Math 510 **Fall 2022**
Notes for Homework 2 **Drew Armstrong**

1. Matrix Multiplication

In the last section we talked about individual vector spaces such as $\mathbb{R}^n$ and $L^2[0,1]$. Each of these has an inner product, hence it also has a vector norm and a metric. Now we discuss linear functions between different vector spaces. In the finite dimensional case we can encode such functions as matrices. But matrix arithmetic does more than just encode linear functions; it is an extremely powerful language that gives out much more than we put in.

I assume you already know the definition of matrix multiplication. Here is a reminder.

Definition of Matrix Multiplication. Consider two matrices

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & & \vdots \\ a_{\ell 1} & \cdots & a_{\ell m} \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & & \vdots \\ b_{m1} & \cdots & b_{mn} \end{pmatrix}.$$

We say that A has shape $\ell \times m$ and B has shape $m \times n$. (The number of rows comes first.) Since the number of columns of A equals the number of rows of B (they both equal m), we can define the product matrix AB, which has shape $\ell \times n$:

$$AB = \begin{pmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & & \vdots \\ c_{\ell 1} & \cdots & c_{\ell n} \end{pmatrix}.$$

The entries of A, B and AB are related as follows:

$$\boxed{c_{ij} = \sum_{k=1}^{m} a_{ik} b_{kj}.}$$

I could have postponed this gory definition until it emerged naturally from the theory. But, as I said, the mechanics of matrix arithmetic is more than the sum of its parts, so I wanted to explore the mechanics first.

Row Times Column = Dot Product. Suppose that $\ell = n = 1$, so that

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1m} \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} b_{11} \\ \vdots \\ b_{m1} \end{pmatrix}.$$

Then the matrix product AB has shape 1×1 (it is just a scalar) and corresponds to the dot product of vectors:

$$\begin{pmatrix} a_{11} & \cdots & a_{1m} \end{pmatrix} \begin{pmatrix} b_{11} \\ \vdots \\ b_{m1} \end{pmatrix} = a_{11}b_{11} + a_{12}b_{21} + \cdots + a_{1m}b_{m1}.$$

From now on we will identify vectors $\mathbf{v} = (v_1, \ldots, v_n) \in \mathbb{R}^n$ with **column vectors**:

$$\mathbf{v} = (v_1, \ldots, v_n) = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}.$$

To talk about **row vectors** we will use the operation of *transposition*:

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & & \vdots \\ a_{\ell 1} & \cdots & a_{\ell m} \end{pmatrix} \quad \rightsquigarrow \quad A^T := \begin{pmatrix} a_{11} & \cdots & a_{\ell 1} \\ \vdots & & \vdots \\ a_{1m} & \cdots & a_{\ell m} \end{pmatrix}.$$

Thus the transpose of a column vector is a row vector:

$$\mathbf{v}^T = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}^T = \begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix}.$$

Finally, we can express the dot product of any two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ in terms of matrix multiplication:

$$\mathbf{u}^T \mathbf{v} = \begin{pmatrix} u_1 & \cdots & u_n \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} u_1 v_1 + \cdots + u_n v_n = \mathbf{u} \bullet \mathbf{v}.$$

Column Times Row = Something Else. Warning. A column times a row is not a scalar; it is a matrix of any shape that we want. That is, for any $\mathbf{u} \in \mathbb{R}^m$ and $\mathbf{v} \in \mathbb{R}^n$ we obtain an $m \times n$ matrix[1]

$$\mathbf{u}\mathbf{v}^T = \begin{pmatrix} u_1 \\ \vdots \\ u_m \end{pmatrix} \begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix} = \begin{pmatrix} u_1 v_1 & \cdots & u_1 v_n \\ \vdots & & \vdots \\ u_m v_1 & \cdots & u_m v_n \end{pmatrix}.$$

Row times column and column times row are the two basic examples. In between there are many different ways to think about matrix multiplication. For example:

$$\begin{aligned} (ij \text{ entry of } AB) &= (i\text{th row of } A)(j\text{th col of } B), \\ (i\text{th row of } AB) &= (i\text{th row of } A)B, \\ (j\text{th col of } AB) &= A(j\text{th col of } B). \end{aligned}$$

If A has shape $\ell \times m$ and B has shape $m \times n$ then we also have

$$AB = \sum_{k=1}^{m} (k\text{th col of } A)(k\text{th row of } B),$$

where the right hand side is a sum of m matrices, each of shape $\ell \times n$.

All of these rules are examples of a very general recursive property of matrix multiplication.

