Matrix Math

Randy S

Abstract This article gives a brief review of matrix algebra, introduces the concept of the exponential of a matrix, and derives a useful identity for the derivative of the inverse or determinant of a matrix with respect to one of its components.

Contents

1 A note about language

The plural form of index is indices, pronouned in-dih-sees, but the singular form is still *index*. Indice ("in-dih-see") is **not a word**.

Simiarly, the plural form of matrix is matrices, pronouned may-trih-sees, but the singular form is still *matrix*. Matrice ("may-trih-see") is **not a word**.

2 A linear transformation in terms of a matrix

Let U and V be two vectors spaces, both over the same field F, with either $\mathcal{F} = \mathbb{R}$ or $\mathcal{F} = \mathbb{C}$. In this context, elements of $\mathcal{F}$ are called **scalars**. Recall that a map $\sigma : U \to V$ is called a **linear transformation** if

$$\sigma(u + u') = \sigma(u) + \sigma(u') \qquad\qquad \sigma(zu) = z\sigma(u)$$

for all $u, u' \in U$ and all $z \in \mathcal{F}$. If the vector spaces are finite-dimensional, then such a transformation can be represented by a matrix. Let u_a denote the components of $u \in U$ in some fixed basis for U, and let v_a denote the components of $v \in V$ in some fixed basis for V. Then the components of $v \equiv \sigma(u)$ may be written

$$v_a = \sum_b A_{ab} u_b \tag{1}$$

for some fixed set of coefficients $A_{ab} \in \mathcal{F}$ that depends only on σ, not on u. The **matrix** with components A_{ab} is denoted A, and equation (1) is abbreviated

$$v = Au.$$

This can be iterated. Given a sequence of two linear transformations, $U \to V \to W$, the components w_a of the output are related to the components u_a of the input by

$$w_a = \sum_{b,c} B_{ab} A_{bc} u_c \tag{2}$$

where A is the matrix representing the first map $U \to V$, and B is the matrix representing the second map $V \to W$. Equation (2) is abbreviated

$$w = BAu.$$

The overall map $U \to V \to W$ is represented by the matrix BA, called the **product** of B and A, defined by

$$(BA)_{ac} \equiv \sum_b B_{ab} A_{bc}. \tag{3}$$

3 Rows and columns

The components A_{ab} of a matrix A can be written explicitly as an array, with the first index a specifying the row and the second index b specifying the column. Example: a matrix A of size 2×3 is written

$$A = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \end{bmatrix}.$$

A vector can be regarded as a matrix of size $N \times 1$. Example: a vector v with three components is written

$$v = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}.$$

Equation (3) says that the component in the the jth row and kth column of the product AB is calculated by multiplying the components in the jth row of A by the corresponding components in the kth column of B, and summing those products.

4 Transpose and hermitian conjugate

Given a matrix A with components A_{ab}, its **transpose** A^T is the matrix with components

$$(A^T)_{ab} = A_{ba}.$$

In words: the transpose exchanges the roles of rows and columns. This definition implies that the transpose of a product is the product of the transposes in reverse order:

$$(AB)^T = B^T A^T.$$

A matrix A is called **symmetric** if $A^T = A$, or **antisymmetric** if $A^T = -A$.

Similarly, given a matrix A with components A_{ab}, its **hermitian conjugate** $A^\dagger$ is the matrix with components

$$(A^\dagger)_{ab} = (A_{ba})^*,$$

where the asterisk denotes complex conjugation. If the components are all real-valued, then the hermitian conjugate is the same as the transpose, but they are different when the components are complex-valued. This definition implies

$$(AB)^\dagger = B^\dagger A^\dagger.$$

A matrix A is called **hermitian** if $A^\dagger = A$, or **antihermitian** if $A^\dagger = -A$.

