A Theorem on Products of Matrices

Marc Artzrouni
Department of Mathematics
Drexel University
Philadelphia, Pennsylvania 19104

Submitted by Richard A. Brualdi

ABSTRACT

A new result on products of matrices is proved in the following theorem: let M_i $(i=1,2,\dots)$ be a bounded sequence of square matrices, and K be the l.u.b. of the spectral radii $\rho(M_i)$. Then for any positive number ε there is a constant A and an ordering $p(j)$ $(j=1,2,\dots)$ of the matrices such that

$$\left\|\prod_{j=1}^{n} M_{p(j)}\right\| \leqslant A\cdot(K+\varepsilon)^n \qquad (n=1,2,\dots).$$

The ordering is well defined by $p(j)$, a one-to-one mapping on the set of positive integers. In general the inequality does not hold for any ordering $p(j)$ (a counterexample is provided); however, some sufficient conditions are given for the result to remain true irrespective of the order of the matrices.

$M_n(\mathbb{C})$ is the normed metric space of all n-square matrices over $\mathbb{C}$. The norm of a matrix $X=(x_{ij})$ will be $\|X\|$ defined by

$$\|X\| = \max_i \sum_{j=1}^{n} |x_{ij}|. \tag{1}$$

We now prove the following theorem:

THEOREM. *Let M_i $(i=1,2,\dots)$ be a bounded sequence of matrices belonging to $M_n(\mathbb{C})$, and K be the l.u.b. of the spectral radii $\rho(M_i)$. Then, for any number $\varepsilon>0$ there is a constant $A>0$ and an ordering $p(j)$ $(j=1,2,\dots)$*

52 Vanderbilt Ave., New York, NY 10017 0024-3795/83/010153-07$3.00

of the matrices M_i *such that*

$$\left\| \prod_{j=1}^{n} M_{p(j)} \right\| \leqslant A \cdot (K+\varepsilon)^n \qquad (n=1,2,\ldots). \tag{2}$$

The ordering of the matrices is defined by $p(n)$, *a one-to-one mapping of* $N-\{0\}$ (*the set of positive integers*) *onto itself.*

Proof. Let $S=\cup_i M_i$; $\bar{S}$ is the closure of S in $M_n(\mathbb{C})$, and clearly if $X \in \bar{S}$ then $\rho(X) \leqslant K$, since $\rho(M_i) \leqslant K$ for all i. $\bar{S}$ is closed and bounded; it is therefore a compact set of $M_n(\mathbb{C})$.

If $X \in \bar{S}$ and ε is a positive number, then by Ostrowski's theorem [1, p. 143] there exist $a_1(X,\varepsilon)>0$ and $a_2(X,\varepsilon)>0$, depending only on X and ε, such that if a sequence of matrices $Z_j (j=1,2,\ldots)$ belongs to the open ball of center X and radius $a_2(X,\varepsilon)$ [i.e. $Z_j \in B(X, a_2(X,\varepsilon))$], we will have, irrespective of the order of the matrices Z_j,

$$\left\| \prod_{j=1}^{n} Z_j \right\| \leqslant a_1(X,\varepsilon)\left[\rho(X)+\frac{\varepsilon}{2}\right]^n \qquad (n=1,2,\ldots), \tag{3}$$

ε being fixed, for each matrix $X \in \bar{S}$ there is a number $a_2(X,\varepsilon)$ [and a number $a_1(X,\varepsilon)$] such that Equation (3) holds. Form now $E=\cup_{X\in\bar{S}} B(X, a_2(X,\varepsilon))$; E is an open covering of $\bar{S}$. Now $\bar{S} \subset E$, and by definition of a compact set there is a finite subfamily $E'=\cup_{i=1}^{m} B(X_i, a_2(X_i,\varepsilon))$ of E which is a covering of $\bar{S}$: $\bar{S} \subset E'$. Let $B_i = B(X_i, a_2(X_i,\varepsilon))$. Define now the m sets S'_j $(j=1,2,\ldots,m)$ in the following way:

$$S'_1 = S \cap B_1,$$
$$S'_2 = S \cap B_2 - S'_1,$$
$$S'_3 = S \cap B_3 - (S'_1 \cup S'_2),$$
$$\vdots$$
$$S'_m = S \cap B_m - (S'_1 \cup S'_2 \cup \cdots \cup S'_{m-1}).$$

The sets S'_j are disjoint and we have:

$$S = \bigcup_{j=1}^{m} S'_j \quad \text{and} \quad S'_j \subset B_j \qquad (j=1,2,\ldots,m).$$

$S = \cup_i M_i$ is an infinite countable set of matrices, and therefore there is at least one set S'_j containing infinitely many matrices M_i. Let $T_1 = \cup_{k=1}^{m_1} S'_k$ be the union of the m_1 ($m_1 \geqslant 1$) sets containing (each one) infinitely many matrices of S. We will now disregard the set $T_2 = \cup_{k=m_1+1}^{m} S'_k$ containing only finitely many matrices, because they will not affect the nature of the product appearing in the inequality (2). The basic idea of the proof is to order the matrices M_i so that those belonging to a given set B_i will be lumped together which will then make Ostrowski's result [Equation (3)] applicable in each of the m_1 sets $S'_1, S'_2, \dots, S'_{m_1}$. We define m_1 increasing sequences of integers in the following way:

$$p_k(1) = \{\inf n | M_n \in S'_k\},$$

$$p_k(j) = \{\inf n | n > p_k(j-1), M_n \in S'_k\} \qquad (j = 2,3,\dots, \quad k = 1,2,\dots,m_1).$$

These m_1 sequences are well defined because the sets S'_j are disjoint and each contains infinitely many matrices M_i.