[1] Later we will call these *rank one matrices*.

Theorem: Block Multiplication. Suppose that we partition two matrices into submatrices by inserting vertical and horizontal lines:

$$A = \left(\begin{array}{c|c|c} A_{11} & \cdots & A_{1m} \\ \hline \vdots & & \vdots \\ \hline A_{\ell 1} & \cdots & A_{\ell m} \end{array}\right) \quad \text{and} \quad B = \left(\begin{array}{c|c|c} B_{11} & \cdots & B_{1n} \\ \hline \vdots & & \vdots \\ \hline B_{m1} & \cdots & B_{mn}. \end{array}\right)$$

Let's say that each submatrix A_{ij} has shape $\lambda_i \times \mu_j$ and each B_{ij} has shape $\mu_i \times \nu_j$, so

$$\begin{aligned} \#(\text{rows of } A) &= \lambda_1 + \cdots + \lambda_\ell, \\ \#(\text{cols of } A) &= \mu_1 + \cdots + \mu_m, \\ \#(\text{rows of } B) &= \mu_1 + \cdots + \mu_m, \\ \#(\text{cols of } B) &= \nu_1 + \cdots + \nu_n. \end{aligned}$$

Then I claim that that the product matrix AB can be partitioned as

$$AB = \left(\begin{array}{c|c|c} C_{11} & \cdots & C_{1n} \\ \hline \vdots & & \vdots \\ \hline C_{\ell 1} & \cdots & C_{\ell n} \end{array}\right),$$

where the submatrix C_{ij} is given by

$$\boxed{C_{ij} = \sum_{k=1}^{m} A_{ik} B_{kj}.}$$

Note that $\#(\text{cols of } A_{ik}) = \mu_k = \#(\text{rows of } B_{kj})$ so that each matrix product $A_{ik}B_{kj}$ is defined and has shape $\ell_i \times \nu_j$. Thus C_{ij} is a sum of m matrices, each of shape $\lambda_i \times \nu_j$. In particular, C_{ij} has shape $\lambda_i \times \mu_j$. Note that the standard formula for unpartitioned matrices corresponds to the case when each submatrix A_{ij} and B_{ij} has size 1×1.

I won't prove right this now because the notation is too hairy.[2] Instead let's see some examples illustrating the few rules that we stated above. Let

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

Multiplying rows of A by columns of B gives

$$\left(\begin{array}{ccc} 1 & 1 & 1 \\ \hline 1 & 2 & 3 \end{array}\right) \left(\begin{array}{c|c} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{array}\right) = \left(\begin{array}{c|c} \begin{pmatrix} 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} & \begin{pmatrix} 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \\ \hline \begin{pmatrix} 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} & \begin{pmatrix} 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \end{array}\right) = \left(\begin{array}{c|c} 2 & 2 \\ \hline 3 & 5 \end{array}\right).$$

[2] Later it will follow easily from properties of linear functions between direct sums of vector spaces.

Multiplying rows of A by B gives

$$\left(\begin{array}{ccc} 1 & 1 & 1 \\ \hline 1 & 2 & 3 \end{array}\right)\begin{pmatrix}1 & 0\\ 1 & 1\\ 0 & 1\end{pmatrix} = \left(\begin{array}{c} \begin{pmatrix}1 & 1 & 1\end{pmatrix}\begin{pmatrix}1 & 0\\ 1 & 1\\ 0 & 1\end{pmatrix} \\ \hline \begin{pmatrix}1 & 2 & 3\end{pmatrix}\begin{pmatrix}1 & 0\\ 1 & 1\\ 0 & 1\end{pmatrix}\end{array}\right) = \left(\begin{array}{cc} 2 & 2 \\ \hline 3 & 5\end{array}\right).$$

Multiplying A by columns of B gives

$$\begin{pmatrix}1 & 1 & 1\\ 1 & 2 & 3\end{pmatrix}\left(\begin{array}{c|c} 1 & 0\\ 1 & 1\\ 0 & 1\end{array}\right) = \left(\begin{array}{c|c} \begin{pmatrix}1 & 1 & 1\\ 1 & 2 & 3\end{pmatrix}\begin{pmatrix}1\\1\\0\end{pmatrix} & \begin{pmatrix}1 & 1 & 1\\ 1 & 2 & 3\end{pmatrix}\begin{pmatrix}0\\1\\1\end{pmatrix}\end{array}\right) = \left(\begin{array}{c|c} 2 & 2\\ 3 & 5\end{array}\right).$$