The notation A^* typically denotes the matrix with components

$$(A^*)_{ab} = (A_{ab})^*,$$

without switching the order of the indices. In words: A^* is obtained from A by replacing each component with its complex conjugate, without taking the transpose. However, the notation A^* is also often used for something different, namely the **adjoint** of a linear operator A, especially in the math literature. If a linear operator is being represented by a matrix, then the adjoint A^* is represented by the hermitian conjugate $A^\dagger$, which *does* involve a transpose. To avoid misunderstanding, you should always pay close attention to the context when interpreting the notation A^*.

5 Matrix algebra

Let V be an N-dimensional vector space, and consider linear transformations from V to itself. Such a transformation is represented by a **square matrix** of size $N \times N$. The set of all square matrices of this size will be denoted $\mathcal{M}_N$. For all $A, B \in \mathcal{M}_N$, the sum $A + B$ and product AB are both defined

- The sum $A + B$ is the matrix with components $(A + B)_{ab} = A_{ab} + B_{ab}$.
- The product AB was defined in section 2. The product is **associative**, which means $(AB)C = A(BC)$. However, the product is not commutative: $AB \neq BA$ for most $A, B \in \mathcal{M}_N$. Two matrices A, B for which $AB = BA$ are said to **commute** with each other.

Two members of $\mathcal{M}_N$ have special names:

- The **zero matrix** 0 is defined by the property $0 + A = A + 0 = A$ for every $A \in \mathcal{M}_N$. This implies that the components of the zero matrix are all zero.[1] The zero matrix also satisfies $0A = A0 = 0$ for every $A \in \mathcal{M}_N$.
- The **identity matrix** I is defined by the property $IA = AI = A$ for every $A \in \mathcal{M}_N$. This implies that the diagonal components I_{aa} are all 1 and the other components are all zero.[2]

The product of a scalar $z \in \mathcal{F}$ and a matrix $A \in \mathcal{M}_N$ is another matrix $zA \in \mathcal{M}_N$, whose components are $(zA)_{ab} = zA_{ab}$. In words: when multiplying a matrix by a scalar, every component of the matrix is multiplied by the scalar.

Altogether, $\mathcal{M}_N$ is an **algebra**. Its elements can be added to each other, multiplied by each other, and multiplied by scalars. The algebra is associative but not commutative.

[1] Using the same symbol 0 both for the zero matrix and for the individual number zero is relatively safe, because the distinction is usually clear from the context if it matters at all.

[2] Sometimes the identity matrix I is denoted by the symbol 1, the same symbol we use for the individual number 1. That's relatively safe, but this article uses the distinct symbol I for the identity matrix anyway.

6 Inverse, eigenvectors, and determinant

A given matrix $A \in \mathcal{M}_N$ may or may not have an **inverse**, which is a matrix A^{-1} defined by the condition

$$A^{-1}A = AA^{-1} = I.$$

A matrix A is called **invertible** if it has an inverse.

A nonzero vector v is called an **eigenvector** of A if

$$Av = \lambda v$$

for some number $\lambda \in \mathcal{F}$ called the **eigenvalue**.[3] A matrix $A \in \mathcal{M}_N$ may have up to N linearly independent eigenvectors, or it may have fewer. Example: the only eigenvectors of the matrix

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

are $v \propto (1, 0)$. The corresponding eigenvalue is 1.

The **determinant** of a matrix A, denoted $\det A$, is defined in article 81674 using the **wedge product**. For $A \in \mathcal{M}_N$, the definition is

$$(\det A)\gamma_1 \wedge \gamma_2 \wedge \cdots \wedge \gamma_N = (A\gamma_1) \wedge (A\gamma_2) \wedge \cdots \wedge (A\gamma_N), \tag{4}$$

where $\gamma_1, \gamma_2, ..., \gamma_N$ is any set of N linearly independent vectors. If a matrix A has N linearly independent eigenvectors, then we can take the γ_ks to be those eigenvectors to see that the determinant of A is the product of the eigenvalues. We can also use this definition to show that a matrix is invertible if and only if its determinant is nonzero, using the fact that a matrix is invertible if and only if it maps any set of linearly independent vectors to another set of linearly independent vectors.

