1 Matrix Algebra

A necessary component of understanding linear models is to understand how we scale up from the simplistic slope-intercept formula to a tool capable of ingesting millions of data points and spitting out results. It's impossible to gain this understanding without inevitably discussing the fundamental matrix algebra that allows such a massive leap in computational power; here we begin the discussion of such mathematics.

1.1 Scalars, Vectors, and Matrices

To begin, we discuss the three relevant "objects" of interest to linear algebra: scalars, vectors, and matrices. A *matrix* is best thought of as a grid, or box, of numbers. A *vector* is a matrix with only one column, while a *scalar* is any individual real number. We notate matrices with bold, capital letters; vectors with bold lower case letters; and scalars with unbolded lower case letters.

$$\text{Matrix: } \mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \qquad \text{Vector: } \mathbf{x} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \qquad \text{Scalar: } a = 5$$

To proceed in an efficient manner we must assume that the general methods of scalar algebra are well understood by the reader, (addition, subtraction, multiplication, division, exponentiation, etc.). Of more interest to us is the analogous operations for our "new objects", vectors and matrices.

1.2 Sums and Products

In order to sum two matrices, **A** and **B**, we simply sum the elements in **A** with the elements in **B** that share positions:

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$$

$$\mathbf{A} + \mathbf{B} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 4 & 4 \end{bmatrix}$$

This operation does require that the two matrices are equal in their number of *columns* and *rows*. This idea can be easily extended to vectors as they are single column matrices. As long as two vectors have equal row lengths they can be summed using the same process.

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

$$\text{Columns: } \begin{cases} \text{Column 1: } (a_{11}, a_{21}) \\ \text{Column 2: } (a_{12}, a_{22}) \end{cases} \qquad \text{Rows: } \begin{cases} \text{Row 1: } (a_{11}, a_{12}) \\ \text{Row 2: } (a_{21}, a_{22}) \end{cases}$$

Any matrix or vector can be multiplied by a scalar, resulting in every element being "scaled", (hence the name), by the multiple of that scalar:

$$3 \times \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 3 & 6 \\ 9 & 12 \end{bmatrix} \qquad 10 \times \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 10 \\ 20 \\ 30 \end{bmatrix}$$

To proceed with multiplication of vectors and matrices with one another we'll need to understand an additional operation: Transposes. Given a matrix, $\mathbf{A}$, the *transpose*, $\mathbf{A}'$, is the original matrix with the columns and rows swapped:

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \qquad \mathbf{A}' = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix}$$

An amusing result is that the transpose *of* a transpose is equal to the original matrix:

$$\mathbf{A}' = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix} \qquad (\mathbf{A}')' = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}$$

The reason we introduce this operation is due to a requirement of matrix multiplication: two matrices, $\mathbf{A}$ and $\mathbf{B}$, can only be multiplied in the form of $\mathbf{AB}$ if the columns in $\mathbf{A}$ are of *equal length* to the rows in $\mathbf{B}$. This means that if we wanted to multiply $\mathbf{A}$ by itself our available options would be $\mathbf{AA}'$ or $\mathbf{A}'\mathbf{A}$. The method of multiplying two matrices involves taking the product of row elements in the first matrix and the column elements of the second matrix, then computing their sum. We repeat this process until each row in the first matrix has been matched with each column in the second, multiplied, and summed.

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

$$\begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} = \begin{bmatrix} (1 \cdot 1) + (3 \cdot 3) + (5 \cdot 5) & (1 \cdot 2) + (3 \cdot 4) + (5 \cdot 6) \\ (2 \cdot 1) + (4 \cdot 3) + (6 \cdot 5) & (2 \cdot 2) + (4 \cdot 4) + (6 \cdot 6) \end{bmatrix} = \begin{bmatrix} 35 & 44 \\ 44 & 56 \end{bmatrix}$$

This requirement of "dimensional matching" means that matrix multiplication is not *commutative*, so our standard algebraic methods for scalars won't consistently work. That said, matrix multiplication is *distributive* over addition and subtraction:

$$\mathbf{A}(\mathbf{B} \pm \mathbf{C}) = \mathbf{AB} \pm \mathbf{AC}$$

Matrix multiplication can also be expanded the same way that any college algebra course would show with scalars. Assume matrices $\mathbf{A}$ and $\mathbf{B}$ are both $n \times n$ dimensional. Then $(\mathbf{A}+\mathbf{B})^2$ can be expanded to:

$$(\mathbf{A}+\mathbf{B})^2 = (\mathbf{A}+\mathbf{B})(\mathbf{A}+\mathbf{B}) = \mathbf{AA}+\mathbf{AB}+\mathbf{AB}+\mathbf{BB}$$

The product of two matrices will always contain the row length of the first matrix and the column length of the second matrix. We often notate the dimensions of a matrix as subscripts separated by "$\times$", ordered as row then column:

$$\mathbf{A}_{n\times p}\mathbf{B}_{p\times n} = \mathbf{C}_{n\times n} \qquad \mathbf{B}_{p\times n}\mathbf{A}_{n\times p} = \mathbf{D}_{p\times p}$$

We can clearly see in this case that $\mathbf{AB} \neq \mathbf{BA}$ but if the dimension of each matrix becomes $n \times n$ then $\mathbf{AB} = \mathbf{BA}$.

