THE GROWTH OF ITERATES OF MULTIVARIATE GENERATING FUNCTIONS

J.D. BIGGINS

ABSTRACT. The vector-valued function $m(\theta)$ of a p-vector θ has components $m_1(\theta), m_2(\theta), \ldots, m_p(\theta)$. For each i, $\exp(m_i(-\theta))$ is the (multivariate) Laplace transform of a discrete measure concentrated on $[0, \infty)^p$ with only a finite number of atoms. The main objective is to give conditions for the functional iterates $m^{(n)}$ of m to grow like ρ^n for a suitable $\rho > 1$. The initial stimulus was provided by results of Miller and O'Sullivan (1992) on enumeration issues in 'context free languages', results which can be improved using the theory developed here. The theory also allow certain results in Jones (2004) on multitype branching to be proved under significantly weaker conditions.

1. INTRODUCTION

Let μ be a vector of discrete measures, with components $\mu_1, \mu_2, \ldots, \mu_p$, each with all its mass on $[0, \infty)^p$ and only a finite number of atoms. The (cumulant) generating function of the collection μ is the vector-valued function m with components $m_1, m_2, \ldots, m_p$ given by

$$m_i(\theta) = \log\left(\int_{[0,\infty)^p} \exp(b^T\theta)\mu_i(db)\right),$$

where θ is a p-vector and $b^T\theta$ is the usual inner product, $\sum b_i\theta_i$; here, and throughout, vectors are column vectors and v_i is the ith component of the vector v. Hence, $\exp(m_i(-\theta))$ is the (multivariate) Laplace transform of the measure μ_i. Thus $m(\theta)$ is finite for all θ, because each component of μ has compact support, and so its nth iterate, $m^{(n)}$, is well defined and given by

$$m^{(n)}(\theta) = m(m^{(n-1)}(\theta)).$$

The objective is to seek information on the growth rate of the sequence $\{m^{(n)}(\theta) : n\}$ when $m^{(n)}(\theta)$ goes to infinity with n. Crudely, the main result gives conditions for $m^{(n)}(\theta)$ to grow like ρ^n for a suitable constant ρ. The statement of the main results and an outline of the structure of the paper are contained in the next section.

In the univariate case, $\exp(m(\theta)) = \sum_b \mu(b)e^{\theta b}$ and ρ is just the highest power of e^θ that occurs. Then, the limit result is easy to obtain, as can be seen in Section 5. In the multivariate case there is no direct analogue of the highest power, resulting in real increase in the difficulty of establishing limiting behaviour.

The estimation of iterates of generating functions has been a central topic in the study of branching processes, where the components of μ are probability measures on vectors with non-negative integer components, that is on $\mathbb{Z}_+^p$. The relationship

---

between branching processes and fractals, in particular, diffusions on fractals, has kindled interest in tail estimates for the limit of the branching process, which, in their turn, are intimately related to the kind of generating function results derived here. Describing these connections properly would require an extensive discussion: the papers by Barlow and Perkins [3], Kumagi [12], Biggins and Bingham [4], Athreya and Vidyashankar [2] and Jones [8, 9] give a flavour of the issues.

There is some overlap in ideas used here with those in the two papers by Jones [8, 9], which are on similar problems. Reference is made there to unpublished work of mine, which has matured into this published version. In particular, under certain restrictions, Jones [9] shows the existence of the Harris function (defined later, in Theorem 3.1) associated with a multitype Galton-Watson process with a bounded family size. This function is the natural analogue of a function identified for the one-type process in the pioneering work of Harris [6], which explains the name. The results here allow the existence of the multitype Harris function to be demonstrated under natural conditions, without the restrictions Jones imposes. This application is discussed more fully in Section 3.

The paper by Miller and O'Sullivan [14], discussing enumeration issues associated with 'context free languages', provided the original motivation for this study. They assume, in their Definition 9, when translated into generating functions and with 0 interpreted as the zero vector, that $m_i^{(n+1)}(0)/m_i^{(n)}(0)$ has a limit as n goes to infinity and then they give a variational identification of that limit. They conjecture that, in their context, the limit always exists. In fact, because of periodicity effects, this is not always so. Their real interest was in establishing and identifying the geometric growth rate of $m_i^{(n)}(0)$: the results here do this. Section 4 gives further details.

2. Statement of main results

As usual, inequalities should be interpreted componentwise where necessary. The inequality sign $\gg$ will be used when strict inequality holds in all components. Upper case letters will be used for matrices and, usually, lower case ones for vectors and scalars; the exception is that the ith row of the matrix B is denoted by B_i and the jth component of this row is $B_{i,j}$. The supremum norm will be used for vectors and matrices, so $\|v\| = \max |v_i|$ and $\|A\| = \max |A_{ij}|$.

In considering the iterates of m it is natural to impose a condition that ensures that no smaller-dimensional generating function could be studied. Recall that, when M is a square matrix of non-negative entries, M is primitive (or positively regular) if there is an n such that every entry in M^n is positive, M is irreducible if for every (i,j) there is an n such that $(M^n)_{i,j}$ is strictly positive, and an irreducible matrix M has a period d' such that $M^{d'}$ is block diagonal with each block being primitive — see Seneta [16, 17, §1.1 and §1.3]. Let $I(\mu)$ be the incidence matrix of $\int b^T \mu(db)$; thus, $I(\mu)$ has a one in the (i,j)th place when $\mu_i(b) > 0$ for some b with $b_j > 0$ and a zero otherwise. The measures μ and the corresponding function m will be called irreducible when the matrix $I(\mu)$ is irreducible. It will be assumed throughout that m is irreducible.

Let $\mathcal{P}$ be the index set $\{1, 2, \ldots, p\}$. Let $\mathcal{B}$ be the set of matrices defined by

$$\mathcal{B} = \{B : \mu_i(B_i^T) > 0 \text{ for all } i \in \mathcal{P}\};$$

thus, when $B \in \mathcal{B}$, the transpose of its ith row is the position of an atom of μ_i for each i. Any $B \in \mathcal{B}$ is a non-negative matrix, but it need not be irreducible. Nonetheless, Perron-Frobenius theory guarantees that its largest eigenvalue is non-negative with associated left and right eigenvectors that have non-negative, but not necessarily strictly positive, components — see Seneta [16, 17, Exercise 1.12] or Lancaster and Tismenetsky [13, Theorem 15.5.1].

Let ρ be the largest eigenvalue obtained as B ranges over $\mathcal{B}$; that is

$$(2.1) \qquad \rho = \max\{\rho' : \rho' \text{ is the largest eigenvalue of } B; B \in \mathcal{B}\}.$$

It is assumed that $\rho > 1$ throughout.

Proposition 2.1. *For each n, the function $m^{(n)}$ is increasing and convex in θ. Let $\mathfrak{T} = \{\theta : \|m^{(n)}(\theta)\| \to \infty\}$. If $\tilde{\theta} \in \mathfrak{T}$ then $\{\theta : \theta \geq \tilde{\theta}\} \subset \mathfrak{T}$.*

Theorem 2.2. *Suppose m is irreducible and $\rho > 1$. There exists a period d such that*

$$\frac{m^{(nd+r)}(\theta)}{\rho^{nd+r}} \to l^{(r)}(\theta)$$

as $n \to \infty$, on $\mathfrak{T}$, where, for each r, $l^{(r)}(\theta)$ is finite, strictly positive, convex and increasing in θ, and $l^{(r)}(\theta) = l^{(r+d)}(\theta)$. Furthermore, the univariate case $(p = 1)$ is always aperiodic $(d = 1)$ with, necessarily, a single limit, $l^{(0)}(\theta)$.

It is more complicated to give general conditions for aperiodicity in the multivariate case — Corollary 2.6 below gives one result. Turning to immediate consequences of Theorem 2.2, the convexity of $m^{(n)}$ and $l^{(r)}$, together with Theorems 10.1 and 10.8 in Rockafellar [15], imply the next corollary.

Corollary 2.3. *The convergence in Theorem 2.2 is uniform on closed bounded convex subsets of the interior of $\mathfrak{T}$, and $l^{(r)}(\theta)$ is continuous on the interior of $\mathfrak{T}$.*

Functional equations for the limits follow immediately, and are recorded in the next result. There is an extensive literature, to which this is therefore a contribution, on functional iteration and functional equations for the associated limits — see, for example, Kuczma et al. [11].

Corollary 2.4. *The limits satisfy:* (i) $\rho^d l^{(r)}(\theta) = l^{(r)}(m^{(d)}(\theta))$; (ii) $\rho l^{(r)}(\theta) = l^{(r-1)}(m(\theta))$.

In the univariate case, ρ is the highest power of e^θ. In the multivariate case, as might be guessed from the definition (2.1), there may be several matrices that might take the role of the highest power, which is a significant complication. Furthermore, in natural examples, these matrices need not be irreducible, even when $I(\mu)$ is primitive. Restricting μ further by forcing these 'highest power' matrices to be primitive would limit greatly the scope of the results, though they would then have easier proofs.

For a vector $s \gg 0$, let $\log s$ be the vector with ith entry $\log s_i$. When all the atoms of μ_i have components that are non-negative integers, and so lie in $\mathbb{Z}_+^p \subset [0, \infty)^p$, the more usual definition of the generating function is

$$f_i(s) = \exp(m_i(\log s)) = \int \exp((\log s)^T b)\mu_i(db) \left(= \sum_{b \in \mathbb{Z}_+^p} s^b \mu_i(b) \right);$$

this iterates consistently with m, so that its nth iterate, $f^{(n)}$, is given by $f^{(n)}(s) = \exp(m^{(n)}(\log s))$. Results on the growth of $m^{(n)}$ translate to results on the growth of $f^{(n)}$; $m^{(n)}$ seems to be the more natural object to study in this problem. Note that, in general, $m^{(n)}$ is not a generating function, that is to say $\exp(m_i^{(n)}(-\theta))$ need not be a Laplace transform — but it is when each μ_i is concentrated on $\mathbb{Z}_+^p$. Thus, once the period has been identified as d, concentrating on the integer-valued case would allow direct consideration of the sequence $\{m^{(nd)}(\theta) : n = 1, 2, \dots\}$, since its first member has the right form, simplifying a few of the arguments.

It is natural to wonder to what extent the results carry over to other vectors of transforms, or to even more general functions, but the methods used here rely heavily on the form assumed for μ.

For $B \in \mathcal{B}$ let the vector $q(B)$ have ith component $\log(\mu_i(B_i))$. Now, let $g(B; \theta)$ be defined for $B \in \mathcal{B}$ by

$$(2.2) \qquad m(\theta) = B\theta + q(B) + g(B; \theta).$$

Then the components of g must be non-negative because, substituting for m_i and q_i in (2.2) and rearranging,

$$\exp(g_i(B; \theta)) = \frac{\exp(-B_i\theta)}{\mu_i(B_i)} \left(\sum_b \mu_i(b) e^{b^T \theta} \right)$$

$$= 1 + \left(\sum_{b \neq B_i} \frac{\mu_i(b)}{\mu_i(B_i)} \exp((b^T - B_i)\theta) \right) \geq 1.$$

Suppose, for a moment, that there is a matrix $A \in \mathcal{B}$ such that, for all $B \in \mathcal{B}$ with $B \neq A$, $A - B \gg 0$. Then A will be the unique member of $\mathcal{B}$ with eigenvalue ρ, A is automatically primitive, and $g(A; \theta) \to 0$ and $m(\theta) \approx A\theta$ as $\min \theta_i \to \infty$. The matrix A is a proper analogue of the highest power in the one-dimensional case. Combining $m(\theta) \approx A\theta$ with Theorem 2.2 suggests that

$$\rho l^{(r+1)}(\theta) = \lim_{n \to \infty} \frac{m(m^{(nd+r)}(\theta))}{\rho^{nd+r}} = \lim_{n \to \infty} A \frac{m^{(nd+r)}(\theta)}{\rho^{nd+r}} = A l^{(r)}(\theta).$$

Something similar, but more complicated, holds in general and is described in the next two theorems. When A is primitive, $\rho^{-n} A^n x \to Lx$ where L is the idempotent matrix projecting onto the (unique) direction of the eigenvector of A corresponding to ρ — Seneta [16, 17, Theorem 1.2]. Thus, $m^{(n+\nu)}(\theta) \approx A^n m^{(\nu)}(\theta) \approx \rho^n L m^{(\nu)}(\theta)$, which suggests $d = 1$ and $A l^{(0)}(\theta) = \rho l^{(0)}(\theta)$. The corollary to the next theorem confirms this speculation.

When $\mathcal{J}$ and $\mathcal{E}$ are subsets of $\mathcal{P}$, let $v_{\mathcal{J}}$ be the restriction of the vector v to $\mathcal{J}$ and let $A_{\mathcal{J},\mathcal{E}}$ be the sub-matrix of A with rows indexed by $\mathcal{J}$ and columns by $\mathcal{E}$.

Theorem 2.5. *Let R be any irreducible matrix with period d' and eigenvalue ρ that arises as $B_{\mathcal{J},\mathcal{J}}$ for some $B \in \mathcal{B}$. Then, for $\theta \in \mathfrak{T}$,*

$$\frac{m_{\mathcal{J}}^{(nd'+r)}(\theta)}{\rho^{nd'+r}} \to l_{\mathcal{J}}^{(r)}(\theta).$$

Furthermore

$$R l_{\mathcal{J}}^{(r)}(\theta) = \rho l_{\mathcal{J}}^{(r+1)}(\theta).$$

The minimal d in Theorem 2.2 either divides or is divisible by d'.

