LINEAR ALGEBRA AND ITS APPLICATIONS

ELSEVIER Linear Algebra and its Applications 287 (1999) 307–314

The convergence of general products of matrices and the weak ergodicity of Markov chains [1]

Michael Neumann [a], Hans Schneider [b,*]

[a] Department of Mathematics, University of Connecticut, Storrs, CT 06268, USA
[b] Department of Mathematics, University of Wisconsin, Madison, WI 53706, USA

Received 5 November 1997; accepted 14 September 1998

Submitted by V. Mehrmann

Dedicated to our good friend and valued colleague Ludwig Elsner on the occasion of his 60th birthday

Abstract

We determine a sufficient condition for the convergence to 0 of general products formed from a sequence of real or complex matrices. Our result is applied to obtain a condition for the weak ergodicity of an inhomogeneous Markov chain. We make some remarks comparing coefficients of ergodicity and we give a method for constructing these.

1. Introduction

Recently there has been much interest in conditions for the convergence of infinite products of real or complex matrices. Several investigations have concentrated on products taken in one direction – left or right, see for example the recent papers by Beyn and Elsner [2] and Hartfiel and Rothblum [5]. However, in this paper, we are concerned with *general products* formed from a

* Corresponding author. Tel.: 608 271 7252; fax: 608 271 8477; e-mail: hans@math.wisc.edu.

[1] Work supported by NSF Grant DMS-9306357.

0024-3795/99/$ – see front matter
PII: S0024-3795(98)10196-9

given infinite sequence of matrices. These are defined further on in the paper and they have previously been considered for nonnegative and stochastic matrices by Seneta in [15], Chs. 3.1 and 4.6. Such products include products taken in one direction.

Our principal result is a sufficient condition for the convergence to 0 of an infinite general product of matrices. Our hypothesis is on the norms of the matrices of the given sequence, and our proof uses classical results on the convergence and divergence of products of positive real numbers.

Our result is motivated by the theory of inhomogeneous Markov chains. We are here concerned with the weak ergodicity of such chains, see [14] and [15], Ch. 4 for references to the early history of this topic and for background material, see also [3,10,11,4].

Weak ergodicity of an inhomogeneous Markov chain is equivalent to the convergence to 0 of a sequence of stochastic matrices considered as operators on a common invariant space of codimension 1. The corresponding operator norms are called coefficients of ergodicity. Thus by applying our result on the convergence to 0 of a product of matrices to ergodicity coefficients, we obtain a sufficient condition for weak ergodicity.

The l_1 coefficient of ergodicity due to Bauer et al. [1] plays a special role in the theory of Markov chains as it is the only ergodicity coefficient associated with a norm that is less than or equal to 1 for all stochastic matrices, see [7] or [9]. However, we shall define coefficients of ergodicity associated with norms that are less than the ℓ_1 coefficient for many stochastic matrices, see [12] for a different technique for finding such coefficients. Thus it appears to be useful to prove results on weak ergodicity, such as ours, which allow ergodicity coefficients to exceed 1.

2. Convergence of infinite products

In this paper $\mathbb{F}$ will stand for the real field $\mathbb{R}$ or the complex field $\mathbb{C}$.

In this section we develop our main results concerning the convergence of products of complex matrices taken in an arbitrary order from an infinite sequence of matrices. Such products were considered (in a slightly less general form) by Seneta [15], Section 4.6) in the case of stochastic matrices, see also [8,13].

Let $A_1, A_2, \ldots$ be a sequence of complex matrices. We shall consider products of matrices obtained from the sequence in the following manner: First choose some permutation of the given infinite sequence to obtain a sequence $B_1, B_2, \ldots$ Then form the products $C_{p,r}$ of the matrices $B_{p+1}, \ldots, B_{p+r}$ in some order. We shall call $C_{p,r}$ a *general product* from the sequence $A_1, A_2, \ldots$ and we shall consider the existence of $\lim_{r\to\infty} C_{p,r}$. If this limit is 0, for all permutations of

$A_1, A_2, \ldots$ and all p, $p \geqslant 0$, then we shall say that all general products from the the sequence $A_1, A_2, \ldots$ converge to 0.

