Section 3.2. Multiplication of Matrices and Multiplication of Vectors and Matrices

Note. In this section, we define the product of matrices, elementary matrices, and explore how these interact with trace and determinant. We also introduce inner products of matrices.

Definition. Let $A = [a_{ij}]$ be an $m \times n$ matrix and let $B = [b_{ij}]$ be an $n \times s$ matrix. The *matrix product* (or "Cayley product") is the $m \times s$ matrix $C = [c_{ij}]$ where $c_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj}$, denoted $C = AB$. When matrices A and B are of dimensions so that the product AB is defined, then A and B are *conformable for multiplication*.

Theorem 3.2.1. Properties of Matrix Multiplication.

(1) Let $A = [a_{ij}]$ be $m \times n$ and let $B = [b_{ij}]$ be $n \times s$. Then $(AB)^T = B^T A^T$.

(2) Let $A = [a_{ij}]$ be $m \times n$, $B = [b_{ij}]$ be $n \times s$, and $C = [c_{ij}]$ be $s \times t$. Then $A(BC) = (AB)C$. That is, matrix multiplication is associative.

(3) Let $A = [a_{ij}]$ be $m \times n$ and let $B = [b_{ij}]$ and $C = [c_{ij}]$ be $n \times s$. Then $A(B+C) = AB + AC$. Let $A = [a_{ij}]$ be $m \times n$ and let $B = [b_{ij}]$ and $C = [c_{ij}]$ be $n \times m$ matrices. Then $(B+C)A = BA+CA$. That is, matrix multiplication distributes over matrix addition.

(4) Let $A = [a_{ij}]$ and $B = [b_{ij}]$ be $n \times n$ matrices. If A and B are diagonal then AB is diagonal. If A and B are upper triangular then AB is upper triangular. If A and B are lower triangular then AB is lower triangular.

Definition. The $n \times n$ matrix which is diagonal with all diagonal entries of 1 is the *identity matrix* of order n, denoted I_n or just I.

Note. If A is $n \times m$ then $I_n A = A I_m = A$.

Definition. If $p = \sum_{k=0}^{n} b_k x^k$ is a polynomial and A is a square matrix then define

$$p(A) = b_0 I + b_1 A + b_2 A^2 + \cdots + b_n A^n.$$

Note. In the study of systems of linear differential equations with constant coefficients of the form $\vec{x}\,'(t) = A\vec{x}(t)$, where t is the variable, the solution is $\vec{x}(t) = \vec{x}_0 e^{tA}$ and the exponentiation of a matrix is defined in terms of the power series for the exponential function (which is a limit of a polynomial); of course convergence becomes an issue. This is one reason to consider the diagonalization of a matrix, which we consider in Section 3.8.

Theorem 3.2.2. Consider partitioned matrices $\begin{bmatrix} A & B \\ C & D \end{bmatrix}$ and $\begin{bmatrix} E & F \\ G & H \end{bmatrix}$ where $A = [a_{ij}]$ is $k \times \ell$, $B = [b_{ij}]$ is $k \times m$, $C = [c_{ij}]$ is $n \times \ell$, $D = [d_{ij}]$ is $n \times m$, $E = [e_{ij}]$ is $\ell \times p$, $F = [f_{ij}]$ is $\ell \times q$, $G = [g_{ij}]$ is $m \times p$, and $H = [h_{ij}]$ is $m \times q$. Then the product of the partitioned matrices is the partitioned matrix

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} E & F \\ G & H \end{bmatrix} = \begin{bmatrix} AE + BG & AF + BH \\ CE + DG & CF + DH \end{bmatrix}.$$

Notice that the dimensions of the matrices insure that all matrix products involve matrices conformable for multiplication.

Definition. For a given matrix A, we may perform the following operations:

Row Interchange: Form matrix B by interchanging row i and row j of matrix A, denoted $A \overset{R_i \leftrightarrow R_j}{\widetilde{\quad}} B$.

Row Scaling: Form matrix B by multiplying the ith row of A be a nonzero scalar s, denoted $A \overset{R_i \to sR_i}{\widetilde{\quad}} B$.