Corollary 2.6. *Suppose there is a primitive matrix $A \in \mathcal{B}$ with eigenvalue ρ. Let a be a right eigenvector of A corresponding to ρ (which is unique up to a scale factor). Then $d = 1$ and $l^{(0)}(\theta) = k(\theta)a$ for some scalar $k(\theta)$. Furthermore, k satisfies the functional equation $\rho k(\theta) = k(m(\theta))$.*

Theorem 2.7. *Let $\theta \in \mathfrak{T}$. There are sets of indices $\mathfrak{I}$ and $\mathcal{E}$ that partition $\mathcal{P}$ and matrices $A^{(r)}(\theta) \in \mathcal{B}$, $r = 0, 1, \ldots, d-1$ such that the following hold.*

(i) *In every $B \in \mathcal{B}$, the maximum eigenvalue of the sub-matrix $B_{\mathcal{E},\mathcal{E}}$ is strictly less than ρ.*

(ii)

$$A^{(r)}(\theta) = \begin{pmatrix} C & 0 \\ A^{(r)}_{\mathcal{E},\mathfrak{I}}(\theta) & A^{(r)}_{\mathcal{E},\mathcal{E}}(\theta) \end{pmatrix},$$

where C is a square matrix, indexed by $\mathfrak{I} \times \mathfrak{I}$, which is block diagonal, with each block an irreducible matrix with eigenvalue ρ, and with the lowest common multiple of the period of these blocks being d.

(iii) *For each $i \in \mathcal{P}$*

$$A^{(r)}_i(\theta)l^{(r)}(\theta) = \max\{B_i l^{(r)}(\theta) : B \in \mathcal{B}\}$$

and

(2.3) $$A^{(r)}(\theta)l^{(r)}(\theta) = \rho l^{(r+1)}(\theta).$$

(iv) *$l^{(0)}(\theta)$ is an eigenvector of the matrix product $A^{(d-1)}(\theta)A^{(d-2)}(\theta)\cdots A^{(0)}(\theta)$, with eigenvalue ρ^d.*

It is worth stressing that the d identified in Theorem 2.7 may not be the minimal period for which Theorem 2.2 holds, but it must be divisible by the minimal period.

A natural use of the limit information in Theorem 2.2 is through an application of Gärtner-Ellis large deviation theory, as is done in the one-dimensional (or one-type) case by Biggins and Bingham [4]. For the use of that theory to go smoothly, the limit needs to be differentiable, which motivates the next two results. A little more is established in the course of the proof. I believe differentiability holds more widely than the conditions here might suggest.

Theorem 2.8. *If θ is in the interior of $\mathfrak{T}$ and, for each r, there is just one matrix $A^{(r)}(\theta)$ that satisfies (2.3) then the limits $l^{(r)}(\theta)$ have a continuous derivative in a neighbourhood of θ.*

Corollary 2.9. *When there is a unique matrix in $\mathcal{B}$ with eigenvalue ρ, $l^{(r)}$ has a continuous derivative throughout the interior of $\mathfrak{T}$.*

The next two sections discuss the main examples. Then, the comparatively straightforward one-dimensional case, for which much stronger results can be obtained without much difficulty, is discussed. Motivated by this, an upper bound on the growth rate of multivariate iterates is established in Section 6, which brings out the need to study products of non-negative matrices drawn from the set $\mathcal{B}$; relevant results are supplied by Kennedy [10]. After an excursion into matrix theory, establishing an upper bound on certain matrix products and describing some aspects of the structure of matrices with eigenvalue ρ, the main proof resumes in Section 8, where convergence is established for certain indices — essentially, for $\mathfrak{I}$ in Theorem 2.5. Convergence for a suitable set of indices arising in this way induces convergence for the remaining indices and completes the proof of Theorem 2.2. Then, in

Section 10, various properties of the limit are obtained, giving Theorems 2.7 and
2.8 and Corollary 2.9.

3. The multitype branching process

This section considers the multitype Galton-Watson process. Well-known facts
about this process, providing enough background for the next three paragraphs,
can be found in Athreya and Ney [2, Chapter V].

Suppose f is the probability generating function, of finite degree, of an irreducible
multitype branching process. This means there is an upper bound on the family
size, all positive moments of f are finite and the mean matrix M is irreducible. Let
$m(\theta) = \log f(e^\theta)$. The irreducibility of the branching process corresponds to the
irreducibility of m as defined at the start of previous section.

It is worth starting with a couple of simple examples. The first illustrates peri-
odicity. Consider a two-type process with

$$f(s) = \frac{1}{2} \left(\begin{array}{c} s_1^2 s_2^2 + s_2^5 \\ s_1 s_2^3 + s_1^5 \end{array} \right).$$

The idea is that the highest powers in each component determine the growth rate
of $m^{(n)}$, but these involve switching back and forth between types, and hence a
periodic behaviour, even though the mean matrix is primitive. Here

$$\mathcal{B} = \left\{ \left(\begin{array}{cc} 2 & 2 \\ 1 & 3 \end{array} \right), \left(\begin{array}{cc} 2 & 2 \\ 5 & 0 \end{array} \right), \left(\begin{array}{cc} 0 & 5 \\ 1 & 3 \end{array} \right), \left(\begin{array}{cc} 0 & 5 \\ 5 & 0 \end{array} \right) \right\}.$$

It is easy to see that $\rho = 5$ and that there is just one member of $\mathcal{B}$ with this eigen-
value. Therefore Corollary 2.9 applies, so the limiting functions are differentiable
in θ. It is not too hard to program an iterative scheme, based on equation (8.1)
below, to estimate $\rho^{-n} m^{(n)}(\theta)$ for large n. Doing this confirms numerically that,
in examples with $\theta \gg 0$ and $\theta_1 \neq \theta_2$, convergence is periodic, with period two. For
contrast, the generating function

$$f(s) = \frac{1}{2} \left(\begin{array}{c} s_1^2 s_2^2 + s_1^5 \\ s_1 s_2^3 + s_2^5 \end{array} \right),$$

which again has $\rho = 5$, has three matrices in $\mathcal{B}$ with this eigenvalue:

$$\left(\begin{array}{cc} 5 & 0 \\ 0 & 5 \end{array} \right), \left(\begin{array}{cc} 5 & 0 \\ 1 & 3 \end{array} \right), \left(\begin{array}{cc} 2 & 2 \\ 0 & 5 \end{array} \right).$$

Theorem 2.5 implies this is an aperiodic case. The first matrix here must be $A^{(0)}(\theta)$
in Theorem 2.7. The first part of Theorem 2.7(iii) constrains the direction of $l^{(0)}(\theta)$:
it must lie between that of the right eigenvectors of the second and third matrices,
that is $1/2 \leq l_2^{(0)}(\theta)/l_1^{(0)}(\theta) \leq 3/2$. Numerical experiment confirms that values over
this range occur as θ varies.

Let $\tilde{\rho}$ be the maximum eigenvalue of the mean matrix M and let $\tilde{r}$ be the
associated right eigenvector. Assume that the process is supercritical, that is that
$\tilde{\rho} > 1$ — the two examples just described are supercritical. Jensen's inequality gives
$m(\theta) \geq M\theta$, which implies that $m^{(n)}(\theta) \to \infty$ when $\theta \gg 0$; hence, $\{\theta : \theta \gg 0\} \subset \mathfrak{T}$.
Now, let Z_n be the vector of numbers of the various types at generation n, which
has probability generating function $f^{(n)}$. There is a non-negative random variable
W, which is not identically zero, such that $\tilde{\rho}^{-n} Z_n^T \tilde{r} \to W$, almost surely. Let E_i

be the expectation when the initial ancestor has type $i \in \mathcal{P}$ and let the (cumulant) generating function ϕ_i be defined by

$$\phi_i(\psi) = \log \left(E_i \left[e^{\psi W} \right] \right).$$

These satisfy the Poincaré functional equation: $\phi(\tilde{\rho}\psi) = m(\phi(\psi))$. Also, $\phi(\psi) \gg 0$ for all $\psi > 0$. It is worth commenting that these results only need M to be irreducible, not primitive.

Theorem 2.2 gives that there are d, ρ and $l^{(0)}$ such that

$$\rho^{-nd} m^{(nd)}(\theta) \to l^{(0)}(\theta).$$

For this section only, let $\gamma = \log \rho / \log \tilde{\rho}$, so that $\tilde{\rho}^{\gamma} = \rho$.

Theorem 3.1. *There is a positive, bounded, continuous function $H : (0, \infty) \to \mathbb{R}^p_+$ — the Harris function — which is multiplicatively periodic with period $\tilde{\rho}^d$ (i.e. $H(\psi) = H(\tilde{\rho}^d \psi)$), such that*

(3.1) $$\|\psi^{-\gamma} \phi(\psi) - H(\psi)\| \to 0 \text{ as } \psi \to \infty.$$

Furthermore, H is continuously differentiable when $l^{(0)}$ in Theorem 2.2 if continuously differentiable on $\{\theta : \theta \gg 0\}$.

Proof. Theorem 2.2, with $r = 0$, and the Poincaré equation together give

$$\rho^{-nd} \phi(\tilde{\rho}^{nd} \psi) = \rho^{-nd} m^{(nd)}(\phi(\psi)) \to l^{(0)}(\phi(\psi)),$$

as $n \to \infty$. Now, for $\psi \in (0, \infty)$, let

$$H(\psi) = \psi^{-\gamma} l^{(0)}(\phi(\psi)).$$

Then H is strictly positive and continuous because $l^{(0)}$ is. Also,

$$\begin{aligned} H(\tilde{\rho}^d \psi) &= \psi^{-\gamma} \rho^{-d} l^{(0)}(\phi(\tilde{\rho}^d \psi)) \\ &= \psi^{-\gamma} \rho^{-d} l^{(0)}(m^{(d)}(\phi(\psi))) \\ &= \psi^{-\gamma} \rho^{-d} \rho^d l^{(0)}(\phi(\psi)) \\ &= H(\psi), \end{aligned}$$

where the first equality uses the definitions of H and γ, the second uses the Poincaré functional equation and the third uses Corollary 2.4(i). This shows that H is multiplicatively periodic. Since $l^{(0)}(\theta)$ is increasing in θ and $\phi(\psi)$ is increasing in ψ, $\psi^{\gamma} H(\psi)$ is increasing in ψ and so, using $H(\psi) = H(\tilde{\rho}^d \psi)$, $H(\psi) \leq \rho^d H(1)$. Thus H is bounded. Using Theorem 2.2, with $r = 0$, and Corollary 2.3,

$$(\tilde{\rho}^{nd} \psi)^{-\gamma} \phi(\tilde{\rho}^{nd} \psi) = \psi^{-\gamma} \rho^{-nd} \phi(\tilde{\rho}^{nd} \psi) \to \psi^{-\gamma} l^{(0)}(\phi(\psi)) = H(\psi),$$

uniformly in $\psi \in [1, \tilde{\rho}^d]$, which gives (3.1).

The components of $\phi(\psi)$ are generating functions and hence continuously differentiable on $\psi > 0$. Hence H is continuously differentiable at $\psi > 0$ when $l^{(0)}$ is at $\phi(\psi)$. $\qquad\square$

In the one-type case, stronger results than (3.1) go back to Harris [6]: he showed that $\phi(\psi) - \psi^{\gamma} H(\psi)$ is bounded as $\psi \to \infty$. The last part of Theorem 5.1 in Section 5 goes a little further, showing that $\phi(\psi) - \psi^{\gamma} H(\psi)$ has a limit.

Note that Corollary 2.4(ii) gives $\rho^r l^{(r)}(\theta) = l^{(0)}(m^{(r)}(\theta))$ which can be combined with the Poincaré equation to confirm that $H(\tilde{\rho}^r \psi) = \psi^{-\gamma} l^{(r)}(\phi(\psi))$. Thus,

as might be anticipated, no extra information about H is obtained by applying Theorem 2.2 with r different from 0.

Jones [9] (who used $\hat{H}$ for H) obtained (3.1) under significantly stronger conditions. He assumes that the mean matrix of the branching process is primitive, rather than irreducible, but that is a well-understood and unimportant restriction. He also assumes that that there is a unique $B \in \mathcal{B}$ with eigenvalue ρ, that this B has a right eigenvector that is strictly positive in all components and that $\rho^{-n}B^n$ converges as $n \to \infty$. All this holds in the first example above except that $\rho^{-n}B^n$ does not converge. Uniqueness fails in the second example. Theorem 3.1 applies to both examples to give (3.1); Corollary 2.9 implies that H is also differentiable in the first one.

Biggins and Bingham [4] use the one-type Harris function in describing right-tail behaviour associated with the branching process and Jones [9], in his Theorem 9, gives a similar result for the multitype case. Differentiability of H is important in the application of large deviation theory to obtain these results.

4. Application to enumeration

As noted already, the original motivation for this study was the paper by Miller and O'Sullivan [14]. Interest here centres on the results in Sections 5 and 6 of that paper, which the authors describe as studying 'the combinatorics of the set of all trees that can be generated from context-free substitution rules'. A general introduction to context-free languages can be found in Hopcroft et al. [7, Chapter 5]. This framework has a rich vocabulary of its own, which will only be reproduced in enough detail to place it in the context of the present discussion.

The basic objects are 'variables', labelled by the finite set $\mathcal{P}$. A variable can be re-expressed, in terms of the same set of variables, according to a finite number of rewrite (or substitution, or production) rules. There may be several possible permitted substitutions for a variable. A single rewrite rule, associated with a particular variable, is a finite string from $\mathcal{P}$. Such a rule can be summarised by a vector b, indexed by $\mathcal{P}$ and recording how many times each of the variables occurs in the substituting string; note that b has integer components. It is possible, and allowed, that different rewrite rules produce the same b.

The generating function of the set of rules can now be defined by letting $\mu_i(b)$ be the number of times there are rules replacing the variable labelled i by a string summarised by b. Then, with 0 interpreted as the zero vector, $\exp(m_i(0))$ gives the number of rewrite rules when the initiating variable is i.

The rewrite rules are used recursively, generating a tree of possibilities. Miller and O'Sullivan [14] call the rules strongly connected if there is an n such that, starting from any variable, and making the substitutions repeatedly, every variable occurs in the tree at level n. This corresponds to the matrix $I(\mu)$ being primitive (positive regular) and, a fortiori, irreducible, and hence to m being irreducible.

