Transpose

Reading: Strang 2.7

Learning objective: Be able to take the transpose of a matrix and identify symmetric matrices. Understand the relation of transpose to dot products.

Transpose

The transpose of a m-by-n matrix A is a n-by-m matrix denoted A^T and defined as

$$A^T(i,j) = A(j,i)$$

A^T is the matrix A reflected about its diagonal.

$$A = \begin{bmatrix} 1 & 5 & -2 \\ 6 & -4 & 8 \\ 7 & 3 & -1 \end{bmatrix} \qquad A^T = \begin{bmatrix} 1 & 6 & 7 \\ 5 & -4 & 3 \\ -2 & 8 & -1 \end{bmatrix}$$

The columns of A become the rows of A^T.

More Examples

The transpose of a column vector becomes a row vector.

$$\vec{v} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \qquad \vec{v}^T = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$$

The transpose of a 3-by-2 matrix is a 2-by-3 matrix.

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

Question

What is $(A^T)^T$?

Symmetric Matrix

A matrix A satisfying $A = A^T$ is called symmetric.

The "Pascal Matrix" from the LU decomposition example is a symmetric matrix.

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 1 & 3 & 6 & 10 \\ 1 & 4 & 10 & 20 \end{bmatrix}$$

Question

Say that D is a diagonal matrix:

$$D(i, j) = 0 \text{ whenever } i \neq j$$

Is D symmetric?

Transpose and Addition

Now let's look at how the transpose interacts with matrix operations.

First, how the transpose affects a sum of two matrices:

$$(A + B)^T = A^T + B^T$$

Transpose and Product

More interesting is how transpose affects a product.

$$\boxed{(AB)^T = B^T A^T}$$

This formula is reminiscent of the one for the inverse of a product: we write the product in reverse order and take the transpose of each term.

This formula connects the row and column pictures of matrix multiplication.

Transpose and Product

We prove $(AB)^T = B^T A^T$ by starting out with the simpler case where B has a single column:

$$(A\vec{b})^T = \vec{b}^T A^T$$

The left hand side is a linear combination of the columns of A, transposed into a row vector.

The right hand side is a linear combination of the rows of A^T, which are the columns of A written as row vectors.

$$(A\vec{b})^T = \vec{b}^T A^T$$

Say that A is a m-by-n matrix and $\vec{b}$ is an n-dimensional column vector.

$$A\vec{b} = b_1 \begin{bmatrix} \vdots \\ A(:,1) \\ \vdots \end{bmatrix} + b_2 \begin{bmatrix} \vdots \\ A(:,2) \\ \vdots \end{bmatrix} + \cdots + b_n \begin{bmatrix} \vdots \\ A(:,n) \\ \vdots \end{bmatrix}$$

Using the rule for the transpose of a sum:

$$(A\vec{b})^T = b_1 \begin{bmatrix} \vdots \\ A(:,1) \\ \vdots \end{bmatrix}^T + b_2 \begin{bmatrix} \vdots \\ A(:,2) \\ \vdots \end{bmatrix}^T + \cdots + b_n \begin{bmatrix} \vdots \\ A(:,n) \\ \vdots \end{bmatrix}^T$$

$$(A\vec{b})^T = b_1 \begin{bmatrix} \vdots \\ A(:,1) \\ \vdots \end{bmatrix}^T + b_2 \begin{bmatrix} \vdots \\ A(:,2) \\ \vdots \end{bmatrix}^T + \cdots + b_n \begin{bmatrix} \vdots \\ A(:,n) \\ \vdots \end{bmatrix}^T$$

Now note that $A(:,1)^T = A^T(1,:)$.

When transposed, the first column of A becomes the first row of A^T.

This gives

$$\begin{aligned} (A\vec{b})^T &= b_1 \left[\cdots A^T(1,:) \cdots\right] + \cdots + b_n \left[\cdots A^T(n,:) \cdots\right] \\ &= b^T A^T. \end{aligned}$$

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

Now we are ready to prove this.

What is the i^{th} row on the left hand side?

$$\begin{aligned} (AB(:,i))^T &= B(:,i)^T A^T \\ &= B^T(i,:) A^T \end{aligned}$$

$$(A\vec{b})^T = \vec{b}^T A^T$$

$$B(:,i)^T = B^T(i,:)$$

The last expression is the i^{th} row of the right hand side.