All of these statements can be trivially extended to the multiplication of vectors with matrices and vice versa. The vector $\mathbf{v}$ can be multiplied with the matrix $\mathbf{M}$ if and only if the length of $\mathbf{v}$ is equal to the row length of $\mathbf{M}$. The opposite direction, $\mathbf{Mv}$ can only be done if the column length of $\mathbf{M}$ is equal to the length of $\mathbf{v}$.

There are a few helpful properties of these operations to keep in mind when solving more complex problems. For convenience we'll observe them as a list, with no particular order.

The transpose of the sum of two matrices is equal to the sum of those matrices tranposed:

$$(\mathbf{A}+\mathbf{B})' = \mathbf{A}'+\mathbf{B}'$$

The transpose of the product of two matrices is equal to the product of those two matrices, transposed, in reverse order:

$$(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$$

If $\mathbf{v}$ is a vector with length n, then $\mathbf{v}'\mathbf{v}$ is the sum of squares and $\mathbf{vv}'$ is a matrix with dimensions $n \times n$:

$$\mathbf{v}'\mathbf{v} = v_1^2 + v_2^2 + ... + v_n^2 \qquad \mathbf{vv}' = \begin{bmatrix} v_1^2 & v_1v_2 & ... & v_1v_n \\ v_2v_1 & v_2^2 & ... & v_2v_n \\ \vdots & \vdots & & \vdots \\ v_nv_1 & v_nv_2 & ... & v_n^2 \end{bmatrix}$$

The square root of $\mathbf{v}'\mathbf{v}$ is the distance from the origin to the point $\mathbf{v}$, known as the length of $\mathbf{v}$:

$$\sqrt{\mathbf{v}'\mathbf{v}} = \sqrt{\sum_{i=1}^{n} v_i^2}$$

1.3 Special Matrices

It's useful to develop some definitions of certain matrices with unique properties that will reoccur throughout discussions of linear models and matrix algebra. One of these such matrices is the *symmetric* matrix. The matrix $\mathbf{A}$ is symmetric is $\mathbf{A}' = \mathbf{A}$:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{bmatrix}' = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{bmatrix}$$

A matrix with equal row and column lengths is described as a*square* matrix; we can easily prove that all symmetric matrices must be square but not all square matrices are symmetric.

Remark 1. If a matrix is symmetric, it must be square.

Proof. Let $\mathbf{A}$ be a matrix with n rows and p columns. Let $\mathbf{B} \equiv \mathbf{A}'$.

Suppose $\mathbf{A}$ is symmetric but not square. Then $\mathbf{A} = \mathbf{B}$ and $n \neq p$ by definition. This implies that $\mathbf{A}_{n\times p} = \mathbf{A}'_{n\times p} = \mathbf{A}_{p\times n}$. Since $n \neq p$ this is a contradiction. Thus $\mathbf{A} = \mathbf{B}$ if and only if $n = p$.

□

Remark 2. If a matrix is square, it is not necessarily symmetric.

Proof. Consider the matrix $\mathbf{C}$ and it's transpose, $\mathbf{C}'$, below:

$$\mathbf{C} = \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{bmatrix} \qquad \mathbf{C}' = \begin{bmatrix} c_{11} & c_{21} & c_{31} \\ c_{12} & c_{22} & c_{32} \\ c_{13} & c_{23} & c_{33} \end{bmatrix}$$

Since $c_{13} \neq c_{31}$, $\mathbf{C}$ can be square without being symmetric. □

A special case of the symmetric matrix is the *diagonal* matrix:

$$\mathbf{D}_1 = \begin{bmatrix} d_{11} & 0 & 0 \\ 0 & d_{22} & 0 \\ 0 & 0 & d_{33} \end{bmatrix} \qquad \mathbf{D}_2 = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 5 \end{bmatrix}$$

With this we can also define the diagonal of any matrix to be the elements where the row and column position of the matrix are equal. We can notate a diagonal matrix with elements equal to another matrix as diag($\mathbf{A}$).

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \qquad \text{diag}(\mathbf{A}) = \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}$$

Another case of the diagonal matrix is the *identity* matrix, notated as $\mathbf{I}$, a diagonal matrix of 1s.

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

In general, the identity matrix acts similarly to 1 in scalar algebra:

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

We close our discussion of special form matrices by defining $\mathbf{j}$, $\mathbf{J}$, $\mathbf{0}$, and $\mathbf{O}$; the vectors and matrices filled with 1s and 0s respectively:

$$\mathbf{j} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \qquad \mathbf{J} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \qquad \mathbf{0} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \qquad \mathbf{O} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

1.4 Linear Dependence

Linear dependence can be best thought of as a synonym for "linear redundancy", which is what we'll build our definition from. Let's define two vectors, $\mathbf{v}$ and $\mathbf{w}$, as two *linearly independent* vectors. On an x, y coordinate plane these two vectors separate from one another, (Figure 1).