The growth of the number of expressions generated by making all possible substitutions n times (or derivations of n levels) is one of the quantities studied by Miller and O'Sullivan [14]. It is easy to confirm, by induction, that $m^{(n)}(0)$ is the vector of the logarithm of these numbers for the various initial variables. Therefore, in the notation established here, Miller and O'Sullivan assume, in Definition 9, that $m_i^{(n+1)}(0)/m_i^{(n)}(0)$ has a limit as n goes to infinity. When the limit does exist, they characterise it in their Theorem 7 in the following way. It is the largest eigenvalue

possible for the mean matrix of a multivariate probability generating functions obtained by replacing the measures μ_i by probability distributions with the same supports. The probability generating functions define a multitype Galton-Watson process and so the limit is the fastest growth rate possible for such a process with these supports.

Theorem 2.2 shows that $m_i^{(n+1)}(0)/m_i^{(n)}(0)$ does indeed have a limit, as Miller and O'Sullivan assume, in the aperiodic case (i.e. when d is one) and then the limit is ρ, which is described, more transparently than in Miller and O'Sullivan, by (2.1). Clearly, the limit does not exist in the periodic case ($d > 1$). However, when $d > 1$, it is natural to replace the original rewrite rules by all those obtained by making d successive substitutions all at one go, moving straight to generation d. Then Theorem 2.2 again shows that the desired limit exists. In fact, the real interest is in the growth of $m^{(n)}(0)$ — the concern with the ratio of successive values is actually secondary. Theorem 2.2 supplies directly significantly more refined information on the growth than is given in Miller and O'Sullivan [14].

There is one further important ingredient that must be added. What have been called variables, labelled by $\mathcal{P}$, are more properly called non-terminal variables. Rewrite rules also contain terminal symbols, which, by definition, have no genuine rewrite rules associated with them. It is convenient to associate with each terminal the trivial rewrite rule of replacing a symbol with itself. Now the rewrite rules are summarised by a generating function $\widehat{m}((\theta, \vartheta)^T)$, where θ and ϑ code for the non-terminal variables and terminal symbols respectively. Thus, the m already defined ignores the terminal symbols. Because terminal symbols produce only terminal symbols and so $\widehat{m}$ is not irreducible, it seems the theory developed needs to be extended, but in fact this is not necessary.

Let $\widetilde{m}(\theta) = \widehat{m}_{\mathcal{P}}((\theta, 0)^T)$ and note that, because terminals' rewrite rules are trivial,

$$\widehat{m}\left((\theta, 0)^T\right) = (\widetilde{m}(\theta), 0)^T.$$

Then, by induction,

$$\widehat{m}_{\mathcal{P}}^{(n)}\left((\theta, 0)^T\right) = \widetilde{m}^{(n)}(\theta).$$

The function $\exp(\widetilde{m}(\theta))$ differs from $\exp(m(\theta))$ only by the addition of constants to the components. Thus $\widetilde{m}$ is irreducible, because m is, and the 'highest power' matrices are unchanged. Consequently, both generating functions have the same ρ, which encompasses the Corollary to Theorem 7 in Miller and O'Sullivan [14]. Now, $\widetilde{m}^{(n)}(0)$ gives the logarithm of the number of expressions after n substitutions and is covered by Theorem 2.2. Furthermore, with $-\infty$ interpreted as the vector with all components $-\infty$, $\widetilde{m}^{(n)}(-\infty)$ gives the logarithm of the number of expressions made up only of terminal symbols after n substitutions. To apply Theorem 2.2 to this, first compute $\widetilde{m}^{(\nu)}(-\infty)$ directly for a suitable ν and then consider the asymptotics of $\widetilde{m}^{(n-\nu)}(\widetilde{m}^{(\nu)}(-\infty))$. The example given next illustrates these ideas.

Miller and O'Sullivan [14] discuss the Arithmetic Expression Grammar. In this there are three non-terminal variables

(4.1) $$\{\text{expression, term, factor}\}$$

and the terminal symbols

$$\{+, -, \times, /, (,), \text{letter}\}$$

— here the terminal symbol 'letter' aggregates a finite alphabet of ℓ possiblities (e.g. $a, A, b, B, \dots$). The eight non-trivial rewrite rules are:

$$\begin{aligned}
\text{expression} &\longrightarrow \text{expression} + \text{term} \\
\text{expression} &\longrightarrow \text{expression} - \text{term} \\
\text{expression} &\longrightarrow \text{term}
\end{aligned}$$

$$\begin{aligned}
\text{term} &\longrightarrow \text{term} \times \text{factor} \\
\text{term} &\longrightarrow \text{term} \, / \, \text{factor} \\
\text{term} &\longrightarrow \text{factor}
\end{aligned}$$

$$\begin{aligned}
\text{factor} &\longrightarrow (\text{expression}) \\
\text{factor} &\longrightarrow \text{letter}
\end{aligned}$$

Note that the first two rules differ only in the terminal symbols they include, as do the fourth and fifth, and so will produce the same b when computing m. The final rule is the only one that produces a terminal symbol but no non-terminal variables. Labelling the three non-terminals with integers in the order they appear in (4.1), these rules translate into the generating function for non-terminals:

$$\begin{aligned}
\exp(m_1(\log s)) &= 2s_1 s_2 + s_2; \\
\exp(m_2(\log s)) &= 2s_2 s_3 + s_3; \\
\exp(m_3(\log s)) &= s_1.
\end{aligned}$$

Thus the matrices in $\mathcal{B}$ have a first row that is either $(1,1,0)$ or $(0,1,0)$ a second row that is either $(0,1,1)$ or $(0,0,1)$ and a third row that is $(1,0,0)$. The one with largest eigenvalue is the primitive matrix

$$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \end{pmatrix},$$

which gives $\rho \ (= 1.754\dots)$ as its largest eigenvalue, as already noted by Miller and O'Sullivan [14], with corresponding eigenvector

$$a = \begin{pmatrix} 1 \\ \rho - 1 \\ 1/\rho \end{pmatrix} = \begin{pmatrix} 1 \\ 0.754\dots \\ 0.569\dots \end{pmatrix}.$$

Since the matrix is primitive, Corollary 2.6 applies. Thus the period is one and the limit, $l^{(0)}(\theta)$, has the form $k(\theta)a$ where $k(\theta)$ is a scalar function. Furthermore, Corollary 2.9 implies that k is differentiable on the interior of $\mathfrak{T}$, which certainly contains $\{\theta : \theta \geq 0\}$. It is possible to compute, or, more precisely, estimate, $l^{(0)}$, and hence k, numerically using an iterative scheme for $\rho^{-n} m^{(n)}(\theta)$, based on equation (8.1) below (reassured by the knowledge the limit exists). This gave $k(0) = 1.1567\dots$. It was also confirmed numerically that $l^{(0)}$ is a scalar multiple of a and that $k(m(0)) = \rho k(0) = 2.0299\dots$.

Turning to the inclusion of non-terminals, the relevant generating function when there are ℓ letters is

$$\widetilde{m}(\log s) = \log \begin{pmatrix} 2s_1 s_2 + s_2 \\ 2s_2 s_3 + s_3 \\ s_1 + \ell \end{pmatrix}.$$

Let $\tilde{k}_\ell(\theta)$ be the k, from Corollary 2.6, associated with this generating function. Starting from a single initial non-terminal variable i, the number of expressions

after n substitutions is $\exp(\widetilde{m}_i^{(n)}(0))$ and, using Theorem 2.2,

$$\widetilde{m}_i^{(n)}(0) \sim \left[\tilde{k}_\ell(0)\right]\rho^n a_i.$$

Let $t^* = \widetilde{m}^{(3)}(-\infty)$. Substitution gives

$$t^* = \widetilde{m}\left(\widetilde{m}\left(\widetilde{m}\left(\log\begin{pmatrix}0\\0\\0\end{pmatrix}\right)\right)\right) = \log\begin{pmatrix}\ell\\2\ell^2+\ell\\\ell\end{pmatrix} \geq 0.$$

Starting from a single initial non-terminal variable i, the number of expressions with only terminals after n substitutions is $\exp(\widetilde{m}_i^{(n)}(-\infty))$ and Theorem 2.2 gives

$$\widetilde{m}_i^{(n)}(-\infty) = \widetilde{m}_i^{(n-3)}(t^*) \sim \tilde{k}_\ell(t^*)\rho^{n-3}a_i = \left[\rho^{-3}\tilde{k}_\ell(t^*)\right]\rho^n a_i.$$

Since there are more possibilities when expressions include also non-terminal variables, it follows that $\rho^{-3}\tilde{k}_\ell(t^*) \leq \tilde{k}_\ell(0)$.

The iterative calculation already used to obtain $k(0)$ can be used to obtain the following values, given to three decimal places — they confirm that $\rho^{-3}\tilde{k}_\ell(t^*) \leq \tilde{k}_\ell(0)$.

ℓ	0	1	2	4	8	16	32	64	128	256
$\tilde{k}_\ell(0)$	1.157	1.307	1.403	1.531	1.689	1.870	2.069	2.282	2.503	2.729
$\rho^{-3}\tilde{k}_\ell(t^*)$	—	0.329	0.524	0.738	0.962	1.189	1.420	1.651	1.883	2.115

5. The one-dimensional case

The case where $p = 1$ is examined first. The results themselves have little novelty, but the proofs indicate the line that will be taken when $p > 1$. As has been mentioned already, here $\rho = \max\{b : \mu(b) > 0\}$, which is the highest power of e^θ that occurs in $e^{m(\theta)}$. Let $q = \log\mu(\rho)$, so that the coefficient of $e^{\rho\theta}$ in $e^{m(\theta)}$ is e^q. Now let the function g be defined by

(5.1) $$m(\theta) = \rho\theta + q + g(\theta) = \rho\theta + \log\mu(\rho) + g(\theta),$$

which is the one-dimensional version of (2.2).

Theorem 5.1. *On $\mathfrak{T} = \{\theta : m^{(n)}(\theta) \to \infty\}$, $\rho^{-n}m^{(n)}(\theta)$ converges to a limit $l(\theta)$, given by*

$$l(\theta) = \theta + \frac{q}{(\rho-1)} + \sum_{i=1}^{\infty}\frac{g(m^{(i-1)}(\theta))}{\rho^i}.$$

This limit is strictly positive, finite, continuous and convex on $\mathfrak{T}$ and satisfies the functional equation $\rho l(\theta) = l(m(\theta))$. Furthermore,

$$(m^{(n)}(\theta) - \rho^n l(\theta)) \to q/(\rho-1),$$

uniformly on any $[a, \infty) \subset \mathfrak{T}$.

Proof. Note first that $m(-\theta)$, as the logarithm of the Laplace transform of a measure on $[0, \infty)$, is a decreasing convex function of θ. It is easily checked that because m is convex and increasing $m^{(n)}$ is too. Therefore, for some s, $\mathfrak{T}$ is either (s, ∞) or $[s, \infty)$.

With the convention that $m^{(0)}(\theta) = \theta$, (5.1) gives, for any positive integer n,

$$\frac{m^{(n)}(\theta)}{\rho^n} = \frac{m^{(n-1)}(\theta)}{\rho^{n-1}} + \frac{q}{\rho^n} + \frac{g(m^{(n-1)}(\theta))}{\rho^n}.$$

Iterating gives, for $n \geq \nu \geq 0$,

$$(5.2) \qquad \frac{m^{(n)}(\theta)}{\rho^n} = \frac{m^{(\nu)}(\theta)}{\rho^\nu} + \sum_{i=\nu+1}^{n} \left(\frac{q}{\rho^i} + \frac{g(m^{(i-1)}(\theta))}{\rho^i} \right).$$

Also, rearranging (5.1) and substituting for m,

$$e^{g(\theta)} = 1 + \sum_{b < \rho} \frac{\mu(b)}{\mu(\rho)} e^{-\theta(\rho-b)},$$

and so $g(\theta)$ is finite, non-negative and tends to zero as $\theta \to \infty$. Since g is non-negative and (by standing assumption) $\rho > 1$, (5.2) gives

$$(5.3) \qquad \frac{m^{(n)}(\theta)}{\rho^n} \to \frac{m^{(\nu)}(\theta)}{\rho^\nu} + \frac{q}{\rho^\nu(\rho-1)} + \sum_{i=\nu+1}^{\infty} \frac{g(m^{(i-1)}(\theta))}{\rho^i} = l(\theta).$$

Taking $\nu = 0$ in (5.3) gives the required representation of l. For $\theta \in \mathfrak{T}$, the sequence $\{m^{(n)}(\theta) : n \in \mathbb{Z}_+\}$ is bounded below, and so $\{g(m^{(n)}(\theta)) : n \in \mathbb{Z}_+\}$ is bounded above, which forces l to be finite. Furthermore, there is some ν such that $m^{(\nu)}(\theta) + q(\rho-1)^{-1}$ is strictly positive, and so $l(\theta)$ is too.

Let $[a, \infty) \subset \mathfrak{T}$, so that $m^{(n)}(a) \to \infty$, Then, since $m^{(n)}(\theta)$ is increasing in θ and $g(\theta)$ tends to zero as $\theta \to \infty$,

$$(5.4) \qquad \sup\{g(m^{(r)}(\theta)) : \theta \geq a, r \geq n\} \to 0$$

as $n \to \infty$. Consequently,

$$\sum_{i=n+1}^{\infty} \frac{g(m^{(i-1)}(\theta))}{\rho^i} \to 0$$

uniformly in $\theta \geq a$ and then the continuous, convex functions $\rho^{-n} m^{(n)}(\theta)$ converge uniformly to $l(\theta)$. Hence, l is continuous and convex on $\mathfrak{T}$. Letting n go to infinity in $\rho \rho^{-n} m^{(n)}(\theta) = \rho^{-(n-1)} m^{(n-1)}(m(\theta))$ gives the functional equation.