Transpose and Inverse

The last result also implies the following:

If A is invertible then A^T is as well and

$$(A^T)^{-1} = (A^{-1})^T$$

Proof: Since the identity matrix is symmetric

$$\begin{aligned} I = I^T &= (AA^{-1})^T \\ &= (A^{-1})^T A^T \end{aligned}$$

Thus A^T has a left inverse and is invertible.

Dot Product

We can use the transpose to express the dot product.

Let $\vec{u} = (1, 4, 9), \vec{v} = (1, 1, 1)$ be two column vectors.

Then

$$\langle \vec{u}, \vec{v} \rangle = \vec{u}^T \vec{v} = \begin{bmatrix} 1 & 4 & 9 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

Dot Product

The dot product is where the transpose finds its mathematical meaning.

Let A be an n-by-n matrix.

The transpose gives the answer to the question:

For what matrix B does

$$\langle \vec{u}, A\vec{v} \rangle = \langle B\vec{u}, \vec{v} \rangle,$$

for all vectors $\vec{u}, \vec{v}$?

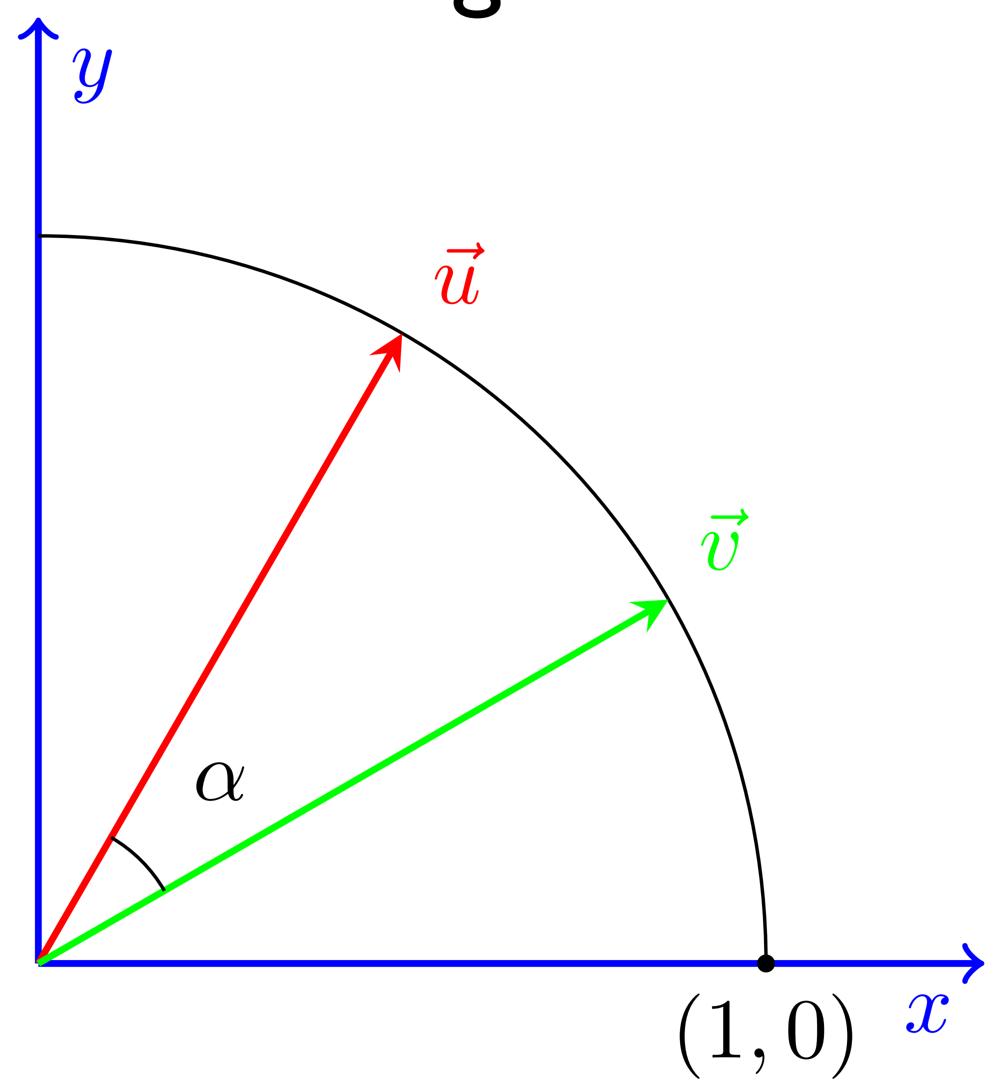
Example

Let A be the matrix that rotates 2D vectors counter-clockwise by θ.

