Math. Proc. Camb. Phil. Soc. (1979), **86**, 345
Printed in Great Britain

Random evolutions and the spectral radius of a non-negative matrix

BY JOEL E. COHEN
The Rockefeller University, New York

(*Received* 14 *November* 1978)

1. *Introduction and summary.* This paper offers yet another example of what probability theory can do for analysis. Using a Feynman–Kac formula derived in the theory of random evolutions (5), we find an expression (1) for the spectral radius $r(A)$ of a finite square non-negative matrix A. This expression makes it very easy to study how $r(A)$ behaves as a function of the diagonal elements of A.

Kac (7) derived an expression of the same form as (1) for the principal eigenvalue of a second-order ordinary differential equation, using a Feynman–Kac formula for Brownian motion rather than for a finite-state Markov chain. His result has been extensively generalized (Donsker and Varadhan (3)).

A direct derivation of (1) for non-negative matrices and the two main consequences of (1) derived in section 2 (inequalities (7) and (9)) may be new. Inequality (7) is a lower bound for $r(A)$ when A is irreducible. Inequality (9) asserts that, whether A is irreducible or not, $r(A)$ is a convex function of the main diagonal of A.

Section 3 reviews alternative, partially successful approaches to the same results.

This paper substantially generalizes the major results of (2) and provides much easier proofs. It is mathematically independent of (2) but does not contain some of the special results developed there for demographic applications of non-negative matrices.

For general background and definitions not provided below, see (**4, 9, 11, 14**).

2. *Feynman–Kac meet Perron–Frobenius.* Let A be a matrix of order

$$n \times n \quad (1 < n < \infty)$$

with finite possibly complex elements a_{ij}. The spectral radius $r(A)$ of A is the largest of the magnitudes $|\lambda_i|$ of the eigenvalues $\lambda_i, i = 1, \ldots, n$ of A. If $\|.\|$ is any matrix norm, then $r(A) = \lim_{t\to\infty} \|A^t\|^{1/t}$. We shall abbreviate $\lim_{t\to\infty}$ to $\lim_t$. We adopt the row sum norm $\|A\| := \max_i \sum_j |a_{ij}|$ (where ':=' means equality by definition).

Henceforth let A be non-negative, i.e. every $a_{ij} \geq 0$; we write $A \geq 0$. By the Perron–Frobenius theorem at least one of the eigenvalues of A, say λ_1, is real and equal to $r(A) \geq 0$. There exists an n-vector $u = (u_i)$, $u_i \geq 0$, such that $Au = \lambda_1 u$. If $A \geq 0$, then $e^A \geq 0$ and $r(e^A) = e^{r(A)}$. If A is irreducible, then $r(A) = \lambda_1 > 0$, λ_1 is simple (no other eigenvalue of A equals λ_1), and all $u_i > 0$. We do not assume A is irreducible unless we say so explicitly.

Let $s_i := \sum_j a_{ij}$ and $S := (s_{ij})$, where $s_{ii} := s_i, s_{ij} := 0$ if $i \neq j$; S is $n \times n$. Then $Q := A - S$ has zero row sums and non-negative off-diagonal elements. Thus Q defines the intensity matrix or infinitesimal generator of a continuous-time homogeneous

0305-0041/79/0000-6480 $03.50

Markov chain $W(t)$, $t \geqslant 0$, with state space $X = \{1, \ldots, n\}$. If A is irreducible, then Q is irreducible also and 0 is a simple eigenvalue of Q. Let E_i denote the (conditional) expectation over all sample paths of the Markov chain $W(t)$ which are initially in state i, i.e. such that $W(0) = i$. Let $g_j(t)$ be the occupation time in the jth state up to time t. In general, given $W(0) = i$, it is possible that $g_j(t) = t$ for all t (e.g. if $i = j$ and $a_{ik} := q_{ik} = 0$ for $k \neq i$) or that $g_j(t) = 0$ for all t (e.g. if state j is isolated from state i) or that $0 < g_j(t) < t$. If A is irreducible, then there exists a positive n-vector $\pi := (\pi_i)$ such that for almost every sample path of the chain, $\lim_t g_j(t)/t = \pi_j > 0$. Chung((1), p. 93) proves this limit theorem for a discrete-parameter chain and leaves the continuous-parameter analogue 'as a long exercise for the interested reader' (p. 228).