As an example of a sequence of general product suppose the chosen order is $A_{43}, A_9, A_7, A_5, A_{14}, A_2, \ldots$ Then the sequence of $(C_{2,1}, C_{2,2}, \ldots)$ may begin thus: $C_{2,1} = A_7$, $C_{2,2} = A_7A_5$, $C_{2,3} = A_5A_2A_7$, $C_{2,4} = A_2A_7A_{14}A_5$. Note that, for a given sequence $C_{p,1}, C_{p,2}, \ldots$ of general products each factor of $C_{p,r}$ occurs in $C_{p,r+1}$, but the order in which the factors occur in $C_{p,r}$ is arbitrary.

Let μ be a matrix norm (viz. a submultiplicative norm on $\mathbb{F}^{nn}$) and denote

$$\mu^+(P) = \max(\mu(P), 1) \quad \text{and} \quad \mu^-(P) = \min(\mu(P), 1).$$

Now let $A_1, A_2, \ldots$ be a sequence of matrices in $\mathbb{F}^{nn}$ and let μ be a matrix norm. We now define two conditions:

Condition (C). We say that the sequence $A_1, A_2, \ldots,$ satisfies Condition (C) for the norm μ if

$$\sum_{i=1}^{\infty} (\mu^+(A_i) - 1) \text{ converges.} \tag{1}$$

Condition (D). We say that the sequence $A_1, A_2, \ldots$ satisfies Condition (D) for the norm μ if

$$\sum_{j=1}^{\infty} (1 - \mu^-(A_i)) \text{ diverges.} \tag{2}$$

We are now ready to prove the following result.

Proposition 2.1. *Let $A_1, A_2, \ldots$ be a sequence of matrices in $\mathbb{F}^{nn}$. Let μ be a matrix norm on $\mathbb{F}^{nn}$. Suppose that the sequence $A_1, A_2, \ldots$ satisfies Condition* (C) *for the norm μ. Then all general products from $A_1, A_2, \ldots$ are bounded.*

Proof. Let $B_1, B_2, \ldots$ be a permutation of $A_1, A_2, \ldots$ and let $C_{p,r}$ be a product of $B_{p+1}, \ldots, B_{p+r}$ in some order. By Condition (C) and [6], Theorem 14, $\sum_{i=1}^{\infty} (\mu^+(B_i) - 1)$ converges and hence $\sum_{i=1}^{\infty} (\mu^+(B_{p+i}) - 1)$ also converges. Thus, by [6], Theorem 51, the product $\prod_{i=1}^{\infty} \mu^+(B_{p+i})$ converges and so there exists a positive constant M such that $\prod_{i=1}^{r} \mu^+(B_{p+i}) \leqslant M$, for each $r \in \{1, 2, \ldots\}$. It follows that

$$\mu(C_{p,r}) \leqslant \mu(B_{p+1}) \cdots \mu(B_{p+r}) = \left[\prod_{i=1}^{r} \mu^-(B_{p+i})\right] \left[\prod_{i=1}^{r} \mu^+(B_{p+i})\right]$$

$$\leqslant M \left[\prod_{i=1}^{r} \mu^-(B_{p+i})\right] \leqslant M. \qquad \square \tag{3}$$

The above proposition allows us to prove a stronger result under an additional condition. Note that in the theory of infinite products of nonnegative numbers it is customary to speak of *divergence* to 0 (see e.g. [6], p. 93).

Theorem 2.2. *Let $A_1, A_2, \ldots$ be a sequence of matrices in $\mathbb{F}^{nn}$. Let μ be a matrix norm on $\mathbb{F}^{nn}$. Suppose that the sequence $A_1, A_2, \ldots$ satisfies Conditions* (C) *and* (D) *for the norm μ. Then all general products from $A_1, A_2, \ldots$ converge to* 0.

Proof. Let $B_1, B_2, \ldots$ be a permutation of $A_1, A_2, \ldots$ and let $C_{p,r}$ be a product of $B_{p+1}, \ldots, B_{p+r}$ in some order. As in the proof of Proposition 2.1, we have that

$$\mu(C_{p,r}) \leqslant M\left[\prod_{i=1}^{r} \mu^{-}(B_{p+i})\right]. \tag{4}$$

By Condition (D), the sum $\sum_{i=1}^{\infty}(1 - \mu^{-}(A_i))$ diverges and so, by [6], Theorem 14, $\sum_{i=1}^{\infty}(1 - \mu^{-}(B_i))$ diverges. Thus $\sum_{i=1}^{\infty}(1 - \mu^{-}(B_{p+i}))$ also diverges.