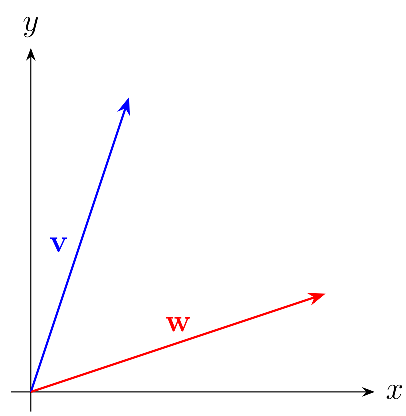


Figure 1: Two vectors, $\mathbf{v}$ and $\mathbf{w}$ in $\mathbb{R}^2$ are linearly independent.

This means that the two vectors span separate "trajectories" of the coordinate plane, which is interesting to both mathematicians and statisticians alike. Now let's consider a third vector, $\mathbf{u}$, which is *linearly dependent* of $\mathbf{v}$. By our definition of "linear redundancy", $\mathbf{u}$ is on the same trajectory as $\mathbf{v}$ so we could express it *using* $\mathbf{v}$ (Figure 2).

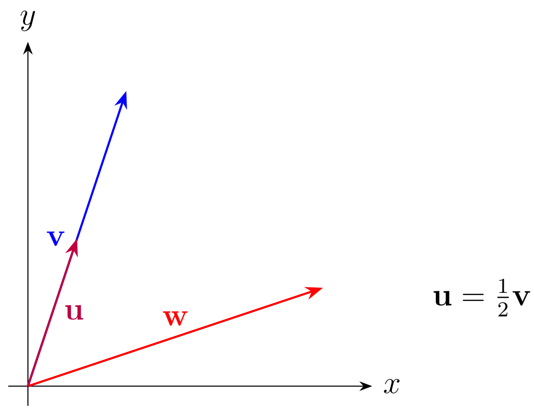


Figure 2: $\mathbf{u}$ is linearly dependent of $\mathbf{v}$, while $\mathbf{v}$ and $\mathbf{w}$ are linearly independent.

While this has a nice interpretation in 2 dimensions, (and even in 3 dimensions with some extra effort), it's not very convenient when considering the dimensions we often see in statistics. For this reason we can instead check the condition of whether a given vector can be expressed as a linear combination of another vector — just as was done in Figure 2 when $\mathbf{u}$ was shown to be $\frac{1}{2}\mathbf{v}$.

By definition, a set of vectors $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v_n}$ are linearly dependent if scalars $s_1, s_2, ..., s_n$ that satisfy $s_i \neq 0$ can be found such that

$$s_1\mathbf{v}_1 + s_2\mathbf{v}_2 + ... + s_n\mathbf{v}_n = \mathbf{0}$$

The methods for finding these scalars won't be discussed in depth here, but a common and popular method is "row reduction" or Gaussian Elimination which has extensive and widely available literature on it's steps.

The *rank* of a matrix is the number of linearly independent columns or rows of that matrix. We'll find that the number of linearly independent columns and rows of any matrix will always be equal — so it doesn't matter which we determine first. By this logic the rank of a matrix can never be greater than the smallest length between the rows and columns. Better put, suppose $\mathbf{A}$ is a matrix with n rows and p columns.

$$\text{rank}(\mathbf{A}) \leq \min(n, p)$$

If the matrix, $\mathbf{A}$, above is of rank p and $p < n$ then $\mathbf{A}$ is said to be *full rank*. That is to say that every column is linearly independent. There are many useful features of rank in matrix algebra, but many won't show themselves until much later. One we can observe now is that, given the matrix $\mathbf{A}$ is full rank, then the expression $\mathbf{AB} = \mathbf{AC}$ implies $\mathbf{B} = \mathbf{C}$. Additionally:

1. If $\mathbf{A}$ and $\mathbf{B}$ can be multiplied, then $\text{rank}(\mathbf{AB}) \leq \text{rank}(\mathbf{A})$ and $\text{rank}(\mathbf{AB}) \leq \text{rank}(\mathbf{B})$.
2. If $\mathbf{B}$ and $\mathbf{C}$ are full-rank square matrices then $\text{rank}(\mathbf{AB}) = \text{rank}(\mathbf{CA}) = \text{rank}(\mathbf{A})$.
3. For any matrix $\mathbf{A}$, $\text{rank}(\mathbf{A}'\mathbf{A}) = \text{rank}(\mathbf{AA}') = \text{rank}(\mathbf{A}') = \text{rank}(\mathbf{A})$.

1.5 Inverse

The inverse of a function is traditionally taught as a reflection of the function on the opposite axis from the original. While this is useful for teaching grade school it makes the discussion of matrix inversions difficult. When we invert a matrix we are better suited thinking about reflecting the plane with which the span of the vectors contained in the matrix rest upon. An intuition of the univariate version of this concept can be seen in Figure 3.