Now, observe that

$$m^{(n)}(\theta) - \rho^n l(\theta) = \rho^n \sum_{i=n+1}^{\infty} \left(\frac{q}{\rho^i} + \frac{g(m^{(i-1)}(\theta))}{\rho^i} \right)$$

$$= \frac{q}{(\rho-1)} + \sum_{i=1}^{\infty} \frac{g(m^{(n+i-1)}(\theta))}{\rho^i} \to \frac{q}{(\rho-1)},$$

uniformly in $\theta \geq a$, using (5.4). $\qquad \square$

Theorem 5.2. *The limit l is differentiable on the interior of $\mathfrak{T}$.*

Proof. The limit is differentiable if the term-by-term derivative of the limit is uniformly convergent, Apostol [1, Theorem 13.14], that is if

$$\sum_{i=1}^{\infty} \frac{g'(m^{(i-1)}(\theta))}{\rho} \frac{(m^{(i-1)}(\theta))'}{\rho^{i-1}}$$

is uniformly convergent. Let $\mathfrak{C}$ be a compact subset of the interior of $\mathfrak{T}$. The functions $\{m^{(n)}(\theta)/\rho^n\}$ are convex and converge to a finite limit, l, on the interior of $\mathfrak{T}$; hence, they must converge uniformly to l on $\mathfrak{C}$ and the modulus of their derivatives must be bounded on $\mathfrak{C}$. Thus, for a suitable N, for all $n \geq N$ there is

a finite K_1 and positive K_2 such that $|(m^{(n)}(\theta))'/\rho^n| \le K_1$ and $m^{(n)}(\theta) \ge K_2\rho^n$ throughout $\mathfrak{C}$. In addition, using the definition of g,

$$0 \le -g'(\theta) = e^{-g(\theta)} \sum_{b<\rho} (\rho - b)\frac{\mu(b)}{\mu(\rho)}e^{-\theta(\rho-b)} \le \sum_{b<\rho}(\rho-b)\frac{\mu(b)}{\mu(\rho)}e^{-\theta(\rho-b)}.$$

Since μ is discrete, $\inf\{(\rho - b) : b < \rho, \ \mu(b) > 0\} > 0$ and so, for a suitable finite κ_1 and positive κ_2,

$$0 \le -g'(\theta) \le \kappa_1 e^{-\kappa_2\theta}.$$

Combining all this, for $n \ge N$ and $\theta \in \mathfrak{C}$,

$$\left| \frac{g'(m^{(n)}(\theta))}{\rho} \frac{(m^{(n)}(\theta))'}{\rho^n} \right| \le K_1\kappa_1 \exp(-\kappa_2 m^{(n)}(\theta)) \le K_1\kappa_1 \exp(-\kappa_2 K_2\rho^n),$$

which, rather comfortably, gives the desired uniform convergence. $\qquad\square$

It is possible to go further, considering the convergence of $m^{(n)}$ in a complex neighbourhood of θ, but that would be a distraction from the main development here and will be considered elsewhere. The results in this section do not rely that heavily on the way m is defined. The important feature is that there is a $\rho > 1$ such that the function g defined by (5.1) has the right properties. For example, in Theorem 5.1, μ could be any measure on $[0, \rho]$ (or even $(-\infty, \rho]$, if $\mathfrak{T}$ is restricted to the θ where m is increasing) with weight $e^q > 0$ at $\rho > 1$, and then Theorem 5.2 will also hold if there is an $\kappa > 1$ such that, for all large θ, $\int(\rho - b)e^{-\theta(\rho-b)}\mu(db) \le (\log\theta)^{-\kappa}$.

6. A bounding lemma

Recall that $\mathcal{B}$ is the set of matrices defined by

$$\mathcal{B} = \{B : \mu_i(B_i^T) > 0, i \in \mathcal{P}\}$$

so that when $B \in \mathcal{B}$, its ith row is the position of an atom of μ_i for each i.

A set of matrices will be said to be *closed under row substitution* if when A and B are in the set all other matrices obtained by substituting rows in B for the corresponding ones in A are also in the set. The inclusion criterion for a row in the construction of a member of $\mathcal{B}$ is independent of the other rows; hence $\mathcal{B}$ is closed under row substitution. Note too that $\mathcal{B}$ is finite because μ has only finitely many atoms.

Suppose $\mathcal{S}$ is a set of non-negative matrices. Kennedy [10] defines $\mathcal{S}$ to be irreducible if for any i and j there is some product of matrices from $\mathcal{S}$ that has a strictly positive (i, j)th entry, and he defines $\mathcal{S}$ to have the *optimal choice property* when for every non-negative vector x there is a $B \in \mathcal{S}$ such that $Bx = \sup\{Ax : A \in \mathcal{S}\}$. Related material can be found in Seneta [16, §3.1] (but not in Seneta [17]) and in Grey [5]. It is easy to check that the set $\mathcal{B}$ being irreducible in Kennedy's sense is equivalent to the matrix $I(\mu)$, introduced at the start of Section 2, being irreducible. The finiteness of $\mathcal{B}$ and its row-substitution property ensure that it has the *strong optimal choice property*, that for every vector x, not just non-negative ones, there is a $B \in \mathcal{B}$ such that $Bx = \sup\{Ax : A \in \mathcal{B}\}$. The theory developed in Kennedy [10] on the products of matrices drawn from a set with the optimal choice property will be useful here. The reason for this soon emerges once a proof along the lines of that given for Theorem 5.1, where $p = 1$, is attempted.

Recall that, by (2.2), $m(\theta) = B\theta + q(B) + g(B; \theta)$ and that g has non-negative components. However, in a significant contrast to the one-dimensional case, there may be no fixed B such that $g(B; \theta)$ is bounded on $\{\theta : \theta \geq c\}$. For the initial estimates, this will necessitate allowing the B used in the expansion (2.2) to depend on θ.

The strong optimal choice property means that, for every θ there is a $B(\theta) \in \mathcal{B}$ such that
$$B(\theta)\theta = \sup\{B\theta : B \in \mathcal{B}\}.$$
Obviously the idea is that, when the argument of m is a θ with strictly positive components, $B(\theta)$ supplies the analogue of the highest power in one dimension. Note that only the direction of θ matters, not its magnitude. (It is perhaps also worth noting that for θ with all components negative $B(\theta)$ is the analogue of the lowest power.)

In the next lemma, and elsewhere, the matrix $\prod_{i=1}^{n} B^{(i)}$ should be interpreted as the backward product $B^{(n)} \cdots B^{(2)} B^{(1)}$ (rather than $B^{(1)} B^{(2)} \cdots B^{(n)}$). The matrices in the lemma are drawn from $\mathcal{B}$. Once a suitable bound on the growth of a matrix product from $\mathcal{B}$ has been obtained, a maximum growth rate for $m^{(n)}(\theta)$ will follow from this result.

Lemma 6.1. *Let* $m^{(0)}(\theta) = \theta$, $B^{(n)} = B(m^{(n-1)}(\theta))$ *and* $g^{(n)} = g(B^{(n)}; m^{(n-1)}(\theta))$.
(i) *Suppressing the argument* θ,
$$m^{(n)} = \left(\prod_{i=1}^{n} B^{(i)}\right) m^{(0)} + \sum_{k=1}^{n} \left(\prod_{i=k+1}^{n} B^{(i)}\right) \left(q\left(B^{(k)}\right) + g^{(k)}\right).$$

(ii) *The sequences* $\{q(B^{(k)})\}$ *and* $\{g^{(k)}\}$ *are bounded uniformly in* θ. *More specifically, let* t_i *be the logarithm of the total mass of* μ_i *and let* u_i *be the logarithm of the smallest non-zero value of* $\mu_i(b)$; *then, for any* $B \in \mathcal{B}$, $u \leq q(B) \leq t$ *and* $0 \leq g(B(\theta); \theta) \leq t - u$.

Proof. Take B to be $B^{(n)}$ in (2.2) to give
$$m^{(n)} = B^{(n)} m^{(n-1)} + q\left(B^{(n)}\right) + g^{(n)}.$$

Induction now gives the required representation.

The bounds for $q(B)$ are immediate and the non-negativity of g has already been noted. The definition of $B(\theta)$ implies that $b^T \theta \leq B(\theta)_i \theta$ whenever $\mu_i(b) > 0$ and so
$$m_i(\theta) = \log\left(\sum_b \mu_i(b) \exp(b^T \theta)\right) \leq \log\left(\exp(B(\theta)_i \theta) \sum_b \mu_i(b)\right) = t_i + B(\theta)_i \theta.$$

Hence,
$$g(B(\theta); \theta) = m(\theta) - B(\theta)\theta - q(B(\theta)) \leq t - u. \qquad \square$$

7. MATRIX PROPERTIES

Lemma 6.1 focuses attention on the growth of the product of matrices drawn from an irreducible collection with the row substitution property or with the, weaker, optimal choice property. However, before turning to the results needed about collections of matrices, a few simple results about a single non-negative matrix are recorded.

The discussion must consider irreducible sub-matrices, which requires some care over terminology. For a non-negative square matrix B, say 'index i leads to index j in B' if some positive integer power of B has a strictly positive (i, j)th entry; also, by convention, i leads to itself. The self-communicating classes defined by a matrix B are collections of indices in which every one leads to every other. A matrix B is irreducible if its indices form a single self-communicating class in B. If $\mathcal{I}$ is a self-communicating class in B then $B_{\mathcal{I},\mathcal{I}}$ is irreducible. Furthermore, if $B_{\mathcal{I},\mathcal{I}}$ is irreducible then $\mathcal{I}$ must be a subset of the indices of a self-communicating class in B.

Lemma 7.1. *The largest eigenvalue of a non-negative matrix is the maximum of the largest eigenvalues of the sub-matrices indexed by its self-communicating classes.*

Proof. Observe that when

$$B = \left(\begin{array}{cc} B(1) & 0 \\ B(3) & B(4) \end{array} \right)$$

the eigenvalues of B are exactly those of $B(1)$ and $B(4)$. Applying this and the canonical form of a non-negative matrix, described in Seneta [16, 17, §1.2], gives the result. $\square$

Lemma 7.2. *Let C be block diagonal, with each block an irreducible matrix with eigenvalue η and let $d \, (= d(C))$ be the lowest common multiple of the periods of the blocks in C. There is a block diagonal idempotent matrix $\tilde{L} \, (= \tilde{L}(C))$ such that*

$$(7.1) \qquad \eta^{-nd} C^{nd+r} \to C^r \tilde{L} = \tilde{L} C^r \text{ as } n \to \infty$$

and $\tilde{L} C^d = \eta^d \tilde{L}$. Furthermore, there is an $\zeta < \eta$ such that

$$(7.2) \qquad \zeta^{-n} \| C^{nd+r} - \eta^{nd} \tilde{L} C^r \| \to 0 \text{ as } n \to \infty.$$

Proof. Consider first an irreducible matrix R with period d'. The matrix $R^{d'}$ consists of d' square matrices on its diagonal, each of which is primitive (i.e. has some power with all entries positive), and zeros elsewhere; see the proof of Theorem 1.4 in Seneta [16, 17]. Theorem 1.2 in Seneta [16, 17] now gives the convergence, geometrically quickly, of each diagonal block in $\eta^{-nd'} R^{nd'}$, as $n \to \infty$ to a strictly-positive (rank one) idempotent matrix. Applying this to each diagonal block in C gives all the claimed results. $\square$

A non-negative matrix B with maximal eigenvalue $\eta > 0$ will be said to be *structured simply* when, for suitable sets of indices $\mathcal{I}$ and $\mathcal{E}$, B has the form

$$(7.3) \qquad B = \left(\begin{array}{cc} B_{\mathcal{I},\mathcal{I}} & B_{\mathcal{I},\mathcal{E}} \\ B_{\mathcal{E},\mathcal{I}} & B_{\mathcal{E},\mathcal{E}} \end{array} \right) = \left(\begin{array}{cc} C & 0 \\ F & D \end{array} \right),$$

where C is block diagonal, with each block an irreducible matrix with eigenvalue η, and D has maximum eigenvalue strictly less than η.

Lemma 7.3. *Suppose B is structured simply. Let d and $\tilde{L}$ be associated with $C \, (= B_{\mathcal{I},\mathcal{I}})$ as in Lemma 7.2. There is an idempotent matrix $L \, (= L(B))$ given by*

$$L = \left(\begin{array}{cc} \tilde{L} & 0 \\ \sum_{l=1}^{\infty} \sum_{c=0}^{d-1} D^{ld-c-1} F \tilde{L} C^c & 0 \end{array} \right)$$

and a $\zeta < \eta$ such that

$$\zeta^{-n} \| B^{nd+r} - \eta^{nd} L B^r \| \to 0$$

as $n \to \infty$, *and* $\eta^{-d} B^d L = L$.

Proof. Using the form of B,

$$B^k = \left(\begin{array}{cc} C & 0 \\ F & D \end{array} \right)^k = \left(\begin{array}{cc} C^k & 0 \\ \sum_{l=0}^{k-1} D^l F C^{k-l} & D^k \end{array} \right).$$

Applying (7.2), $\|\eta^{-nd} C^{nd} - \tilde{L}\| \to 0$, geometrically quickly, as $n \to \infty$. Then the fact that the maximum eigenvalue of D is less than η and simple bounding give the limiting behaviour of B^{nd}. $\qquad\square$

Let $\mathcal{S}$ be a finite set of non-negative matrices that is irreducible and has the optimal choice property. The next theorem, which is extracted from Section 3 of Kennedy [10], is the starting point for the study of such collections of matrices. Kennedy defines the constant $1/\gamma$ as the radius of convergence of certain power series, but that will not matter here. It is worth pointing out straightaway that the vector w and the matrix A in this result need not be unique.