$$A = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

Say that $\vec{u}, \vec{v}$ are unit vectors that make an angle α (as in the picture).

$$\langle \vec{u}, \vec{v} \rangle = \cos(\alpha)$$



Example

Let A be the matrix that rotates 2D vectors counter-clockwise by θ.

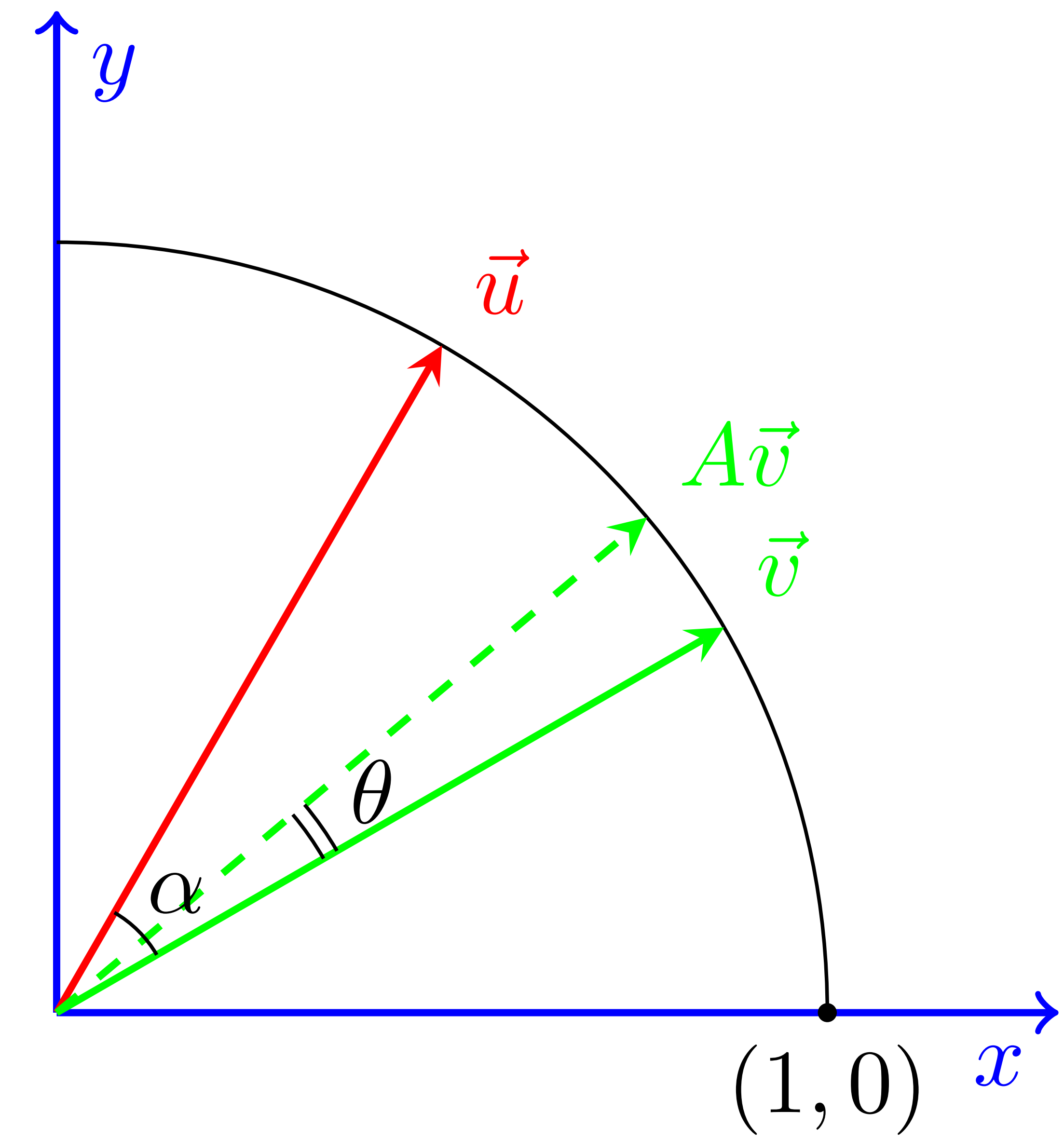
$$A = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