Row Addition: Form matrix B by adding to the ith row of A s times the jth row of A, denoted $A \overset{R_i \to R_i + sR_j}{\widetilde{\quad}} B$.

These three operations on matrix A are the *elementary row operations*. We can define the *elementary column operations* similarly by forming matrix B from matrix A by column manipulations.

Note. We will see that each elementary row operation can be performed on a matrix by multiplying it on the left by an appropriate ("elementary") matrix. Similarly, elementary column operations can be performed in a matrix by multiplying it on the right be an appropriate matrix. Gentle calls multiplying a matrix on the left "premultiplication" and multiplying on the right "postmultiplication" (see page 62). In these notes, we use the terminology "multiply on the left/right."

Definition. An *elementary matrix* (or *elementary transformation matrix* or *elementary operator matrix*) is an $n \times n$ matrix which is formed by performing one elementary row operation or one elementary column operation on the $n \times n$ identity matrix I_n. If the operation is row or column interchange, the resulting matrix is an *elementary permutation matrix*; if the pth and qth row or (equivalently) column of I_n have been interchanged then the elementary permutation matrix is denoted E_{pq}. A product of elementary permutation matrices is a *permutation matrix*.

Theorem 3.2.3. Each of the three elementary row operations on $n \times m$ matrix A can be accomplished by multiplication on the left by an elementary matrix which is formed by performing the same elementary row operation on the $n \times n$ identity matrix. Each of the three elementary column operations on $n \times m$ matrix A can be accomplished by multiplication on the right by an elementary matrix which is formed by performing the same elementary column operation on the $m \times m$ identity matrix.

Note. We can use the results of Section 3.1 to find the determinants of each of the three types of elementary matrices. Recall that $\det(I_n) = 1$. For row interchange, we have $I_n \overset{R_p \leftrightarrow R_q}{\widetilde{}} E_{pq}$, and so $\det(E_{pq}) = -1$ by Theorem 3.1.C. For row scaling, we have $I_n \overset{R_p \to sR_p}{\widetilde{}} E_{sp}$, and so $\det(E_{sp}) = s$ by Theorem 3.1.B. For row addition, we have $I_n \overset{R_p \to R_p + sR_q}{\widetilde{}} E_{psq}$, and so $\det(E_{psq}) = 1$ by Theorem 3.1.E.

Theorem 3.2.4. For $n \times n$ matrices A and B, $\det(AB) = \det(A)\det(B)$.

Note. By convention we now consider vectors in $\mathbb{R}^n$ as **column vectors** (or $n \times 1$ matrices). So the vector x will represent the matrix $[x_1, x_2, \ldots, x_n]^T$. In this way, for $n \times m$ matrix A and vector $x \in \mathbb{R}^m$, Ax is defined and is $n \times 1$. For $n \times m$ matrix A and vector $x \in \mathbb{R}^n$, $x^T A$ is defined and is $1 \times m$.

Note. If we denote the m columns of $n \times m$ matrix A as $a_1, a_2, \ldots, a_m$ and consider $x \in \mathbb{R}^m$ as a vector of scalars then $Ax = \sum_{i=1}^m x_i a_i$ is a linear combination of the columns of A with coefficients of the entries of vector x. That is, Ax is an element of the column space of A. So searching for solutions x to the equation $Ax = b$ is equivalent to determining if b is in the column space of A.

Definition. For vectors $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$, the *outer product* is the $n \times m$ matrix xy^T.

Note. Unlike the inner product, the outer product is not in general commutative (consider the case $m \neq n$). If we consider xx^T for $x \in \mathbb{R}^n$, we have an $n \times n$ matrix with (i, j)th entry

$$\sum_{k=1}^n x_{ik} x_{kj}^t = \sum_{k=1}^n x_{ik} x_{jk} = \sum_{k=1}^n x_{jk} x_{ik} = \sum_{k=1}^n x_{jk} x_{ki}^t$$

which is the (j, i)th entry of the $n \times n$ matrix (where $x_{kj}^t = x_{jk}$); that is, xx^T is a symmetric matrix.