Theorem 7.4. *Suppose $\mathcal{S}$ is a finite collection of non-negative matrices that is irreducible and has the optimal choice property. There is a vector $w \gg 0$, an $A \in \mathcal{S}$ and a positive γ such that*

$$\gamma w = A w = \sup_{B \in \mathcal{S}} B w.$$

Corollary 7.5. *For any non-negative matrix B, let $\eta(B)$ be its 'Perron-Frobenius' eigenvalue. Then*

$$\gamma = \gamma(\mathcal{S}) = \max_{B \in \mathcal{S}} \eta(B).$$

Proof. For any $B \in \mathcal{S}$, $\gamma w \geq Bw$. Multiplying this by a non-negative left eigenvector corresponding to the maximum eigenvalue of B gives $\gamma \geq \eta(B)$ and $\gamma = \eta(A)$. $\qquad\square$

Theorem 7.4 immediately yields a simple bound on matrix products, given next, which will be critical for bounding the growth of $m^{(n)}$. The important point about this bound is that it does not depend on the actual terms in the product but only on their number.

Corollary 7.6. *Suppose $\{B^{(i)}\}$ are in $\mathcal{S}$. Let E be the matrix with every entry equal to one. There is a finite scalar κ such that*

$$\prod_{i=1}^{n} B^{(i)} \leq \kappa \gamma^n E.$$

Proof. By Theorem 7.4 and Corollary 7.5, $\prod_{i=1}^{n} B^{(i)} w \leq \gamma^n w$. Now let κ be the ratio of the maximum and minimum components of w, which is finite because w is strictly positive. $\qquad\square$

In the main proof, products drawn from collections of non-negative matrices that are not irreducible will also need to be considered. Sufficient information follows easily from the irreducible case and is recorded next.

Corollary 7.7. *Let $\widetilde{\mathcal{S}}$ be any finite set of non-negative matrices with the optimal choice property and maximum eigenvalue $\tilde{\gamma}$ and let $\epsilon > 0$.*

(i) *There is a vector $w(\epsilon) \gg 0$ such that*

$$\sup_{B \in \mathcal{S}} Bw(\epsilon) \leq (\tilde{\gamma} + \epsilon)w(\epsilon).$$

(ii) *The maximum eigenvalue of the product of n members of $\widetilde{\mathcal{S}}$ is no more than $\tilde{\gamma}^n$.*

Proof. For $\delta > 0$, let $\widetilde{\mathcal{S}}_\delta$ be the set of matrices obtained by adding δE to every matrix in $\widetilde{\mathcal{S}}$. This collection is irreducible and has the optimal choice property. The maximal eigenvalue of $\widetilde{\mathcal{S}}_\delta$ goes down to $\tilde{\gamma}$ as $\delta \downarrow 0$ and for any non-negative vector x, $Bx \leq (B + \delta E)x$. Thus, applying Theorem 7.4 to $\widetilde{\mathcal{S}}_\delta$ for a suitably small δ gives (i). Now let B^* be the product of n matrices from $\widetilde{\mathcal{S}}$. Using (i), for small enough δ,

$$(B^* + \delta E)\, w(\epsilon) \leq ((\tilde{\gamma} + \epsilon)^n + \epsilon)w(\epsilon).$$

Therefore, by Theorem 1.6 in Seneta [16, 17], $B^* + \delta E$ has maximum eigenvalue at most $(\tilde{\gamma} + \epsilon)^n + \epsilon$. Since $0 \leq B^* \leq B^* + \delta E$ the same is true of B^*. $\square$

The row-substitution property imposes constraints on the form that matrices in $\mathcal{S}$ with eigenvalue γ can take. The next result, which considers this, is critical for the main results developed here.

Proposition 7.8. *Suppose $\mathcal{S}$ is a finite collection of non-negative matrices that is irreducible, closed under row substitution, and has maximum eigenvalue γ.*

(i) *Every $B \in \mathcal{S}$ with eigenvalue γ is structured simply (as defined around (7.3)).*

(ii) *There is an $A \in \mathcal{S}$ that is structured simply with eigenvalue γ and with $\mathcal{E}$ having the property that $B_{\mathcal{E},\mathcal{E}}$ has a maximum eigenvalue strictly less than γ for every $B \in \mathcal{S}$ (not just for A).*

Proof. By Lemma 7.1, when B has eigenvalue γ it has the form

$$B = \begin{pmatrix} R & B_{\mathcal{I}',\mathcal{E}'} \\ B_{\mathcal{E}',\mathcal{I}'} & B_{\mathcal{E}',\mathcal{E}'} \end{pmatrix},$$

with R irreducible and with eigenvalue γ. Suppose, for a contradiction, that for some $i \in \mathcal{I}'$ and $j \in \mathcal{E}'$ the entry $B_{i,j}$ is positive. Since $\mathcal{S}$ is irreducible, there is a sequence of distinct indices $j(1), j(2), \ldots, j(k)$ with $j(1) = j$, $j(l) \in \mathcal{E}'$ for $l = 1, 2 \ldots, k-1$ and $j(k) \in \mathcal{I}'$, and matrices $B(l) \in \mathcal{S}$, such that $B(l)_{j(l),j(l+1)} > 0$. Use the row-substitution property to replace the $j(l)$th row of B by the $j(l)$th row of $B(l)$ to obtain another matrix in $\mathcal{S}$. The new matrix has R as a sub-matrix of a strictly larger irreducible matrix (indexed by $\mathcal{I}'$ and, at least, $j(1), j(2), \ldots, j(k-1)$), which must therefore have eigenvalue strictly greater than γ. This contradicts the definition of γ, and so $B_{\mathcal{I}',\mathcal{E}'}$ must contain only zeros. The same argument applies to any irreducible sub-matrix of $B_{\mathcal{E}',\mathcal{E}'}$ with eigenvalue γ. Collecting all these gives C in (7.3), and proves (i).

Now suppose A' is structured simply with eigenvalue γ but does not satisfy (ii), so that there are indices $\mathcal{I}_1 \subset \mathcal{E}$ and some matrix $B \in \mathcal{S}$ such that $B_{\mathcal{I}_1,\mathcal{I}_1}$ is irreducible with (maximum) eigenvalue γ. Construct $A'' \in \mathcal{S}$ by substituting the rows of B indexed by $\mathcal{I}_1$ for those in A'. Then A'' also has the structure (7.3), but with an additional block in C and hence a smaller set $\mathcal{E}$. Repeating all this a finite number of times will lead to a suitable A. $\square$

Clearly, A in part (ii) is, typically, not unique since the row-substitution property can be used to replace $A_{\mathcal{E},\mathcal{P}}$. It is not hard to go further and construct an $\mathcal{S}$ in which $\mathcal{E}$ in (ii) is also not unique.

8. Convergence for some indices

The first lemma, from which Proposition 2.1 follows directly, records some simple properties of $m^{(n)}$.

Lemma 8.1. *For each n, the function $m^{(n)}$ is increasing and convex in θ. Let $\mathfrak{T} = \{\theta : \|m^{(n)}(\theta)\| \to \infty \text{ as } n \to \infty\}$. If $\theta \in \mathfrak{T}$ then $m_i^{(n)}(\theta) \to \infty$ as $n \to \infty$ for all $i \in \mathcal{P}$. For every $\theta \in \mathfrak{T}$, there is an ν such that $m^{(n)}(\theta) \gg 0$ for all $n \geq \nu$.*

Proof. It is easy to confirm that the iterates $m^{(n)}$ are increasing and convex in θ if m is; now note that, for each i, $m_i(-\theta)$ is the logarithm of the Laplace transform of a measure on $[0,\infty)^p$ and so is a decreasing convex function.

Let $\mathfrak{T}_i = \{\theta : m_i^{(n)}(\theta) \to \infty\}$. When $\mu_i(b) > 0$ for some b with $b_j > 0$,

$$\exp\left(m_i^{(n+1)}(\theta)\right) = \exp\left(m_i(m^{(n)}(\theta))\right) > \mu_i(b)\exp\left(b_j m_j^{(n)}(\theta)\right),$$

which implies that $\mathfrak{T}_j \subset \mathfrak{T}_i$. Hence $\mathfrak{T} = \mathfrak{T}_i$ for every i, since, by assumption, μ is irreducible. The final assertion follows immediately. $\square$

The set of non-negative matrices $\mathcal{B}$ is finite and closed under row substitution, and so the results in Section 7 apply with $\gamma = \gamma(\mathcal{B}) = \rho$. The first result on the growth of $m^{(n)}$ is a simple, but important, bound.

Lemma 8.2. *For any bounded set $\mathfrak{B}$,*

$$\max_i \sup_{\theta \in \mathfrak{B}} \sup_n \frac{m_i^{(n)}(\theta)}{\rho^n} < \infty.$$

In particular, once its existence is established, the limit in Theorem 2.2 must be bounded above.

Proof. Dividing the representation for $m^{(n)}$ given in Lemma 6.1(i) through by ρ^n and using Corollary 7.6 gives, for some finite scalar κ,

$$\frac{m^{(n)}}{\rho^n} \leq \kappa E\left(m^{(0)} + \sum_{k=1}^n \frac{q(B^{(k)})}{\rho^k} + \sum_{k=1}^n \frac{g^{(k)}}{\rho^k}\right).$$

By assumption, $\rho > 1$. By Lemma 6.1(ii), $\{g^{(k)}\}$ and $\{q(B^{(k)})\}$ are uniformly bounded in k (and the suppressed argument θ). Finally, $m^{(0)}(\theta) = \theta$ and so is bounded on $\mathfrak{B}$. $\square$

A fixed B will be used in the rest of the section and so the argument B will be dropped from $q(B)$ and $g(B;\theta)$ to ease the notation. With this convention

$$(8.1) \qquad \frac{m^{(n)}(\theta)}{\rho^n} = \frac{B}{\rho}\frac{m^{(n-1)}(\theta)}{\rho^{n-1}} + \frac{q}{\rho^n} + \frac{g(m^{(n-1)}(\theta))}{\rho^n},$$

and so, iterating this, taking ν from Lemma 8.1,

$$(8.2) \qquad \frac{m^{(n)}(\theta)}{\rho^n} = \frac{B^{n-\nu}}{\rho^{n-\nu}}\frac{m^{(\nu)}(\theta)}{\rho^\nu} + \sum_{l=\nu+1}^n \frac{B^{n-l}}{\rho^{n-l}}\left(\frac{q}{\rho^l} + \frac{g(m^{(l-1)}(\theta))}{\rho^l}\right).$$

Clearly, (5.2) is the one-dimensional version of (8.2). In (5.2), understanding the tail sum was the key to studying what happened as $n \to \infty$. Here, in addition, the behaviour of $\{\rho^{-n}B^n\}$ has to be taken into account. Since $\mathcal{B}$ is a finite set of non-negative matrices and closed under row substitution, Proposition 7.8 helps with this.

Lemma 8.3. *Take $B \in \mathcal{B}$ with eigenvalue ρ. By Proposition 7.8(i), B is structured simply. Then, with $g(\theta) = g(B; \theta)$ and $\mathfrak{I}$ from (7.3),*

$$\sum_{i \in \mathfrak{I}} \sum_{l=1}^{\infty} \frac{g_i(m^{(l-1)}(\theta))}{\rho^l} < \infty.$$

Proof. By Lemma 8.2, the left-hand side of (8.2) is bounded above, and so the right-hand side is also. Let $\tilde{v} \gg 0$ be a left eigenvector of C for the eigenvalue ρ and let $v_i = \tilde{v}_i$ for $i \in \mathfrak{I}$ and $v_i = 0$ for all other i. Then v is a left eigenvector of B. Taking the inner product of (8.2) with v and then letting $n \to \infty$ shows that

$$\sum_{l=\nu+1}^{\infty} v^T \frac{g(m^{(l-1)}(\theta))}{\rho^l} = \sum_{i \in \mathfrak{I}} \sum_{l=\nu+1}^{\infty} \tilde{v}_i \frac{g_i(m^{(l-1)}(\theta))}{\rho^l}$$

must be finite. Since $\tilde{v} \gg 0$, the result follows. $\qquad\square$

For any vector v, recall that $v_{\mathfrak{I}}$ is its restriction to the index set $\mathfrak{I}$.

Proposition 8.4. *Take $B \in \mathcal{B}$ with eigenvalue ρ. By Proposition 7.8(i), B is structured simply, with associated $\mathfrak{I}$ and C from (7.3). Let d and $\tilde{L}$ arise from C as in Lemma 7.2. Then*

$$\frac{m_{\mathfrak{I}}^{(nd+r)}(\theta)}{\rho^{nd+r}} \to l_{\mathfrak{I}}^{(r)}(\theta)$$

as $n \to \infty$, where $l_{\mathfrak{I}}^{(r)}(\theta)$ is finite, convex and strictly positive on $\mathfrak{T}$. Furthermore,

$$l_{\mathfrak{I}}^{(r)}(\theta) = \rho^{-r} C^r \tilde{L} l_{\mathfrak{I}}(\theta)$$

where

$$(8.3) \qquad l_{\mathfrak{I}}(\theta) = \theta_{\mathfrak{I}} + \frac{1}{\rho^d - 1} \left(\sum_{c=0}^{d-1} C^c q_{\mathfrak{I}} \right) + \sum_{l=1}^{\infty} \rho^{-ld} \left(\sum_{c=0}^{d-1} C^c g_{\mathfrak{I}}(m^{(ld-c-1)}(\theta)) \right),$$

and $C l_{\mathfrak{I}}^{(r)}(\theta) = \rho l_{\mathfrak{I}}^{(r+1)}(\theta)$.

