THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Matrix multiplication is determined by orthogonality and trace

OPEN ACCESS

MATRIX MULTIPLICATION IS DETERMINED BY ORTHOGONALITY AND TRACE

CHRIS HEUNEN AND CLARE HORSMAN

ABSTRACT. Any associative bilinear multiplication on the set of n-by-n matrices over some field of characteristic not two, that makes the same vectors orthogonal and has the same trace as ordinary matrix multiplication, must be ordinary matrix multiplication or its opposite.

1. INTRODUCTION

Matrix multiplication is the fundamental operation of linear algebra, modelling composition of linear maps in terms of coordinates. This article characterises matrix multiplication simply in terms of orthogonality and trace, in the following way. Suppose there were another way to combine two matrices $x, y \in \mathbb{M}_n$ into a new one $x \star y \in \mathbb{M}_n$, that resembles composition of linear maps in the sense that it is associative, bilinear, and respects the identity matrix:

(A) "*Associativity*": $x \star (y \star z) = (x \star y) \star z$ for all $x, y, z \in \mathbb{M}_n$;
(B) "*Bilinearity*": $(\lambda x) \star y = \lambda(x \star y) = x \star (\lambda y)$, $(x + y) \star z = (x \star z) + (y \star z)$, and $x \star (y + z) = (x \star y) + (x \star z)$ for scalars λ and $x, y, z \in \mathbb{M}_n$;
(I) "*Identity*": $x \star 1 = x$ for all $x \in \mathbb{M}_n$, where $1 \in \mathbb{M}_n$ is the identity matrix.

To these basic requirements we add two properties that, at first sight, do not fix much information about $x \star y$. First, we require $x \star y$ to have the same trace as ordinary matrix multiplication xy:

(T) "*Trace*": $\mathrm{Tr}(x \star y) = \mathrm{Tr}(xy)$ for all $x, y \in \mathbb{M}_n$.

Second, we require that $\star$ makes the same vectors orthogonal as ordinary matrix multiplication. Formulated algebraically:

(O) "*Orthogonality*": $x \star y = 0$ when $xy = yx = 0$, $xx = x$, and $yy = y$, for $x, y \in \mathbb{M}_n$ that have rank one.

These assumptions already imply that $x \star y$ must equal the ordinary matrix multiplication, or its opposite, after all. That is, we will prove the following theorem, for any scalar field $\mathbb{F}$ of characteristic not two.

Theorem 1. *A function* $\star \colon \mathbb{M}_n \times \mathbb{M}_n \to \mathbb{M}_n$ *satisfies (A), (B), (I), (T) and (O) if and only if either* $x \star y = xy$ *for all* $x, y \in \mathbb{M}_n$*, or* $x \star y = yx$ *for all* $x, y \in \mathbb{M}_n$*.*

This fits in the recent programme of results surrounding linear preserver problems, *i.e.* linear maps that preserve zero products, commutativity, etc. [4, 6, 1, 7]. Indeed, we will rely on one of those results [1]. Our original motivation came from quantum theory, where the above properties arise as desiderata for a possible noncommutative extension of Bayesian inference [5, 3].

2. The main result

Proposition 2. *A map $\star\colon \mathbb{M}_n \times \mathbb{M}_n \to \mathbb{M}_n$ meets (B), (I), and (O) if and only if*

$$(1) \qquad\qquad x \star y = xy + g(xy - yx)$$

for some linear map $g\colon \mathbb{M}_n \to \mathbb{M}_n$ and all $x, y \in \mathbb{M}_n$.

Proof. Assume (B), (I) and (O). Then [1, Theorem 2.2] applies, giving linear maps $f, g\colon \mathbb{M}_n \to \mathbb{M}_n$ with $x \star y = f(xy) + g(yx)$. Furthermore, $x = x \star 1 = f(x1) + g(1x)$, and so $f(x) = x - g(x)$. Therefore $x \star y = f(xy) + g(yx) = xy + g(yx - xy)$. Taking the negative of g now gives (1).

