Matrix algebra

Matrix algebra

One-by-one matrices

Let us motivate what we want to achieve with matrices. What do real-valued linear mappings look like? A linear function of real numbers that you have seen in calculus is of the form

$$f(x) = mx + b.$$

However, the properties of linear mappings discussed in the previous section are that

$$f(x + y) = f(x) + f(y) \qquad f(ax) = af(x).$$

Plugging in the definition from above gives that

$$f(x + y) = m(x + y) + b = mx + my + b$$
$$f(ax) = m(ax) + b = a(mx) + b$$

and neither of these match up appropriately, since

$$f(x) + f(y) = mx + b + my + b = mx + my + 2b$$
$$af(x) = a(mx + b) = a(mx) + ab$$
.

In order for these to work, we need to have $b = 0$. Therefore, real-valued linear mappings of the real line, linear functions that eat numbers and spit out numbers, are just multiplications by a number.

Consider a mapping defined by multiplying by a number. Let's call this number α. The mapping then takes x to αx. What we can do is to *add* such mappings. If we have another mapping β, then

$$\alpha x + \beta x = (\alpha + \beta)x.$$

We get a new mapping $\alpha + \beta$ that multiplies x by, well, $\alpha + \beta$. If D is a mapping that doubles things, $Dx = 2x$, and T is a mapping that triples, $Tx = 3x$, then $D + T$ is a mapping that multiplies by 5, $(D + T)x = 5x$.

Similarly we can *compose* such mappings, that is, we could apply one and then the other. We take x, we run it through the first mapping α to get α times x, then we run αx through the second mapping β. In other words,

$$\beta(\alpha x) = (\beta\alpha)x.$$

We just multiply those two numbers. Using our doubling and tripling mappings, if we double and then triple, that is $T(Dx)$ then we obtain $3(2x) = 6x$. The composition TD is the mapping that multiplies by 6. For larger matrices, composition also ends up being a kind of multiplication.

Matrix addition and scalar multiplication

The mappings that multiply numbers by numbers are just 1×1 matrices. The number α above could be written as a matrix $[\alpha]$. So perhaps we would want to do the same things to all matrices that we did to those 1×1 matrices at the start of this section above. First, let us add matrices. If we have a matrix A and a matrix B that are of the same size, say $m \times n$, then they are mappings from $\mathbb{R}^n$ to $\mathbb{R}^m$. The mapping $A + B$ should also be a mapping from $\mathbb{R}^n$ to $\mathbb{R}^m$, and it should do the following to vectors:

$$(A + B)\vec{x} = A\vec{x} + B\vec{x}.$$

Learning outcomes: Perform addition and multiplication operations on matrices Compute inverses of 2×2 matrices Identify triangular, diagonal, and symmetric matrices.

Author(s): Matthew Charnley and Jason Nowell

It turns out you just add the matrices element-wise: If the ij^{th} entry of A is a_{ij}, and the ij^{th} entry of B is b_{ij}, then the ij^{th} entry of $A + B$ is $a_{ij} + b_{ij}$. If

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \qquad \text{and} \qquad B = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{bmatrix},$$

then

$$A + B = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} & a_{13} + b_{13} \\ a_{21} + b_{21} & a_{22} + b_{22} & a_{23} + b_{23} \end{bmatrix}.$$

Let us illustrate on a more concrete example:

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} + \begin{bmatrix} 7 & 8 \\ 9 & 10 \\ 11 & -1 \end{bmatrix} = \begin{bmatrix} 1+7 & 2+8 \\ 3+9 & 4+10 \\ 5+11 & 6-1 \end{bmatrix} = \begin{bmatrix} 8 & 10 \\ 12 & 14 \\ 16 & 5 \end{bmatrix}.$$

Let's check that this does the right thing to a vector. Let's use some of the vector algebra that we already know, and regroup things:

