 4a_n^2x_3x_4x_5 + 4a_n^3x_3x_4x_5 + a_n^2x_1x_3x_4x_5 - a_n^4x_2x_4^2x_5\big)\big)^{-1},\\
&f^5_{n-1}(x_1, x_2, x_3, x_4, x_5)\\
&\quad = \big(a_n\big(3 + 18a_n + 19a_n^2\big)(1 + 3a_n + x_3)x_5(x_2x_3 + x_5 + a_nx_5)\big)\\
&\qquad \times\big((-2-3a_n)\big(-2-20a_n-72a_n^2-108a_n^3-54a_n^4+x_1x_3+6a_nx_1x_3\\
&\qquad\qquad +9a_n^2x_1x_3+2x_3^2+8a_nx_3^2+6a_n^2x_3^2-x_1x_3^3+2a_n^2x_2x_4\\
&\qquad\qquad + 12a_n^3x_2x_4 + 18a_n^4x_2x_4 + 3a_n^2x_2x_3x_4 + 9a_n^3x_2x_3x_4 + a_n^2x_2x_3^2x_4\\
&\qquad\qquad + 6a_n^2x_4x_5 + 24a_n^3x_4x_5 + 18a_n^4x_4x_5 + a_n^2x_1x_4x_5 + 3a_n^3x_1x_4x_5\\
&\qquad\qquad + 4a_n^2x_3x_4x_5 + 4a_n^3x_3x_4x_5 + a_n^2x_1x_3x_4x_5 - a_n^4x_2x_4^2x_5\big)\big)^{-1}.
\end{aligned}
$$

Acknowledgments. The author thanks John Biggins for his many valuable comments on the Ph.D. dissertation upon which this paper is based and Harry Cohn for his helpful conversation on branching processes and for telling me about his work on continuity.

REFERENCES

ASMUSSEN, S. and HERING, H. (1983). *Branching Processes*. Birkhäuser, Boston.

ATHREYA, K. B. and NEY, P. E. (1972). *Branching Processes*. Springer, New York.

BARLOW, M. T. (1993). Random walks, electrical resistance and nested fractals. In *Asymptotic Problems in Probability Theory: Stochastic Models and Diffusions on Fractals* (K. D. Elworthy and N. Ikeda, eds.) 131–157. Pitman, Montreal.

BARLOW, M. T. and PERKINS, E. A. (1988). Brownian motion on the Sierpinski gasket. *Probab. Theory Related Fields* **79** 543–623.

BIRKHOFF, G. (1957). Extensions of Jentzsch's theorem. *Trans. Amer. Math. Soc.* **85** 219–227.

COHEN, J. E. (1979). Contractive inhomogeneous products of nonnegative matrices. *Math. Proc. Cambridge Philos. Soc.* **86** 351–364.

COHN, H. (1989). On the growth of the multitype supercritical branching process in a random environment. *Ann. Probab.* **17** 1118–1123.

COHN, H. (1993). Supercritical branching processes: a unified approach. Research Report 9, Dept. Statistics, Univ. Melbourne, Australia.

COHN, H. and JAGERS, P. (1994). General branching processes in varying environment. *Ann. Appl. Probab.* **4** 184–193.

COHN, H. and NERMAN, O. (1990). On products of nonnegative matrices. *Ann. Probab.* **18** 1806–1815.

FEARN, D. (1971). Galton–Watson processes with generation dependence. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **4** 159–172. Univ. California Press, Berkeley.

HAJNAL, J. (1976). On products of nonnegative matrices. *Math. Proc. Cambridge Philos. Soc.* **79** 521–530.

HARRIS, T. E. (1963). *The Theory of Branching Processes*. Springer, Berlin.

HATTORI, K., HATTORI, T. and WATANABE, H. (1994). Asymptotically one-dimensional diffusions on the Sierpinski gasket and the *abc*-gaskets. *Probab. Theory Related Fields* **100** 85–116.

HATTORI, T. (1994). Asymptotically one-dimensional diffusions on scale-irregular gaskets. Technical Report UTUP-105, Faculty of Engineering, Utsunomiya Univ., Japan.

HATTORI, T. and WATANABE, H. (1993). On a limit theorem for nonstationary branching processes. In *Seminar on Stochastic Processes, 1992* (E. Çinlar, ed.) 173–187. Birkhäuser, Boston.

JAGERS, P. (1974). Galton–Watson processes in varying environments. *J. Appl. Probab.* **11** 174–178.

JONES, O. D. (1995). Random walks on prefractals and branching processes. PhD. dissertation, Statistical Lab., Cambridge Univ.

KUMAGAI, T. (1993). Construction and some properties of a class of nonsymmetric diffusion processes on the Sierpinski gasket. In *Asymptotic Problems in Probability Theory: Stochastic Models and Diffusions on Fractals* (K. D. Elworthy and N. Ikeda, eds.) 219–247. Pitman, Montreal.

MARCUS, M. and MINC, H. (1964). *A Survey of Matrix Theory and Matrix Inequalities*. Allyn and Bacon, Boston.

SENETA, E. (1981). *Nonnegative Matrices and Markov Chains*, 2nd ed. Springer, New York.

SENETA, E. and SHERIDAN, S. (1981). Strong ergodicity of nonnegative matrix products. *Linear Algebra Appl.* **37** 277–292.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF QUEENSLAND
ST. LUCIA, QLD 4072
AUSTRALIA
E-MAIL: odj@maths.uq.oz.au