If $\star$ is of the form (1), then it is easy to show that (B), (I), and (O) hold. $\qquad\square$

Lemma 3. *Suppose $\star$ satisfies (B), (I), and (O). Then (T) holds if and only if g sends traceless matrices to traceless matrices.*

Proof. Property (T) is equivalent to $\mathrm{Tr}(g(xy - yx)) = 0$ for all $x, y \in \mathbb{M}_n$. But matrices of the form $xy - yx$ are precisely those with trace zero [2]. $\qquad\square$

Write e_{ij} for the standard matrix units, so that $e_{ij}e_{kl} = \delta_{jk}e_{il}$. Because of bilinearity, property (A) holds precisely when it holds for $x = e_{ab}$, $y = e_{cd}$, $z = e_{ef}$ for all $a, b, c, d, e, f \in \{1, \ldots, n\}$. Via Proposition 2, property (A) comes down to

$$(2) \qquad \begin{aligned} &- \delta_{de}\delta_{af}g(e_{cb}) + \delta_{de}e_{ab} \star g(e_{cf}) - \delta_{cf}e_{ab} \star g(e_{ed}) \\ &= - \delta_{af}\delta_{bc}g(e_{ed}) + \delta_{bc}g(e_{ad}) \star e_{ef} - \delta_{ad}g(e_{cb}) \star e_{ef} \end{aligned}$$

for all $a, b, c, d, e, f \in \{1, \ldots, n\}$. We will use this formula very often below.

By linearity of g, we may write $g(e_{ij}) = \sum_{k,l=1}^n G_{kl,ij}e_{lk}$ for entries $G_{kl,ij} \in \mathbb{F}$. By convention, we will write $g(ij)_{kl}$ for $G_{kl,ij}$, and $g(ii - jj)_{kl}$ for $G_{kl,ii} - G_{kl,jj}$.

Lemma 4. *For distinct $i, j, k, l \in \{1, \ldots, n\}$:*

$$(3) \qquad\qquad g(ij)_{kl} = 0,$$
$$(4) \qquad\qquad g(ii - jj)_{kl} = 0.$$

For distinct $i, j, k \in \{1, \ldots, n\}$:

$$(5) \qquad\qquad g(ij)_{jk} = 0,$$
$$(6) \qquad\qquad g(ij)_{kj} = 0,$$
$$(7) \qquad\qquad g(ij)_{ik} = 0,$$
$$(8) \qquad\qquad g(ij)_{ki} = 0,$$
$$(9) \qquad\qquad g(ij)_{kk} = 0,$$
$$(10) \qquad\qquad g(ii - jj)_{kk} = 0,$$
$$(11) \qquad\qquad g(ii - jj)_{ik} = 0,$$
$$(12) \qquad\qquad g(ii - jj)_{ki} = 0,$$
$$(13) \qquad\qquad g(ii - jj)_{kj} = 0,$$
$$(14) \qquad\qquad g(ii - jj)_{jk} = 0.$$

For distinct $i, j \in \{1, \ldots, n\}$:

$$(15) \qquad\qquad\qquad g(ij)_{jj} = 0,$$

$$(16) \qquad\qquad\qquad g(ij)_{ii} = 0,$$

$$(17) \qquad\qquad\qquad g(ij)_{ij} = 0,$$

$$(18) \qquad\qquad\qquad g(ii - jj)_{ij} = 0,$$

$$(19) \qquad\qquad\qquad g(ii - jj)_{ji} = 0.$$

Proof. All these equations are derived following the same pattern. Let i, j, k, l be distinct. Taking $a = k$, $b = l$, $c = i$, and $d = e = f = j$ in (2) results in $e_{kl} \star g(e_{ij}) = 0$. Taking the trace of both sides and using property (T) now shows $g(ij)_{kl} = \mathrm{Tr}(g(e_{ij})e_{kl}) = \mathrm{Tr}(e_{kl}g(e_{ij})) = \mathrm{Tr}(e_{kl} \star g(e_{ij})) = 0$, establishing (3).