Proof. Take kd larger than the ν identified in Lemma 8.1. Restricting (8.2) to $\mathfrak{I}$ gives

$$\frac{m_{\mathfrak{I}}^{(n)}(\theta)}{\rho^n} = \frac{C^{n-kd}}{\rho^{n-kd}} \frac{m_{\mathfrak{I}}^{(kd)}(\theta)}{\rho^{kd}} + \sum_{l=kd+1}^{n} \frac{C^{n-l}}{\rho^{n-l}} \left(\frac{q_{\mathfrak{I}} + g_{\mathfrak{I}}(m^{(l-1)}(\theta))}{\rho^l} \right).$$

Hence, using dominated convergence, justified by Lemma 8.3, and the limiting behaviour of $\rho^{-n}C^n$, described in Lemma 7.2,

$$\frac{m_{\mathfrak{J}}^{(nd+r)}(\theta)}{\rho^{nd+r}} = \frac{C^{nd+r-kd}}{\rho^{nd+r-kd}} \frac{m_{\mathfrak{J}}^{(kd)}(\theta)}{\rho^{kd}} + \sum_{l=kd+1}^{nd+r} \frac{C^{nd+r-l}}{\rho^{nd+r-l}} \left(\frac{q_{\mathfrak{J}} + g_{\mathfrak{J}}(m^{(l-1)}(\theta))}{\rho^l} \right)$$

$$\to \rho^{-r}C^r \tilde{L} \frac{m_{\mathfrak{J}}^{(kd)}(\theta)}{\rho^{kd}} + \sum_{l=k+1}^{\infty} \sum_{c=0}^{d-1} \rho^{-r}C^r \tilde{L} C^c \left(\frac{q_{\mathfrak{J}} + g_{\mathfrak{J}}(m^{(ld-c-1)}(\theta))}{\rho^{ld}} \right)$$

$$= \rho^{-r}C^r \tilde{L} \left\{ \frac{m_{\mathfrak{J}}^{(kd)}(\theta)}{\rho^{kd}} + \sum_{l=k+1}^{\infty} \rho^{-ld} \sum_{c=0}^{d-1} C^c \left(q_{\mathfrak{J}} + g_{\mathfrak{J}}(m^{(ld-c-1)}(\theta)) \right) \right\}$$

as $n \to \infty$. Since $g_{\mathfrak{J}}$ is non-negative, taking k large enough that

$$\left(m_{\mathfrak{J}}^{(kd)}(\theta) + \sum_{l=1}^{\infty} \rho^{-ld} \left(\sum_{c=0}^{d-1} C^c q_{\mathfrak{J}} \right) \right) \gg 0$$

shows that the limit must be strictly positive. By Lemma 8.1, the components of $m^{(n)}$ are convex and increasing in θ; thus, the limit is too. Now, take $k = 0$ to give $l_{\mathfrak{J}}^{(r)}(\theta)$ and the formula for $l_{\mathfrak{J}}(\theta)$. Finally,

$$C l_{\mathfrak{J}}^{(r)}(\theta) = C \left(\rho^{-r}C^r \tilde{L} l_{\mathfrak{J}}(\theta) \right) = \rho l_{\mathfrak{J}}^{(r+1)}(\theta),$$

as claimed.   □

The matrix B used in Proposition 8.4 need not be unique. When it is not, another choice can produce a different set of indices $\mathfrak{J}$ on which convergence occurs. For the main proof, this route to convergence of other indices is irrelevant.

9. CONVERGENCE FOR THE REMAINING INDICES

Starting from any matrix $B \in \mathcal{B}$ with eigenvalue ρ, which is necessarily structured simply, Proposition 8.4 establishes the required limiting behaviour for indices in $\mathfrak{J}$. To consider the remaining indices, instead of starting from any such B, the argument starts with a matrix A identified using Proposition 7.8(ii). For this section, one such A is used, and $\mathfrak{J}$ and $\mathcal{E}$ are derived from it. The basic idea is to show that the limiting behaviour in $\mathcal{E}$ is induced by the already known behaviour for the indices in $\mathfrak{J}$.

Write any $B \in \mathcal{B}$ in the form

$$B = \begin{pmatrix} B_{\mathfrak{J},\mathfrak{J}} & B_{\mathfrak{J},\mathcal{E}} \\ B_{\mathcal{E},\mathfrak{J}} & B_{\mathcal{E},\mathcal{E}} \end{pmatrix}.$$

Let

$$\mathcal{G}_{\mathcal{E},\mathcal{P}} = \{\rho^{-1}(B_{\mathcal{E},\mathfrak{J}}, B_{\mathcal{E},\mathcal{E}}) : B \in \mathcal{B}\} \text{ and } \mathcal{G}_{\mathcal{E},\mathcal{E}} = \{\rho^{-1}B_{\mathcal{E},\mathcal{E}} : B \in \mathcal{B}\}.$$

The choice of A, through Proposition 7.8(ii), means that the maximum eigenvalue of the collection $\{D : D \in \mathcal{G}_{\mathcal{E},\mathcal{E}}\}$ is δ' with $\delta' < 1$.

For vectors a and b, indexed by $\mathfrak{J}$ and $\mathcal{E}$ respectively, and for $i \in \mathcal{E}$, let

$$h_i(a, b) = \max\{F_i a + D_i b : (F, D) \in \mathcal{G}_{\mathcal{E},\mathcal{P}}\}.$$

Since $\mathcal{B}$ is closed under row substitution, so is $\mathcal{G}_{\mathcal{E},\mathcal{P}}$. Hence, the rows of (F, D) giving the maxima here can be chosen independently, and so it makes sense to write the definition of h in vector form:

$$h(a, b) = \max\{Fa + Db : (F, D) \in \mathcal{G}_{\mathcal{E},\mathcal{P}}\}.$$

Furthermore, the maximum is attained by some $(F, D) \in \mathcal{G}_{\mathcal{E},\mathcal{P}}$ and h is (convex and) continuous in (a, b).

The next lemma indicates why h is important in describing the growth of $m_{\mathcal{E}}^{(n)}(\theta)$. To simplify the notation, fix θ and let

(9.1) $$\alpha^{(n)} = \frac{m_{\mathcal{J}}^{(n)}(\theta)}{\rho^n} \text{ and } \beta^{(n)} = \frac{m_{\mathcal{E}}^{(n)}(\theta)}{\rho^n}.$$

The behaviour of $\alpha^{(n)}$ is given by Proposition 8.4, that of $\beta^{(n)}$ is to be established. Note that, in this lemma and the following one, the inequalities are between vectors.

Lemma 9.1. *With u and t as defined in Lemma* 6.1,

$$h\left(\alpha^{(n-1)}, \beta^{(n-1)}\right) + \frac{1}{\rho^n}u_{\mathcal{E}} \leq \beta^{(n)} \leq h\left(\alpha^{(n-1)}, \beta^{(n-1)}\right) + \frac{1}{\rho^n}t_{\mathcal{E}}.$$

Proof. By definition, for $i \in \mathcal{E}$,

$$\beta_i^{(n)} = \frac{1}{\rho^n}m_i^{(n)}(\theta)$$

$$= \frac{1}{\rho^n}\log\left(\sum_b \mu_i(b)\exp\left(b^T m^{(n-1)}(\theta)\right)\right)$$

$$= \frac{1}{\rho^n}\log\left(\sum_b \mu_i(b)\exp\left(\rho^{n-1}\left(b_{\mathcal{J}}^T\alpha^{(n-1)} + b_{\mathcal{E}}^T\beta^{(n-1)}\right)\right)\right)$$

$$\geq \frac{1}{\rho^n}u_i + \frac{1}{\rho}\max\left\{b_{\mathcal{J}}^T\alpha^{(n-1)} + b_{\mathcal{E}}^T\beta^{(n-1)} : b \text{ with } \mu_i(b) > 0\right\}$$

$$= \frac{1}{\rho^n}u_i + h_i\left(\alpha^{(n-1)}, \beta^{(n-1)}\right),$$

using the definition of h. Similarly,

$$\beta_i^{(n)} = \frac{1}{\rho^n}\log\left(\sum_b \mu_i(b)\exp\left(\rho^{n-1}\left(b_{\mathcal{J}}^T\alpha^{(n-1)} + b_{\mathcal{E}}^T\beta^{(n-1)}\right)\right)\right)$$

$$\leq \frac{1}{\rho^n}t_i + h_i\left(\alpha^{(n-1)}, \beta^{(n-1)}\right),$$

giving the upper bound. $\square$

These bounds suggest strongly that when $\alpha^{(n)}$ tends to a the sequence $\beta^{(n)}$ should tend to a solution, b, of $b = h(a, b)$. They also suggest that a proof might work by showing that: (i) if the sequence $v^{(n)}$ is given by $v^{(n+1)} = h(\alpha^{(n)}, v^{(n)})$ then the difference between $v^{(n)}$ and $\beta^{(n)}$ goes to zero as n goes to infinity; (ii) for a given a, the solution of $b = h(a, b)$ is unique; (iii) the sequence $v^{(n)}$ converges to b. This is the line that will be followed, but allowing for periodicity in the limiting behaviour of $\alpha^{(n)}$ needs a little care. Before considering this further, a simple lemma shows how differences in the values of h can be related to differences in the argument.

Lemma 9.2. *Suppose $(F(1), D(1))$ and $(F(2), D(2))$ in $\mathcal{G}_{\mathcal{E}, \mathcal{P}}$ are such that*

$$h(a, b) = F(1)a + D(1)b \text{ and } h(x, y) = F(2)x + D(2)y.$$

Then

$$F(2)(a - x) + D(2)(b - y) \le h(a, b) - h(x, y) \le F(1)(a - x) + D(1)(b - y).$$

Proof. Since $h(x, y) = \max\{Fx + Dy : (F, D) \in \mathcal{G}_{\mathcal{E}, \mathcal{P}}\}$ with the maximum being attained at $(F(2), D(2))$,

$$\begin{aligned}
h(a, b) - h(x, y) &= F(1)a + D(1)b - (F(2)x + D(2)y) \\
&\le F(1)a + D(1)b - (F(1)x + D(1)y) \\
&= F(1)(a - x) + D(1)(b - y).
\end{aligned}$$

The other inequality is derived similarly. $\qquad\qquad\qquad\qquad\qquad\qquad\square$

Lemma 9.3. *Let $\tilde{\beta}^{(0)} = v$ and define $\tilde{\beta}^{(n+1)}$ by*

$$\tilde{\beta}^{(n+1)} = h(\alpha^{(n)}, \tilde{\beta}^{(n)}) \text{ for } n \in \mathbb{Z}_+.$$

Then, whatever the value of v,

$$\|\beta^{(n)} - \tilde{\beta}^{(n)}\| \to 0$$

as $n \to \infty$.

Proof. Take $\delta < 1$ but large enough that $\delta\rho > 1$ and that the maximum eigenvalue of the collection $\{D : D \in \mathcal{G}_{\mathcal{E}, \mathcal{E}}\}$ is strictly less than δ. Corollary 7.7(i) applies to $\mathcal{G}_{\mathcal{E}, \mathcal{E}}$, to show there is a $w = w(\delta) \gg 0$, normalised to have $\|w\| = 1$, such that

$$\sup_{D \in \mathcal{G}_{\mathcal{E}, \mathcal{E}}} Dw \le \delta w.$$

Now, choose c so large that $t_{\mathcal{E}} \le cw$ and k_0 so large that

$$-k_0 \delta w \le (\beta^{(0)} - v) \le k_0 \delta w.$$

Suppose that

$$-k_n \delta^n w \le (\beta^{(n)} - \tilde{\beta}^{(n)}) \le k_n \delta^n w;$$

then, using the upper bounds in Lemmas 9.1 and 9.2, for some $D \in \mathcal{G}_{\mathcal{E}, \mathcal{E}}$,

$$\begin{aligned}
\beta^{(n+1)} - \tilde{\beta}^{(n+1)} &\le h\left(\alpha^{(n)}, \beta^{(n)}\right) - \tilde{\beta}^{(n+1)} + \frac{1}{\rho^{n+1}} t_{\mathcal{E}} \\
&\le h\left(\alpha^{(n)}, \beta^{(n)}\right) - h\left(\alpha^{(n)}, \tilde{\beta}^{(n)}\right) + \frac{1}{\rho^{n+1}} t_{\mathcal{E}} \\
&\le D\left(\beta^{(n)} - \tilde{\beta}^{(n)}\right) + \frac{1}{\rho^{n+1}} t_{\mathcal{E}} \\
&\le \left(k_n \delta^{n+1} + \frac{c}{\rho^{n+1}}\right) w.
\end{aligned}$$

Similarly, using the lower bounds in Lemmas 9.1 and 9.2,

$$\beta^{(n+1)} - \tilde{\beta}^{(n+1)} \ge -k_n \delta^{n+1} w \ge -\left(k_n \delta^{n+1} + \frac{c}{\rho^{n+1}}\right) w.$$

Hence, we can take

$$k_{n+1} = k_n + \frac{c}{(\delta\rho)^{n+1}} \le k_0 + \sum_{r=1}^{\infty} \frac{c}{(\delta\rho)^r} = k_0 + \frac{c}{\delta\rho - 1},$$

and so

$$\|\beta^{(n)} - \tilde{\beta}^{(n)}\| \leq \left(k_0 + \frac{c}{\delta\rho - 1}\right)\delta^n \to 0. \qquad \square$$

Lemma 9.4. *Given* $(a^{(0)}, a^{(1)}, \ldots)$, *let* $\tilde{\beta}^{(0)}(v) = v$ *and define* $\tilde{\beta}^{(n+1)}(v)$ *by*

$$\tilde{\beta}^{(n+1)}(v) = h\left(a^{(n)}, \tilde{\beta}^{(n)}(v)\right) \text{ for } n \in \mathbb{Z}_+.$$

Then, for every v_1 *and* v_2,

$$\|\tilde{\beta}^{(n)}(v_1) - \tilde{\beta}^{(n)}(v_2)\| \to 0$$

as $n \to \infty$.