Let $\max_i := \max_{i \in X}$, $\min_i := \min_{i \in X}$.

THEOREM 1. *If* $A \geqslant 0$,

$$r(A) = \lim_t (1/t) \log \max_i E_i[\exp(\Sigma_j s_j g_j(t))]. \tag{1}$$

Proof. $r(A) = \log r(e^A) = \log \lim_t \|e^{At}\|^{1/t} = \lim_t \log \|e^{At}\|^{1/t} = \lim_t (1/t) \log \max_i$ (ith row sum of e^{At}). Let $u(t) := (u_i(t))$ be an n-vector function of time t such that

$$du/dt = Au, \quad u_i(0) = 1. \tag{2}$$

The unique solution of (2) is

$$u(t) = e^{At} u(0), \tag{3}$$

i.e. $u_i(t) = i$th row sum of e^{At}.

We now define a 'random evolution' (5) whose 'expectation semigroup' is also a solution of (2). This example of a random evolution is more elementary than any of those considered in (5).

Since almost every sample path $W(t)$ is a step function, let t_k be the time of the kth ($k \geqslant 1$) jump or change of state of $W(t)$: $0 < t_1 < t_2 \ldots$, and let $N(t)$ be the number of jumps up to time t. $W(t_k)$ is the state entered after the kth jump. If $W(0) = i$, let

$$M(t) = \exp(s_i t_1) \exp(s_{W(t_1)}(t_2 - t_1)) \ldots \exp(s_{W(N(t))}(t - t_{N(t)})). \tag{4}$$

At any given time except a jump point t_j, $M(t)$ is an exponentially growing quantity, initially ($t = 0$) equal to 1, in which the growth rate $s_{W(t)}$ is chosen by the state of the Markov chain $W(t)$. Since the scalars s_j, $j \in X$, all commute, $E[M(t) | W(0) = i] = E_i[\exp(\Sigma_j s_j g_j(t))]$. Griego and Hersh ((5), p. 411) proved a generalized Feynman–Kac formula; namely

$$u_i(t) = E_i[\exp(\Sigma_j s_j g_j(t))] \tag{5}$$

solves (2). Since any solution such as (5) is also the ith row sum of e^{At}, substitution in the expression developed for $r(A)$ gives Theorem 1.

A consequence of Theorem 1, originally due to Frobenius, is well known ((4), pp. 63, 68).

COROLLARY 1. *Label the rows and columns of* $A \geqslant 0$ *so that* $s_1 = \max_i s_i$ *and* $s_n = \min_i s_i$. *Then*

$$s_n \leqslant r(A) \leqslant s_1. \tag{6}$$

Proof. Since $\Sigma_j g_j(t) = t$, we have from (1) that

$$r(A) \leqslant \lim_t (1/t) \log \max_i E_i[e^{s_1 t}] = s_1;$$

similarly

$$r(A) \geqslant \lim_t (1/t) \log \max_i E_i[e^{s_n t}] = s_n.$$

COROLLARY 2. *If $A \geqslant 0$ is irreducible,*

$$s_n < \Sigma_j \pi_j s_j \leqslant r(A) \tag{7}$$

except when $s_1 = s_n$, in which case $r(A) = s_n$.

Proof. Since e^x is convex in x, we have, for every t,

$$E_i[\exp(\Sigma_j s_j g_j(t))] \geqslant \exp(\Sigma_j s_j E_i(g_j(t))),$$

hence
$$\log \max_i E_i[\exp(\Sigma_j s_j g_j(t))] \geqslant \log \max_i \exp(\Sigma_j s_j E_i(g_j(t)))$$

and
$$(1/t) \log \max_i E_i[\exp(\Sigma_j s_j g_j(t))] \geqslant \max_i \Sigma_j s_j E_i[g_j(t)/t].$$

Passing to the limit, since $\lim_t E_i[g_j(t)/t] = \pi_j > 0$ for all i and j, we have, from Theorem 1, $r(A) \geqslant \Sigma_j \pi_j s_j$. Since all $\pi_j > 0$, $\Sigma_j \pi_j s_j > s_n$ except when $s_1 = s_n$.

It would be desirable, but seems difficult, similarly to derive from (1) an upper bound on $r(A)$ which improves on that in (6) when $A \geqslant 0$ is irreducible.