We choose a number $A > 1$ such that $A \geqslant \max_{j \leqslant m_1} a_1(X_j, \varepsilon)$ and a sufficiently large integer r such that

$$A^{1/r}\left(K + \frac{\varepsilon}{2}\right) \leqslant (K + \varepsilon). \tag{4}$$

The reason for this choice of r will appear later. We now construct the matrix W shown in Figure 1, having m_1 rows and infinitely many columns. The m_1 sequences p_j are arranged in the m_1 rows of W, and the ordering $p(j)$ of the theorem is determined by the sequence of p's following each other in the order defined by the arrows which go from block W_1, to block W_2, etc. In other words, we define the first elements $p(j)$ by

$$p(1) = p_1(1), \qquad p(2) = p_1(2), \dots, \qquad p(r) = p_1(r),$$

$$p(r+1) = p_2(1), \qquad p(r+2) = p_2(2), \dots .$$

To find a general expression for $p(n)$ we note there is only one way of writing any positive integer n as

$$n = (um_1 + v)r + t$$

where u, v, and t are integers and

$$0 \leqslant u, \qquad 0 \leqslant v < m_1, \quad \text{and} \quad 0 < t \leqslant r.$$

$p(n)$ is then, by definition,

$$p(n) \stackrel{\text{def}}{=} p_{v+1}(ru+t). \tag{5}$$

The sequence $p(n)$ has been constructed by formally lining up all the rows of block W_1, then all the rows of block W_2, etc. By doing so, as n increases, $M_{p(n)}$ sweeps all the matrices M_i, and it is easy to see that $p(n)$ is a well-defined one-to-one function on $N - \{0\}$.

For a given $n = (um_1 + v)r + t$ we will have, bearing in mind the restriction indicated just below

$$\prod_{j=1}^{n} M_{p(j)} = \prod_{j_1=0}^{u} \prod_{j_2=1}^{m_1} \prod_{j_3=1}^{r} M_{p_{j_2}(j_3+rj_1)}. \tag{6}$$

The above product is defined with the following restriction: when j_1 reaches u, then j_2 stops at $v+1$, and when $j_1 = u$ and $j_2 = v+1$, then j_3 stops at t. We now remember that all the matrices we are considering belong to one of the balls B_i $(i = 1,2,\ldots,m_1)$. Therefore, since A is an upper bound for the m_1 values $a_1(X_i, \varepsilon)$, and $\rho(M_i) \leqslant K$, the inequality (3) implies, for all values of j_1 and all values of $j_2 \leqslant m_1$,

$$\left\| \prod_{j_3=1}^{r} M_{p_{j_2}(j_3+rj_1)} \right\| \leqslant A \cdot \left(K + \frac{\varepsilon}{2} \right)^r. \tag{7}$$

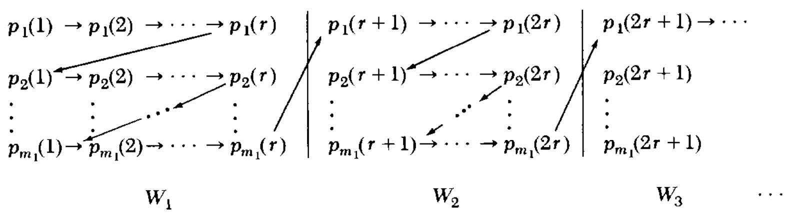

FIG. 1.

If we define $C = K + \varepsilon/2$, then the equation (6) and the inequality (7) imply

$$\left\| \prod_{j=1}^{n} M_{p(j)} \right\| \leqslant \prod_{j_1=0}^{u} \prod_{j_2=1}^{m_1} \left\| \prod_{j_3=1}^{r} M_{p_{j_2}(j_3+rj_1)} \right\|$$

$$\leqslant \underbrace{A^{m_1 u} C^{r m_1 u}}_{j_1 < u} \cdot \underbrace{A^{v} C^{rv}}_{\substack{j_1 = u \\ j_2 \leqslant v}} \cdot \underbrace{A C^{t}}_{\substack{j_1 = u \\ j_2 = v+1 \\ j_3 \leqslant t}}$$

$$= A \cdot A^{-t/r} \left[A^{1/r} C \right]^{(u m_1 + v) r + t}. \tag{8}$$

Because $t/r > 0$ and $A > 1$, we have $A^{-t/r} < 1$, and r was chosen so as to have $A^{1/r} C \leqslant (K + \varepsilon)$; we finally have

$$\left\| \prod_{j=1}^{n} M_{p(j)} \right\| \leqslant A \cdot (K + \varepsilon)^{n}. \qquad \blacksquare$$

In this theorem the sharpest bounds for the product are obtained when ε is very small, and in particular if $K < 1$ it is possible to find ε such that $K + \varepsilon < 1$, and the product then converges to 0. In this context we now discuss the question of the ordering of the matrices M_i.