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \end{bmatrix} + \begin{bmatrix} 7 & 8 \\ 9 & 10 \\ 11 & -1 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \end{bmatrix} = \left(2 \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix} - \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix} \right) + \left(2 \begin{bmatrix} 7 \\ 9 \\ 11 \end{bmatrix} - \begin{bmatrix} 8 \\ 10 \\ -1 \end{bmatrix} \right)$$

$$= 2 \left(\begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix} + \begin{bmatrix} 7 \\ 9 \\ 11 \end{bmatrix} \right) - \left(\begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix} + \begin{bmatrix} 8 \\ 10 \\ -1 \end{bmatrix} \right)$$

$$= 2 \begin{bmatrix} 1+7 \\ 3+9 \\ 5+11 \end{bmatrix} - \begin{bmatrix} 2+8 \\ 4+10 \\ 6-1 \end{bmatrix} = 2 \begin{bmatrix} 8 \\ 12 \\ 16 \end{bmatrix} - \begin{bmatrix} 10 \\ 14 \\ 5 \end{bmatrix}$$

$$= \begin{bmatrix} 8 & 10 \\ 12 & 14 \\ 16 & 5 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \end{bmatrix} \qquad \left(= \begin{bmatrix} 2(8) - 10 \\ 2(12) - 14 \\ 2(16) - 5 \end{bmatrix} = \begin{bmatrix} 6 \\ 10 \\ 27 \end{bmatrix} \right).$$

If we replaced the numbers by letters that would constitute a proof! You'll notice that we didn't really have to even compute what the result is to convince ourselves that the two expressions were equal.

If the sizes of the matrices do not match, then addition is not defined. If A is 3×2 and B is 2×5, then we cannot add these matrices. We don't know what that could possibly mean.

It is also useful to have a matrix that when added to any other matrix does nothing. This is the zero matrix, the matrix of all zeros:

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}.$$

We often denote the zero matrix by 0 without specifying size. We would then just write $A + 0$, where we just assume that 0 is the zero matrix of the same size as A.

There are really two things we can multiply matrices by. We can multiply matrices by scalars or we can multiply by other matrices. Let us first consider multiplication by scalars. For a matrix A and a scalar α we want αA to be the matrix that accomplishes

$$(\alpha A)\vec{x} = \alpha(A\vec{x}).$$

That is just scaling the result by α. If you think about it, scaling every term in A by α accomplishes just that: If

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}, \qquad \text{then} \qquad \alpha A = \begin{bmatrix} \alpha a_{11} & \alpha a_{12} & \alpha a_{13} \\ \alpha a_{21} & \alpha a_{22} & \alpha a_{23} \end{bmatrix}.$$

For example,

$$2 \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} = \begin{bmatrix} 2 & 4 & 6 \\ 8 & 10 & 12 \end{bmatrix}.$$

Let us list some properties of matrix addition and scalar multiplication. Denote by 0 the zero matrix, by α, β scalars,

and by A, B, C matrices. Then:

$$A + 0 = A = 0 + A,$$
$$A + B = B + A,$$
$$(A + B) + C = A + (B + C),$$
$$\alpha(A + B) = \alpha A + \alpha B,$$
$$(\alpha + \beta)A = \alpha A + \beta A.$$

These rules should look very familiar.

Matrix multiplication

As we mentioned above, composition of linear mappings is also a multiplication of matrices. Suppose A is an $m \times n$ matrix, that is, A takes $\mathbb{R}^n$ to $\mathbb{R}^m$, and B is an $n \times p$ matrix, that is, B takes $\mathbb{R}^p$ to $\mathbb{R}^n$. The composition AB should work as follows

$$AB\vec{x} = A(B\vec{x}).$$

First, a vector $\vec{x}$ in $\mathbb{R}^p$ gets taken to the vector $B\vec{x}$ in $\mathbb{R}^n$. Then the mapping A takes it to the vector $A(B\vec{x})$ in $\mathbb{R}^m$. In other words, the composition AB should be an $m \times p$ matrix. In terms of sizes we should have

$$\text{`` } [m \times n]\,[n \times p] = [m \times p]. \text{ ''}$$

Notice how the middle size must match.