The angle between $\vec{u}$ and $A\vec{v}$ is $\alpha - \theta$.

$$\langle \vec{u}, A\vec{v} \rangle = \cos(\alpha - \theta)$$

What should B do so that

$$\langle B\vec{u}, \vec{v} \rangle = \cos(\alpha - \theta) \quad ?$$



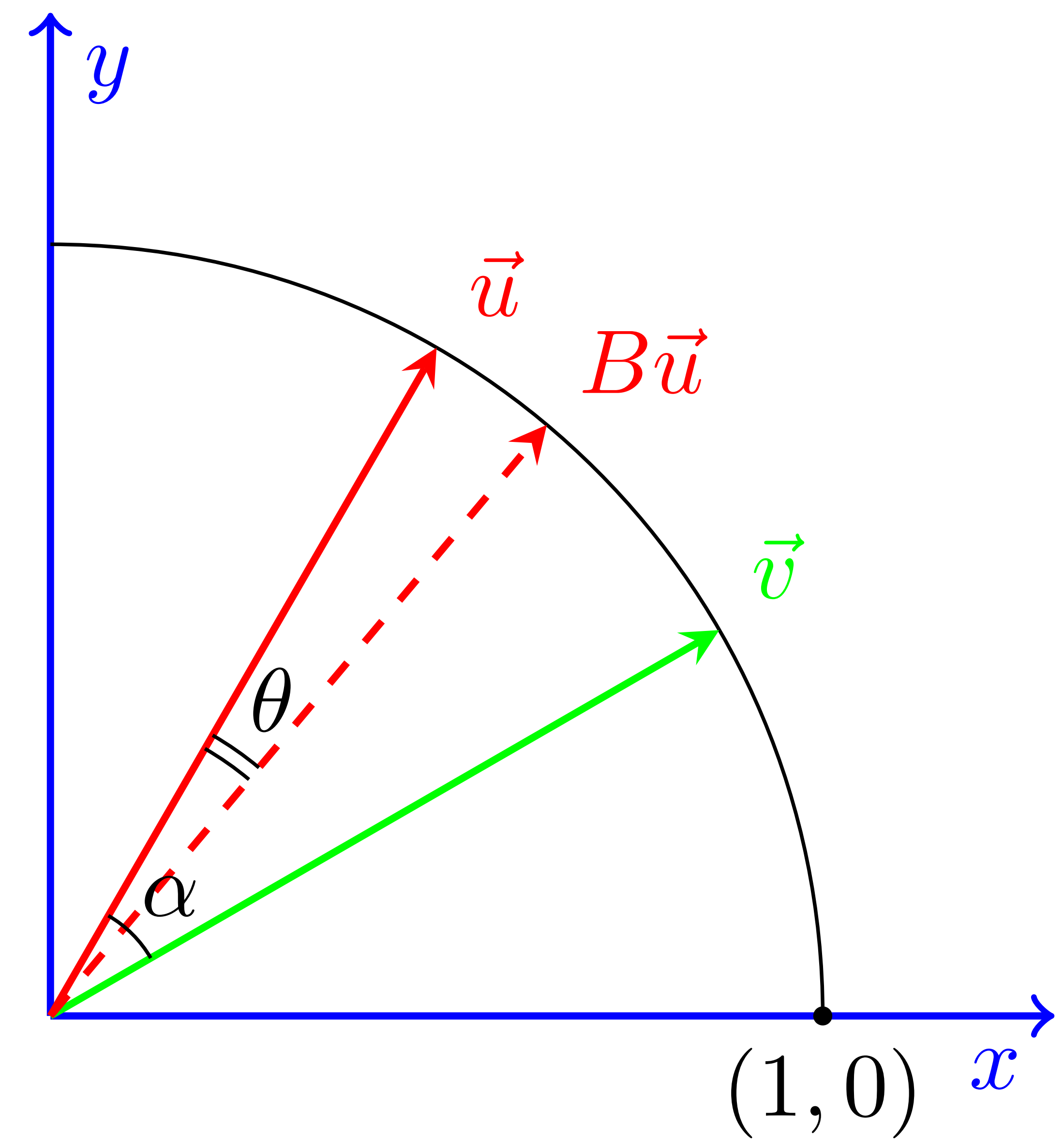
Example

Let A be the matrix that rotates 2D vectors counter-clockwise by θ.