Finally, multiplying columns of A by rows of B gives

$$\begin{aligned}\left(\begin{array}{c|c|c} 1 & 1 & 1\\ 1 & 2 & 3\end{array}\right)\left(\begin{array}{cc} 1 & 0\\ \hline 1 & 1\\ \hline 0 & 1\end{array}\right) &= \begin{pmatrix}1\\1\end{pmatrix}\begin{pmatrix}1 & 0\end{pmatrix} + \begin{pmatrix}1\\2\end{pmatrix}\begin{pmatrix}1 & 1\end{pmatrix} + \begin{pmatrix}1\\3\end{pmatrix}\begin{pmatrix}0 & 1\end{pmatrix}\\ &= \begin{pmatrix}1 & 0\\ 1 & 0\end{pmatrix} + \begin{pmatrix}1 & 1\\ 2 & 2\end{pmatrix} + \begin{pmatrix}0 & 1\\ 0 & 3\end{pmatrix}\\ &= \begin{pmatrix}2 & 2\\ 3 & 5\end{pmatrix}.\end{aligned}$$

Each of these kinds of multiplication is useful for a different purpose. It is important to know them all.

2. Linear Functions

The ultimate goal of matrices is to hide all of the details of matrix arithmetic behind uppercase Roman letters. This lets us ignore irrelevant details to focus on higher level structure. The magic property that makes this work is the *associative property of matrix multiplication*.

Magic: Associativity of Matrix Multiplication. Consider matrices A, B, C of sizes $\ell \times m$, $m \times n$ and $n \times p$, respectively. Then the matrices AB, BC, $A(BC)$ and $(AB)C$ are defined, and we have

$$A(BC) = (AB)C.$$

This is not at all obvious from the definitions given above. A brute force proof is possible, but not enlightening. There is a much more conceptual explanation.

Definition of Linear Functions. Consider vector spaces V and W over $\mathbb{R}$ (or $\mathbb{C}$). A function $T : V \to W$ is called *linear* when it satisfies the following three properties:

- $T(\mathbf{0}) = \mathbf{0}$,
- $T(\alpha\mathbf{v}) = \alpha T(\mathbf{v})$,
- $T(\mathbf{v}_1 + \mathbf{v}_2) = T(\mathbf{v}_1) + T(\mathbf{v}_2)$.

In other words, a linear function preserves the vector space operations of addition and scalar multiplication. We can also summarize these properties in one step by saying that T preserves *linear combinations*:

$$T(\alpha_1\mathbf{v}_1 + \cdots + \alpha_n\mathbf{v}_n) = \alpha_1 T(\mathbf{v}_1) + \cdots + \alpha_n T(\mathbf{v}_n).$$

Why? Many natural operations are linear:

- Differention between suitable spaces of functions is linear.
- Integration from a suitable space of functions to $\mathbb{R}$ is linear.
- An inner product $\langle -, - \rangle$ on V over $\mathbb{R}$ is *bilinear*. That is, for any $\mathbf{v} \in V$, each of the following two functions is linear:
$$\langle \mathbf{v}, - \rangle : V \to \mathbb{R} \quad \text{and} \quad \langle -, \mathbf{v} \rangle : V \to \mathbb{R}$$
- A Hermitian inner product $\langle -, - \rangle$ on V over $\mathbb{C}$ is *sesquilinear* (one and a half times linear). This means that for each fixed $\mathbf{v} \in V$, the function $V \to \mathbb{C}$ defined by $\mathbf{u} \mapsto \langle \mathbf{v}, \mathbf{u} \rangle$ is linear, but the function $V \to \mathbb{C}$ defined by $\mathbf{u} \mapsto \langle \mathbf{u}, \mathbf{v} \rangle$ is *conjugate linear*:
$$\langle \alpha_1 \mathbf{u}_1 + \cdots + \alpha_n \mathbf{u}_n, \mathbf{v} \rangle = \alpha_1^* \langle \mathbf{u}_1, \mathbf{v} \rangle + \cdots + \alpha_n^* \langle \mathbf{u}_n, \mathbf{v} \rangle.$$

If V and W are finite dimensional with $\dim V = n$ and $\dim W = m$, then choosing bases turns linear transformations $T : V \to W$ into $m \times n$ matrices. To keep things simple, for now we will work with Euclidean space and standard bases. Here is the big idea:

$$\begin{array}{c} \text{linear functions} \\ T : \mathbb{R}^n \to \mathbb{R}^m \end{array} \quad \leftrightsquigarrow \quad m \times n \text{ matrices}$$

