Extended Perron–Frobenius results

J.D. Biggins*
The University of Sheffield, UK

A. Rahimzadeh Sani†
Teacher Training University of Tehran, Iran

Abstract

Matrices with entries that are smooth functions of a variable $\lambda \in L$ and non-negative when $\lambda \in L_- \subset L$ are considered. Perron-Frobenius theory applies when the entries are non-negative. Here, analogous results are shown to hold in neighbourhood of L_-, with the various quantities varying smoothly. The main examples are matrices with entries that are (i) Fourier-Stieltjes transforms of finite measures and (ii) Laplace-Stieltjes transforms, which arise from Markov-additive processes.

Key words. non-negative matrices, maximum eigenvalue

AMS subject classification. 15A48

1 Introduction

Powers of matrices with entries that are Laplace-Stieltjes transforms occur in Applied Probability; see for example [1], [4], [8], [9], [11], [13], [14] and [15]. Mainly, these matrices arise from Markov additive processes, as defined in [14], for example. For real values of the argument these matrices have non-negative entries to which standard Perron–Frobenius theory applies. Results about these matrices for arguments that are complex, but sufficiently close to real, are needed to prove local limit theorems and refined large deviation theorems for Markov additive processes. Furthermore, for those proofs, various estimates need to be uniform in sufficiently small neighbourhoods. The same issue arises in giving a multitype version of the arguments in [2], described in [3]. The aim of this note is to establish matrix results of the appropriate kind.

It is natural to allow the Laplace-Stieltjes transforms to be multivariate, which means that their argument is taken from $\mathbb{C}^d = \mathbb{C} \times \ldots \times \mathbb{C}$, where $\mathbb{C}$ are the complex numbers. The results will also cover multivariate Fourier-Stieltjes transforms, where the argument is taken from $\mathbb{R}^d$.

2 Notation and main results

A matrix $M = \{m_{ij}\}_{p\times p}$, has nth power M^n with entries denoted by m^n_{ij}. The eigenvalues of M are the zeros of the characteristic polynomial $q(z) = \det(zI - M)$ which is of degree p. Denote the distinct roots of $q(z)$ by $\rho_1, \ldots, \rho_s$, with multiplicities $m_1, \ldots, m_s$, respectively; then $q(z) = \prod_{i=1}^{s}(z - \rho_i)^{m_i}$. The roots are listed in order, $|\rho_1| \geq \ldots \geq |\rho_s|$; furthermore, they are arranged with $m_i \geq m_{i+1}$ when $|\rho_i| = |\rho_{i+1}|$. By analogy with the case when the entries are non-negative, the eigenvalue ρ_1 of M is called the *maximum-modulus eigenvalue* if it is a simple root of the characteristic polynomial, that is if $m_1 = 1$, and $|\rho_1| > |\rho_2|$. Note that when a maximum-modulus eigenvalue exists it is automatically unique.

A non-negative square matrix $A = \{a_{ij}\}$ is called *positive regular* if all its entries are finite and, for some n and all i and j, $a^n_{ij} > 0$ (see [12] or [16]). The following version of Perron-Frobenius Theorem which is paraphrased from [16] sets the scene for the results obtained here.

Perron-Frobenius Theorem *Suppose $A = \{a_{ij}\}$ is a $p \times p$ positive regular matrix. Then there exists a maximum modulus eigenvalue ρ $(= \rho_1)$ which is positive. The associated left (row) and right (column) eigenvectors $u = (u_1, \ldots, u_p)$ and $v = (v_1, \ldots, v_p)$ have strictly positive components. When the*

*j.biggins@sheffield.ac.uk
†RAHIMSAN@saba.tmu.ac.ir

eigenvectors are normalized so that $\sum_{j=1}^{p} u_j = 1$ *and* $\sum_{j=1}^{p} u_j v_j = 1$,

$$A^n = \rho^n vu + O\left(n^{(m_2-1)}|\rho_2|^n\right) \quad \textit{(elementwise)},$$

as $n \to \infty$.

Let the entries $m_{ij}(\lambda)$ of the matrix $M(\lambda) = \{m_{ij}(\lambda)\}_{p\times p}$ be functions of $\lambda \in L$. Clearly the eigenvalues and their multiplicities all depend on λ. The essential characteristics we need from $M(\lambda)$ are summarised in the following condition. In it, and the associated results, the brackets give an alternative version.

