JOURNAL OF ALGEBRA 109, 14–24 (1987)

Markov Chains and Tensor Multiplications

ROBERT GRONE*

Department of Mathematics, Auburn University, Auburn, Alabama 36849

Communicated by Marshall Hall, Jr.

Received June 1985; revised October 7, 1985

An m-dimensional matrix of order n over a field F is an array $A=[a_{i_1}, a_{i_2,..., i_m}]$; $1 \leqslant i_j \leqslant n$; $1 \leqslant j \leqslant m$, of n^m elements of F. This definition coincides with the notion of an m-fold tensor over F^n, and also of an $n \times n$ matrix when $m=2$. In this paper two multiplications on such objects are examined. The first is nonassociative and is motivated by a generalization of a Markov chain. The second is associative, and related to the first by a certain generalized transpose operator. Spectral and unitary properties of the associative multiplication are discussed, as well as connections with block-diagonal matrix multiplication. A connection between graphs and the nonassociative multiplication is exhibited.

1. INTRODUCTION

Let F^n denote the space of n-tuples over a field F. By an *m-dimensional matrix of order n* over F, we mean an array

$$A=[a_\gamma]; \qquad a_\gamma \in F,$$

where

$$\gamma \in \Gamma(m, n)=\{\gamma=(\gamma(1),..., \gamma(m)) \mid 1 \leqslant \gamma(i) \leqslant n, \text{ all } i\}.$$

We will let $M_n^m(F)$ denote the set of such objects, which is a vector space of dimension n^m over F with respect to the obvious addition and scalar multiplication. When $m=2$ and 3, these arrays are square or cubic. For any m, the space of such objects may be regarded as:

(i) the mth tensor or Kronecker product of F^n, $\otimes^m F^n$;

(ii) the dual space of the space of m-linear functionals on $\overset{m}{X} F^n$; or

(iii) the mth homogeneous component of $F[x_1,..., x_n]$, where the x_i's are noncommuting indeterminates over F.

* Current address: Department of Mathematical Sciences, San Diego State University, San Diego, CA 92182.

0021-8693/87 $3.00

In the case $m=2$, these objects are $n \times n$ matrices, which are afforded several other interpretations. The interpretation of $n \times n$ matrices as linear transformations on F^n affords an associative multiplication on such objects (and vice versa).

The question arises whether there is some "natural" way to define a multiplication on m-dimensional matrices for $m > 2$. In [9], Yamada is able to motivate a multiplication from probabilistic considerations. In this paper we motivate a quite different multiplication on m-dimensional matrices by also resorting to probability theory. In doing this a generalization of the idea of a Markov chain is developed which is of independent interest. We also point out that though this multiplication is nonassociative when $m \geqslant 3$, it is canonically related to a certain associative multiplication, which is different from the associative multiplication of Yamada. When $m=2$, both the multiplications turn out to be usual matrix multiplication and so are both valid generalizations. We relate the associative multiplication to usual matrix multiplication and exploit this relation to prove various algebraic properties of the associative multiplication. The nonassociative multiplication is given a graph theoretic interpretation which serves as a further motivation for the nonassociative multiplication. We conclude with suggestions for further work.

2. A Generalization of Markov Chains

We begin by motivating and defining the stochastic m-dimensional real matrices. For a system that may be in any one of n distinct states at any time $t = 1, 2, \ldots$, we define, for each $\gamma \in \Gamma(m, n)$; $k = 1, 2, \ldots$, the

k-step transition probabilities:

$p_\gamma^k =$ the probability that state $\gamma(m)$ occurs at $t = m + k + j - 1$ given that state $\gamma(i)$ occurred at $t = i + j$, for all $i = 1, 2, \ldots, m-1$, and all $j = 0, 1, \ldots$.