We again apply Theorem 51 of [6] to obtain that $\prod_{i=1}^{\infty} \mu^{-}(B_{p+i})$ diverges. But since $\mu^{-}(B_i) \leqslant 1$, the last product must diverge to 0 and the proof is done. □

3. Applications to stochastic matrices

In this section we apply the foregoing results to stochastic matrices. In order to be consistent with our previous section we consider *column* stochastic matrices. Thus "stochastic matrix" will mean "column stochastic matrix".

Let $e = (1, \ldots, 1)^{\mathrm{T}} \in \mathbb{R}^n$ and let

$$H = \{x \in \mathbb{R}^n : e^{\mathrm{T}}x = 0\}. \tag{5}$$

If A is a stochastic matrix in $\mathbb{R}^{n,n}$, then H is invariant under A. If ν is a norm on $\mathbb{R}^n$ and A is a stochastic matrix, then corresponding *coefficient of ergodicity* is defined by

$$\nu_{\mathrm{e}}(A) = \sup_{0 \neq x \in H} \frac{\nu(Ax)}{\nu(x)}, \tag{6}$$

as is usual in the literature on Markov chains, see for example [13]. We may extend the definition of ν_{e} to all matrices A in $\mathbb{R}^{nn}$ which leave H invariant. Evidently ν_{e} is the (submutliplicative) operator norm induced by ν on the algebra of matrices which leave H invariant.

The ℓ_1 norm on $\mathbb{R}^n$ plays a special role in the theory of Markov chains and we shall denote it henceforth by ω. The corresponding coefficient of ergodicity was apparently first computed by Bauer et al. [1] – see also [16] – and equals

$$\omega_e(A) = (1/2) \max_{i,k \in \{1,\dots,n\}} \sum_{j=1}^{n} |a_{ij} - a_{kj}|.$$

However, this quantity appeared previously in the equivalent form

$$1 - \min_{i,k} \sum_{j=1}^{n} \min(a_{ij}, a_{kj})$$

in [4]; see [15], Theorem 2.10 for a proof of the equality of the two expressions. It is known that $\omega_e(A) \leqslant 1$ for all stochastic matrices A and ω_e is the only coefficient of ergodicity that satisfies this inequality [7,9], but see also [13].

Definition 3.1. Let $P_1, P_2, \dots$ be a sequence of $n \times n$ stochastic matrices. We shall say that all general products formed from this sequence are *weakly ergodic* if for all general products $B_{p,1}, B_{p,2}, \dots$, we have that

$$\lim_{r \to \infty} B_{p,r} x = 0 \quad \text{for all } x \in H. \tag{7}$$

Since every $x \in H$ can be written $x = c(u - v)$, where $u, v \in \mathbb{R}^n$ are nonnegative and $e^T u = e^T v = 1$ and $c \in \mathbb{R}$, it is easily seen that, for each product considered, our definition is equivalent to that in [4,11], or [15], Defn. 3.3.

By Theorem 2.2 we now immediately obtain the following.

Theorem 3.2. *Let v be a norm on $\mathbb{R}^n$ and let v_e be the corresponding coefficient of ergodicity. Let $P_1, P_2, \dots$ be a sequence of $n \times n$ stochastic matrices. Then all general products formed from this sequence are weakly ergodic if*

$$\sum_{i=1}^{\infty} (v_e^+(P_i) - 1) \text{ converges} \tag{8}$$

and

$$\sum_{i=1}^{\infty} (1 - v_e^-(P_i)) \text{ diverges.} \tag{9}$$

Note that Eq. (8) is automatically satisfied if $v = \omega$, the ℓ_1–norm. This special case of Theorem 3.2 is contained in [4], Theorem 3, see also [15], Exercise 4.36. By means of an inequality found in [15], (4.6), this result in turn implies [10], Theorem 3, see also [15], Theorem 4.9. Further results related to the special case $v = \omega$, of Theorem 3.2 (and which therefore involve only Eq. (9) explicitly) are to be found in [14], Theorem 1, and in [11]. The theorem in the latter paper is there illustrated by an example of a sequence of stochastic matrices that satisfies Eq. (9) for the norm ω see [11], p. 333.

The following corollary to Theorem 3.2 is due to Rhodius ([13], Thm. 3, Part I) in the case of $v = \omega$, see [8], Thm. A (i), for a related result.