PF: *For all* i, j, $m_{ij}(\lambda)$ *are* q (≥ 1) *times continuously differentiable (analytic) functions in* $\lambda \in L$, *where* L *is an open set in* $\mathbb{R}^d$ *(in* $\mathbb{C}^d$*), and for any* $\theta \in L_- \subset L$, $M(\theta)$ *is positive regular.*

The first result, which is routine to establish, gives the existence of the maximum modulus eigenvalue and properties of its eigenvectors, provided λ is near enough to L_-.

Proposition 1 *Let* $M(\lambda)$ *satisfy the condition* **PF** *on the open set* L. *Then there is an open set* $\Omega \subset L$ *containing* L_- *such that for* $\lambda \in \Omega$ *the following hold.*

(i) $M(\lambda)$ *has a unique maximum-modulus eigenvalue,* $\rho(\lambda)$ *that is* q*-times continuously differentiable (analytic) in* λ.

(ii) The left and right eigenvectors associated with $\rho(\lambda)$, $u(\lambda)$ *and* $v(\lambda)$, *normalised so that* $\sum_{i=1}^{p} u_i(\lambda) = 1$ *and* $u(\lambda)v(\lambda) = \sum_{i=1}^{p} u_i(\lambda)v_i(\lambda) = 1$, *are* q*-times continuously differentiable (analytic) in* λ *and, for all* i, $u_i(\lambda) \neq 0$ *and* $v_i(\lambda) \neq 0$.

The analyticity of $\rho(\theta)$ was established when the matrix entries are a function of a single variable, in [13]. For matrices of Laplace-Stieltjes transforms, nearly all the conclusions of this theorem are contained in Theorem 4.1 of [14]; see also Theorem 11.5.1 of [12] and Proposition 4.8(iii) in [10].

The second result, which is the main one described here, gives an asymptotic estimate, equation (1), of $M^n(\lambda)$, as $n \to \infty$. The key point, for the applications, is that (1) is a uniform bound on the rate of coinvergence in a suitable neighbourhood of any point in L_-. This uniformity has no direct parallel in the Perron-Frobenius Theorem given above. Although it is 'obvious' uniformity should hold it proved quite tricky to establish, which may explain why refined large deviations for Markov additive processes governed by a finite Markov chain have not yet been given, though extensions in other directions of large deviation results for these processes have been very successful — [8], [15].

Theorem 2 *Let the matrix* $M(\lambda)$ *satisfy the condition* **PF** *on the open set* L, *let* $\Omega \subset L$ *be the set introduced in Proposition* 1 *and let* $\Xi^{(n)}(\lambda) = \{\xi_{ij}^{(n)}(\lambda)\}_{p\times p}$ *be defined by*

$$M(\lambda)^n = v(\lambda)u(\lambda)\rho(\lambda)^n + \Xi^{(n)}(\lambda).$$

(Note that $\Xi^{(n)}$ *is not the nth power of* $\Xi^{(1)}$*.) For any* $\vartheta \in L_-$, *there is a neighbourhood of* ϑ, *say* B, *with* $B \subset \Omega$ *and a constant* $\gamma \in (0, 1)$ *such that*

$$\gamma^{-n} \max_{i,j} \sup_{\lambda\in B} \left|\xi_{ij}^{(n)}(\lambda)\rho(\lambda)^{-n}\right| \to 0, \quad \textit{as } n \to \infty. \tag{1}$$

Suppose M is a matrix of Laplace-Stieltjes transforms of measures so that

$$m_{ij}(\lambda) = \int e^{\lambda x} dF_{ij}(x); \quad \lambda \in \mathbb{C}^d,$$

which is analytic on the interior of the set where it is finite. Then we can use

$$L = \bigcap_{i,j} \text{int}\{\lambda = \theta + \mathbf{i}\eta \in \mathbb{C}^d \ : \ m_{ij}(\theta) < \infty\}$$

and

$$L_- = \{\lambda \in L : \lambda = \theta + \mathbf{i}\eta \text{ with } \eta = 0\}.$$

Instead, suppose M is a matrix of Fourier-Stieltjes transforms of finite measures so that

$$m_{ij}(\lambda) = \int e^{\mathbf{i}\lambda x} dF_{ij}(x); \quad \lambda \in \mathbb{R}^d.$$

If in addition

$$\int |x|^q dF_{ij}(x) < \infty$$

then these are q-times continuously differentiable functions and we can use

$$L = \mathbb{R}^d \text{ and } L_- = \{0\}.$$

3 Proofs

The first lemma deals with the existence of a continuously differentiable (or analytic) maximum-modulus eigenvalue. For $\lambda \in \mathbb{R}^d$ (or $\mathbb{C}^d$), $B(\lambda, \epsilon)$ is the open ball of radius ϵ centred on λ.