[3] Every square matrix over the complex numbers $\mathbb{C}$ has at least one eigenvalue in $\mathbb{C}$ (Axler (1995)), but a square matrix over $\mathbb{R}$ may fail to have any eigenvalues in $\mathbb{R}$ (example: the generator of rotations in two-dimensional space). Sometimes an *eigenvalue* of A is defined to be a complex number λ for which $A - \lambda I$ is not invertible (Axler (1995)). The definitions are interchangeable for a matrix over $\mathbb{C}$ of finite size, but when dealing with infinite-dimensional Hilbert spaces, the name *eigenvalue* is usually reserved for elements of the spectrum that have corresponding eigenvectors (article 74088).

7 Trace

The **trace** of a matrix A is the sum of its diagonal components:

$$\text{tr}A = \sum_a A_{aa}. \tag{5}$$

For some purposes expressing $\text{tr}A$ in terms of wedge products is convenient, like the definition of $\det A$ in equation (4). If $\gamma_1, \gamma_2, ..., \gamma_N$ is any set of N linearly independent vectors, then

$$\begin{aligned}(\text{tr}A)\gamma_1 \wedge \gamma_2 \wedge \cdots \wedge \gamma_N &= (A\gamma_1) \wedge \gamma_2 \wedge \cdots \wedge \gamma_N \\ &\quad + \gamma_1 \wedge (A\gamma_2) \wedge \cdots \wedge \gamma_N \\ &\quad + \gamma_1 \wedge \gamma_2 \wedge \cdots \wedge (A\gamma_N).\end{aligned} \tag{6}$$

Proof: For each k, take the vector γ_k to be the vector whose kth component is 1 and whose other components are zero. Then the kth term on the right-hand side of (6) is equal to

$$(A_{kk})\gamma_1 \wedge \gamma_2 \wedge \cdots \wedge \gamma_N.$$

To prove that equation (6) still works for any other set of N linearly independent vectors, write $\gamma_k = S\mu_k$ for any invertible linear transformation S, and then use the definition of $\det S$ (equation (4)) to see that (6) implies $\text{tr}A = \text{tr}(S^{-1}AS)$ after canceling a factor of $\det S$ from both sides of the equation.

If A has N linearly independent eigenvectors, then we can use these as the vectors γ_k in equation (6) to deduce that $\text{tr}A$ is the sum of the eigenvalues of A.

8 Diagonal and block-diagonal

A matrix A is called **diagonal** if $A_{ab} = 0$ whenever $a \neq b$. The components with $a = b$ are called the **diagonal components**. A diagonal matrix is sometimes specified by listing its diagonal components, like this:

$$A = \text{diag}(A_{11},\, A_{22},\, ...,\, A_{NN}).$$

Diagonal matrices commute with each other.

More generally, suppose that the list of allowed index-values is partitioned into subsets, each subset consisting of a list of consecutive values. Then a matrix A is called **block-diagonal** if $A_{ab} = 0$ whenever a and b belong to different subsets.

9 The derivative of an inverse

The identity

$$\boxed{\frac{\partial}{\partial A_{ab}}(A^{-1})_{cd} = -(A^{-1})_{ca}(A^{-1})_{bd}} \tag{7}$$

holds for any invertible matrix A. To derive this, apply $\partial/\partial A_{ab}$ to both sides of the identity $I = A^{-1}A$ to get

$$0 = \left(\frac{\partial}{\partial A_{ab}}A^{-1}\right)A + A^{-1}\frac{\partial}{\partial A_{ab}}A,$$

and then multiply both sides on the right by A^{-1} to get

$$0 = \left(\frac{\partial}{\partial A_{ab}}A^{-1}\right) + A^{-1}\left(\frac{\partial}{\partial A_{ab}}A\right)A^{-1}.$$

The result (7) is an easy consequence of this.