We illustrate the numerical power of Corollary 2 with an example which Marcus and Minc ((11), p. 158) used to compare localization theorems for the spectral radius of a positive square matrix. Let

$$A = \begin{pmatrix} 1 & 1 & 2 \\ 2 & 1 & 3 \\ 2 & 3 & 5 \end{pmatrix}.$$

Then
$$Q = \begin{pmatrix} -3 & 1 & 2 \\ 2 & -5 & 3 \\ 2 & 3 & -5 \end{pmatrix}$$

and $\pi^T Q = 0$ is satisfied by $\pi^T = (\frac{16}{40}, \frac{11}{40}, \frac{13}{40})$. From Corollary 2, $6{\cdot}5 \leqslant r(A)$. Since $r(A^T) = r(A)$, we may also apply Corollary 2 to A^T, noting that the Q matrix of A^T is not the transpose of the Q matrix of A. For Q derived from A^T, $\pi^T Q = 0$ is satisfied by $\pi^T = (\frac{11}{41}, \frac{16}{41}, \frac{14}{41})$. From Corollary 2, $6{\cdot}707 \leqslant r(A^T)$. Combining the results for $r(A)$ and $r(A^T)$ gives $6{\cdot}707 \leqslant r(A)$. According to (11), $r(A) = 7{\cdot}531$. The best lower bound in (11) is $5{\cdot}162 \leqslant r(A)$. This bound, due to a theorem of Ostrowski, requires knowledge of the least element of A in addition to the row sums and a positive eigenvector corresponding to $r(A)$.

Lemma 1 of (3) gives an upper bound on $r(A)$,

$$r(A) \leqslant \max_i \left(s_i + (1/u_i) \sum_{j=1}^{n} q_{ij} u_j \right),$$

where $Q = (q_{ij})$ and $u = (u_i)$ is any n-vector with all $u_i > 0$. Richard Griego has observed that, in this numerical example, with $u_i = i$, one obtains $r(A) \leqslant 9$. This bound is better than the best upper bound, $9{\cdot}359$, given by Marcus and Minc ((11), p. 158).

THEOREM 2. *Let $V := (v_{ij})$ be an $n \times n$ non-negative diagonal matrix ($v_{ij} = 0$ for $i \neq j$) such that for at least one i, $0 \leqslant v_{ii} < v := \max_j v_{jj} < \infty$. If $A \geqslant 0$, then*

$$r(A) \leqslant r(A+V) \leqslant r(A) + v. \tag{8}$$

If A is irreducible, both inequalities in (8) are strict.

Proof. From (1), $r(A+V) = \lim_t (1/t) \log \max_i E_i[\exp(\Sigma_j s_j g_j(t)) \exp(\Sigma_j v_{jj} g_j(t))]$. For every sample path and every t, $1 \leq \exp(\Sigma_j v_{jj} g_j(t)) \leq e^{vt}$. Equation (8) follows.

If A is irreducible, so are $A+V$ and $A+vI$, where I is the $n \times n$ identity matrix. From (1) it is apparent that $r(A)+v = r(A+vI)$. Since at least one element of $A+V$ exceeds the corresponding element of A, and at least one element of $A+vI$ exceeds the corresponding element of $A+V$, it follows from a well-known lemma of Wielandt ((4), p. 57) that $r(A) < r(A+V) < r(A+vI)$, proving Theorem 2.

THEOREM 3. *Let V be any $n \times n$ non-negative diagonal matrix. For $A \geq 0$ and $0 < h < 1$,*

