On the determinant of the product of two square matrices

On one level, this note deals with the determinants of square matrices and their product and is about a proof of the theorem that the determinant of a product equals the product of the determinants of the factors. On another level, this note is about notation, and it is the latter issue that I shall address first.

* * *

In contrast to my current habits, I shall use sub- and superscripts, and shall even do so in a nested fashion when the need arises. Also, multiplication will be denoted by juxtaposition.

I assume that the reader is willing to recognize in

$$\langle \Sigma i : 0 \leq i < n : A_i B^i \rangle$$

the sum of n products of an A and a B each. Suppose now, that we have to deal with many of such sums, but all with the same range for the dummy. (This is, for instance, a very common situation in geometrical considerations, in which the above n equals the number of dimensions of the space in question.) In that situation it pays to state the range once and for all, and not to repeat it in each summation. The above expression can then be abbreviated to

$$\langle \Sigma i :: A_i B^i \rangle \quad .$$

(I have adopted the habit of maintaining both colons as an indication that the range is implicitly understood.)

In the above, the variable i is declared to be the dummy by its occurrence after Σ, but there are other ways of identifying the dummy. In what is known as "the Einstein convention", which is tailored to the summation of products, the above is abbreviated to

$$A_i B^i \quad ,$$

where the dummy i is identified as such by the fact that it occurs once as subscript of a factor of the product and once as superscript. This convention not only eliminates the "$\Sigma i ::$", but also the scoping parenthesis pair "$\langle \, \rangle$".

When I was introduced to the Einstein convention, no one mentioned the elimination of the scoping parentheses, for they had not been introduced yet: people would just write

$$\sum_i A_i B^i + C$$

without making explicit the convention that would disambiguate the above.

Similarly, the identification of the dummy by its "double occurrence" raises the question "double occurrence where?". For instance, would summation be implied in

$$A_i + B^i \qquad ?$$

Since, in all their applications, dimensional considerations would rule out formulae like the above, the physicists of those days did not bother.

We, however, live in different times, and state explicitly that summation is implied by the double occurrence, once as subscript and once as superscript of factors of the same product, which is then the term to be summed.

We would like to point out that the omission of the scoping parentheses is justified by the observation of

$$\langle \Sigma i :: A_i\, B^i \rangle \langle \Sigma j :: C_j\, D^j \rangle = \langle \Sigma i, j :: A_i\, B^i\, C_j\, D^j \rangle \quad ,$$

an equality which is a consequence of the fact that multiplication distributes over addition. With the Einstein convention —and the As, Bs, Cs and Ds being real or complex— the above expression is written

$$A_i\, B^i\, C_j\, D^j \quad \text{or} \quad A_i\, C_j\, B^i\, D^j$$

or any other of the 24 permutations. The omission of parentheses usually reflects the associativity of an (infix) operator; in the case of the Einstein convention, the omission of the parentheses also subsumes a distribution law.

For various reasons I shall introduce and explain my next notational convention in terms of a most classical notation, in fact the notation which was deemed respectable (and in which I was educated) half a century ago.

In the following, A and B are $n \times n$ matrices with real or complex elements. Their rows and their columns are indexed from 1 through n . The element in row i_1 and column j_1 of matrix A is denoted by

$$\underline{A}_{i_1}^{j_1} \quad ,$$

and similarly for the elements of matrix B. (The purpose of the underlining will become clear in a moment.) Identifiers h, i, j, k, l will be used to denote index vectors of length n and the elements of h are denoted by

$$h_1 \ h_2 \ \ldots \ h_n \quad ,$$

and similarly for i, j, k and l . The individual indices in the above enumeration have a range from 1 through n .