10 The derivative of a determinant

For any invertible matrix A, the identity

$$\boxed{\frac{\partial}{\partial A_{ab}}\det A = (A^{-1})_{ab}\det A} \tag{8}$$

holds. To prove this, suppose that A is an $N \times N$ matrix, and let $\gamma_1, \gamma_2, ..., \gamma_N$ be a set of N linearly independent vectors. The trace of a matrix is equal to the sum of its diagonal components, so the identity

$$\text{tr}\left(\frac{\partial A}{\partial A_{ab}}B\right) = B_{ab}$$

holds for all matrices A, B. In particular,

$$\text{tr}\left(\frac{\partial A}{\partial A_{ab}}A^{-1}\right) = (A^{-1})_{ab}.$$

Now, recall that the determinant of A is defined by

$$(\det A)\gamma_1 \wedge \gamma_2 \wedge \cdots \wedge \gamma_N = (A\gamma_1) \wedge (A\gamma_2) \wedge \cdots \wedge (A\gamma_N),$$

Apply $\partial/\partial A_{ab}$ to both sides to get

$$\begin{aligned}\frac{\partial}{\partial A_{ab}}(\det A)\gamma_1 \wedge \gamma_2 \wedge \cdots \wedge \gamma_N &= (MA\gamma_1) \wedge (A\gamma_2) \wedge \cdots \wedge (A\gamma_N)\\ &+ (A\gamma_1) \wedge (MA\gamma_2) \wedge \cdots \wedge (A\gamma_N)\\ &+ (A\gamma_1) \wedge (A\gamma_2) \wedge \cdots \wedge (MA\gamma_N)\end{aligned} \tag{9}$$

with $M \equiv \frac{\partial A}{\partial A_{ab}}A^{-1}$. The right-hand side of (9) is equal to

$$(\text{tr}M)(A\gamma_1) \wedge (A\gamma_2) \wedge \cdots \wedge (A\gamma_N) = (\text{tr}A)(\det A)\gamma_1 \wedge \gamma_2 \wedge \cdots \wedge \gamma_N.$$

Combining these ingredients gives the result (8).

11 The expontential of a matrix

If the components of a matrix A are functions of a real variable θ, then we can say that the matrix itself is a function of θ, denoted $A(\theta)$. The derivative of $A(\theta)$ with respect to θ is defined component-wise:

$$\left(\frac{d}{d\theta}A(\theta)\right)_{ab} \equiv \frac{d}{d\theta}A_{ab}(\theta).$$

Given any matrix $B \in \mathcal{M}_N$, the **exponential** function $A(\theta) = \exp(\theta B)$ is a matrix $A(\theta)$ whose components are functions of θ, defined by the conditions

$$\frac{d}{d\theta}\exp(\theta B) = B\exp(\theta B) \qquad\qquad \exp(\theta B)\big|_{\theta=0} = 1. \tag{10}$$

The definition is unambiguous because this system of N^2 first-order differential equations has a unique solution.[4] The matrix $\exp(\theta B)$ is also denoted $e^{\theta B}$ or $e^{B\theta}$.

As an example, consider the matrix

$$B = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix},$$

which satisfies $B^2 = -I$. Then the definition (10) is satisfied by

$$e^{B\theta} = I\cos\theta + B\sin\theta = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}. \tag{11}$$

We recognize this as a rotation matrix. The matrix B is called the **generator** of these rotations. This is analogous to **Euler's formula** $e^{i\theta} = \cos\theta + i\sin\theta$, where i is the imaginary unit $i^2 = -1$.

The definition implies $e^{\theta B}e^{\phi B} = e^{(\theta+\phi)B}$, just like the exponential of a real variable. If B and C commute with each other, then we also have $e^{B\theta}e^{C\theta} = e^{(B+C)\theta}$. However, beware that $e^{B\theta}e^{C\theta} \neq e^{(B+C)\theta}$ for most matrices $B, C \in \mathcal{M}_N$.