$$r((1-h)A + h(A+V)) \leq (1-h)r(A) + hr(A+V). \tag{9}$$

Proof. Using Theorem 1,

$$\begin{aligned}
&(1-h)r(A) + hr(A+V)\\
&\quad = (1-h)\lim_t (1/t)\log \max_i E_i[\exp(\Sigma_j s_j g_j(t))]\\
&\qquad + h \lim_t (1/t) \log \max_i E_i[\exp(\Sigma_j (s_j + v_{jj}) g_j(t))]\\
&\quad = \lim_t (1/t) \log (\max_i E_i[\exp(\Sigma_j s_j g_j(t))])^{1-h}\\
&\qquad + \lim_t (1/t) \log (\max_k E_k[\exp(\Sigma_j (s_j + v_{jj}) g_j(t))])^h\\
&\quad = \lim_t (1/t) \log \max_i (E_i[\exp(\Sigma_j s_j g_j(t))])^{1-h}\\
&\qquad + \lim_t (1/t) \log \max_k (E_k[\exp(\Sigma_j (s_j + v_{jj}) g_j(t))])^h\\
&\quad = \lim_t (1/t) \log [\max_i (E_i[\exp(\Sigma_j s_j g_j(t))])^{1-h} \max_k (E_k[\exp(\Sigma_j (s_j + v_{jj}) g_j(t))])^h]\\
&\quad \geq \lim_t (1/t) \log \max_i [(E_i[\exp(\Sigma_j s_j g_j(t))])^{1-h} (E_i[\exp(\Sigma_j (s_j + v_{jj}) g_j(t))])^h] \qquad (10)\\
&\quad \geq \lim_t (1/t) \log \max_i E_i[\exp((1-h)(\Sigma_j s_j g_j(t)) + h\Sigma_j (s_j + v_{jj}) g_j(t))] \qquad (11)\\
&\quad = \lim_t (1/t) \log \max_i E_i[\exp \Sigma_j ((1-h) s_j + h(s_j + v_{jj})) g_j(t)]\\
&\quad = r((1-h)A + h(A+V)).
\end{aligned}$$

The step from (10) to (11) follows from ((13), p. 68, § 81·2). This proves Theorem 3.

We conjecture that if, as in Theorem 2, V is not a scalar multiple of an identity matrix, and if A is irreducible, then the inequality in (9) is strict. This conjecture is known to be true if only a single element of V is positive (2). In general, consider the step from (10) to (11) for any fixed t, that is, before passing to $\lim_t$. Then the inequality between (10) and (11) is an equality if and only if ((6), p. 22, theorem 11) $\exp(\Sigma_j s_j g_j(t))$ is proportional to $\exp(\Sigma_j (s_j + v_{jj}) g_j(t))$ almost everywhere (a.e.) with respect to the probability measure on sample paths such that $W(0) = i$. Thus equality holds if and only if there exists $c > 0$ such that $c \exp(\Sigma_j s_j g_j(t)) = \exp(\Sigma_j s_j g_j(t)) \exp(\Sigma_j v_{jj} g_j(t))$ a.e., that is, such that $c = \exp(\Sigma_j v_{jj} g_j(t))$ a.e. This condition could be satisfied, e.g. if $v_{jj} = v$ for all j or if for some j, $g_j(t) = t$ a.e. But when A is irreducible and $V \neq vI$, it is not true that $c = \exp(\Sigma_j v_{jj} g_j(t))$ a.e. so that the inequality is strict. It must be shown that strict inequality persists in the limit.

Unlike the methods of proof in (2), the representation of the spectral radius and the methods used here appear to generalize immediately to countably infinite non-negative matrices A for which the associated matrix Q is the infinitesimal generator of a reasonably well behaved Markov chain. However, we shall not make this claim precise here.

3. *Perturbation theory and determinantal identities.* A natural approach to the conclusions of Theorems 2 and 3 would use perturbation theory for linear operators (8). If $A \geqslant 0$ is primitive, i.e. if, for some k, every element of A^k is positive, then the Perron–Frobenius theorem assures the existence of positive right and left eigenvectors corresponding to the unique (real) eigenvalue of modulus $r(A)$. Equation (8), with both inequalities strict, then follows by combining several results from ((8), p. 75, II (2·3) and II (2·6); p. 78, II (2·21); p. 80, II (2·34)).

To prove the convexity of $r(A)$ as a function of, say, a_{11} only, all other elements of A held constant (which is a special case of the conclusion of our Theorem 3), one must show $s_{11} < 0$ where now $S := (s_{ij})$ is the value at $r(A)$ of the reduced resolvent of A with respect to $r(A)$ ((8), p. 76, II (2·10); p. 40, I (5·27)). S is expressed in terms of the eigenprojections and nilpotents of A (which is T in (8)) in ((8), p. 40, I (5·32)). No direct demonstration that $s_{11} < 0$ has been found.