For each k, we define an m-dimensional matrix of order n referred to as the *k-step transition matrix* via:

$$P^{(k)} = [p_\gamma^k].$$

It is clear that these matrices have nonnegative entries and satisfy

$$\sum_{s=1}^{n} p^k(\gamma(1), \ldots, \gamma(m-1), s) = 1 \qquad \text{all} \quad \gamma \in \Gamma(m, n). \tag{1}$$

We will refer to any m-dimensional matrix with nonnegative entries which satisfies (1) as *stochastic.*

If we examine the coefficients of $P^{(1)}$ and $P^{(2)}$, we find from probabilistic considerations that

$$(P^2)_\gamma = \sum_{s=1}^{n} p^1(\gamma(1),..., \gamma(m-1), s)^{p^1}(\gamma(2),..., \gamma(m-1), s, \gamma(m))$$

all γ. This generalization of Markov chains motivates the multiplication defined via:

$$(A \cdot B)_\gamma = \sum_{s=1}^{n} a_{(\gamma(1),...,\gamma(m-1),s)} b_{(\gamma(2),...,\gamma(m-1),s,\gamma(m))}. \tag{2}$$

Straightforward computations yield the following:

THEOREM 1. *The multiplication on m-dimensional matrices defined via* (2):

(i) *satisfies* $A \cdot (B + C) = A \cdot B + A \cdot C$;

(ii) *satisfies* $(A + B) \cdot C = A \cdot C + B \cdot C$;

(iii) *has a unique right multiplicative identity* $E = [\delta_{\gamma(m-1),\gamma(m)}]$;

(iv) *is non-associative when* $m > 2$, $n > 1$, *even on powers* (*i.e.*, $(A \cdot A) \cdot A \neq A \cdot (A \cdot A)$ *in general*);

(v) *satisfies* $P^{(k)} = P^{(1)} \cdot P^{(k-1)}$, *all* $k = 2, 3,...$ *for the previously defined transition matrices.*

When $m = 2$ this notion of a k-step transition matrix coincides with the usual definition of such objects in the theory of Markov chains and the multiplication which it motivates is the usual matrix multiplication. When $m \geqslant 3$ Theorem 1 ensures that this multiplication is not even power-associative.

3. AN EXAMPLE

Suppose that a set of teams involved in a particular sport forms a league which arranges games at regular intervals between its members. Also suppose that for a given team there are two possible outcomes for a game: a win (outcome 1), or a loss (outcome 2).

If we apply the notion of a (2-dimensional) Markov chain to this situation we arrive at a transition matrix

$$P=\begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix},$$

where

$p_{11}=$ the probability of a win given the previous game was a win,

$p_{12}=$ the probability of a loss given the previous game was a win,

$p_{21}=$ the probability of a win given the previous game was a loss,

$p_{22}=$ the probability of a loss given the previous game was a loss.

Once the matrix P is obtained, one can calculate the k-step transition matrix in the usual way. This disadvantage of this approach is that the entries of the matrix P may all be very close to 2^{-1}, in which case not much information is obtained. In many sports it does seem in fact that not much can be said about a team's chances of winning by looking solely at the outcome of the previous game.

If we apply the generalized notion of a Markov chain we obtain an m-dimensional transition matrix P with entries:

$p_{(1,\dots,1)}=$ the probability of a win given that the previous $m-1$ games were won,

$\vdots$

$p_{(2,\dots,2)}=$ the probability of a loss given that the previous $m-1$ games were lost.

With this example in mind, it seems likely that the entries of the m-dimensional transition matrix would vary much more than the entries of the 2-dimensional transition matrix. In other words; a team's performance can be more accurately predicted from its last $m-1$ games than its last game.

4. A Related Associative Multiplication

For any $s=1,\dots,n$, $\gamma\in\Gamma$, let

$$(\gamma:s)=(\gamma(1),\dots,\gamma(m-1),s),$$

$$(s:\gamma)=(s,\gamma(2),\dots,\gamma(m)).$$

In an attempt to modify the multiplication defined in (2), consider the multiplication defined via:

$$(AB)_\gamma = \sum_{s=1}^{n} a_{(\gamma:s)} b_{(s:\gamma)}. \tag{3}$$

The properties of this multiplication are given in the following:

THEOREM 2. *The multiplication on m-dimensional matrices defined by* (3):

(i) *satisfies* $A(B+C)=AB+AC$;
(ii) *satisfies* $(A+B)C=AC+BC$;
(iii) *has a two-sided identity* $I=[\delta_{\gamma(1),\gamma(m)}]$;
(iv) *coincides with the usual definition when* $m=2$;
(v) *satisfies* $(AB)C=A(BC)$.