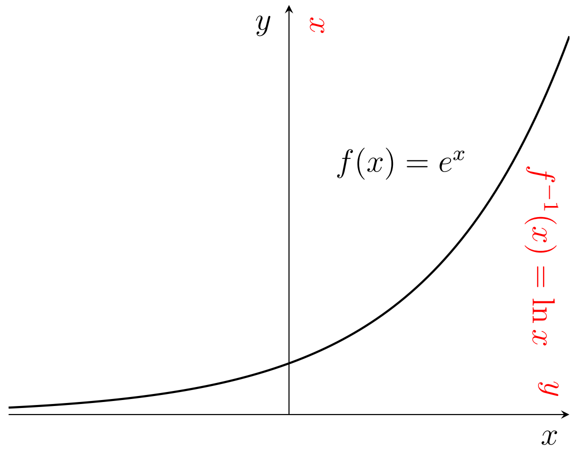


Figure 3: The function $f(x)$ can be inverted by interchanging the x and y axes in the plane, (shown in red), rather than reflecting the function off the y-axis.

We'll need this intuition to appropriately justify the operation of *matrix inversion*. Consider the 2×2 matrix, $\mathbf{A}$, below:

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

A useful interpretation of $\mathbf{A}$ is that it is a set of instructions by which we rotate an x, y coordinate plane $90°$ counterclockwise, shown in Figure 4.

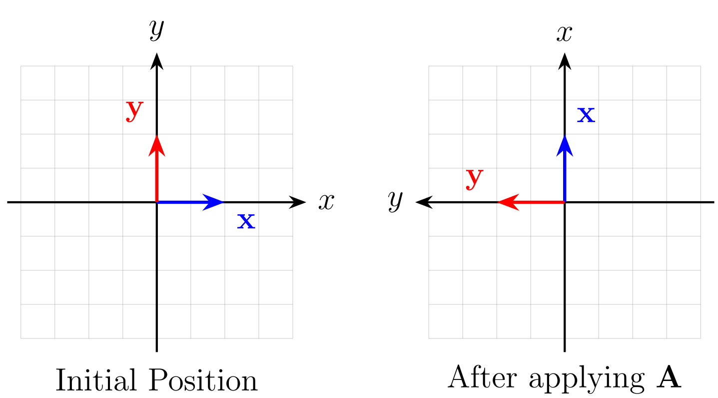


Figure 4: The matrix, $\mathbf{A}$ is a set of instructions to rotate the plane $90°$ counterclockwise, moving the basis vectors $\mathbf{x}$ and $\mathbf{y}$, (commonly notated $\hat{\imath}$ and $\hat{\jmath}$ in linear algebra classrooms).

The inverse of $\mathbf{A}$, notated $\mathbf{A}^{-1}$, is whatever matrix contains the set of instructions to reverse, or antagonize, the instructions of $\mathbf{A}$. In this case, that is the matrix telling us to rotate the plane *clockwise* by 90°, (Figure 5).

$$\mathbf{A} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \qquad \mathbf{A}^{-1} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

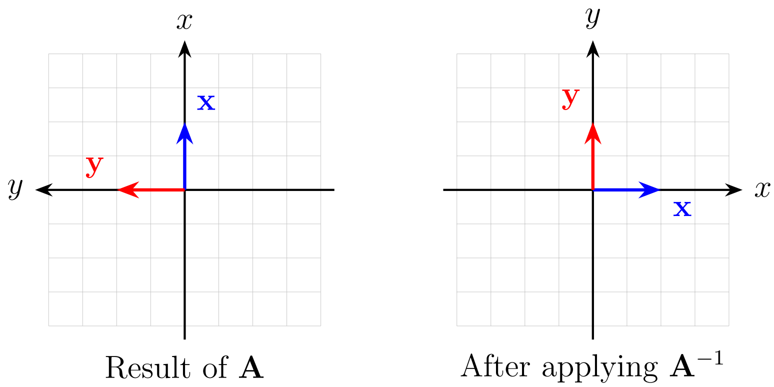


Figure 5: The matrix $\mathbf{A}^{-1}$ informs us to reverse the instructions of $\mathbf{A}$, placing the x and y axes back into their original positions.

This action of "do something then do the opposite" is the logical equivalent of "do nothing". While this has an intuitive enough geometric interpretation to avoid figures, it *also* has a nice interpretation mathematically; the identity matrix.

$$\mathbf{A}\mathbf{A}^{-1} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \mathbf{I}$$

This lends itself to a satisfying interpretation of linear independence, rank, and the fundamental rules behind matrix inervsion. The matrix, $\mathbf{A}$, is only invertible if matrix is a full-rank square matrix, also known as being *nonsingular*.

What this is saying in our new context of matrices is "We can only reverse a set of instructions is the instructions are complete and wholly unique". If the matrix is not full-rank, that means its instructions are redundant in some way. Hence, there's no appropriate way to reverse those instructions. Likewise, if the matrix is rectangular, rather than square, then the instructions are incomplete; the same incompatability with reversing the instructions occurs.