Still another way to show that $r(A)$ is a convex function of the main diagonal of $A \geqslant 0$ is to prove that the Hessian $H := (h_{ij})$, where $h_{ij} = \partial^2 r/\partial a_{ii}\partial a_{jj}$, and every principal submatrix of H are positive semi-definite (i.e. non-negative definite). If A is 2×2, this conclusion is immediate from ((2), p. 185, theorem 2). If A is 3×3, the conclusion has been proved directly by a long but elementary calculation of the determinants of the principal submatrices of H, obtaining h_{ij} by implicit differentiation of the characteristic polynomial of A. In fact the principal 1×1 and 2×2 submatrices of A are positive definite when $A_{3\times 3}$ is irreducible.

For an irreducible $n \times n$ matrix $A \geqslant 0$, since $\Sigma_{i=1}^{n} \partial r/\partial a_{ii} = 1$ ((2), p. 184), each row sum of H is 0. If it were true that $h_{ij} \leqslant 0$ whenever $i \neq j$, it would follow readily that H is positive semi-definite. However, there exist positive 3×3 matrices for which $h_{12} > 0$. It does follow that $\det(H) = 0$, where det := determinant. Moreover, since H is also symmetric, $\det(H(1)) = \ldots = \det(H(n))$, where $H(j)$ is the matrix formed by deleting the jth row and column of H ((12), p. 372, § 389), so the sign and magnitude of all primary minors of H are determined by the sign and magnitude of any one of them.

Since the characteristic equation of A implicitly defines $r(A)$ as a function of $(a_{11}, \ldots, a_{nn})$, a classical formula ((12), p. 660, ex. 11) for the determinant of the Hessian of an implicitly defined function could, in principle, be used to check the sign of the principal minors of H.

I am grateful for helpful comments from James Glimm, Richard Griego, Arthur Jaffe, Mark Kac, Samuel Karlin, Bruce Knight, George Papanicolaou, Gordon D. Simons, Jr., Burton Singer, and S. R. S. Varadhan. I thank members of the Department of Statistics, University of North Carolina, Chapel Hill, and of the Department of Mathematics, University of Wisconsin, Madison, for their hospitality during 1977 and 1978. National Science Foundation grant DEB74-13276 provided partial financial support for this work. I thank Anne Whittaker for typing the manuscript.

Added in proof: S. Friedland (manuscript, February 1979) has proved the conjecture that follows Theorem 3.

REFERENCES

(1) CHUNG, K. L. *Markov chains with stationary transition probabilities* (New York, Springer-Verlag, 1967).

(2) COHEN, J. E. Derivatives of the spectral radius as a function of non-negative matrix elements. *Math. Proc. Cambridge Philos. Soc.* **83** (1978), 183–190.

(3) DONSKER, M. D. and VARADHAN, S. R. S. On a variational formula for the principal eigenvalue for operators with maximum principle. *Proc. Nat. Acad. Sci. U.S.A.* **72** (1975), 780–783.

(4) GANTMACHER, F. R. *The theory of matrices*, vol. 2 (New York, Chelsea, 1960).

(5) GRIEGO, R. and HERSH, R. Theory of random evolutions with applications to partial differential equations. *Trans. Amer. Math. Soc.* **156** (1971), 405–418.

(6) HARDY, G. H., LITTLEWOOD, J. E. and PÓLYA, G. *Inequalities*, 2nd ed. (Cambridge University Press, 1952).

(7) KAC, M. On some connections between probability theory and differential and integral equations. *2nd Berkeley Symp. Math. Stat. Probab.*, ed. J. Neyman (Los Angeles, University of California Press, 1951), pp. 189–215.

(8) KATO, T. *Perturbation theory for linear operators*, 2nd ed. (New York, Springer-Verlag, 1976).

(9) LANCASTER, P. *Theory of matrices* (New York, Academic Press, 1969; reproduced by the author, Calgary, 1977).

(10) LOÈVE, M. *Probability theory*, 3rd ed. (Princeton, Van Nostrand, 1963).

(11) MARCUS, M. and MINC, H. *A survey of matrix theory and matrix inequalities* (Boston, Allyn and Bacon, 1964).

(12) MUIR, T. *A treatise on the theory of determinants*, rev. W. H. Metzler (privately published, Albany, New York; reprinted by Dover, New York).

(13) PÓLYA, G. and SZEGÖ, G. *Problems and theorems in analysis*, vol. 1, 4th ed. (New York, Springer-Verlag, 1972).

(14) SENETA, E. *Non-negative matrices; an introduction to theory and applications* (London, George Allen and Unwin, 1973).