In Ostrowski's theorem the product is majorized irrespective of the order of the matrices, and we may wonder if this remains true in the more general situation discussed here. A simple counterexample will show that this is not the case: Let

$$X = \begin{pmatrix} -0.6 & 1.9 \\ -0.3 & 0 \end{pmatrix} \quad \text{and} \quad Y = \begin{pmatrix} -0.5 & -0.5 \\ 0.5 & -0.3 \end{pmatrix}.$$

Both matrices have spectral radius strictly less than 1, but the spectral radius of the product is larger than 1: $\rho(X) = 0.75$, $\rho(Y) = 0.63$, and $\rho(XY) = 1.21$. Consider now the following sequence of matrices:

$$X, Y, X, Y, X, Y, X, Y, \dots . \tag{A}$$

For any given integer r define the product $P_r(n)$:

$$P_r(n) = \left[X^r Y^r \right]^n. \tag{9}$$

Given any number $d < 1$, for a sufficiently large r_0 we will have

$$\|X^{r_0}\| \leqslant d \qquad \text{and} \quad \|Y^{r_0}\| \leqslant d$$

because the spectral radii are < 1. Then

$$\|P_{r_0}(n)\| \leqslant d^{2n}$$

and the corresponding ordering of the matrices is:

$$\underbrace{XXXXXXX \cdots X}_{r_0} \cdot \underbrace{YYYYYYY \cdots Y}_{r_0} \cdot \underbrace{XXXXXXX \cdots X}_{r_0} \cdots . \tag{B}$$

For this ordering (B) of the matrices the product goes to 0, but with ordering (A) the product will go to infinity because $\rho(XY) > 1$: $\lim_n \|P_1(n)\| = \infty$. To summarize the situation, (B) is the ordering of the theorem, for which the product goes to 0, but there is also the ordering (A), for which, on the contrary, the product goes to infinity. Hence, the ordering of the matrices is essential to the theorem. We may, however, ask what extra conditions on the matrices should be added for the product to be bounded in the indicated way irrespective of the order of the matrices. We now briefly turn to this question and give a couple of simple conditions for this to be true.

(1) The most trivial condition for the product to be bounded irrespective of the order of the terms is of course for the matrices to be commutative. This is however, a very strong condition.

(2) Ostrowski's theorem is based on the fact that for any matrix $X \in M_n(\mathbb{C})$ and any $\varepsilon > 0$ there is an invertible matrix $P(X, \varepsilon)$, depending on X and ε, such that

$$\|P(X, \varepsilon) X P^{-1}(X, \varepsilon)\| \leqslant \rho(X) + \varepsilon. \tag{10}$$

We will apply this result to each matrix M_i of our theorem and majorize $\rho(M_i)$ by K. We can then define, for each i, the nonempty set $Q_i \subset M_n(\mathbb{C})$ in the following way:

$$Q_i = \left\{ P(M_i, \varepsilon) \middle| \|P(M_i, \varepsilon) M_i P^{-1}(M_i, \varepsilon)\| \leqslant K + \varepsilon \right\}. \tag{11}$$

Q_i is simply the set of matrices $P(M_i, \varepsilon)$ for which the indicated inequality holds. If we define $Q = \bigcap_i Q_i$, the hypothesis that will enable us to prove the result we were aiming at will simply be that Q is not empty: $Q \neq \varnothing$. Indeed, if all the sets Q_i have an invertible matrix P_0 in common, we have

$$\left\|\prod_{j=1}^{n} M_j\right\| = \left\|P_0^{-1}\left[\prod_{j=1}^{n} P_0 M_j P_0^{-1}\right]P_0\right\| \leqslant \|P_0^{-1}\|\|P_0\|(K+\varepsilon)^n. \tag{12}$$

So if $A = \|P_0^{-1}\|\,\|P_0\|$, the result follows:

$$\left\|\prod_{j=1}^{n} M_j\right\| \leqslant A(K+\varepsilon)^n. \tag{13}$$

This majorization is of course independent of the ordering of the matrices M_i. $Q \neq \varnothing$ is therefore a sufficient condition for the result to hold, irrespective of the order of the matrices. It is not known whether this condition is necessary. Finally there is nothing essential about the norm defined in Equation (1), since all norms are equivalent in a finite-dimensional space.

REFERENCES

1 A. M. Ostrowski, *Solutions of Equations in Euclidian and Banach Spaces*, Academic, 1973.

Received 19 February 1982; revised 11 June 1982