[4]To prove this, suppose it had two solutions, say $A(\theta)$ and $\tilde{A}(\theta)$. Then their difference would be a solution of the first of equations (10) that is equal zero when $\theta = 0$, which implies that it equals zero for all θ.

12 More exampes of matrix exponentials

The previous section showed how to use a matrix exponential to represent a rotation in a two-dimensional vector space ($N = 2$). More generally, we can use a matrix exponential to represent a rotation in an arbitrary plane in an N-dimensional vector space. Let a, b be two linearly indepenent vectors, so that they define a plane. We can represent each vector as a matrix of size $N \times 1$, and then the plane itself is naturally represented by the antisymmetric matrix $B = ab^T - ba^T$. This matrix satisfies $B^3 \propto -B$, and we can normalize it so that $B^3 = -B$. With that normalization, the definition (10) is satisfied by

$$e^{B\theta} = I + (1 - \cos\theta)B^2 + B\sin\theta.$$

This generalizes the earlier example (11). Here's an example in three-dimensional space: if $a = (1, 0, 0)$ and $b = (0, 1, 0)$, then

$$B = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad \Rightarrow \qquad e^{B\theta} = \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

A similar idea can be used to represent Lorentz boosts (article 77597). For this, we need the generator B to be symmetric and satisfy $B^3 = B$, without the minus sign. Then the definition (10) is satisfied by

$$e^{B\theta} = I + (\cosh\theta - 1)B^2 + B\sinh\theta.$$

The hyperbolic functions $\cosh\theta$ and $\sinh\theta$ are defined in article 77597. Example:

$$B = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad \Rightarrow \qquad e^{B\theta} = \begin{bmatrix} \cosh\theta & \sinh\theta & 0 \\ \sinh\theta & \cosh\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

In addition to rotations in space-space planes and boosts in time-space planes, the Lorentz group also includes **null rotations** in a plane that contains exactly one lightlike direction. A null rotation is the borderline between a rotation and a boost. The null rotation

$$e^{B\theta} = \begin{bmatrix} 1+\theta^2/2 & -\theta^2/2 & \theta & 0 \\ \theta^2/2 & 1-\theta^2/2 & \theta & 0 \\ \theta & -\theta & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

is generated by the sum of a rotation-generator and a boost-generator:

$$B = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

One way to prove this is to use the fact that this generator satisfies $B^3 = 0$, so the definition (10) is satisfied by

$$e^{B\theta} = I + \theta B + \frac{\theta^2}{2} B^2.$$

13 The determinant of an exponential

The identity

$$\boxed{\det\exp(\theta B) = \exp(\theta\,\mathrm{tr}B)} \tag{12}$$

holds for any matrix B. To prove this, start with the definition of the determinant of $A \equiv \exp(\theta B)$, as in equation (4). Take the derivative of that definition with respect to θ and use (6) to deduce

$$\frac{d}{d\theta}\det\exp(\theta B) = (\mathrm{tr}B)\det\exp(\theta B),$$

whose unique solution with $\det\exp(\theta B)\big|_{\theta=0} = 1$ is given by equation (12). To help make (12) more memorable, notice that it's obvious when B is diagonal, because then

$$\det\exp(\theta B) = \prod_n \exp\big(\theta B_{nn}\big) \tag{13}$$

$$\exp(\theta\,\mathrm{tr}B) = \exp\left(\theta\sum_n B_{nn}\right). \tag{14}$$

14 References

Axler, 1995. "Down With Determinants!" *American Mathematical Monthly* **102**: 139-154, https://www.maa.org/sites/default/files/pdf/awards/Axler-Ford-1996.pdf

15 References in this series

Article **74088** (https://cphysics.org/article/74088):
"Linear Operators on a Hilbert Space" (version 2022-10-23)

Article **77597** (https://cphysics.org/article/77597):
"Energy and Momentum at All Speeds" (version 2022-02-18)

Article **81674** (https://cphysics.org/article/81674):
"Can the Cross Product be Generalized to Higher-Dimensional Space?" (version 2022-02-06)