Proof. Properties (i)–(iv) are routine, and to prove (v), we calculate

$$\begin{aligned}
[(AB)C]_\gamma &= \sum_{s=1}^{n} (AB)_{(\gamma:s)} C_{(s:\gamma)} \\
&= \sum_{s,t=1}^{n} A_{((\gamma:s):t)} B_{(t:(\gamma:s))} C_{(s:\gamma)} \\
&= \sum_{s,t=1}^{n} A_{(\gamma:t)} B_{((t:\gamma):s)} C_{(s:\gamma)} \\
&= \sum_{s,t=1}^{n} A_{(\gamma:t)} B_{((t:\gamma):s)} C_{(s:(t:\gamma))} \\
&= \sum_{t=1}^{n} A_{(\gamma:t)} (BC)_{(t:\gamma)} \\
&= [A(BC)]_\gamma.
\end{aligned}$$

If $\sigma \in S_m$, then we define the *permutation operator*, or *generalized transpose operator* afforded by σ on the m-dimensional matrices in the following way. If $\sigma = (\sigma(1),..., \sigma(n))$ and $\gamma \in \Gamma(m, n)$, then

$$\gamma\sigma = (\gamma(\sigma(1)),..., \gamma(\sigma(m)))$$

and

$$(A^{T(\sigma)})_\gamma = (A)_{\gamma\sigma}.$$

THEOREM 3. *The multiplications in* (2) *and* (3) *are related in the following way. If* $\sigma_0 \in S_m$ *is given by*

$$\sigma_0 = (m-1, 1, \ldots, m-2, m),$$

then

$$A \cdot B = AB^{T(\sigma_0)}.$$

5. SPECTRAL AND UNITARY PROPERTIES OF THE ASSOCIATIVE MULTIPLICATION

For any fixed $A \in M_n^m(F)$, the set

$$\{p(x) \in F[x] \mid p(A) = 0\}$$

is an ideal in $F[x]$. Hence A has a unique (monic) *minimal polynomial*, $m_A(x)$, whose roots will be referred to as the *characteristic values* of A. This definition partially refutes an assertion of Barter [1] that no analogs of characteristic values exist for homogeneous vector functions of degree greater than one.

EXAMPLE. If a matrix A has 0's in the positions where the identify matrix has 0's, then the characteristic values of A are just the set of distinct elements appearing in those positions where the identity matrix has 1's. Since there are n^{m-1} such positions, we can see that

$$\deg(m_A(x)) \leqslant n^{m-1}$$

with the upper bound clearly attainable.

Since the dimension of $M_n^m(F)$ is n^m, it is clear that the degree of the minimal polynomial, and hence the number of characteristic values, is at most $n^m - 1$. In fact, the maximum is actually n^{m-1}, as in the previous example. This will be made clear by the discussion in Section 6.

In addition to characteristic values, we may define notions of similarity, singularity, nilpotence, etc., in the obvious fashion, and such theorems as the following examples are easily proven.

THEOREM 4. *A is a left zero-divisor*

iff *A is a right zero-divisor*
iff *there exists* $0 \neq B = p(A)$ *such that* $0 = AB$
iff *0 is a characteristic value of A.*

THEOREM 5. *If A and B are similar, then $m_A(x) = m_B(x)$.*

We may also define the *rank* of A to be

$$\rho(A) = \frac{1}{n} \dim\{AB \mid B \in M_n^m(F)\},$$

and such facts as:

(i) If A is nonsingular $\rho(AB) = \rho(B)$, all B,

(ii) $\rho(AB) \leqslant \rho(B)$,

(iii) $\rho(A^{T(\sigma)}) = \rho(A)$, all $\sigma \in S_m$,

(iv) $\rho(A+B) \leqslant \rho(A) + \rho(B)$,

are readily verifiable. Note that this definition is not the same as the definition of *tensor rank*. (See, e.g., [3, 6, 7] for a discussion of tensor rank.)

If $F = C$, the field of complex numbers, we make the following definitions.

(i) The adjoint of A, A^*, is defined by

$$(A^*)_\gamma = [\bar{a}_{(\gamma(m), \gamma(2), \ldots, \gamma(m-1), \gamma(1))}],$$

(ii) A is *unitary* if $A^* = A^{-1}$,

(iii) A is *Hermitian* if $A^* = A$,

(iv) A is *normal* if $A^*A = AA^*$,

(v) A is *unitarily diagonalizable* if there exists a unitary U such that U^*AU has 0's in all positions in which the identity matrix does.

6. RELATIONS TO BLOCK-MATRIX MULTIPLICATION

In this section we make two observations which greatly simplify problems arising concerning the associative multiplication on $M_n^m(F)$.