Proof. Using the upper bound in Lemma 9.2,

$$\tilde{\beta}^{(n+1)}(v_1) - \tilde{\beta}^{(n+1)}(v_2) \leq h\left(a^{(n)}, \tilde{\beta}^{(n)}(v_1)\right) - h\left(a^{(n)}, \tilde{\beta}^{(n)}(v_2)\right)$$
$$\leq D\left(\tilde{\beta}^{(n)}(v_1) - \tilde{\beta}^{(n)}(v_2)\right),$$

for some $D \in \mathcal{G}_{\mathcal{E},\mathcal{E}}$, and there is a similar lower bound. Hence, much as in the proof of Lemma 9.3, there is a finite c such that

$$\|\tilde{\beta}^{(n+1)}(v_1) - \tilde{\beta}^{(n+1)}(v_2)\| \leq c\delta^n. \qquad \square$$

Lemma 9.5. *Given* $\left(a^{(0)}, a^{(1)}, \ldots, a^{(d-1)}\right)$, *let* $\widehat{\beta}^{(0)} = v$ *and define* $\widehat{\beta}^{(n)}$ *by*

$$\widehat{\beta}^{(nd+r+1)} = h\left(a^{(r)}, \widehat{\beta}^{(nd+r)}\right) \text{ for } r = 0, 1, \ldots, d-1; n \in \mathbb{Z}_+.$$

Then

$$\widehat{\beta}^{(nd+r)} \to b^{(r)} \text{ as } n \to \infty,$$

where, with $b^{(d)} = b^{(0)}$, $\left(b^{(0)}, b^{(1)}, \ldots, b^{(d-1)}\right)$ *is the unique solution to*

(9.2) $$b^{(r)} = h\left(a^{(r-1)}, b^{(r-1)}\right) \text{ for } r = 1, 2, \ldots, d.$$

Furthermore, when all of $\left(a^{(0)}, a^{(1)}, \ldots, a^{(d-1)}\right)$ *have all components strictly positive, so do* $\left(b^{(0)}, b^{(1)}, \ldots, b^{(d-1)}\right)$.

Proof. Take δ and w as in the proof of Lemma 9.3 and let $a^{(nd+r)} = a^{(r)}$ for $n \in \mathbb{Z}_+$. Also, take k_0 to be large enough that

$$-k_0 w \leq \left(\widehat{\beta}^{(d)} - v\right) \leq k_0 w.$$

Now, suppose

$$-k_n w \leq \left(\widehat{\beta}^{(n+d)} - \widehat{\beta}^{(n)}\right) \leq k_n w.$$

Then, for some $D \in \mathcal{G}_{\mathcal{E},\mathcal{E}}$,

$$\widehat{\beta}^{(n+d+1)} - \widehat{\beta}^{(n+1)} = h\left(a^{(n)}, \widehat{\beta}^{(n+d)}\right) - h\left(a^{(n)}, \widehat{\beta}^{(n)}\right)$$
$$\leq D\left(\widehat{\beta}^{(n+d)} - \widehat{\beta}^{(n)}\right)$$
$$\leq k_n \delta w,$$

and, similarly,

$$\widehat{\beta}^{(n+d+1)} - \widehat{\beta}^{(n+1)} \geq -k_n \delta w.$$

Thus we can take

$$k_{n+1} = k_n \delta = k_0 \delta^{n+1},$$

which implies that

$$\|\widehat{\beta}^{(n+d)} - \widehat{\beta}^{(n)}\| \leq k_0 \delta^n.$$

Since this has a finite sum, the convergence of $\widehat{\beta}^{(nd+r)}$ as $n \to \infty$ follows. The continuity of h implies that the limits satisfy the 'fixed-point' equations. If there were distinct solutions to these equations, taking v_1 and v_2 to be the first member of each, Lemma 9.4 shows they must in fact be equal.

Let u be such that $a^{(i)} \gg u \gg 0$ for $i = 0, 1, \ldots, d-1$. When $v = 0$, it is easy to see that $\widehat{\beta}^{(n)} \geq 0$ for every n; hence, $b^{(r)} \geq 0$ for $r = 0, 1, \ldots, d-1$. Also, for any $(F, D) \in \mathcal{G}_{\mathcal{E}, \mathcal{P}}$,

$$b^{(i+1)} = h\left(a^{(i)}, b^{(i)}\right) \geq Fu + Db^{(i)}.$$

Iterating this, for any $(F(j), D(j)) \in \mathcal{G}_{\mathcal{E}, \mathcal{P}}$,

$$b^{(i+1)} \geq D(1) \cdots D(k-1) F(k) u.$$

Now, by irreducibility, k and the matrices can be selected to make any chosen component of the right-hand side positive. □

Lemma 9.6. Let $\left(a^{(0)}, a^{(1)}, \ldots, a^{(d-1)}\right)$ be such that

$$\alpha^{(nd+r)} \to a^{(r)}$$

as $n \to \infty$. As in Lemma 9.4, let $\tilde{\beta}^{(0)} = v$ and define $\tilde{\beta}^{(n+1)}$ by

$$\tilde{\beta}^{(n+1)} = h\left(\alpha^{(n)}, \tilde{\beta}^{(n)}\right) \text{ for } n \in \mathbb{Z}_+.$$

Take $\left(b^{(0)}, b^{(1)}, \ldots, b^{(d-1)}\right)$ to be the solution to (9.2). Let $b^{(nd+r)} = b^{(r)}$ for $n \in \mathbb{Z}_+$. Then

$$\|\tilde{\beta}^{(n+1)} - b^{(n+1)}\| \to 0$$

as $n \to \infty$.

Proof. Take δ and w as in the proof of Lemma 9.3 and let $a^{(nd+r)} = a^{(r)}$ for $n = 1, 2, \ldots$. For $\epsilon > 0$, choose n' so large that for all $n \geq n'$ and all $(F, D) \in \mathcal{G}_{\mathcal{E}, \mathcal{P}}$,

$$-\epsilon w \leq F\left(\alpha^{(n)} - a^{(n)}\right) \leq \epsilon w.$$

Now, suppose

$$-k_n w \leq \left(\tilde{\beta}^{(n)} - b^{(n)}\right) \leq k_n w.$$

Then, using Lemma 9.2, for some $(F, D) \in \mathcal{G}_{\mathcal{E}, \mathcal{P}}$,

$$\tilde{\beta}^{(n+1)} - b^{(n+1)} = h\left(\alpha^{(n)}, \tilde{\beta}^{(n)}\right) - h\left(a^{(n)}, b^{(n)}\right)$$
$$\leq F\left(\alpha^{(n+d)} - a^{(n)}\right) + D\left(\tilde{\beta}^{(n+d)} - b^{(n)}\right)$$
$$\leq (\epsilon + k_n \delta) w$$

for $n \geq n'$, and, similarly,

$$\tilde{\beta}^{(n+1)} - b^{(n+1)} \geq -(\epsilon + k_n \delta) w.$$

Thus we can take

$$k_{n+1} = \epsilon + k_n \delta = \epsilon \sum_{r=0}^{n-n'} \delta^r + k_{n'} \delta^{n-n'+1}.$$

Hence,

$$\limsup_n \|\tilde{\beta}^{(n+1)} - b^{(n+1)}\| \leq \frac{\epsilon}{1-\delta} + \limsup_n k_{n'} \delta^{n-n'+1} = \frac{\epsilon}{1-\delta}.$$

Since ϵ is arbitrary this proves the result. $\qquad\square$

Proposition 9.7. *Let $A \in \mathcal{B}$ be chosen using Proposition 7.8(ii). This A gives the index set $\mathcal{E}$ and the period d. Then*

$$\frac{m_{\mathcal{E}}^{(nd+r)}(\theta)}{\rho^{nd+r}} \to l_{\mathcal{E}}^{(r)}(\theta)$$

as $n \to \infty$, where $l_{\mathcal{E}}^{(r)}(\theta)$ is finite, convex, increasing and strictly positive on $\mathfrak{T}$. Furthermore, $\left(l_{\mathcal{E}}^{(0)}(\theta), l_{\mathcal{E}}^{(1)}(\theta), \ldots, l_{\mathcal{E}}^{(d-1)}(\theta) \right)$ is the unique solution to (9.2) when $a^{(r)}$ is given by $l_{\mathcal{J}}^{(r)}(\theta)$ from Proposition 8.4.

Proof. Recall the definitions of $\alpha^{(n)}$ and $\beta^{(n)}$ at (9.1). Then, by Proposition 8.4, $\alpha^{(nd+r)}$ converges to a limit $a^{(r)}$, which is $l_{\mathcal{J}}^{(r)}(\theta)$, as $n \to \infty$. By Lemma 9.6, $\tilde{\beta}^{(nd+r)}$ defined through $\tilde{\beta}^{(n+1)} = h(\alpha^{(n)}, \tilde{\beta}^{(n)})$ converges to $b^{(r)}$, where $(b^{(0)}, b^{(1)}, \ldots, b^{(d-1)})$ is the unique solution to (9.2). Lemma 9.3 implies that $\beta^{(nd+r)} - \tilde{\beta}^{(nd+r)}$ goes to zero as $n \to \infty$. Combining these gives the required convergence. Lemma 9.5 shows that all components of each of $(b^{(0)}, b^{(1)}, \ldots, b^{(d-1)})$ are positive. Finally, the convexity and monotonicity of $m_i^{(n)}$ carry over to the limits. $\qquad\square$

Proof of Theorem 2.2. Let $A \in \mathcal{B}$ be chosen using Proposition 7.8(ii). With this choice, Propositions 8.4 and 9.7 combine to give the main result. Theorem 5.1 covers the aperiodicity of the univariate case. $\qquad\square$

Note that, in principle, different choices of A can produce different periods. The minimal period possible in Theorem 2.2 will divide all of these.

Proof of Theorem 2.5. The argument used in Proposition 8.4 applies to each block in C for every suitable B; R is such a block. This gives convergence on $\mathcal{J}$, with period d'. The limits, and the convergence, must be consistent with Theorem 2.2. Hence the smaller of d' and d must divide the larger. $\qquad\square$

Proof of Corollary 2.6. By Theorem 2.5,

$$A^d l^{(0)}(\theta) = \rho A^{d-1} l^{(1)}(\theta) = \cdots = \rho^d l^{(d)}(\theta) = \rho^d l^{(0)}(\theta).$$

Thus $l^{(0)}(\theta)$ is an eigenvector of A^d. Since A^d is primitive, its right eigenvector corresponding to the eigenvalue ρ^d is unique, up to scaling, and therefore must be also be the eigenvector of A corresponding to ρ. Hence $l^{(0)}(\theta) = k(\theta)a$. The functional equation for k follows directly, or from Corollary 2.4(i). $\qquad\square$

10. PROPERTIES OF THE LIMIT

Propositions 8.4 and 9.7 contain significant information on the limit, which the next result exploits.

Lemma 10.1. *For a fixed $\theta \in \mathfrak{T}$, there is an $A^{(r)} = A^{(r)}(\theta)$ that is structured simply, with $A_{\mathcal{J},\mathcal{J}}^{(r)} = C$, independent of r and with*

$$(10.1) \qquad\qquad A^{(r)} l^{(r)}(\theta) = \rho l^{(r+1)}(\theta).$$

Furthermore, with this $\mathfrak{I}$ *and* C, *and* $l_{\mathfrak{I}}(\theta)$ *given by the formula* (8.3) *in Proposition* 8.4, *there is an idempotent matrix* L' *such that*

$$(10.2) \qquad\qquad l^{(0)}(\theta) = L'_{\mathcal{P},\mathfrak{I}} l_{\mathfrak{I}}(\theta),$$

and

$$(10.3) \qquad\qquad l^{(r)}(\theta) = \rho^{-r} \left(\left(\prod_{j=0}^{r-1} A^{(j)} \right) L'_{\mathcal{P},\mathfrak{I}} \right) l_{\mathfrak{I}}(\theta).$$

Proof. Proposition 8.4 gives $C l_{\mathfrak{I}}^{(r)}(\theta) = \rho l_{\mathfrak{I}}^{(r+1)}(\theta)$. By Proposition 9.7

$$l_{\mathcal{E}}^{(r+1)}(\theta) = h(l_{\mathfrak{I}}^{(r)}(\theta), l_{\mathcal{E}}^{(r)}(\theta)).$$

There is some $(F, D) \in \mathcal{G}_{\mathcal{E},\mathcal{P}}$ giving the maximum in the definition of h and so there must be an $(F^{(r)}, D^{(r)}) \in \mathcal{G}_{\mathcal{E},\mathcal{P}}$ such that

$$l_{\mathcal{E}}^{(r+1)}(\theta) = F^{(r)} l_{\mathfrak{I}}^{(r)}(\theta) + D^{(r)} l_{\mathcal{E}}^{(r)}(\theta).$$

Then

$$A^{(r)} = \left(\begin{array}{cc} C & 0 \\ \rho F^{(r)} & \rho D^{(r)} \end{array} \right)$$

has the required properties.