Corollary 3.3. *Let $P_1, P_2, \ldots$ be a sequence of stochastic matrices. If $v_e(P_i) \leqslant 1$ for all $i, i = 1, 2, \ldots$, and there exists a point of accumulation c of the sequence $v_e(P_1), v_e(P_2), \ldots$ such that $c < 1$, then all general products of the sequence are weakly ergodic.*

Proof. Clearly condition (8) holds and there is an infinite subsequence of indices $j_1, j_2, \ldots$ such that $v_e(P_{i_j}) < (1 + c)/2 < 1$, $j = 1, 2, \ldots$ Then condition (9) holds for this subsequence. The result follows from Theorem 3.2. □

Another corollary of Theorem 3.2 is ([8], Thm. A(ii)):

Corollary 3.4. *Let $P_1, P_2, \ldots$ be a sequence of stochastic matrices and let v be a norm on $\mathbb{R}^n$. If all points of accumulation c of the sequence $v_e(P_1)$, $v_e(P_2), \ldots$ satisfy $c < 1$, then all general products of the sequence are weakly ergodic.*

Proof. The sequence $v_e(P_1), v_e(P_2), \ldots$, is bounded, since all elements of the stochastic matrices $P_j, j = 1, 2, \ldots$, are bounded above by 1. Since the set of accumulation points of a bounded sequence is compact, there exists $d < 1$ such that only a finite number of terms of the above sequence of ergodicity coefficients exceed d. Hence Eqs. (8) and (9) hold for this sequence and the corollary follows from Theorem 3.2. □

4. Comparisons of ergodic coefficients

Detailed comparison of our results above to some results in [15], Ch. 4 is difficult since in the latter the coefficient of ergodicity is defined in [15], Definition 4.6 as a continuous function τ on the set of stochastic matrices satisfying $0 \leqslant \tau \leqslant 1$, and theorems there require the hypothesis that this function is submultiplicative, see also [14] for a similar approach. On the other hand we define the coefficient in terms of a norm. It may be noted that in our proofs we have not used the subadditive property of a matrix norm, only its submultiplicative property.

The ℓ_1 coefficient of ergodicy ω_e is advantageous for obtaining theoretical results since $\omega_e(A) \leqslant 1$ for all stochastic matrices A. However, as we pointed out in the introduction, there are natural and useful coefficients of ergodicity v_e associated with a norm v such that $v_e(A) > 1$ for some stochastic matrix A. In order to provide a class of such examples, we state and prove a proposition.

Proposition 4.1. *Let $n > 1$ and let $1 \leqslant k \leqslant n$. For any $A \in \mathbb{R}^{nn}$ which leaves H invariant, let $\tilde{A}_k \in \mathbb{R}^{n-1,n-1}$ be defined by*

$$\tilde{a}_{ij} = a_{ij} - a_{ik}, \quad i, j = 1, \ldots, n, \quad i, j \neq k. \tag{10}$$

Let $\tilde{v}$ be any norm on $\mathbb{R}^{n-1}$. Then there exists a norm v on $\mathbb{R}^n$ such that the corresponding coefficient of ergodicity satifies $v_e(A) = \tilde{v}^0(\tilde{A}_k)$ for all stochastic matrices $A \in \mathbb{R}^{nn}$, where *$\tilde{v}^0$ is the operator norm on $\mathbb{R}^{n-1,n-1}$ induced by $\tilde{v}$.*

Proof. Without loss of generality, we may assume that $k = 1$ and we put $\tilde{A}_1 = \tilde{A}$. For $x \in \mathbb{R}^n$ define $\tilde{x} = (x_2, \dots, x_n)^{\mathrm{T}} \in \mathbb{R}^{n-1}$. Note that $\tilde{e} = (1, \dots, 1)^{\mathrm{T}} \in \mathbb{R}^{n-1}$ and define the norm v on $\mathbb{R}^n$ by

$$v(x) = |x_1 + \tilde{e}^{\mathrm{T}}\tilde{x}| + \tilde{v}(\tilde{x}), \quad x \in \mathbb{R}^n. \tag{11}$$

Let $x \in H$. Then $x_1 = -\tilde{e}^{\mathrm{T}}\tilde{x}$ and hence $v(x) = \tilde{v}(\tilde{x})$. If A leaves H invariant (and, in particular if A is stochastic) we therefore have $v(Ax)/v(x) = \tilde{v}(\widetilde{Ax})/\tilde{v}(\tilde{x})$ when $x \neq 0$. But it easily checked that $(\widetilde{Ax}) = \tilde{A}\tilde{x}$ for $x \in H$ since $x^{\mathrm{T}} = (-\tilde{e}^{\mathrm{T}}\tilde{x}, \tilde{x}^{\mathrm{T}})$. The result follows. □

The special case of Proposition 4.1, where $\tilde{v}$ is the ℓ_1 norm on $\mathbb{R}^{n-1}$ may be found in [12], Example 1 with a different proof. Observe that the coefficient of ergodicity so obtained is in general *not* equal to ω_e.