Our new convention gives a meaning to A_i^j – note that both sub- and superscripts are vectors – ; A_i^j stands for a product of n elements of matrix A, to be precise

(0) $\quad A_i^j = \underline{A}_{i_1}^{j_1} \ \underline{A}_{i_2}^{j_2} \ldots \ \underline{A}_{i_n}^{j_n} \quad .$

Denoting the matrix product by an infix $\bullet$, we now give

<u>Lemma 0</u> $\quad (A \bullet B)_i^j = A_i^k B_k^j \quad .$

<u>Proof</u> We observe

$$(A \bullet B)_i^j$$

= {expansion of index vectors}

$$(\underline{A \bullet B})_{i_1}^{j_1} \; (\underline{A \bullet B})_{i_2}^{j_2} \; \ldots \; (\underline{A \bullet B})_{i_n}^{j_n}$$

= {definition of matrix product}

$$(\underline{A}_{i_1}^{k_1} \, \underline{B}_{k_1}^{j_1}) \; (\underline{A}_{i_2}^{k_2} \, \underline{B}_{k_2}^{j_2}) \; \ldots \; (\underline{A}_{i_n}^{k_n} \, \underline{B}_{k_n}^{j_n})$$

= {complex multiplication is commutative}

$$(\underline{A}_{i_1}^{k_1} \, \underline{A}_{i_2}^{k_2} \ldots \underline{A}_{i_n}^{k_n}) \; (\underline{B}_{k_1}^{j_1} \, \underline{B}_{k_2}^{j_2} \ldots \underline{B}_{k_n}^{j_n})$$

= {contraction of index vectors}

$$A_i^k \, B_k^j$$

(End of Proof)

* * *

Next we define for each i the two constants ε_i and ε^i ; since for all i

$$\varepsilon_i = \varepsilon^i$$

we need to define only one of them. With the vector nat given by

$$nat = 1 \;\; 2 \; \ldots \; n$$

we have

$\varepsilon_i = 0$ if i is <u>not</u> a permutation of nat (in which case i contains at least 1 duplicate)

$\varepsilon_i = +1$ if i is an even permutation of nat

$\varepsilon_i = -1$ if i is an odd permutation of nat.

Without taking into account when permutations are even or odd, the above implies

<u>Lemma 1</u> $\varepsilon_i \; \varepsilon^i = n!$

because n distinct objects admit $n!$ distinct sequential arrangements.

Let for the sake of the following discussion, the operator $'$ be associated with two positions in index vectors and denote the interchange of the elements in those two positions. Since the interchange of two elements turns an even permutation into an odd one and vice versa, ε satisfies

(1) $\varepsilon^i = -\varepsilon^{i'}$.

We now observe

$\varepsilon^{i} A_{i}^{j}$

$=$ {if i ranges over all n^n possible values, so does i'}

$\varepsilon^{i'} A_{i'}^{j}$

$=$ { $\varepsilon^{i'} = -\varepsilon^{i}$, see (1); $A_{i'}^{j} = A_{i}^{j'}$, see (0)}

$-\varepsilon^{i} A_{i}^{j'}$,

i.e. we have established

(2) $\varepsilon^{i} A_{i}^{j} = -\varepsilon^{i} A_{i}^{j'}$.

From (1) and (2) we establish

<u>Lemma 2</u> $\varepsilon^{i} A_{i}^{k} \varepsilon^{\ell} = \varepsilon^{i} A_{i}^{\ell} \varepsilon^{k}$

<u>Proof</u> On account of (2) and (1), the transformation of $\varepsilon^{i} A_{i}^{k}$ to $\varepsilon^{i} A_{i}^{\ell}$ introduces as many sign changes as the transformation from ε^{ℓ} to ε^{k}, so they cancel.
(End of Proof.)

The traditional definition of Det. A, i.e. "the determinant of A", is

(3) $\mathrm{Det.}A = \varepsilon^{i} A_{i}^{nat}$

but I prefer the more symmetric

(4) $\mathrm{Det.}A = \varepsilon^{i} A_{i}^{j} \varepsilon_{j} / n!$

The equivalence of these two definitions depends on

(5) $\quad \varepsilon^{nat} = +1$,

which enables us to derive:

$\varepsilon^i A_i^{nat}$

$=\quad$ {Lemma 1}

$\varepsilon^i A_i^{nat} \varepsilon^j \varepsilon_j / n!$

$=\quad$ {Lemma 2}

$\varepsilon^i A_i^j \varepsilon^{nat} \varepsilon_j / n!$

$=\quad$ {(5)}

$\varepsilon^i A_i^j \varepsilon_j / n!$

from which observation the equivalence of (3) and (4) follows.