Lemma 3 *Let the matrix $M(\lambda)$ satisfy the condition* **PF** *on the open set L and let $\theta \in L_-$. Then there is a neighbourhood of θ, $B = B(\theta, \delta) \subset L$, such that for all $\lambda \in B$, $M(\lambda)$ has a maximum-modulus eigenvalue, $\rho(\lambda)$, which is q-times continuously differentiable (analytic).*

Proof. For $\theta \in L_-$, $M(\theta)$ is a non-negative matrix. Its Perron-Frobenius eigenvalue is $\rho(\theta)$ and the other (distinct) eigenvalues are $\rho_2(\theta), \ldots, \rho_s(\theta)$. For $\lambda \in L$, the characteristic polynomial of $M(\lambda)$, given by $f(z, \lambda) = \det(zI - M(\lambda))$ is q-times continuously differentiable. (In the analytic case, $f(z, \lambda) = \det(zI - M(\lambda))$ is analytic in each of the variables z, $\lambda_1, \ldots \lambda_d$ when the others are fixed. Hence it is analytic as a function of several complex variables — see [6, Theorem 2.2.8].) Since $\rho(\theta)$ is a simple root (i.e. with multiplicity one) of the equation $f(z, \theta) = 0$, f satisfies the conditions of the implicit function theorem at $(\rho(\theta), \theta)$; see [5, 10.2.2–10.2.4]. Thus there is a neighbourhood $B = B(\theta, \delta) \subset L$ in which there is a uniquely determined q-times continuously differentiable (analytic) function $r(\lambda)$ taking the value $\rho(\theta)$ at θ and satisfying $f(r(\lambda), \lambda) = 0$ for all $\lambda \in B$. Thus $r(\lambda)$ is a eigenvalue of $M(\lambda)$ for $\lambda \in B$.

It remains to show that for δ sufficiently small $r(\lambda)$ is simple and largest in magnitude. Let 3ϵ be less than $(\rho(\theta) - |\rho_2(\theta)|)$ and small enough that balls of radius 3ϵ centred on the distinct eigenvalues of $M(\theta)$ are disjoint. Then any point within ϵ of one of $\rho_2(\theta), \ldots, \rho_s(\theta)$ is smaller in magnitude than every point in the ϵ ball centred on $\rho(\theta)$.

Let δ be small enough to ensure that $|r(\lambda) - \rho(\theta)| < \epsilon$ for all $\lambda \in B$. The eigenvalues of $M(\lambda)$ are given by the p roots of the polynomial (in z) $\det(zI - M(\lambda))$. This polynomial has coefficients that are continuous in λ. By taking δ small enough the maximum distance between the roots at θ and those at $\lambda \in B(\theta, \delta)$, after a suitable pairing of the two sets of roots, is less than ϵ (see Appendix D of [7]). Then there can only be one root within ϵ of $\rho(\theta)$, which therefore must be $r(\lambda)$, and that root must be larger in magnitude than all others, by the choice of ϵ. □