Now we define the inverse of a matrix as follows; the inverse of the matrix, $\mathbf{A}$, is the matrix that satisfies $\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$. The inverse then satisfies the following by nature:

$$(\mathbf{A}^{-1})^{-1} = \mathbf{A}$$

The method of finding inverse matrices is almost exclusively left to computers. These inversions become very cumbersome to perform beyond 4×4 matrices and rely on complex algorithms or creative mathematical simplifications. As such, we'll stick to remarks and theorems about inverses and move along.

1) If $\mathbf{A}$ is nonsingular, the system of equations $\mathbf{Ax} = \mathbf{c}$ has the unique solution:

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{c}.$$

2) If $\mathbf{A}$ is nonsingular, $\mathbf{A}'$ is also nonsingular and the following holds:

$$(\mathbf{A}')^{-1} = (\mathbf{A}^{-1})'.$$

3) If $\mathbf{A}$ and $\mathbf{B}$ are nonsingular matrices of equal dimensions, then $\mathbf{AB}$ is nonsingular and:

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}.$$

1.6 Partitioned Matrices

An exceedingly useful tool for linear model theory is the partitioning of matrices. A matrix can be partitioned into a set of smaller submatrices, not necessarily of equal size, to allow for a reduction in complexity of certain matrix operations. The matrix $\mathbf{A}$ below has been partitioned into four separate submatrices, each notated as matrix "elements" of the larger parent matrix:

$$\mathbf{A} = \left[\begin{array}{cc|cc} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ \hline a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{array}\right] = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}.$$

If two matrices, $\mathbf{A}$ and $\mathbf{B}$ can be multiplied, and are partitioned into submatrices that can be multiplied with their respective submatrix counterparts, the product $\mathbf{AB}$ is computed with the same method as standard matrix multiplication:

$$\mathbf{AB} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{11}\mathbf{B}_{11} + \mathbf{A}_{12}\mathbf{B}_{21} & \mathbf{A}_{11}\mathbf{B}_{11} + \mathbf{A}_{12}\mathbf{B}_{22} \\ \mathbf{A}_{21}\mathbf{B}_{11} + \mathbf{A}_{22}\mathbf{B}_{21} & \mathbf{A}_{21}\mathbf{B}_{11} + \mathbf{A}_{22}\mathbf{B}_{22} \end{bmatrix}.$$

The same logic applies with vectors. We can partition the vector, $\mathbf{b}$ into two subvectors and multiply them with the parition of $\mathbf{A}$ as two submatrices:

$$\mathbf{Ab} = [\mathbf{A}_1, \mathbf{A}_2] \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix} = \mathbf{A}_1\mathbf{b}_1 + \mathbf{A}_2\mathbf{b}_2.$$

A vital result of partitioned matrices is their inverse. Consider, first, the symmetric, nonsingular matrix $\mathbf{A}$. $\mathbf{A}$ is partitioned into four submatrices:

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix},$$

Define $\mathbf{A}_{22\cdot1} = \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}$. The inverse of $\mathbf{A}$ is then:

$$\mathbf{A}^{-1} = \begin{bmatrix} \mathbf{A}_{11}^{-1} + \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{A}_{22\cdot1}^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{A}_{22\cdot1}^{-1} \\ -\mathbf{A}_{22\cdot1}^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{A}_{22\cdot1}^{-1} \end{bmatrix}.$$

The importance of this result is primarily confined to theory; albiet very important theory for establishing an understanding of linear models. A common special case of this result is when the matrix has 0s in the off-diagonal:

$$\begin{bmatrix} \mathbf{A}_{11} & \mathbf{O} \\ \mathbf{O} & \mathbf{A}_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}_{11}^{-1} & \mathbf{O} \\ \mathbf{O} & \mathbf{A}_{22}^{-1} \end{bmatrix}.$$

For the sake of completeness we'll look at one more set of results — however as with everything in these partitioned matrices the results are primarily useful for theory.

If a square, nonsingular matrix $\mathbf{D}$ is defined as $\mathbf{D} = (\mathbf{B} + \mathbf{cc}')$, where $\mathbf{c}$ is a vector and $\mathbf{B}$ is a nonsingular matrix, the inverse of $\mathbf{D}$ is given by:

$$\mathbf{D}^{-1} = \mathbf{B}^{-1} - \frac{\mathbf{B}^{-1}\mathbf{cc}'\mathbf{B}^{-1}}{1 + \mathbf{c}'\mathbf{B}^{-1}\mathbf{c}}.$$

This can be generalized further by defining the matrices $\mathbf{A}$ and $\mathbf{B}$ along with $\mathbf{A} + \mathbf{PBQ}$ as nonsingular, allowing for the following result:

$$(\mathbf{A} + \mathbf{PBQ})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{PB}(\mathbf{B} + \mathbf{BQA}^{-1}\mathbf{PB})^{-1}\mathbf{BQA}^{-1}.$$

1.7 Determinants

A "loose" geometric definition of determinants are scalars that describe the degree with which a linear transformation changes the space occupied by the basis vectors. For this explanation we should think back to our counterclockwise rotation. We'll redefine $\mathbf{A}$ to be a *scaled* version of $\mathbf{A}$, from the inverse example, by a factor of 2.