OK, now we know what sizes of matrices we should be able to multiply, and what the product should be. Let us see how to actually compute matrix multiplication. We start with the so-called *dot product* (or *inner product*) of two vectors. Usually this is a row vector multiplied with a column vector of the same size. Dot product multiplies each pair of entries from the first and the second vector and sums these products. The result is a single number. For example,

$$\begin{bmatrix} a_1 & a_2 & a_3 \end{bmatrix} \cdot \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = a_1 b_1 + a_2 b_2 + a_3 b_3.$$

And similarly for larger (or smaller) vectors. A dot product is really a product of two matrices: a $1 \times n$ matrix and an $n \times 1$ matrix resulting in a 1×1 matrix, that is, a number.

Armed with the dot product we define the *product of matrices*. First let us denote by $\text{row}_i(A)$ the ith row of A and by $\text{column}_j(A)$ the jth column of A. For an $m \times n$ matrix A and an $n \times p$ matrix B we can compute the product AB. The matrix AB is an $m \times p$ matrix whose ij^{th} entry is the dot product

$$\text{row}_i(A) \cdot \text{column}_j(B).$$

For example, given a 2×3 and a 3×2 matrix we should end up with a 2×2 matrix:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix} = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} & a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32} \\ a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31} & a_{21}b_{12} + a_{22}b_{22} + a_{23}b_{32} \end{bmatrix}, \tag{1}$$

or with some numbers:

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} -1 & 2 \\ -7 & 0 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 1 \cdot (-1) + 2 \cdot (-7) + 3 \cdot 1 & 1 \cdot 2 + 2 \cdot 0 + 3 \cdot (-1) \\ 4 \cdot (-1) + 5 \cdot (-7) + 6 \cdot 1 & 4 \cdot 2 + 5 \cdot 0 + 6 \cdot (-1) \end{bmatrix} = \begin{bmatrix} -12 & -1 \\ -33 & 2 \end{bmatrix}.$$

A useful consequence of the definition is that the evaluation $A\vec{x}$ for a matrix A and a (column) vector $\vec{x}$ is also matrix multiplication. That is really why we think of vectors as column vectors, or $n \times 1$ matrices. For example,

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \cdot 2 + 2 \cdot (-1) \\ 3 \cdot 2 + 4 \cdot (-1) \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \end{bmatrix}.$$

If you look at the last section, that is precisely the last example we gave.

You should stare at the computation of multiplication of matrices AB and the previous definition of $A\vec{y}$ as a mapping for a moment. What we are doing with matrix multiplication is applying the mapping A to the columns of B. This is usually written as follows. Suppose we write the $n \times p$ matrix $B = [\vec{b}_1 \ \vec{b}_2 \ \cdots \ \vec{b}_p]$, where $\vec{b}_1, \vec{b}_2, \ldots, \vec{b}_p$ are the columns of B. Then for an $m \times n$ matrix A,

$$AB = A[\vec{b}_1 \ \vec{b}_2 \ \cdots \ \vec{b}_p] = [A\vec{b}_1 \ A\vec{b}_2 \ \cdots \ A\vec{b}_p].$$

The columns of the $m \times p$ matrix AB are the vectors $A\vec{b}_1, A\vec{b}_2, \ldots, A\vec{b}_p$. For example, in (1), the columns of

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix}$$

are

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{bmatrix} b_{11} \\ b_{21} \\ b_{31} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{bmatrix} b_{12} \\ b_{22} \\ b_{32} \end{bmatrix}.$$

This is a very useful way to understand what matrix multiplication is. It should also make it easier to remember how to perform matrix multiplication.