Definition. For $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$, and A an $n \times m$ matrix, the product $x^T Ay$ is a *bilinear form* (notice that $x^T Ay$ is 1×1). When $x \in \mathbb{R}^n$ and A is $n \times n$, $x^T Ax$ is a *quadratic form*.

Note. Fraleigh and Beauregard's *Linear Algebra*, 3rd Edition, Addison-Wesley Publishing Company (1995) defines a quadratic form of a vector $x \in \mathbb{R}^n$ as a function of the form $f(x) = \sum_{i \leq j; i,j=1}^{n} u_{ij} x_i x_j$. So with $n = 3$ we get

$$u_{11}x_1^2 + u_{12}x_1x_2 + u_{13}x_1x_3 + u_{22}x_2^2 + u_{23}x_2x_3 + u_{33}x_3^2.$$

They then show that for every quadratic form, there is a symmetric $n \times n$ matrix A such that the quadratic form is given by x^TAx (this is also shown in our Exercise 3.3). As a result, we concentrate on quadratic forms x^TAx for symmetric square matrices. For applications of quadratic forms to 2 and 3 dimensional geometry, see Fraleigh and Beauregard's Section 8.2. Applications to Geometry.

Definition. The $n \times n$ symmetric matrix A is *nonnegative definite* (or commonly *positive semidefinite*) if for each $x \in \mathbb{R}^n$ we have that the quadratic form satisfies $x^TAx \geq 0$. This is denoted $A \succeq 0$. The $n \times n$ symmetric matrix A is *positive definite* if for each $x \in \mathbb{R}^n$ with $x \neq 0$ we have that the quadratic form satisfies $x^TAx > 0$. This is denoted $A \succ 0$. When A and B are $n \times n$ symmetric matrices and $A - B \succeq 0$, we denote this as $A \succeq B$. When A and B are $n \times n$ symmetric matrices and $A - B \succ 0$, we denote this as $A \succ B$.

Note. For $x, y \in \mathbb{R}^n$, $\langle x, y \rangle = x^T I_n y$; technically, the inner product is a scalar and matrix product is a 1×1 matrix, but "we will treat a one by one matrix or a vector with only one element as a scalar whenever it is convenient to do so" (Gentle, page 69). We use this observation to motivate the following definition.

Definition. Vectors $x, y \in \mathbb{R}^n$ are *orthogonal with respect to* $n \times n$ *matrix* A if $x^T A y = 0$ and we say that x and y are A-*conjugate*. The *elliptic norm* (or *conjugate norm*) with respect to symmetric positive definite matrix A is $\|x\|_A = \sqrt{x^T A x}$.

Note. It is shown in Exercise 3.2.C that $\|\cdot\|_A$ is actually a norm. It is called an "elliptic norm" because the unit ball may not be round under this norm, but instead elliptical. Gentle describes this on page 71 in terms of scaling the units of the axes of an n-dimensional space and "anisotropic spaces."

Definition. Let $A = [a_{ij}]$ and $B = [b_{ij}]$ be $n \times m$ matrices. The *Hadamard product* of A and B is the $n \times m$ matrix $[a_{ij}b_{ij}]$. Hadamard multiplication of matrices is sometimes called "array multiplication" or "element-wise multiplication."

Note. The identity matrix under Hadamard multiplication is the $n \times m$ matrix of all 1s. So a matrix has an inverse under Hadamard multiplication if and only if all of its entries are nonzero (this is Exercise 3.10).

Definition. Let A be an $n \times m$ matrix and let B be a $p \times q$ matrix. The *Kronecker product* (or *tensor product*) of A and B is the $np \times mq$ matrix

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1m}B \\ a_{21}B & a_{22}B & \cdots & a_{2m}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1}B & a_{n2}B & \cdots & a_{nm}B \end{bmatrix}.$$

The Kronecker product is also called the "right direct product" or just "direct product."