And now we are ready to prove this note's main theorem.

<u>Theorem</u> $\quad \mathrm{Det.}(A \bullet B) = (\mathrm{Det.}A)(\mathrm{Det.}B)$

<u>Proof</u> We observe

$\mathrm{Det.}(A \bullet B)$

$=\quad$ {(4)}

$$\varepsilon^i (A \bullet B)_i^j \; \varepsilon_j / n!$$

$= \quad \{\text{Lemma } 0\}$

$$\varepsilon^i A_i^k \; B_k^j \; \varepsilon_j / n!$$

$= \quad \{\text{Lemma } 1\}$

$$\varepsilon^i A_i^k \; \varepsilon^\ell \; \varepsilon_\ell \; B_k^j \; \varepsilon_j / (n!)^2$$

$= \quad \{\text{Lemma } 2\}$

$$\varepsilon^i A_i^\ell \; \varepsilon^k \; \varepsilon_\ell \; B_k^j \; \varepsilon_j / (n!)^2$$

$= \quad \{\text{regrouping}\}$

$$(\varepsilon^i A_i^\ell \; \varepsilon_\ell / n!)(\varepsilon^k B_k^j \; \varepsilon_j / n!)$$

$= \quad \{(4), \text{ twice}\}$

$$(\text{Det.} A)(\text{Det.} B)$$

(End of Proof)

* * *

Matrix theory is closely associated with linear transformation of n-dimensional space, the functional composition of two transformations corresponding to the scalar product of their matrices. Since the determinant of the matrix equals the factor by which the transformation expands the n-dimensional volume, linear transformations provide a model in which our theorem is "obvious".

* * *

The above is essentially the proof I constructed as a young student. (For once I had been dissatisfied with what my master Prof. J. Haantjes, had presented.) At the time my calculations took much more paper: firstly I did not know the Einstein convention, secondly I used without shame or hesitation subscripted dummies and the "..." as I have done here in definition (0) and the proof of Lemma (0).

But as soon as one considers subscription as a notational alternative for function application, the "subscripted dummy" becomes an anomaly, for dummies should be "fresh variables" or "fresh identifiers" and not function values. Instead of considering

$$h_1 \quad h_2 \; . \; . \; . \; h_n$$

as n dummies, it is much cleaner to introduce the single dummy h, of type $\text{index} \rightarrow \text{index}$. Note that such a dummy ranges over n^n different values.

With p, q ranging over the type index

and using the infix dot for function application, (0) would have been

(0') $A_i^j = \langle \Pi p :: \underline{A}_{i.p}^{j.p} \rangle$

and the proof of Lemma 0 would have started

$$
\begin{aligned}
& (A \bullet B)_i^j \\
= & \quad \{ (0') \} \\
& \langle \Pi p :: \underline{(A \bullet B)}_{i.p}^{j.p} \rangle \\
= & \quad \{ \text{def. of matrix product} \} \\
& \langle \Pi p :: \langle \Sigma q :: A_{i.p}^q \, B_q^{j.p} \rangle \rangle \\
= & \quad \{ \text{distribution of } \Pi \text{ over } \Sigma \} \\
& \langle \Sigma k :: \langle \Pi p :: A_{i.p}^{k.p} \, B_{k.p}^{j.p} \rangle \rangle
\end{aligned}
$$

etc.

Note how in the last step the n summations over the index q have been replaced by a single summation over the function k.

With thanks to the ETAC, A. Bijlsma in particular.

Austin, 2 November 1997

prof. dr Edsger W. Dijkstra
Department of Computer Sciences
The University of Texas at Austin
Austin, TX 78712-1188, USA