If we wish to find examples of ergodicity coefficients, it follows from the above proposition that we do not need to start with a norm on $\mathbb{R}^n$ and use the definition (6) to compute the coefficient. Instead, we may pick k, $1 \leqslant k \leqslant n$, and a norm $\tilde{v}_k$ on $\mathbb{R}^{n-1}$ and compute $\tilde{v}^0(\tilde{A}_k)$. Also, for many purposes, it is possible to consider $\beta = \min_{k=1,\dots,n} \tilde{v}^0(\tilde{A}_k)$, for example when finding upper bounds on the moduli of eigenvalues of A which do not equal 1, and our MATLAB experiments using the **rand** function show that in most cases $\beta < \omega_e(A)$ when $\tilde{v}$ is the ℓ_2 norm.

We end with the easily proved remark that the mapping $x \to \tilde{x}$ of $\mathbb{R}^n$ onto $\mathbb{R}^{n-1}$ in the proof of Proposition 4.1 is a vector space homomorphism whose restriction to H is an isomorphism. The mapping $A \to \tilde{A}_k$ onto $\mathbb{R}^{n-1,n-1}$ from the algebra of matrices in $\mathbb{R}^{nn}$ that map H into itself is a homomorphism whose kernel consists of all matrices in the algebra of rank r, $r \leqslant 1$.

Acknowledgements

We would like to thank Wenchao Huang for some helpful remarks. We thank Olga Holtz for her careful reading of the manuscript.

References

[1] F.L. Bauer, Eck. Deutsch, J. Stoer, Abschätzungen für die Eigenwerte positiver linearer Operatoren, Lin. Alg. Appl. 2 (1969) 275–301.

[2] W.J. Beyn, L. Elsner, Infinite products and paracontracting matrices, Elec. J. Lin. Alg. 2 (1997) 1–8.

[3] J. Hajnal, The ergodic properties of non-homogeneous finite Markov chains, Proc. Cambridge Phil. Soc. 52 (1956) 67–77.

[4] J. Hajnal, Weak ergodicity in non-homogeneous Markov chains, Proc. Cambridge Phil. Soc. 54 (1958) 52–67.

[5] D.J. Hartfiel, U.G. Rothblum, Convergence of inhomogeneous products of matrices and coefficients of ergodicity, Lin. Alg. Appl. 277 (1998) 1–9.

[6] J.M. Hyslop, Infinite Series, Oliver and Boyd, Edinburgh, 1945.

[7] R. Kühne, A. Rhodius, A characterization of Dobrushin's coefficient of ergodicity, Z. Anal. Anw. 9 (2) (1990) 187–188.

[8] A. Leizarowitz, On infinite products of stochastic matrices, Lin. Alg. Appl. 168 (1992) 189–219.

[9] A. Lešanovský, Coefficients of ergodicity generated by non-symmetrical vector norms, Czechoslovak Math. J. 40 (115) (1990) 284–294.

[10] J.L. Mott, Conditions for the ergodicity of non-homogeneous finite Markov chains, Proc. Roy. Soc. Edinburgh. Section A 64 (1957) 369–380.

[11] J.L. Mott, H. Schneider, Matrix norms applied to weakly ergodic Markov chains, Arch. Math. 8 (1957) 331–333.

[12] A. Rhodius, On explicit forms of ergodicity coefficients, Lin. Alg. Appl. 194 (1993) 71–83.

[13] A. Rhodius, On the maximum ergodicity coefficient, the Dobrushin coefficient and products of stochastic matrices, Lin. Alg. Appl. 253 (1997) 141–157.

[14] E. Seneta, On the historical development of the theory of finite inhomogeneous Markov chains, Proc. Cambridge Phil. Soc. 74 (1973) 507–513.

[15] E. Seneta, Non-negative Matrices and Markov Chains, 2nd ed., Springer, New-York, 1981.

[16] C. Zenger, A comparison for some bounds for the nontrivial eigenvalues of stochastic matrices, Numer. Math. 19 (1972) 209–211.