$$A = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

What should B do so that

$$\langle B\vec{u}, \vec{v} \rangle = \cos(\alpha - \theta) \quad ?$$



It should be a clockwise rotation by θ.

Example

Let A be the matrix that rotates 2D vectors counter-clockwise by θ.

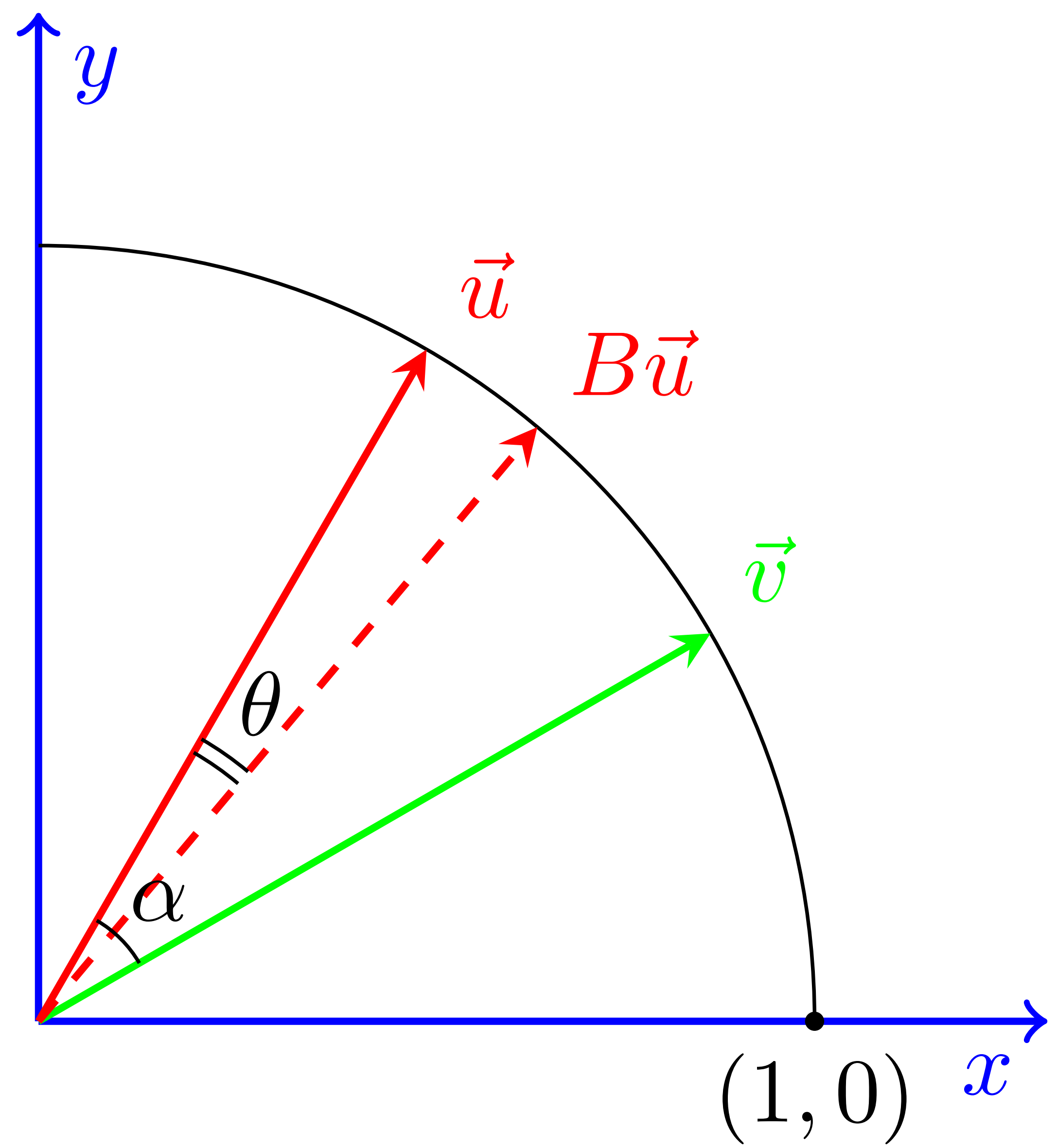
$$A = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

B is a clockwise rotation by θ.