We can go one step farther with this connection. The idea of a linear combination of vectors was defined in § **??**: for a set of vectors $\vec{y}_1, \vec{y}_2, \ldots, \vec{y}_n$, a linear combination of those vectors is a vector of the form $\alpha_1 \vec{y}_1 + \alpha_2 \vec{y}_2 + \cdots \alpha_n \vec{y}_n$ for some real numbers $\alpha_1, \ldots, \alpha_n$. If we write out the product from the end of the last example, we see

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{bmatrix} b_{12} \\ b_{22} \\ b_{32} \end{bmatrix} = \begin{bmatrix} a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32} \\ a_{21}b_{12} + a_{22}b_{22} + a_{23}b_{32} \end{bmatrix} = b_{12} \begin{bmatrix} a_{11} \\ a_{21} \end{bmatrix} + b_{22} \begin{bmatrix} a_{12} \\ a_{22} \end{bmatrix} + b_{32} \begin{bmatrix} a_{13} \\ a_{23} \end{bmatrix}.$$

Thus, computing the product of a matrix and a vector gives a new vector which is a linear combination of the columns of the matrix, where the coefficients are the entries in the vector. This gives two connections between matrix multiplication and linear independence and linear combinations.

(a) If there is a solution to the vector equation $A\vec{x} = \vec{b}$, this means that the vector $\vec{b}$ can be written as a linear combination of the columns of A.

(b) If there is a non-zero solution to the equation $A\vec{x} = \vec{0}$, then the columns of A are linearly dependent, and the solution $\vec{x}$ gives the coefficients necessary to give a linear combination that equals zero.

We'll see a lot more about this, and how to determine it, in § **??**.

Some rules of matrix algebra

For multiplication we want an analogue of a 1. That is, we desire a matrix that just leaves everything as it found it. This analogue is the so-called *identity matrix*. The identity matrix is a square matrix with 1s on the main diagonal and zeros everywhere else. It is usually denoted by I. For each size we have a different identity matrix and so sometimes we may denote the size as a subscript. For example, the I_3 would be the 3×3 identity matrix

$$I = I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Let us see how the matrix works on a smaller example,

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} a_{11} \cdot 1 + a_{12} \cdot 0 & a_{11} \cdot 0 + a_{12} \cdot 1 \\ a_{21} \cdot 1 + a_{22} \cdot 0 & a_{21} \cdot 0 + a_{22} \cdot 1 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}.$$

Multiplication by the identity from the left looks similar, and also does not touch anything.

We have the following rules for matrix multiplication. Suppose that A, B, C are matrices of the correct sizes so that the following make sense. Let α denote a scalar (number). Then

$$\begin{aligned} A(BC) &= (AB)C && \text{(associative law),} \\ A(B + C) &= AB + AC && \text{(distributive law),} \\ (B + C)A &= BA + CA && \text{(distributive law),} \\ \alpha(AB) &= (\alpha A)B = A(\alpha B), \\ IA &= A = AI && \text{(identity).} \end{aligned}$$

Example 1. *Let us demonstrate a couple of these rules. For example, the associative law:*

$$\underbrace{\begin{bmatrix} -3 & 3 \\ 2 & -2 \end{bmatrix}}_{A} \left(\underbrace{\begin{bmatrix} 4 & 4 \\ 1 & -3 \end{bmatrix}}_{B} \underbrace{\begin{bmatrix} -1 & 4 \\ 5 & 2 \end{bmatrix}}_{C} \right) = \underbrace{\begin{bmatrix} -3 & 3 \\ 2 & -2 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} 16 & 24 \\ -16 & -2 \end{bmatrix}}_{BC} = \underbrace{\begin{bmatrix} -96 & -78 \\ 64 & 52 \end{bmatrix}}_{A(BC)},$$

and

$$\left(\underbrace{\begin{bmatrix} -3 & 3 \\ 2 & -2 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} 4 & 4 \\ 1 & -3 \end{bmatrix}}_{B} \right) \underbrace{\begin{bmatrix} -1 & 4 \\ 5 & 2 \end{bmatrix}}_{C} = \underbrace{\begin{bmatrix} -9 & -21 \\ 6 & 14 \end{bmatrix}}_{AB} \underbrace{\begin{bmatrix} -1 & 4 \\ 5 & 2 \end{bmatrix}}_{C} = \underbrace{\begin{bmatrix} -96 & -78 \\ 64 & 52 \end{bmatrix}}_{(AB)C}.$$