Similarly, taking $a = k$, $b = l$, $c = f = i$, and $d = e = j$ in (2) results in $e_{kl} \star g(e_{ii} - e_{jj}) = 0$, from which (4) follows by taking the trace of both sides.

For (5)–(19) we simply list the appropriate choice of indices:

- (5): $c = d = e = i$, $a = f = j$, $b = k$;
- (6): $c = d = e = i$, $b = f = j$, $a = k$;
- (7): $a = c = i$, $d = e = f = j$, $b = k$;
- (8): $a = f = i$, $d = j$, $b = c = e = k$;
- (9): $c = d = e = i$, $f = j$, $a = b = k$;
- (10): $c = f = i$, $d = e = j$, $a = b = k$;
- (11): $a = c = f = i$, $d = e = j$, $b = k$;
- (12): $b = c = f = i$, $d = e = j$, $a = k$; use (5);
- (13): $c = f = i$, $b = d = e = j$, $a = k$;
- (14): $c = f = i$, $a = d = e = j$, $b = k$; use (8);
- (15): $c = d = e = i$, $a = b = f = j$;
- (16): $a = b = c = i$, $d = e = f = j$; use (15);
- (17): $a = c = i$, $b = d = e = f = j$;
- (18): $a = c = f = i$, $b = d = e = j$;
- (19): $b = c = f = i$, $a = d = e = j$. $\square$

Lemma 5. *There is $\lambda \in \{0, -1\}$ with $g(ii - jj)_{ii} = -g(ii - jj)_{jj} = g(ij)_{ji} = \lambda$ for distinct $i, j \in \{1, \ldots, n\}$. Moreover, there is $z \in \mathbb{M}_n$ with $g(ii) = \lambda e_{ii} + z$ for all i.*

Proof. For any $k, l \in \{1, \ldots, n\}$, the entry $g(ij)_{kl}$ can only be nonzero when $k = j$ and $l = i$ by (3), (5)–(9) and (15)–(17). Let $\lambda_{ij} \in \mathbb{F}$ be that entry: $g(ij) = \lambda_{ij} e_{ji}$.

Taking $b = c = e = f = i$ and $a = d = j$ in (2) leads to $g(ii - jj)_{ii} = g(ij)_{ji}$. Similarly, $a = d = i$ and $b = c = e = f = j$ lead to $g(jj - ii)_{jj} = g(ji)_{ij}$. Next, $b = c = d = e = i$ and $a = f = j$ show that $g(ij)_{ji} = g(ji)_{ij}$. Therefore

$$g(ii - jj)_{ii} = g(ij)_{ji} = \lambda_{ij}, \qquad g(ii - jj)_{jj} = -g(ij)_{ji} = -\lambda_{ij}.$$

Combining this with (4), (10)–(14), (18) and (19) shows $g(ii - jj) = \lambda_{ij}(e_{ii} - e_{jj})$.

We may write $g(ii) = \lambda_i e_{ii} + z_i$ for $\lambda_i \in \mathbb{F}$, and $z_i \in \mathbb{M}_n$ linearly independent from e_{ii}. Then $\lambda_{ij}(e_{ii} - e_{jj}) = g(ii - jj) = \lambda_i e_{ii} - \lambda_j e_{jj} + (z_i - z_j)$. It follows that z_i does not depend on i, and we simply write z instead. Similarly, $\lambda_{ij} = \lambda_i = \lambda_j$ does not depend on i or j, and we may simply write λ for λ_{ij}. Hence $g(ii) = \lambda e_{ii} + z$.