This is done by the matrix

$$A^T = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix}$$

$$\langle A^T\vec{u}, \vec{v}\rangle = \langle \vec{u}, A\vec{v}\rangle$$



Dot Product

This is a general result:

$$\boxed{\langle A^T \vec{u}, \vec{v} \rangle = \langle \vec{u}, A\vec{v} \rangle}$$

Reason:

$$\langle A^T \vec{u}, \vec{v} \rangle = (A^T \vec{u})^T \vec{v}$$

$$= \vec{u}^T (A^T)^T \vec{v} \qquad (B\vec{u})^T = \vec{u}^T B^T$$

$$= \vec{u}^T A \vec{v} \qquad A = (A^T)^T$$

Permutation Matrices

Reading: Strang 2.7

Learning objective: Understand the action of a permutation matrix by left and right multiplication. Understand the relationship between the inverse and transpose of a permutation matrix.

Permutation Matrix

A permutation matrix is a square matrix that has all entries equal to 0 or 1.

All columns have exactly one entry equal to 1.

All rows have exactly one entry equal to 1.

Any permutation matrix can be formed by rearranging the rows of the identity matrix.

Example:

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

Action of a Permutation Matrix

When multiplied on the left, a permutation matrix rearranges rows.

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \cdots A(1,:) \cdots \\ \cdots A(2,:) \cdots \\ \cdots A(3,:) \cdots \\ \cdots A(4,:) \cdots \end{bmatrix} = \begin{bmatrix} \cdots A(2,:) \cdots \\ \cdots A(3,:) \cdots \\ \cdots A(4,:) \cdots \\ \cdots A(1,:) \cdots \end{bmatrix}$$

Think of this using the row picture of matrix multiplication.

Action of a Permutation Matrix

When multiplied on the right, a permutation matrix rearranges columns.

$$\begin{bmatrix} \vdots & \vdots & \vdots & \vdots \\ A(:,1) & A(:,2) & A(:,3) & A(:,4) \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} \vdots & \vdots & \vdots & \vdots \\ A(:,4) & A(:,1) & A(:,2) & A(:,3) \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

Think of this using the column picture of matrix multiplication.

Row Swaps

A row swap matrix is a permutation matrix.

Moreover, row swap matrices are the building blocks of permutation matrices.

Every permutation matrix can be written as a product of row swap matrices.

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} R_1 \leftrightarrow R_4 \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} R_1 \leftrightarrow R_3 \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} R_1 \leftrightarrow R_2 \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

Row Swaps

Row swap matrices are symmetric.

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

This matrix swaps rows one and three

The elementary matrix R that swaps rows i and j has:

$$R(i, j) = 1, R(j, i) = 1$$

$$R(k, k) = 1 \text{ for all } 1 \leq k \leq n, k \neq i, j$$

All other entries are zero.

Inverses of Permutation Matrices

Recall also that $RR = I$ for a row swap matrix R.

Row swap matrices can let us figure out the inverse of a permutation matrix.

Let P be a permutation matrix and write it as a product of row swap matrices

$$P = R_k R_{k-1} \cdots R_2 R_1$$

What is its inverse?