Or how about multiplication by scalars:

$$10 \left(\underbrace{\begin{bmatrix} -3 & 3 \\ 2 & -2 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} 4 & 4 \\ 1 & -3 \end{bmatrix}}_{B} \right) = 10 \underbrace{\begin{bmatrix} -9 & -21 \\ 6 & 14 \end{bmatrix}}_{AB} = \underbrace{\begin{bmatrix} -90 & -210 \\ 60 & 140 \end{bmatrix}}_{10(AB)},$$

$$\left(10 \underbrace{\begin{bmatrix} -3 & 3 \\ 2 & -2 \end{bmatrix}}_{A} \right) \underbrace{\begin{bmatrix} 4 & 4 \\ 1 & -3 \end{bmatrix}}_{B} = \underbrace{\begin{bmatrix} -30 & 30 \\ 20 & -20 \end{bmatrix}}_{10A} \underbrace{\begin{bmatrix} 4 & 4 \\ 1 & -3 \end{bmatrix}}_{B} = \underbrace{\begin{bmatrix} -90 & -210 \\ 60 & 140 \end{bmatrix}}_{(10A)B},$$

and

$$\underbrace{\begin{bmatrix} -3 & 3 \\ 2 & -2 \end{bmatrix}}_{A} \left(10 \underbrace{\begin{bmatrix} 4 & 4 \\ 1 & -3 \end{bmatrix}}_{B} \right) = \underbrace{\begin{bmatrix} -3 & 3 \\ 2 & -2 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} 40 & 40 \\ 10 & -30 \end{bmatrix}}_{10B} = \underbrace{\begin{bmatrix} -90 & -210 \\ 60 & 140 \end{bmatrix}}_{A(10B)}.$$

A multiplication rule you have used since primary school on numbers is quite conspicuously missing for matrices. That is, matrix multiplication is not commutative. Firstly, just because AB makes sense, it may be that BA is not even defined. For example, if A is 2×3, and B is 3×4, the we can multiply AB but not BA.

Even if AB and BA are both defined, does not mean that they are equal. For example, take $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$:

$$AB = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix} \qquad \neq \qquad \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = BA.$$

Inverse

A couple of other algebra rules you know for numbers do not quite work on matrices:

(a) $AB = AC$ does not necessarily imply $B = C$, even if A is not 0.

(b) $AB = 0$ does not necessarily mean that $A = 0$ or $B = 0$.

For example:

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 2 \\ 0 & 0 \end{bmatrix}.$$

To make these rules hold, we do not just need one of the matrices to not be zero, we would need to "divide" by a matrix. This is where the *matrix inverse* comes in.

Definition 1. *Suppose that A and B are $n \times n$ matrices such that*

$$AB = I = BA.$$

Then we call B the inverse of A and we denote B by A^{-1}.

If the inverse of A exists, then we say A is invertible. *If A is not invertible, we say A is* singular.

Perhaps not surprisingly, $\left(A^{-1} \right)^{-1} = A$, since if the inverse of A is B, then the inverse of B is A.

If $A = [a]$ is a 1×1 matrix, then A^{-1} is $a^{-1} = \dfrac{1}{a}$. That is where the notation comes from. The computation is not nearly as simple when A is larger.