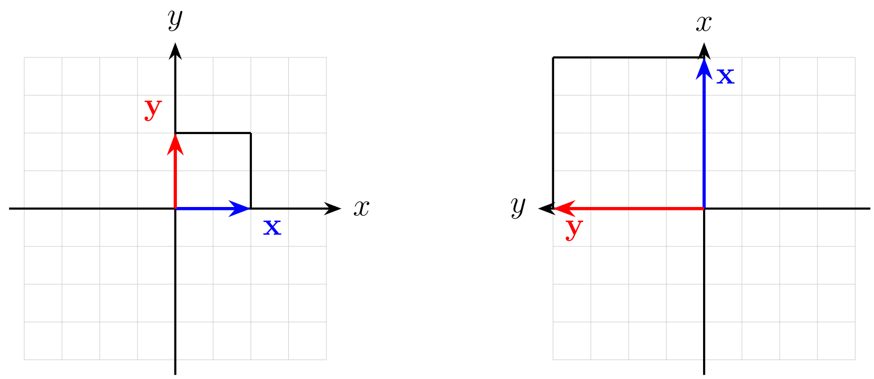
$$\mathbf{A} = \begin{bmatrix} 0 & -2 \\ 2 & 0 \end{bmatrix}$$

The determinant of $\mathbf{A}$ is equal to the subtraction of each pair of diagonal elements:

$$det(\mathbf{A}) = det\left(\begin{bmatrix} 0 & -2 \\ 2 & 0 \end{bmatrix}\right) = ad - bc = (0 \times 0) - (-2 \times 2) = 4$$

This determinant shows a general "degree" or "level" that the linear transformation, $\mathbf{A}$, changes the area covered by the basis vectors (Figure 6).

Like inversions, determinants are commonly left as an excerise for computers after the matrix exceed 3×3. Without needing computation though we can make some broad statements about determinants of special matrices:



Initial Position (Area = 4) After applying $\mathbf{A}$ (Area = 16)

Figure 6: The matrix, $\mathbf{A}$, rotates the plane counterclockwise and expands the space of the original basis vectors by a factor of 2, increasing the area by a factor of 4 — which is the determinant.

(1) For a diagonal matrix, $\mathbf{D} = \text{diag}(d_1, d_2, ...d_n), det(\mathbf{D}) = \prod_{i=1}^{n} d_i$.

(2) For a triangular matrix, the (1) holds.

(3) If $\mathbf{A}$ is singular, $det(\mathbf{A}) = 0$.

(4) If $\mathbf{A}$ is nonsingular, $det(\mathbf{A}) \neq 0$.

(5) If $\mathbf{A}$ is positive definite, $det(\mathbf{A}) > 0$.

(6) $det(\mathbf{A}') = det(\mathbf{A})$.

(7) If $\mathbf{A}$ is nonsingular, $det(\mathbf{A}^{-1}) = \frac{1}{det(\mathbf{A})}$.

Statements (1) and (2) are true by definition. To have a determinant equal to 0 suggests that $\mathbf{Ax} = 0$ holds for a nonzero $\mathbf{x}$. Given this we can prove (3), (4), and (5) rather quickly with a contrapositive:

Proof. Suppose $\mathbf{A}$ is positive definite and $\exists\ \mathbf{x} \neq 0$ such that $\mathbf{Ax} = 0$. Thus $\mathbf{x}'\mathbf{Ax} = 0$.

$\mathbf{A}$ is not positive definite. □

The proofs for (6) and (7) are very direct, but we can also arrive at the proofs geometrically by playing around with the coordinate plane examples in our heads.

The next two remarks are best made by simply showing them:

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

$$det(\mathbf{A}) = det\left(\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}\right) = 2 \times 2 = 4 = 4 \times det(\mathbf{I}) = det(2\mathbf{I}) = \prod_{i=1}^{n} 2 = 2^2 = 4.$$

A diagonal matrix, $\mathbf{D}$, with all elements equal to one another has determinant:

$$det(\mathbf{D}) = det(\text{diag}(c, c, ..., c)) = det(c\mathbf{I}) = \prod_{i=1}^{n} c = c^n.$$

Equivalently, an $n \times n$ matrix multiplied by a scalar, c, has determinant:

$$det(c\mathbf{A}) = c^n det(\mathbf{A}).$$

1.8 Trace

The trace of a matrix is defined as the sum of the diagonal elements of that matrix. Put formally; the trace of and $n \times p$ matrix, $\mathbf{A}$, is:

$$\text{tr}(\mathbf{A}) = \sum_{i=1}^{n} a_{ii}$$

As a result of this definition, our standard summation rules apply with trace. When we consider that the transpose of any matrix *preserves* the position of the diagonal and all quadratic forms of a matrix will *access* the diagonal we arrive at some common sense remarks:

i. $\text{tr}(\mathbf{A} \pm \mathbf{B}) = \text{tr}(\mathbf{A}) \pm \text{tr}(\mathbf{B})$.

ii. $\text{tr}(\mathbf{A}_{n\times p}\mathbf{B}_{p\times n}) = \text{tr}(\mathbf{B}_{p\times n}\mathbf{A}_{n\times p})$.

iii. $\text{tr}(A) = \text{tr}(A')$.