Inverses of Permutation Matrices

$$P = R_k R_{k-1} \cdots R_2 R_1$$

What is its inverse?

$$Q = R_1 R_2 \cdots R_{k-1} R_k$$

$$\begin{aligned} QP &= (R_1 R_2 \cdots R_{k-1} R_k)(R_k R_{k-1} \cdots R_2 R_1) \\ &= R_1 R_2 \cdots R_{k-1} R_{k-1} \cdots R_2 R_1 \\ &\vdots \\ &= R_1 R_1 \\ &= I \end{aligned}$$

Inverses of Permutation Matrices

$$P = R_k R_{k-1} \cdots R_2 R_1$$

What is its inverse?

$$Q = R_1 R_2 \cdots R_{k-1} R_k$$

The inverse is $Q = P^T$.

$$P^T = R_1^T R_2^T \cdots R_{k-1}^T R_k^T$$

$$= R_1 R_2 \cdots R_{k-1} R_k \qquad R_i^T = R_i$$

Permutation Matrices and LU Decomposition

Reading: Strang 2.7

Learning objective: See that the LU decomposition of a matrix does not always exist. Understand why the more general $PA = LU$ decomposition does always exist.

When does LU exist?

We have seen the usefulness of the LU decomposition. When can we find one?

Not always! Consider

one must be zero

$$\begin{bmatrix} 0 & 1 \\ 4 & 5 \end{bmatrix} = \begin{bmatrix} * & 0 \\ * & * \end{bmatrix} \begin{bmatrix} * & * \\ 0 & * \end{bmatrix} = \begin{bmatrix} * & * \\ * & * \end{bmatrix}$$

one must be zero

What goes wrong?

We found the LU decomposition by Gaussian elimination that involved no row swaps.

$$\begin{bmatrix} 0 & 1 \\ 4 & 5 \end{bmatrix}$$

What is first step of Gaussian elimination here?

Row swap!

$$\begin{bmatrix} 4 & 5 \\ 0 & 1 \end{bmatrix}$$

Now we are already upper triangular.

Good Ordering

We can generalize the LU decomposition to a form that exists for every n-by-n matrix.

This is the decomposition $PA = LU$ for a permutation matrix P.

We permute the rows of A first, so that Gaussian elimination on PA involves no row swaps.

Then PA will have an LU decomposition.

Determinants

Reading: Strang 5.1

Learning objective: Be able to interpret the defining properties of the determinant.

Determinant

The determinant is a function that takes as input a square matrix and returns a real number.

This is a single number summary of the matrix.

The determinant of a matrix is nonzero if and only if the matrix is invertible.

Determinants can also be used to solve systems of linear equations and find inverses (although this is not very practical for large systems).

Two-by-Two Case

Suppose we wanted to design a function of two-by-two matrices that is nonzero iff the matrix is invertible.

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

How could we do it?

Well, the usual way we determine if a matrix is invertible is through Gaussian elimination…

Two-by-Two Case

Case 1: $a \neq 0$ $\begin{bmatrix} a & b \\ c & d \end{bmatrix} \xrightarrow[R_2 = R_2 - (c/a)R_1]{} \begin{bmatrix} a & b \\ 0 & d - bc/a \end{bmatrix}$

Invertible iff $a \neq 0, d - bc/a \neq 0$.

Equivalently, $a \neq 0, ad - bc \neq 0$.

Case 2: $a = 0$ $\begin{bmatrix} 0 & b \\ c & d \end{bmatrix} \xrightarrow[R_1 \leftrightarrow R_2]{} \begin{bmatrix} c & d \\ 0 & b \end{bmatrix}$

Invertible iff $c \neq 0, b \neq 0$.

Equivalently, $bc \neq 0$. Or, as $a = 0$, equivalently $ad - bc \neq 0$.

Two-by-Two Case

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

Conclusion: This matrix is invertible iff $ad - bc \neq 0$.

The function $ad - bc$ does what we are looking for.

This is the determinant of a two-by-two matrix.

Notation

We will denote the determinant by replacing the brackets of a matrix by bars.

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

matrix

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

determinant
(a number)

We will also write $\det(A)$ for the determinant of A.

Notation

There is one other notation we will use for determinant.

We will think about the determinant as a function of the rows of a matrix.

To emphasize this, sometimes we write out the rows:

$$\det(A) = \det(A(1,:), A(2,:), \ldots, A(n,:))$$

Definition

In general, we are going to define the determinant function by setting out properties we want it to satisfy.

We will see how these properties lead (via Gaussian elimination) to the main feature of the determinant: it is nonzero iff a matrix is invertible.

Finally, we will see some actual formulas for the determinant (as in the two-by-two case) and that they satisfy the required properties.

Defining Properties

1) The determinant of the n-by-n identity matrix is 1.

We know the identity matrix is invertible, so its determinant should be nonzero.