Note. The right identity for Kronecker multiplication is the 1×1 matrix $[1]$. To find the (i,j)th entry of $A \otimes B$, we need to partition the indices i and j into "pieces." We do so using the greatest integer function $\lfloor x \rfloor$. The ith row of $A \otimes B$ involves $a_{1+\lfloor (i-1)/p \rfloor, k}$ (since B has p rows) and the $(i - p\lfloor (i-1)/p \rfloor)$th row of B (we have that $\lfloor (i-1)/p \rfloor$ is the maximum number of multiples of p such that $p\lfloor (i-1)/p \rfloor \leq i$ and so $i - p\lfloor (i-1)/p \rfloor$ is the remainder when i is divided by p except when the remainder is 0 in which case $i - p\lfloor (i-1)/p \rfloor = p$; so for $i \in \{1, 2, \ldots, np\}$ we have $i - p\lfloor (i-1)/p \rfloor \in \{1, 2, \ldots, p\}$). Similarly, the jth column of $A \otimes B$ involves $a_{k, 1+\lfloor (j-1)/q \rfloor}$ (since B has q columns) and the $(j - q\lfloor (j-1)/q \rfloor)$th column of B. So the (i,j)th entry of $A \otimes B$ is

$$a_{1+\lfloor (i-1)/p \rfloor, 1+\lfloor (j-1)/q \rfloor} b_{i-p\lfloor (i-1)/p \rfloor, j-q\lfloor (j-1)/q \rfloor}. \qquad (*)$$

Notice the error on Gentle's page 73, equation (3.69) which is incorrect when i is a multiple of p or j is a multiple of q.

Theorem 3.2.5. Properties of the Kronecker Product.

Let A, B, C, D be matrices which are conformable for the addition and regular matrix multiplication given below. Then

(1) Associativity of Scalar Multiplication:

$$(aA) \otimes (bB) = ab(A \otimes B) = (abA) \otimes B = A \otimes (abB).$$

(2) Distribution of $\otimes$ Over $+$: $(A + B) \otimes C = A \otimes C + B \otimes C$.

(3) Associativity of $\otimes$: $(A \otimes B) \otimes C = A \otimes (B \otimes C)$.

(4) Transposition of a Kronecker Product: $(A \otimes B)^T = A^T \otimes B^T$.

(5) Interaction of Kronecker Products and Regular Products:

$$(A \otimes B)(C \otimes D) = (AC) \otimes (BD).$$

Note. The proof of Theorem 3.2.5 is to be given in Exercise 3.6.

Theorem 3.2.6. Let A be an $n \times n$ matrix and let B be an $m \times m$ matrix. Then $\det(A \otimes B) = \det(A)^n \det(B)^m$.

Note. Gentle comments: "The proof of [Theorem 3.2.6], like many facts about determinants, is straightforward but involves tedious manipulation of cofactors" (page 73). This time we agree and an exploration of the claim would take us too far from our core topics, so we omit it. However, a proof can be found in David Harville's *Matrix Algebra From a Statistician's Perspective*, Springer-Verlag (1997), pages 343–350.

Theorem 3.2.7. Let A and B be $n \times n$ matrices. Then $\mathrm{tr}(A \otimes B) = \mathrm{tr}(A)\mathrm{tr}(B)$.

Definition. Let A and B be $n \times m$ matrices. Let the columns of A be the vectors $a_1, a_2, \ldots, a_m$ and let the columns of B be the vectors $b_1, b_2, \ldots, b_m$. The *inner product* (or *dot product*) of A and B is the scalar $\langle A, B \rangle = \sum_{j=1}^m a_j^T b_j = \sum_{j=1}^m \langle a_j, b_j \rangle$.

Note. Again, $\langle A, B\rangle$ is technically a 1×1 matrix, but we treat it as a scalar. Notice that if A and B are $m \times 1$ then the inner product of the $\langle A, B\rangle$ reduces to the usual vector inner product.