Given $\mathbf{A}$ is $n \times p$ and a_i is the i^{th} column of $\mathbf{A}$:

iv. $\text{tr}(\mathbf{A}'\mathbf{A}) = \sum_{i=1}^{n} \mathbf{a}_i'\mathbf{a}_i$.

Given $\mathbf{A}$ is $n \times p$ and a_i' is the i^{th} row of $\mathbf{A}$:

v. $\text{tr}(\mathbf{A}\mathbf{A}') = \sum_{i=1}^{n} \mathbf{a}_i\mathbf{a}_i'$.

vi. For any $n \times n$ matrix $\mathbf{A}$ and nonsingular $n \times n$ matrix $\mathbf{P}$, $\text{tr}(\mathbf{P}^{-1}\mathbf{A}\mathbf{P})$.

While trace is a straight forward operation it's a very powerful one that continues to make an appearance throughout the theory of linear models. Many of these remarks are important components of being able to manipulate trace into a form useful for simplifying otherwise monolithic proofs.

1.9 Eigenvalues and Eigenvectors

Consider the marix, $\mathbf{A}$, below:

$$\mathbf{A} = \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix}$$

This matrix corresponds to the movement of the coordinate plane so that the x and y axis are placed onto the trajectory of the vectors $[3,0]'$ and $[1,2]'$ and scaled appropriately. This is a difficult intuition in the context of typical flat coordinate grids, but in Figure 7 we can see the true geometry is actually quite simple.

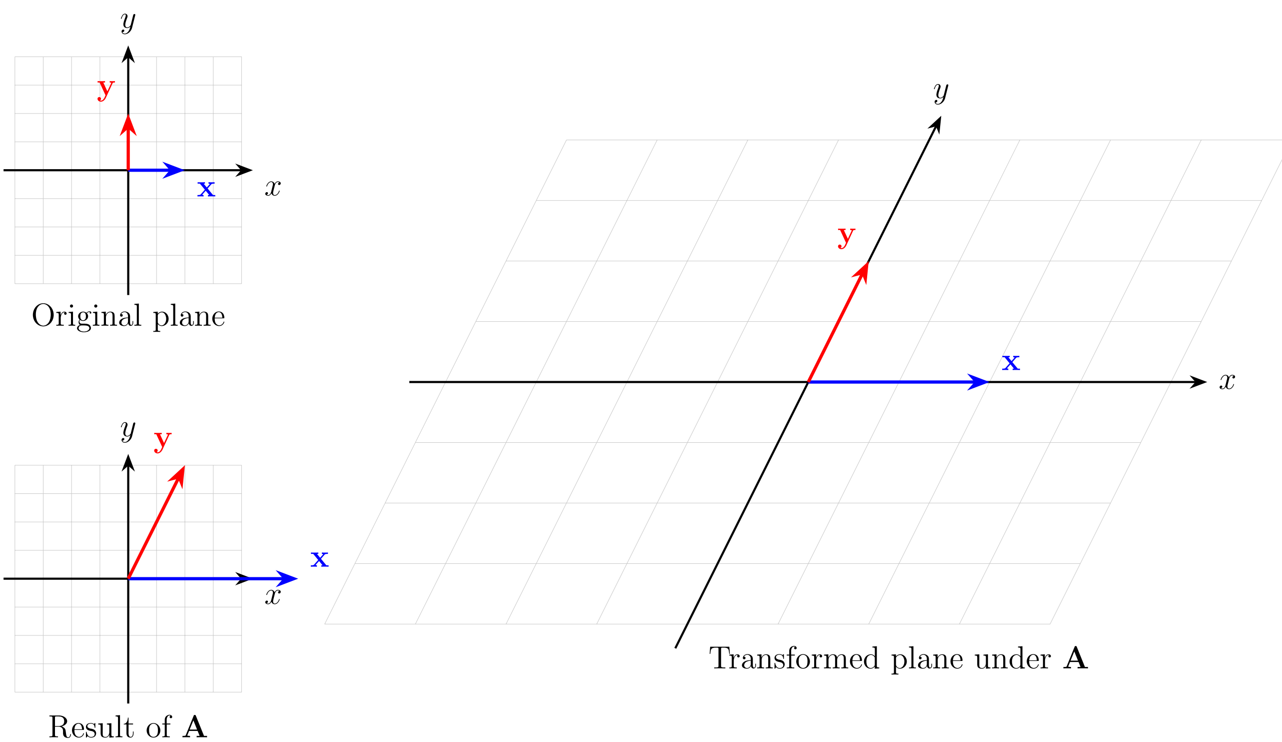


Figure 7: The standard flat coordinate plane shows how the basis vectors $\mathbf{x}$ and $\mathbf{y}$ are moved from their original positions when the transformation, $\mathbf{A}$, is applied. The reality is that $\mathbf{A}$ is stretching the plane so that $\mathbf{x}$ is scaled and $\mathbf{y}$ is *knocked off* its span entirely.