Proof of Proposition 1. Let $\theta \in L_-$; first we prove the theorem in a neighbourhood of θ. By Lemma 3, the maximum-modulus eigenvalue $\rho(\lambda)$ exists for all $\lambda \in B = B(\theta, \delta_1)$ for a suitable δ_1. The related left eigenvector $u(\lambda)$ satisfies the system of p linear equations $u(\lambda)(\rho(\lambda)I - M(\lambda)) = 0$; the characteristic polynomial of $M(\lambda)$ is the determinant of this linear system. Since the maximum-modulus eigenvalue is a simple root of the characteristic equation, the system has rank $p - 1$. Consequently, there is a non-trivial solution with one free variable. The continuity of the entries of $M(\lambda)$ as functions of λ means that for some $\delta \le \delta_1$ there is a submatrix, of order $(p-1) \times (p-1)$, with non-zero determinant in $B(\theta, \delta)$. Then, throughout this neighbourhood the same free variable can be used. Adding an additional equation to the system, namely $\sum_{k=1}^{p} u_k(\lambda) = 1$, produces a new system of $p + 1$ equations with a unique solution, depending on λ, which is a left eigenvector with the prescribed normalisation. Having chosen $u(\lambda)$ the same argument is used for the right eigenvector, $v(\lambda)$, but the additional equation is now $\sum_{k=1}^{p} u_k(\lambda) v_k(\lambda) = 1$. The coefficients in these systems of equations are q-times continuously differentiable (or analytic) and so the eigenvectors are too. Using the continuity of u and v and the fact that they are strictly positive at θ we can choose δ so that $u(\lambda) \neq 0$ and $v(\lambda) \neq 0$ for all $\lambda \in B$. Finally, defining Ω to be $\bigcup_{\theta \in L_-} B(\theta, \delta)$ completes the proof. □

Before starting the proof of Theorem 2, it will be convenient to change notation slightly. For the matrix $M(\lambda)$ satisfying the condition **PF** on the set L we denote the p eigenvalues of $M(\lambda)$ on $\Omega \subset L$ by

$$|\rho_1| > |\rho_2| \ge \ldots \ge |\rho_p|, \tag{2}$$

(so now there is repetition according to multiplicity) with $\rho = \rho_1$. Note that the eigenvalues depend on λ but this is left implicit in (2). At several points in the proof, where it makes the presentation clearer, the arguments of functions will be suppressed in this way.

The resolvent of $M(\lambda)$ is defined by

$$R(z) = \{r_{ij}(z)\}_{p\times p} = (I - zM(\lambda))^{-1},$$

which, for all i and j, has the expansion $r_{ij}(z) = \sum_{n=0}^{\infty} z^n m_{ij}^n(\lambda)$ (see Theorem 11.1.1 in [12]) when $z < ||M(\lambda)||^{-1}$, where $||\cdot||$ is a matrix norm. In the next lemma, we draw on ideas in the proof of Theorem 1.1, in [16] to obtain properties of $r_{ij}(z)$. As will become clear, the work in this lemma is identifying the numerator in the first term of (3).

Lemma 4 *Let $M(\lambda)$ satisfy the condition* **PF** *and let $\theta \in L_-$. Then there is a neighbourhood $B = B(\theta, \delta) \subset \Omega$ and there are finite functions $b_{k,ij}(\lambda)$ defined on B such that, for all $\lambda \in B$ and all i and j,*

$$r_{ij}(z) = \frac{v_i u_j}{1 - z\rho} + \frac{\sum_{k=0}^{p-2} b_{k,ij} z^k}{\prod_{k=2}^{p}(1 - z\rho_k)} \tag{3}$$

for all sufficiently small z, where on the right of (3) the functional dependence (of u, v, the ρs and the bs) on λ has been supressed in the notation.

Proof. For fixed $\theta \in L_-$, let $B = B(\theta, \delta) \subset \Omega$ and M_1 be such that for all $\lambda \in B$, $||M(\lambda)|| < M_1$. We take $|z| < 1/M_1$ and $\lambda \in B$ so that, using [12, Theorem 10.3.1], $|\rho(\lambda)z| < 1$ throughout B. Now, as in (3), supress λ in the notation.

Let $h(z) = \det(I - zM)$, then, using (2), $h(z) = (1 - z\rho)\prod_{k=2}^{p}(1 - z\rho_k)$. Let $\operatorname{adj}(I - zM) = \{d_{ij}(z)\}_{p\times p}$, then each d_{ij} is a polynomial in z, with degree at most $p - 1$, and with coefficients that are continuously differentiable in $\lambda \in B$. Inverting $I - zM$ gives

$$r_{ij}(z) = \frac{d_{ij}(z)}{(1 - z\rho)\prod_{k=2}^{p}(1 - z\rho_k)} = \frac{\sum_{k=0}^{p-1} c_{k,ij} z^k}{(1 - z\rho)\prod_{k=2}^{p}(1 - z\rho_k)}. \tag{4}$$