This property holds for our two-by-two formula.

$$\begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1$$

Defining Properties

2) The determinant changes sign when two rows are exchanged.

$$\det(A(1,:),\ldots,A(i,:),\ldots,A(j,:),\ldots,A(n,:)) =$$
$$-\det(A(1,:),\ldots,A(j,:),\ldots,A(i,:),\ldots,A(n,:))$$

This property holds for our two-by-two formula.

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = -\begin{vmatrix} c & d \\ a & b \end{vmatrix}$$

Defining Properties

3) The determinant is a linear function of each row separately.

Let's spend some time to make sense of this.

We talked about linear functions before. They satisfy two properties:

- $f(t \cdot \vec{u}) = t f(\vec{u})$
- $f(\vec{u} + \vec{v}) = f(\vec{u}) + f(\vec{v})$

Defining Properties

3) The determinant is a linear function of each row separately.

This is saying that if we fix all the rows of a matrix but one, and view the determinant as a function of that single row, it is a linear function.

- $f(t \cdot \vec{u}) = t f(\vec{u})$
- $f(\vec{u} + \vec{v}) = f(\vec{u}) + f(\vec{v})$

Defining Properties

3) The determinant is a linear function of each row separately.

Let's see what this means in terms of our two-by-two formula.

Say we multiply the first row by t and leave the second row alone. What happens to the determinant?

$$\begin{vmatrix} ta & tb \\ c & d \end{vmatrix} = tad - tbc = t(ad - bc)$$

The determinant is multiplied by t .

Defining Properties

3) The determinant is a linear function of each row separately.

The same thing happens if we leave the first row alone and multiply the second row by t.

$$\begin{vmatrix} a & b \\ tc & td \end{vmatrix} = t(ad - bc)$$

Defining Properties

3) The determinant is a linear function of each row separately.

Note that we are only changing one row at a time here. The determinant is not a linear function of the matrix itself.

$$\begin{vmatrix} ta & tb \\ tc & td \end{vmatrix} = t^2(ad - bc)$$

Defining Properties

3) The determinant is a linear function of each row separately.

Say we add a vector to the first row and leave the second row alone. What happens to the determinant?

$$\begin{vmatrix} a + a' & b + b' \\ c & d \end{vmatrix} = (a + a')d - (b + b')c = ad - bc + a'd - b'c$$

$$= \begin{vmatrix} a & b \\ c & d \end{vmatrix} + \begin{vmatrix} a' & b' \\ c & d \end{vmatrix}$$

Defining Properties

3) The determinant is a linear function of each row separately.

Similarly, if we add a vector to the second row and leave the first row alone:

$$\begin{vmatrix} a & b \\ c+c' & d+d' \end{vmatrix} = a(d+d') - b(c+c')$$

$$= \begin{vmatrix} a & b \\ c & d \end{vmatrix} + \begin{vmatrix} a & b \\ c' & d' \end{vmatrix}$$

Defining Properties

3) The determinant is a linear function of each row separately.

In general, this property is saying

§ $\det(A(1,:),\ldots,t\cdot A(i,:),\ldots,A(n,:)) = t\cdot\det(A(1,:),\ldots,A(i,:),\ldots,A(n,:))$

§ $\det(A(1,:),\ldots,A(i,:)+\vec{b}^T,\ldots,A(n,:)) = \det(A(1,:),\ldots,A(i,:),\ldots,A(n,:)) + \det(A(1,:),\ldots,\vec{b}^T,\ldots,A(n,:))$

Multilinear function

3) The determinant is a linear function of each row separately.

Because of this property we say the determinant is a multilinear function.

Questions

When we take this approach of defining a function by its properties, there are a few questions we should ask.

1) Is there a function satisfying the 3 properties?

   In the 2-by-2 case, we have seen a formula that satisfies the properties, but what about in general?

2) Is there more than one function satisfying the 3 properties?

   If we want to talk about the determinant function, there should be a unique function satisfying the properties.

Questions

1) Is there a function satisfying the 3 properties?

2) Is there more than one function satisfying the 3 properties?

I'll go ahead and say the answer: there is a unique function satisfying the 3 properties.

Thus it makes sense to call this function the determinant.

We'll only be able to prove this later when we derive formulas for the determinant.