The first observation is that the multiplication defined in (3) is "essentially 3-dimensional." More precisely, it is possible to define a multiplication on $n \times n_1 \times n$ arrays by means of (3), and if we identify an m-dimensional matrix of order n as an $n \times n^{m-2} \times n$ 3-dimensional matrix by utilizing the usual lexicographic ordering on $(\gamma(2), \gamma(3), \ldots, \gamma(m-1))$, then the multiplication afforded by the 3-dimensional multiplication coincides with the multiplication in (3).

The second observation relates the 3-dimensional multiplication on $n \times$

$n_1 \times n$ arrays to multiplication of (usual) block matrices. Given two $n \times n_1 \times n$ arrays A, B we define $n \times n$ matrices A_s, B_s, via:

$$A_s = [a_{isj}], \qquad s = 1, \dots, n_1.$$
$$B_s = [b_{isj}], \qquad s = 1, \dots, n_1.$$

If we now identify A and B with the matrices in (4) then the 3-dimensional multiplication coincides with the usual multiplication of the block-diagonal matrices in (4), i.e., $\widetilde{AB} = \tilde{A}\tilde{B}$, where

$$\tilde{A} = \begin{bmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_{n_1} \end{bmatrix}, \qquad \tilde{B} = \begin{bmatrix} B_1 & & & \\ & B_2 & & \\ & & \ddots & \\ & & & B_{n_1} \end{bmatrix}. \tag{4}$$

We summarize the previous remarks as follows. For each $\omega \in \Gamma(m-2, n)$ and $A \in M_n^m(F)$ we define the $n \times n$ matrix

$$A_\omega = [a_{(i,\omega,j)}]. \tag{5}$$

We then define the $n^{m-1} \times n^{m-1}$ *expanded matrix* $\tilde{A}$ as the block-diagonal matrix

$$\tilde{A} = \sum_{\omega \in \Gamma(m-2,n)}^{\cdot} A_\omega.$$

THEOREM 6. *For any A, $B \in M_n^m(F)$,*

$$\widetilde{AB} = \tilde{A}\tilde{B}.$$

The following four theorems are examples of corollaries of Theorem 6.

THEOREM 7. *$\rho(A)$ is an integer.*

THEOREM 8. *A is normal iff A is unitarily diagonalizable.*

THEOREM 9. *A normal matrix is Hermitian iff its characteristic roots are real.*

THEOREM 10. *A normal matrix is unitary iff its characteristic roots are of modulus one.*

The expanded matrix A is reminiscent of the $n \times n$ "contracted matrix" of Yamada in [9]. Using the previous notation, the *contracted matrix* in [9] is

$$\operatorname{ctr}(A) = \sum_{\omega \in \Gamma(m-2,n)} A_\omega .$$

We can also use the expanded matrix to define a *determinant* function on $M_m^n(F)$.

$$\det(A) = \det(A) = \prod_{\omega \in \Gamma(m-2,n)} \det(A_\omega). \tag{6}$$

Note that this definition and Theorem 6 yield that

$$\det(AB) = \det(A)\det(B) \qquad \text{all} \quad A, B \in M_n^m(F).$$

If we use this determinant function to define the *characteristic polynomial of* A,

$$\operatorname{ch}_A(x) = \det(xI - A),$$

we can then prove that

$$\min_A(x) \mid \operatorname{ch}_A(x) \qquad \text{all} \quad A \in M_n^m(F).$$

We remark that the determinant function defined by (6) is homogeneous of degree n^{m-1} in the entries of A and is quite different than the following determinant-type function which is homogeneous of degree n (see Sokolov [8]),

$$d(A) = \sum_{\sigma_2,\ldots,\sigma_m} \varepsilon(\sigma_2 \cdots \sigma_m) \prod_{i=1}^{n} a_{(i,\sigma_2(i),\ldots,\sigma_m(i))}. \tag{7}$$

When $m = 2$ the two definitions of course coincide, and so both are generalizations of the usual determinant. Several versions of determinants for m-dimensional matrices have already been studied. See, for example, [2, 5, 8].

7. Graph Theory

In this section we relate the nonassociative multiplication, contractions, and graph theory.