The right hand side can be rewritten, by partial fractions, to give

$$r_{ij}(z) = \frac{a_{ij}}{1 - z\rho} + \frac{\sum_{k=0}^{p-2} b_{k,ij} z^k}{\prod_{k=2}^{p}(1 - z\rho_k)}$$

where

$$a_{ij} = \frac{-\rho d_{ij}(1/\rho)}{h'(1/\rho)}. \tag{5}$$

To evaluate a_{ij}, note first that

$$h(z)\, I = (I - zM)\operatorname{adj}(I - zM) = \operatorname{adj}(I - zM)(I - zM) \tag{6}$$

and $h(1/\rho) = 0$. Thus each row of $\operatorname{adj}(I - (1/\rho)M) = \{d_{ij}\}$ is a left eigenvector of M (corresponding to the eigenvalue ρ) and each of its columns is a right eigenvector. The simple maximum-modulus eigenvalue ρ has non-zero left and right eigenvectors u and v; hence there is a $c \neq 0$ such that, for all i and j, $d_{ij} = cv_i u_j$, that is, $\operatorname{adj}(I - (1/\rho)M) = cvu$. Now differentiating (6) with respect to z and multiplying through by v gives

$$h'(z)\, v = \frac{d(\operatorname{adj}(I - zM))}{dz}(I - zM)\, v - \operatorname{adj}(I - zM)\, Mv;$$

so, setting $z = 1/\rho$ and using the normalisation $uv = 1$,

$$h'(1/\rho)\, v = 0 - \operatorname{adj}(I - (1/\rho)M)(\rho v) = -(cvu)(\rho v) = -\rho cv.$$

Hence $h'(1/\rho) = -\rho c$ and substitution into (5) gives $a_{ij} = v_i u_j$. □

Proof of Theorem 2. For any $\theta \in L_-$, let $B = B(\theta, \delta) \subset \Omega$ be the neighbourhood and M_1 the bound on $||M(\lambda)||$ introduced in Lemma 4. To translate (3) into an asymptotic estimate of $M(\lambda)^n$, information on the boundedness of $b_{k,ij}$ as λ varies is needed. This is considered first.

The function $h(z)$ is a polynomial in z with coefficients that are continuously differentiable in $\lambda \in B$, and $1/\rho$ is a simple root of this polynomial. Hence $h_1(z) = h(z)/(1-\rho z)$ is a polynomial in z of degree $p-1$ with coefficients that are continuously differentiable in $\lambda \in B$ and, as noted earlier, the same is true of $d_{ij}(z)$. From (3) and (4)

$$v_i u_j h_1(z) + \sum_{k=0}^{p-2} b_{k,ij} z^k (1 - z\rho) = d_{ij}(z).$$

Equating powers of z here shows that the $b_{k,ij}$ are continuously differentiable functions of $\lambda \in B$. Hence the supremum of $\left|b_{k,ij}(\lambda)/\rho(\lambda)^k\right|$ over i, j, k and λ in the closed ball of radius $\delta/2$ centred at θ will be finite; denote this finite supremum by K.

Let $\lambda \in B(\theta, \delta)$ be fixed. Note that

$$\frac{1}{\prod_{k=2}^{p}(1 - z\rho_k)} = \sum_{n=0}^{\infty} \left(\sum_{k_2+\ldots+k_p=n} \rho_2^{k_2} \ldots \rho_p^{k_p} \right) z^n,$$

where the inner sum is taken over all partitions of n into the sum of $(p-1)$ non-negative integers. Applying this in (3) gives

$$r_{ij}(z) = \sum_{n=0}^{\infty} \left(v_i u_j \rho^n + \sum_{h=0}^{p-2} b_{h,ij} \sum_{k_2+\ldots+k_p=n-h} \rho_2^{k_2} \ldots \rho_p^{k_p} \right) z^n;$$

but, as noted already, $r_{ij}(z) = \sum_{n=0}^{\infty} z^n m_{ij}^n(\lambda)$ and by definition $\xi_{ij}^{(n)} = m_{ij}^n - v_i u_j \rho^n$. Hence

$$\xi_{ij}^{(n)} = \sum_{h=0}^{p-2} b_{h,ij} \sum_{k_2+\ldots+k_p=n-h} \rho_2^{k_2} \ldots \rho_p^{k_p}.$$