Now, let

$$J = \rho^{-d} A^{(d-1)} \cdots A^{(1)} A^{(0)} = \left(\begin{array}{cc} \rho^{-d} C^d & 0 \\ J_{\mathcal{E},\mathfrak{I}} & J_{\mathcal{E},\mathcal{E}} \end{array} \right),$$

so that $J l^{(0)}(\theta) = l^{(0)}(\theta)$. In J, $J_{\mathcal{E},\mathcal{E}}$ is the product of matrices drawn from $\mathcal{G}_{\mathcal{E},\mathcal{E}}$ and so, by Corollary 7.7 (ii), it has maximum eigenvalue strictly less that one. Hence J is structured simply, and the diagonal blocks in $\rho^{-d} C^d$ are all primitive. Also, (7.1) in Lemma 7.2 gives that $\tilde{L}(\eta^{-d} C^d) = \tilde{L}(C) = \tilde{L}$. Thus Lemma 7.3 applies to J to give that there is an idempotent matrix

$$L' = \left(\begin{array}{cc} L'_{\mathfrak{I},\mathfrak{I}} & 0 \\ L'_{\mathcal{E},\mathfrak{I}} & 0 \end{array} \right) = \left(\begin{array}{cc} \tilde{L} & 0 \\ L'_{\mathcal{E},\mathfrak{I}} & 0 \end{array} \right),$$

and there is a $\zeta < 1$ such that

$$\zeta^{-n} \| J^n - L' \| \to 0.$$

Therefore, Proposition 8.4 gives

$$l_{\mathfrak{I}}^{(0)}(\theta) = \tilde{L} l_{\mathfrak{I}}(\theta) = L'_{\mathfrak{I},\mathfrak{I}} l_{\mathfrak{I}}(\theta).$$

Iterating $J l^{(0)}(\theta) = l^{(0)}(\theta)$ gives

$$l_{\mathcal{E}}^{(0)}(\theta) = L'_{\mathcal{E},\mathfrak{I}} l_{\mathfrak{I}}^{(0)}(\theta) = L'_{\mathcal{E},\mathfrak{I}} \tilde{L} l_{\mathfrak{I}}(\theta) = L'_{\mathcal{E},\mathfrak{I}} l_{\mathfrak{I}}(\theta),$$

which completes the proof of (10.2). The relation (10.3) now follows directly. □

Proof of Theorem 2.7. Most of this follows immediately from the first part of the previous result: the fact that $A^{(r)}$ is structured simply gives (i) and (ii); (10.1) is the second part of (iii); (iv) follows from (iii) and $l^{(0)}(\theta) = l^{(d)}(\theta)$. This leaves the first part of (iii). For any $B \in \mathcal{B}$, (2.2) gives

$$\frac{m^{(nd+r+1)}(\theta)}{\rho^{nd+r+1}} - \frac{B}{\rho} \frac{m^{(nd+r)}(\theta)}{\rho^{nd+r}} = \frac{q(B)}{\rho^{nd+r+1}} + \frac{g(B; m^{(nd+r)}(\theta))}{\rho^{nd+r+1}}.$$

Now let $n \to \infty$ and use the positivity of g to give

$$\rho l^{(r+1)}(\theta) - B l^{(r)}(\theta) \geq 0$$

for every $B \in \mathcal{B}$. Hence, using (10.1), $(A^{(r)} - B)l^{(r)}(\theta) \geq 0$ for every B. $\square$

Lemma 10.2. *Fix x. Let A be such that $Ax = \max\{Bx : B \in \mathcal{B}\}$ and let $g(\theta)$ be $g(\theta; A)$. Suppose A_i is the only possibility giving the maximum for the ith row here and let*

$$\inf\{(A_i - b^T)x : \mu_i(b) > 0, \ b^T \neq A_i\} = \delta_i(x) > 0.$$

Then, for scalar $\lambda > 0$, there are constants κ_1, κ_2, which do not depend on i, λ and v, such that

$$\|g_i'(\lambda v)\| \leq \kappa_1 e^{-\lambda \delta_i(x)} e^{\kappa_2 \lambda \|v - x\|}$$

where g_i' is the vector of derivatives of g_i.

Proof. Note that

$$\exp(g_i(\theta)) - 1 = \left(\sum_{b \neq A_i^T} \frac{\mu_i(b)}{\mu_i(A_i^T)} e^{-(A_i - b^T)\theta} \right).$$

Hence,

$$\frac{\partial g_i(\phi)}{\partial \phi_j} = \frac{\partial}{\partial \phi_j} \log \left(1 + \sum_{b \neq A_i^T} \frac{\mu_i(b)}{\mu_i(A_i^T)} e^{-(A_i - b^T)\phi} \right)$$

$$= \frac{\left(\sum_{b \neq A_i^T} \frac{\mu_i(b)}{\mu_i(A_i^T)} (A_{i,j} - b_j) e^{-(A_i - b^T)\phi} \right)}{1 + \left(\sum_{b \neq A_i^T} \frac{\mu_i(b)}{\mu_i(A_i^T)} e^{-(A_i - b^T)\phi} \right)},$$

and so, with $\kappa_4 = \max\{\|A - B\| : B \neq A, B \in \mathcal{B}\}$,

$$\left| \frac{\partial g_i(\phi)}{\partial \phi_j} \right| \leq \kappa_4 \sum_{b \neq A_i^T} \frac{\mu_i(b)}{\mu_i(A_i^T)} e^{-(A_i - b^T)\phi} = \kappa_4 \left(\exp(g_i(\phi)) - 1 \right).$$

Then

$$0 \leq \exp(g_i(\lambda v)) - 1 = \left(\sum_{b \neq A_i^T} \frac{\mu_i(b)}{\mu_i(A_i^T)} e^{-\lambda(A_i - b^T)(x + v - x)} \right)$$

$$\leq \kappa_3 e^{-\lambda \delta_i(x)} e^{p\kappa_4 \lambda \|v - x\|}.$$

Therefore

$$\left| \left(\frac{\partial g_i(\phi)}{\partial \phi_j} \bigg|_{\phi = \lambda v} \right) \right| \leq \kappa_4 \kappa_3 e^{-\lambda \delta_i(x)} e^{p\kappa_4 \lambda \|v - x\|}. \qquad \square$$

When combined with Theorem 2.7(iii), the condition (10.4) in the next result says that C was the only choice for the rows indexed by $\mathfrak{I}$ in $A^{(r)}$.

Lemma 10.3. *Let θ be in the interior of $\mathfrak{T}$ and let $A^{(r)}(\theta)$, C and $l_{\mathfrak{I}}(\theta)$ be as in Lemma 10.1.*

(i) The components of $l_{\mathfrak{I}}(\theta)$ have continuous derivatives in a neighbourhood of θ when for each $i \in \mathfrak{I}$

$$(10.4) \qquad C_i l_{\mathfrak{I}}^{(r)}(\theta) > \max \left\{ B_i l^{(r)}(\theta) : B \in \mathcal{B}, B_i \neq (C_i, 0) \right\},$$

and then the functions $l_{\mathfrak{J}}^{(r)}$ also have continuous derivatives.

(ii) *The functions $l_{\mathcal{E}}^{(r)}$ have continuous derivatives in a neighbourhood of θ when, on that neighbourhood, $l_{\mathfrak{J}}$ is differentiable and the matrices $A^{(r)}(\phi)$ in Lemma* 10.1 *are constant.*

Proof. By continuity, $B_i l^{(r)}(\phi)$ must be maximised by $B_i = (C_i, 0)$ throughout a neighbourhood of θ. Let $\mathfrak{N}$ be a closed ball centred on θ and comfortably inside this neighbourhood. Take δ to be such that

$$0 < \delta < \min_{i \in \mathfrak{J},\, r} \inf_{\phi \in \mathfrak{N}} \delta_i(l^{(r)}(\phi)),$$

where $\delta_i(x)$ is defined in Lemma 10.2. Then, applying Lemma 10.2,

$$\|g_i'(m^{(nd+r)}(\phi))\| \le \kappa_1 \exp(-\rho^{nd+r}\delta) \exp\left\{ \kappa_2 \rho^{nd+r} \left\| l^{(r)}(\phi) - \frac{m^{(nd+r)}(\phi)}{\rho^{nd+r}} \right\| \right\}.$$

By Corollary 2.3, the convex functions $m^{(nd+r)}(\phi)/\rho^{nd+r}$ converge uniformly to $l^{(r)}(\phi)$ on $\mathfrak{N}$. Thus, for n sufficiently large,

$$\left\| l^{(r)}(\phi) - \rho^{-nd-r} m^{(nd+r)}(\phi) \right\| \le \frac{\delta}{2\kappa_2},$$

and so $\|g_i'(m^{(nd+r)}(\phi))\|$ is ultimately uniformly bounded by $\kappa_1 \exp(-\rho^{nd+r}\delta/2)$ on $\mathfrak{N}$. Also, $m^{(nd+d)}(\phi)/\rho^{nd+d}$ are convex functions converging to a finite convex function, and so Theorem 10.6 in Rockafellar [15] implies that these functions must be equi-Lipschitzian (i.e. Lipschitz at every $x \in \mathfrak{N}$ with a constant independent of x and the particular function). This implies the uniform boundedness on $\mathfrak{N}$ of the sequence of derivatives when they exist, as they do here. Now dominated convergence justifies differentiating (8.3) term by term to show that $l_{\mathfrak{J}}(\theta)$ has a continuous derivative throughout $\mathfrak{N}$.

The formula $l_{\mathfrak{J}}^{(r)}(\theta) = \rho^{-r} C^r L_{\mathfrak{J},\mathfrak{J}} l_{\mathfrak{J}}(\theta)$ holds throughout $\mathfrak{N}$, giving differentiability there. Similarly, when the matrices $A^{(r)}(\phi)$ can be taken to be $A^{(r)} = A^{(r)}(\theta)$ in a neighbourhood of θ,

$$l^{(r)}(\phi) = \rho^{-r} \left(\left(\prod_{j=0}^{r-1} A^{(j)} \right) L_{\mathcal{P},\mathfrak{J}}' \right) l_{\mathfrak{J}}(\phi)$$

holds throughout that neighbourhood, which gives differentiability there. $\square$

Proof of Theorem 2.8. When the collection

$$\{ A^{(r)}(\theta) \in \mathcal{B} : r = 0, 1, \ldots, d - 1 \}$$

that satisfies (2.3) is unique, Lemma 10.1 must hold with these matrices and, by Theorem 2.7(iii),

$$A^{(r)}(\theta) l^{(r)}(\theta) > \max\left\{ B l^{(r)}(\theta) : B \in \mathcal{B}, B \ne A^{(r)}(\theta) \right\}.$$

By continuity, $B l^{(r)}(\phi)$ must be maximised uniquely by $B = A^{(r)}(\theta)$ throughout a neighbourhood of θ. Therefore, both parts of Lemma 10.3 apply. $\square$

Proof of Corollary 2.9. Suppose A is the unique matrix in $\mathcal{B}$ with eigenvalue ρ. Then, since $A^{(r)}(\theta)$ in (2.3) must have eigenvalue ρ, $A^{(r)}(\theta) = A$ for all r and θ. Theorem 2.8 now applies. $\square$

Acknowledgements

Krishna Athreya drew my attention to the paper of Miller and O'Sullivan [14] and discussed the early attempts at a solution while we shared a room at the Institute of Mathematics and its Applications, University of Minnesota, during the programme on 'Emerging Applications of Probability Theory' in 1994. David Grey pointed out the relevance of the literature on dynamic programming, leading me to Kennedy's paper [10]. Owen Jones rekindled my interest by wanting to use some of the ideas. Nick Bingham commented on the manuscript and suggested a restructuring, which has significantly improved the presentation. Finally, the referee made improving comments.

References

[1] Tom M. Apostol, *Mathematical analysis: a modern approach to advanced calculus*, Addison-Wesley Publishing Company, Inc., Reading, Mass., 1957. MR0087718 (19,398e)

[2] K. B. Athreya and A. N. Vidyashankar, *Large deviation rates for branching processes. II. The multitype case*, Ann. Appl. Probab. **5** (1995), no. 2, 566–576. MR1336883 (96k:60214)

[3] Martin T. Barlow and Edwin A. Perkins, *Brownian motion on the Sierpiński gasket*, Probab. Theory Related Fields **79** (1988), no. 4, 543–623. MR966175 (89g:60241)

[4] J. D. Biggins and N. H. Bingham, *Large deviations in the supercritical branching process*, Adv. in Appl. Probab. **25** (1993), no. 4, 757–772. MR1241927 (94i:60101)

[5] D. R. Grey, *Nonnegative matrices, dynamic programming and a harvesting problem*, J. Appl. Probab. **21** (1984), no. 4, 685–694. MR766807 (86c:90126)

[6] T. E. Harris, *Branching processes*, Ann. Math. Statistics **19** (1948), 474–494. MR0027465 (10,311b)

[7] John E. Hopcroft, Rajeev Motwani, and Jeffrey D. Ullman, *Introduction to automata theory, languages, and computation, 2nd edition*, Addison-Wesley, Boston, 2001.

[8] Owen Dafydd Jones, *Multivariate Böttcher equation for polynomials with nonnegative coefficients*, Aequationes Math. **63** (2002), no. 3, 251–265. MR1904719 (2003f:39064)

[9] ———, *Large deviations for supercritical multitype branching processes*, J. Appl. Probab. **41** (2004), no. 3, 703–720. MR2074818 (2005i:60169)

[10] Douglas P. Kennedy, *On sets of countable non-negative matrices and Markov decision processes*, Adv. Appl. Probab. **10** (1978), no. 3, 633–646. MR0491767 (58 #10966)

[11] Marek Kuczma, Bogdan Choczewski, and Roman Ger, *Iterative functional equations*, Encyclopedia of Mathematics and its Applications, vol. 32, Cambridge University Press, Cambridge, 1990. MR1067720 (92f:39002)

[12] Takashi Kumagai, *Estimates of transition densities for Brownian motion on nested fractals*, Probab. Theory Related Fields **96** (1993), no. 2, 205–224. MR1227032 (94e:60068)

[13] Peter Lancaster and Miron Tismenetsky, *The theory of matrices*, Computer Science and Applied Mathematics, Academic Press Inc., Orlando, FL, 1985. MR792300 (87a:15001)

[14] Michael I. Miller and Joseph A. O'Sullivan, *Entropies and combinatorics of random branching processes and context-free languages*, IEEE Trans. Inform. Theory **38** (1992), no. 4, 1292–1310. MR1168750 (93m:68103)

[15] R. Tyrrell Rockafellar, *Convex analysis*, Princeton Mathematical Series, No. 28, Princeton University Press, Princeton, N.J., 1970. MR0274683 (43 #445)

[16] E. Seneta, *Non-negative matrices*, Halsted Press [A division of John Wiley & Sons], New York, 1973. MR0389944 (52 #10773)

[17] ———, *Nonnegative matrices and Markov chains*, Springer Series in Statistics, Springer-Verlag, New York, 1981. MR719544 (85i:60058)

Department of Probability and Statistics, The University of Sheffield, Sheffield, S3 7RH, U.K.

E-mail address: J.Biggins@sheffield.ac.uk

URL: http://biggins.staff.shef.ac.uk/