The correspondence is easy to describe. First of all, let A be an $m \times n$ matrix over $\mathbb{R}$. This defines a function $\mathbb{R}^n \to \mathbb{R}^m$ by multiplying column vectors on the left:
$$\mathbf{v} \in \mathbb{R}^n \quad \rightsquigarrow \quad A\mathbf{v} \in \mathbb{R}^m.$$
Indeed, if $\mathbf{v}$ has shape $n \times 1$ then the product matrix $A\mathbf{v}$ is defined and has shape $m \times 1$. It is straightforward to check that this function is linear:
$$A(\alpha_1 \mathbf{v}_+ \cdots + \alpha_n \mathbf{v}_1) = \alpha_1 A\mathbf{v}_1 + \cdots + \alpha_n A\mathbf{v}_n.$$
Conversely, let $T : \mathbb{R}^n \to \mathbb{R}^m$ be any linear function. In order to create an $m \times n$ matrix from T we consider the n standard basis vectors $\mathbf{e}_1, \ldots, \mathbf{e}_n \in \mathbb{R}^n$. Following our convention we will think of these as column vectors:
$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \ldots, \mathbf{e}_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$
Now each basis vector $\mathbf{e}_i \in \mathbb{R}^n$ gets sent by T to a column vector $T(\mathbf{e}_i)$ in $\mathbb{R}^m$. We will record the n column vectors $T(\mathbf{e}_1), \ldots, T(\mathbf{e}_n) \in \mathbb{R}^m$ as the columns of an $m \times n$ matrix:
$$[T] := \begin{pmatrix} | & & | \\ T(\mathbf{e}_1) & \cdots & T(\mathbf{e}_n) \\ | & & | \end{pmatrix}.$$
Thus the linear function $T : \mathbb{R}^n \to \mathbb{R}^m$ becomes an $m \times n$ matrix $[T]$. Furthermore, the linear function defined by the matrix $[T]$ is the same as the linear function T. To see this, we

consider any vector $\mathbf{v} \in \mathbb{R}^n$:

$$\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = v_1 \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + v_2 \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} + \cdots + v_n \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} = v_1\mathbf{e}_1 + v_2\mathbf{e}_2 + \cdots + v_n\mathbf{e}_n.$$

Then from the definition of $[T]$ and the linearity of T we have

$$\begin{aligned} T(\mathbf{v}) &= T(v_1\mathbf{e}_1 + v_2\mathbf{e}_2 + \cdots + v_n\mathbf{e}_n) \\ &= v_1T(\mathbf{e}_1) + v_2T(\mathbf{e}_2) + \cdots + v_nT(\mathbf{e}_n) \\ &= \sum_j v_jT(\mathbf{e}_j) \\ &= \sum_j v_j(j\text{th col of } [T]) \\ &= [T]\mathbf{v}, \end{aligned}$$

where the last expression $[T]\mathbf{v}$ is a matrix product. To summarize: To each linear function $T : \mathbb{R}^n \to \mathbb{R}^n$ we associate an $m \times n$ matrix $[T]$ with the property that

$$\underbrace{T(\mathbf{v})}_{\text{apply the function}} \quad = \quad \underbrace{[T]\mathbf{v}}_{\text{matrix multiplication}}$$

So far this is slightly interesting. It becomes very interesting when we consider functional composition. Suppose we have linear functions $T : \mathbb{R}^n \to \mathbb{R}^m$ and $S : \mathbb{R}^m \to \mathbb{R}^\ell$:

$$\begin{tikzcd} \mathbb{R}^n \arrow[r, "T"] \arrow[rr, bend right, "S\circ T"'] & \mathbb{R}^m \arrow[r, "S"] & \mathbb{R}^\ell \end{tikzcd}$$

Observe that the composite function $S \circ T : \mathbb{R}^n \to \mathbb{R}^\ell$ is also linear:

$$\begin{aligned} (S \circ T)\left(\sum a_i\mathbf{v}_i\right) &= S\left(T\left(\sum a_i\mathbf{v}_i\right)\right) \\ &= S\left(\sum a_iT(\mathbf{v}_i)\right) \\ &= \sum a_iS\left(T(\mathbf{v}_i)\right) \\ &= \sum a_i(S \circ T)(\mathbf{v}_i). \end{aligned}$$

Hence the function $S \circ T : \mathbb{R}^n \to \mathbb{R}^\ell$ corresponds to an $\ell \times n$ matrix $[S \circ T]$. Now we have three matrices:

$$\begin{aligned} [S] &\text{ has shape } \ell \times m, \\ [T] &\text{ has shape } m \times n, \\ [S \circ T] &\text{ has shape } \ell \times n. \end{aligned}$$

The following theorem is the ultimate reason for the concept of matrix multiplication. This theorem could also be taken as the **definition** of matrix multiplication.