Theorem 3.2.8. Properties of the Inner Product of Matrices.
Let A, B, and C be matrices conformable for the addition and inner products given below. Then

(1) If $A \neq 0$ then $\langle A, A\rangle > 0$ and $\langle 0, A\rangle = \langle A, 0\rangle = \langle 0, 0\rangle = 0$.

(2) $\langle A, B\rangle = \langle B, A\rangle$.

(3) $\langle sA, B\rangle = s\langle A, B\rangle = \langle A, sB\rangle$ for scalar $s \in \mathbb{R}$.

(4) $\langle (A + B), C\rangle = \langle A, C\rangle + \langle B, C\rangle$ and $\langle C, (A + B)\rangle = \langle C, A\rangle + \langle C, B\rangle$.

(5) $\langle A, B\rangle = \mathrm{tr}(A^T B)$.

(6) $\langle A, B\rangle = \langle A^T, B^T\rangle$.

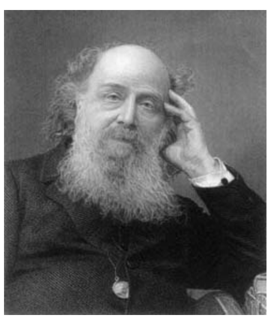
(7) Schwarz Inequality: For $n \times m$ matrices A and B, $|\langle A, B\rangle| \leq \langle A, A\rangle^{1/2}\langle B, B\rangle^{1/2}$.

Note. Theorem 3.2.8 establishes that $\langle \cdot, \cdot\rangle$ actually is an inner product on the vector space of all $n \times m$ matrices $\mathbb{R}^{n\times m}$. The Schwarz Inequality allows us to prove the Triangle Inequality and establish that $\|A\| = \langle A, A\rangle^{1/2}$ is a norm on $\mathbb{R}^{n\times m}$. We can now extend the ideas which are based on inner products from vectors to matrices.

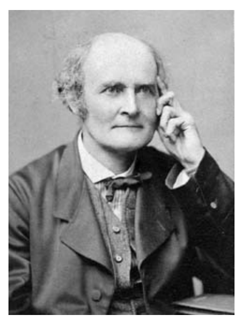
Definition. Two $n \times m$ matrices A and B are *orthogonal* to each other if $\langle A, B \rangle = 0$. A set of $n \times m$ matrices $\{U_1, U_2, \ldots, U_k\}$ is *orthonormal* if $\langle U_i, U_j \rangle = 0$ for $i \neq j$ and $\langle U_i, U_j \rangle = 1$ for $i = j$.

Note. We can use the previous definition to discuss orthonormal bases for $\mathbb{R}^{n \times m}$ and express a vector in the usual way in terms of an orthonormal basis.

Note. The term "matrix" was first used by James Joseph Sylvester (1814–1897) in 1850 in his "Additions to the articles, 'On a New Class of Theorems,' and 'On Pascal's Theorem'," *Philosophical Magazine*, **37**, 363-370 (1850). Arthur Cayley (1821–1895) was the first to concentrate on matrices themselves and the first to publish articles on the properties of matrices. For this reason, Cayley is credited as the creator of the theory of matrices [Morris Kline, *Mathematical Thought from Ancient to Modern Times*, Oxford University Press (1972), page 805].



James Joseph Sylvester



Arthur Cayley

Cayley introduced matrices to simplify the notation in the study of invariants under linear transformations [Kline, page 806] in "Remarques sur la notation des fonctions algébriques," *Journal für die reine und angewandte Mathematik* **50**, 282–285 (1855); a copy is available online, in French (accessed 10/11/2017). In 1858 he published "A Memoir on the Theory of Matrices," *A Philosophical Transactions of the Royal Society of London*, **148**, 17-37 (1858); a copy is available online at archive.org. Cayley defined the sum and product of matrices (hence our use of the term "Cayley product"), and the product of a matrix times a scalar. He introduced the identity matrix, the inverse of a square matrix, and showed how inverse matrices can be used to solve systems of equations [Israel Kleiner, *A History of Abstract Algebra*, Birkhäuser (2007), page 82]. The images are from the MacTutor History of Mathematics archive.

Revised: 6/5/2020