We'll need geometric intuition once again in order to fight our toughest conceptual battle, *eigenvectors*. Eigenvectors are those vectors, like $\mathbf{x}$, that aren't knocked off of their span when a specific transformation is applied. In Figure 8 we can see this as a complete picture, where the vectors $\mathbf{v}_1$ and $\mathbf{v_2}$ represent the totality of the possible eigenvectors of this matrix. Traditionally these vectors would be shown as basis vectors with length 1 being *scaled* based off of the transformation. The degree with which that scaling occurs for each vector is the associated *eigenvalue*.

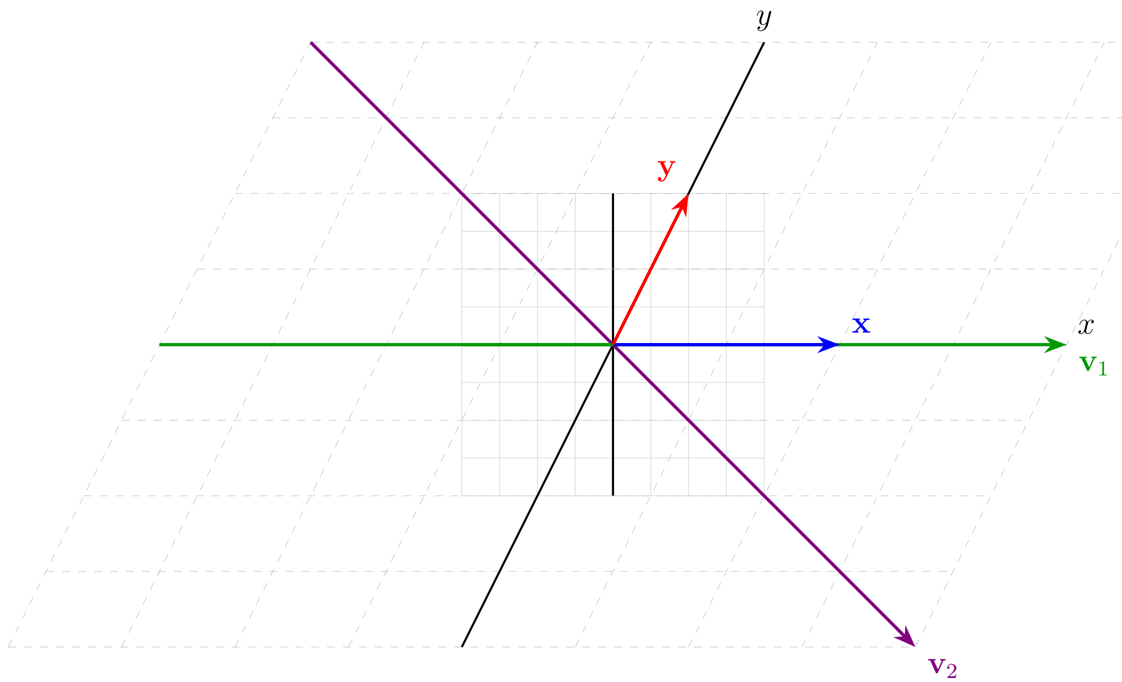


Figure 8: The eigenvectors of $\mathbf{A}$ are the vectors, $\mathbf{v}_1$ and $\mathbf{v}_2$, who's trajectories aren't changed by the transformation.

Figure 8 doesn't show this scaling concept for the sake of image clarity, but if we were to properly measure the scaling of the two eigenvectors we would find the eigenvalues $\lambda_1 = 3$ and $\lambda_2 = 2$ respectively.

Eigenvalues and eigenvectors are quite possibly the most powerful tool that linear algebra has to offer. They act like a form of 'matrix DNA' by holding the fast majority of the information a matrix contains within them. In fact, with only the eigenvalues and eigenvectors one can reliably *reconstruct* the entire matrix they originated from.

For an $n \times n$ matrix, $\mathbf{A}$, the eigenvalues, λ, can be found by solving the characteristic equation:

$$det(\mathbf{A} - \lambda \mathbf{I}) = 0.$$

In practice, this becomes a complex task beyond matrices of size 3×3. For 2×2 matrices a simpler algorithm can isolate the eigenvalues. Consider the 2×2 matrix, $\mathbf{A}$:

$$\mathbf{A} = \begin{bmatrix} 3 & 5 \\ 8 & 2 \end{bmatrix}$$

Let μ be the mean of the diagonal elements of $\mathbf{A}$. Let p be the product of the diagonal elements of $\mathbf{A}$. The eigenvalues of $\mathbf{A}$ are found via:

$$\lambda = \mu \pm \sqrt{\mu^2 - p} = 5 \pm \sqrt{5^2 - 10} = 5 \pm \sqrt{15}$$