The proper formulation of the cancellation rule is:

$$\textit{If } A \textit{ is invertible, then } AB = AC \textit{ implies } B = C.$$

The computation is what you would do in regular algebra with numbers, but you have to be careful never to commute matrices:

$$
\begin{aligned}
AB &= AC, \\
A^{-1}AB &= A^{-1}AC, \\
IB &= IC, \\
B &= C.
\end{aligned}
$$

And similarly for cancellation on the right:

$$\textit{If } A \textit{ is invertible, then } BA = CA \textit{ implies } B = C.$$

The rule says, among other things, that the inverse of a matrix is unique if it exists: If $AB = I = AC$, then A is invertible and $B = C$.

We will see later how to compute an inverse of a matrix in general. For now, let us note that there is a simple formula for the inverse of a 2×2 matrix

$$
\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.
$$

For example:

$$
\begin{bmatrix} 1 & 1 \\ 2 & 4 \end{bmatrix}^{-1} = \frac{1}{1 \cdot 4 - 1 \cdot 2} \begin{bmatrix} 4 & -1 \\ -2 & 1 \end{bmatrix} = \begin{bmatrix} 2 & -1/2 \\ -1 & 1/2 \end{bmatrix}.
$$

Let's try it:

$$
\begin{bmatrix} 1 & 1 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} 2 & -1/2 \\ -1 & 1/2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}
\qquad \text{and} \qquad
\begin{bmatrix} 2 & -1/2 \\ -1 & 1/2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.
$$

Just as we cannot divide by every number, not every matrix is invertible. In the case of matrices however we may have singular matrices that are not zero. For example,

$$
\begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}
$$

is a singular matrix. But didn't we just give a formula for an inverse? Let us try it:

$$
\begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}^{-1} = \frac{1}{1 \cdot 2 - 1 \cdot 2} \begin{bmatrix} 2 & -1 \\ -2 & 1 \end{bmatrix} = \ ?
$$

We get into a bit of trouble; we are trying to divide by zero.

So a 2×2 matrix A is invertible whenever

$$ad - bc \neq 0$$

and otherwise it is singular. The expression $ad - bc$ is called the *determinant* and we will look at it more carefully in a later section. There is a similar expression for a square matrix of any size.

Special types of matrices

A simple (and surprisingly useful) type of a square matrix is a so-called *diagonal matrix*. It is a matrix whose entries are all zero except those on the main diagonal from top left to bottom right. For example a 4×4 diagonal matrix is of the form

$$
\begin{bmatrix}
d_1 & 0 & 0 & 0 \\
0 & d_2 & 0 & 0 \\
0 & 0 & d_3 & 0 \\
0 & 0 & 0 & d_4
\end{bmatrix}.
$$

Such matrices have nice properties when we multiply by them. If we multiply them by a vector, they multiply the kth entry by d_k. For example,

$$
\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} = \begin{bmatrix} 1 \cdot 4 \\ 2 \cdot 5 \\ 3 \cdot 6 \end{bmatrix} = \begin{bmatrix} 4 \\ 10 \\ 18 \end{bmatrix}.
$$

Similarly, when they multiply another matrix from the left, they multiply the kth row by d_k. For example,

$$
\begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & -1 \end{bmatrix}
\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}
=
\begin{bmatrix} 2 & 2 & 2 \\ 3 & 3 & 3 \\ -1 & -1 & -1 \end{bmatrix}.
$$

On the other hand, multiplying on the right, they multiply the columns:

$$
\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}
\begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & -1 \end{bmatrix}
=
\begin{bmatrix} 2 & 3 & -1 \\ 2 & 3 & -1 \\ 2 & 3 & -1 \end{bmatrix}.
$$

And it is really easy to multiply two diagonal matrices together:

$$
\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}
\begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & -1 \end{bmatrix}
=
\begin{bmatrix} 1 \cdot 2 & 0 & 0 \\ 0 & 2 \cdot 3 & 0 \\ 0 & 0 & 3 \cdot (-1) \end{bmatrix}
=
\begin{bmatrix} 2 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & -3 \end{bmatrix}.
$$