It follows from the choice of indices for (18) that $g(ij) = e_{ij} \star g(ii - jj)$, and so

$$g(ij) = \lambda e_{ij} \star (e_{ii} - e_{jj}) = \lambda(-g(ij) - e_{ij} - g(ij)) = -2\lambda g(ij) - \lambda e_{ij}.$$

Hence $(1 + 2\lambda)g(ij) = -\lambda e_{ij}$. But $g(ij) = \lambda e_{ij}$ by definition, so $\lambda \in \{0, -1\}$. $\square$

Proof of Theorem 1. Lemma 5 gives $\lambda \in \{0, -1\}$ and $z \in \mathbb{M}_n$ with $g(ii) = \lambda e_{ii} + z$. Let $x \in \mathbb{M}_n$, say $x = \sum_{i,j=1}^{n} \chi_{ij} e_{ij}$ for $\chi_{ij} \in \mathbb{F}$. It follows from Lemma 4 that

$$g(x) = \sum_{i,j=1}^{n} \chi_{ij} g(ij) = \sum_{i=1}^{n} \chi_{ii} g(ii) + \sum_{i \neq j} \chi_{ij} g(ij)$$

$$= \sum_{i=1}^{n} (\lambda \chi_{ii} e_{ii} + \chi_{ii} z) + \sum_{i \neq j} \lambda \chi_{ij} e_{ij}$$

$$= \lambda \sum_{i,j=1}^{n} \chi_{ij} e_{ij} + \sum_{i=1}^{n} \chi_{ii} z$$

$$= \lambda x + \mathrm{Tr}(x) z.$$

Therefore $x \star y = xy + \lambda(xy - yx)$. $\qquad\qquad\square$

Remark 6. The restriction of the base field $\mathbb{F}$ to characteristic not two is necessary. For example, if $\mathbb{F} = \mathbb{Z}_2$, then

$$(20) \qquad \begin{pmatrix} \alpha & \beta \\ \gamma & \varepsilon \end{pmatrix} \star \begin{pmatrix} \zeta & \eta \\ \theta & \iota \end{pmatrix} = \begin{pmatrix} \alpha\zeta + \gamma\eta & \beta\zeta + \varepsilon\eta \\ \alpha\theta + \gamma\iota & \beta\theta + \varepsilon\iota \end{pmatrix}$$

satisfies (A), (B), (I), (O), and (T), but differs from matrix multiplication or its opposite. In fact, apart from $g = 0, 1$, this is the only one of the form (1), for *e.g.*

$$g \begin{pmatrix} \alpha & \beta \\ \gamma & \varepsilon \end{pmatrix} = \begin{pmatrix} \alpha & \beta \\ \gamma & \alpha \end{pmatrix}.$$

Acknowledgements

References

[1] J. Alaminos, M. Brešar, J. Extremera, and A. R. Villena. On bilinear maps determined by rank one idempotents. *Lin. Alg. Appl.*, 432:738–743, 2010.

[2] A. A. Albert and B. Muckenhoupt. On matrices of trace zero. *Mich. Math. J.*, 4(1):1–3, 1957.

[3] J. Barrett, C. Heunen, C. Horsman, M. Pusey, and R. W. Spekkens. Causal inference in quantum mechanics. *forthcoming*, 2013.

[4] M. Brešar and P. Šemrl. On bilinear maps on matrices with applications to commutativity preservers. *J. Alg.*, 301:803–837, 2006.

[5] M. S. Leifer and R. W. Spekkens. A Bayesian approach to compatibility, improvement, and pooling of quantum states. *arXiv:1110.1085*, 2011.

[6] L. Molnár. *Selected Preserver Problems on Algebraic Structures of Linear Operators and on Function Spaces.* Springer, 2007.

[7] D. Wang, X. Li, and H. Ge. Idempotent elements determined matrix algebras. *Lin. Alg. Appl.*, 435:2889–2895, 2011.

Department of Computer Science, University of Oxford
E-mail address: {chris.heunen,clare.horsman}@cs.ox.ac.uk