Let K_n denote the complete directed graph on n vertices, including loops. By an *m-edge* we mean a directed path of length $m-1$ in K_n, so that the set of m-edges is identified with $\Gamma(m, n)$. By an *m-graph* we mean a collection of m-edges or a subset $\mathscr{G} \subset \Gamma(m, n)$. The *adjacency matrix* of is $A = A(\mathscr{G}) \in M_n^m(R)$ and defined by

$$a_\gamma = \begin{cases} 1 & \text{if} \quad \gamma \in \mathscr{G} \\ 0 & \text{if} \quad \gamma \notin \mathscr{G}, \end{cases}$$

so that the set of m-graphs is in natural correspondence with the set of m-dimensional 0,1-matrices of order n. By an *m-route in* $\mathscr{G}$ we mean a directed path of length at least $m-1$ in K_n having the property that every consecutive subpath of length $m-1$ is an m-edge of $\mathscr{G}$. We say that $\mathscr{G}$ is *m-connected* if any two vertices can be joined by an m-route in $\mathscr{G}$.

Suppose now that $\mathscr{G}$ is an m-graph with adjacency matrix A. We define

$$A^{(1)} = A,$$
$$A^{(k+1)} = A \cdot A^{(k)} \qquad \text{all} \quad k = 1, 2, \ldots.$$

A moment's reflection yields that vertex i can be joined to vertex j by an m-route in of length $m+k-2$ iff the (i, j)th entry of $\operatorname{ctr}(A^{(k)})$ is positive. Hence we have the following theorem, which serves as yet another motivation for the non-associative multiplication.

THEOREM 11. *Is m-connected iff*

$$\operatorname{ctr}(A^{(1)} + \cdots + A^{(k)}) > 0$$

for some k.

8. SUGGESTIONS FOR FURTHER WORK

We say that a transition matrix P has *limiting transition probabilities* if

$$\lim_{k \to \infty} P^{(k)} \qquad \text{exists,}$$

and we say that P is *regular* if

$$P^{(k)} > 0 \qquad \text{for some} \quad k.$$

Investigation into the structure of such matrices would seem to be of interest. For example, do regular transition matrices have limiting transition probabilities?

Assumption on the entries of a transition matrix might be useful. As an elementary example, if p_γ is a function only of $\gamma(m-1)$ and $\gamma(m)$, then P is actually a 2-dimensional Markov process, which is extensively treated in the literature. A less trivial assumption would be that P satisfies

$$P = P^{T(\sigma_0)}, \tag{8}$$

where σ_0 is as in Theorem 3. This would include the case where P_γ was a function only of the multiplicities of $1,..., n$ in $(\gamma(1),..., \gamma(m-1))$. If P does satisfy (8), then Theorem 3 and 6 reduce questions about $P^{(k)}$ to usual matrix multiplication. This makes it trivial to prove that a regular transition matrix satisfying (8) has limiting transition probabilities.

Further work could be done in areas unrelated to Markov chains. Manifestations and properties of the associative multiplication and the determinant functions in Section 6 would be of interest. Inequalities between $\rho(A)$ and the tensor rank of A would be of interest in view of the relation between tensor rank and complexity theory (see, e.g., [3, 6, 7] for a discussion of tensor rank). Related to tensor rank and other questions it would be valuable to consider the structure of normal forms for $A \in M_n^m(F)$ under elementary operations (the case $m = 3$, $n = 2$ has been considered by Kaplansky [4]).

References

1. J. D. Barter, "Univ. of Calif. Publ. Math.," Vol. 1, pp. 321–343, 1920.
2. R. Gouarné et Isaac Samuel, Determinants des matrices à N dimension d'ordre n, *Cah. de Phys.* **16** (1962), pp. 143–152.
3. F. L. Hitchcock, *J. Math. Phys.* **7** (1927), 33–85.
4. I. Kaplansky, colloquium lecture, University of California, 1976.
5. J. B. Kim and J. E. Dowdy, Determinants of n-dimensional matrices, *J. Korean Math. Soc.* **17**(1) (1980), 141–146.
6. M. Marcus, Lengths of tensors, "Proc. Sympos. Wright–Patterson AFB," Academic Press, New York, 1967.
7. M. Marcus, "Finite Dimensional Multilinear Algebra," Dekker, New York, 1973.
8. N. P. Sokolov, "Spatial Matrices and Their Applications," Gosudrstv. Izdat. Fiz.-Mat. Lit., Moscow, 1960.
9. K. Yamada, "Hypermatrix and its Applications," Hitotsubashi J. Arts and Sciences, pp. 34–44, 1965.