For this last reason, they are easy to invert, you simply invert each diagonal element:

$$
\begin{bmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{bmatrix}^{-1}
=
\begin{bmatrix} d_1^{-1} & 0 & 0 \\ 0 & d_2^{-1} & 0 \\ 0 & 0 & d_3^{-1} \end{bmatrix}.
$$

Let us check an example

$$
\underbrace{\begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 4 \end{bmatrix}^{-1}}_{A^{-1}}
\underbrace{\begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 4 \end{bmatrix}}_{A}
=
\underbrace{\begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{4} \end{bmatrix}}_{A^{-1}}
\underbrace{\begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 4 \end{bmatrix}}_{A}
=
\underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{I}.
$$

It is no wonder that the way we solve many problems in linear algebra (and in differential equations) is to try to reduce the problem to the case of diagonal matrices.

Another type of matrix that has similarly nice properties are *triangular* matrices. A matrix is *upper triangular* if all of the entries below the diagonal are zero. For a 3×3 matrix, an upper triangular matrix looks like

$$
\begin{bmatrix} * & * & * \\ 0 & * & * \\ 0 & 0 & * \end{bmatrix}
$$

where the $*$ can be any number. Similarly, a *lower triangular* matrix is one where all of the entries above the diagonal are zero, or, for a 3×3 matrix, something that looks like

$$
\begin{bmatrix} * & 0 & 0 \\ * & * & 0 \\ * & * & * \end{bmatrix}.
$$

A matrix that is both upper and lower triangular is diagonal, because only the entries on the diagonal can be non-zero.

Transpose

Vectors do not always have to be column vectors, that is just a convention. Swapping rows and columns is from time to time needed. The operation that swaps rows and columns is the so-called *transpose*. The transpose of A is denoted by A^T. Example:

$$
\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}^T
=
\begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}.
$$

So transpose takes an $m \times n$ matrix to an $n \times m$ matrix.

A key fact about the transpose is that if the product AB makes sense then $B^T A^T$ also makes sense, at least from the point of view of sizes. In fact, we get precisely the transpose of AB. That is:

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

For example,

$$\left(\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 2 & -2 \end{bmatrix} \right)^T = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & -2 \end{bmatrix} \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}.$$

It is left to the reader to verify that computing the matrix product on the left and then transposing is the same as computing the matrix product on the right.

If we have a column vector $\vec{x}$ to which we apply a matrix A and we transpose the result, then the row vector $\vec{x}^T$ applies to A^T from the left:

$$(A\vec{x})^T = \vec{x}^T A^T.$$

Another place where transpose is useful is when we wish to apply the dot product[1] to two column vectors:

$$\vec{x} \cdot \vec{y} = \vec{y}^T \vec{x}.$$

That is the way that one often writes the dot product in software.

We say a matrix A is *symmetric* if $A = A^T$. For example,

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 5 \\ 3 & 5 & 6 \end{bmatrix}$$

is a symmetric matrix. Notice that a symmetric matrix is always square, that is, $n \times n$. Symmetric matrices have many nice properties[2], and come up quite often in applications.

To end the section, we notice how $A\vec{x}$ can be written more succinctly. Suppose

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \qquad \text{and} \qquad \vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$

Then

$$A\vec{x} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 \end{bmatrix}.$$

For example,

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \cdot 2 + 2 \cdot (-1) \\ 3 \cdot 2 + 4 \cdot (-1) \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \end{bmatrix}.$$

In other words, you take a row of the matrix, you multiply them by the entries in your vector, you add things up, and that's the corresponding entry in the resulting vector.

---

[1] As a side note, mathematicians write $\vec{y}^T \vec{x}$ and physicists write $\vec{x}^T \vec{y}$. Shhh... don't tell anyone, but the physicists are probably right on this.

[2] Although so far we have not learned enough about matrices to really appreciate them.