To complete the proof we use this expression to give an upper bound for the entries in $\Xi^{(n)}\rho^{-n}$ that is uniform in λ. Let $\epsilon > 0$ be as defined in Lemma 3, with δ then chosen at least as small as there. Then, for $\lambda \in B(\theta, \delta)$, and $j = 2, 3, \ldots, p$

$$\left| \frac{\rho_j(\lambda)}{\rho(\lambda)} \right| \leq \frac{|\rho_2(\theta)| + \epsilon}{\rho(\theta) - \epsilon} \leq \frac{\rho(\theta) - 2\epsilon}{\rho(\theta) - \epsilon} = \alpha < 1.$$

Then for all $k \geq 2$, and $\lambda \in B(\theta, \delta/2)$,

$$\left| \rho^{-n} \xi_{ij}^{(n)} \right| \leq K \sum_{h=0}^{p-2} \alpha^{n-h} \sum_{k_2+\ldots+k_p=n-h} 1.$$

In the last sum, each k_j, $(2 \leq j \leq p)$, can take at most $n-h+1$ values, and hence the sum has at most $(n-h+1)^{p-1} \leq (n+1)^{p-1}$ terms. Hence, $|\rho^{-n}\xi_{ij}^{(n)}| \leq C[(n+1)^{p-1}\alpha^n]$, where $C = K(p-1)\alpha^{-p}$ is independent of n and λ. The result claimed now follows directly for any $\gamma \in (\alpha, 1)$. □

References

[1] J. D. Biggins, The first- and last-birth problems for a multitype age-dependent branching process, *Adv. Appl. Probab.*, 8 (1976), 446–459.

[2] J. D. Biggins, Uniform convergence of martingales in the branching random walk, *Ann. Probab.*, 20 (1992), 137–151.

[3] J. D. Biggins, and A. Rahimzadeh Sani, (2003). Convergence of transforms in multitype branching random walk.

[4] M. Bramson, P. Ney, and J. Tao, The population composition of a multitype branching random walk, *Ann. Appl. Probab.*, 2 (1992), 575–596.

[5] J. DIEUDONNÉ, *Foundations of Modern Analysis*, Academic Press, London, 1960.

[6] L. HÖRMANDER, *An Introduction to Complex Analysis in Several Variables*, North-Holland, Amsterdam, 1973.

[7] R. A. HORN AND C. R. JOHNSON, *Matrix Analysis*, Cambridge University Press, Cambridge, 1985.

[8] ISCOE, I.; NEY, P. AND NUMMELIN, E. (1985). Large Deviations of Uniformly recurrent Markov Additive processes. *Adv. Appl. Math.* **6**, 375–412.

[9] J. L. JENSEN, *Saddlepoint Approximations*, Oxford University Press, Oxford, 1995.

[10] Y. KONTOYIANNIS AND S.P. MEYN (2003). Spectral theory and limit theorems for geometrically ergodic Markov processes. *Ann. Appl. Probab.* **13**, 304–362.

[11] A. KYPRIANOU, AND A. RAHIMZADEH SANI, (2001). Martingale convergence and the functional equation in the multi-type branching random walk. *Bernoulli.* **7**, 593–604.

[12] P. LANCASTER AND M. TISMENETSKY, *The Theory of Matrices, second edition with applications*, Academic Press, New York, 1985.

[13] H. D. MILLER, A convexity property in the theory of random variables defined on a finite Markov chain, *Ann. Math. Statist.*, 32 (1961), pp 1260–1270.

[14] NEY, P. AND NUMELLIN, E. (1987). Markov Additive Processes, I. *Ann. Probab.* **15**, 561–592.

[15] NEY, P. AND NUMELLIN, E. (1987). Markov Additive Processes, II. *Ann. Probab.* **15**, 593–609.

[16] E. SENETA, *Non-negative Matrices*, Allen and Unwin, London, 1973.

J.D. BIGGINS
DEPARTMENT OF PROBABILITY AND STATISTICS,
HICKS BUILDING,
THE UNIVERSITY OF SHEFFIELD,
SHEFFIELD, S3 7RH,
U.K.

A. RAHIMZADEH SANI,
DEPARTMENT OF MATHEMATICS,
TEACHER TRAINING UNIVERSITY OF TEHRAN,
49 MOFATTEH AVENUE, TEHRAN, 15614,
IRAN.