Matrix Multiplication = Composition of Linear Functions. For any linear functions $T : \mathbb{R}^n \to \mathbb{R}^m$ and $S : \mathbb{R}^m \to \mathbb{R}^\ell$, the composite $S \circ T : \mathbb{R}^n \to \mathbb{R}^\ell$ is also linear, and we have

$$[S \circ T] = [S][T].$$

Proof. The proof will use the following rule of matrix multiplication:

$$(j\text{th col of } AB) = A(j\text{th col of } B).$$

Our goal is to show that $[S \circ T]$ and $[S][T]$ have the same columns. From the definition of the matrix $[S \circ T]$ we have

$$(j\text{th col of } [S \circ T]) = (S \circ T)(\mathbf{e}_j) = S(T(\mathbf{e}_j)).$$

On the other hand, from the above property of matrix multiplication we have

$$(j\text{th col of } [S][T]) = [S](j\text{th col of } [T]) = [S]T(\mathbf{e}_j) = S(T(\mathbf{e}_j)).$$

$\square$

Remark: It is worth meditating on this proof. When you understand it then you can say that you really understand the concept of matrix multiplication.

Before moving to some examples, we pause to give the *correct* (conceptual) proof that matrix multiplication is associative.

Proof of Associativity. Consider matrices A, B, C of shapes $\ell \times m$, $m \times n$ and $n \times p$, respectively. We can use these to define linear functions $R : \mathbb{R}^m \to \mathbb{R}^\ell$, $S : \mathbb{R}^n \to \mathbb{R}^m$ and $T : \mathbb{R}^p \to \mathbb{R}^n$ by matrix multiplication:

$$\begin{aligned} R(\mathbf{v}) &:= A\mathbf{v} \text{ for } \mathbf{v} \in \mathbb{R}^\ell, \\ S(\mathbf{v}) &:= B\mathbf{v} \text{ for } \mathbf{v} \in \mathbb{R}^n, \\ T(\mathbf{v}) &:= C\mathbf{v} \text{ for } \mathbf{v} \in \mathbb{R}^m. \end{aligned}$$

Then, of course, the corresponding matrices are $[R] = A$, $[S] = B$ and $[T] = C$. Here is a picture of the functions:

$$\begin{tikzcd} \mathbb{R}^p \arrow[r, "T"] \arrow[rr, bend right, "S\circ T"'] & \mathbb{R}^n \arrow[r, "S"] \arrow[rr, bend left, "R\circ S"] & \mathbb{R}^m \arrow[r, "R"] & \mathbb{R}^\ell \end{tikzcd}$$

Recall that composition of functions is naturally associative. That is, for any $\mathbf{v} \in \mathbb{R}^p$ we have

$$(R \circ (S \circ T))(\mathbf{v}) = R(S(T(\mathbf{v}))) = ((R \circ S) \circ T)(\mathbf{v}),$$

which means that $R\circ(S\circ T)=(R\circ S)\circ T$ as functions $\mathbb{R}^p\to\mathbb{R}^\ell$. Then the previous theorem tells us that

$$\begin{aligned}A(BC)&=[R]([S][T])\\&=[R][S\circ T]\\&=[R\circ(S\circ T)]\\&=[(R\circ S)\circ T]\\&=[R\circ S][T]\\&=([R][S])[T]\\&=(AB)C.\end{aligned}$$

Note that we never had to mention the entries of the matrices. Magic! □

3. Matrix Arithmetic

Let's zoom out again. One of the strengths of matrix notation is that we can sometimes solve a problem purely symbolically, without mentioning the entries of the matrices. In fact, by hiding the appropriate details we can sometimes turn a difficult problem into an almost trivial matrix computation.

Here is the context for matrix arithmetic.

Vector Spaces of Matrices. Let $\mathbb{R}^{m\times n}$ denote the set of $m\times n$ with real entries. (We define $\mathbb{C}^{m\times n}$ similarly.) By convention we will write

$$\mathbb{R}^n=\mathbb{R}^{n\times 1}=\text{ the set of } n\times 1 \text{ column vectors.}$$

Matrices can be added and multiplied by scalars in an obvious way. That is, given $m\times n$ matrices $A,B\in\mathbb{R}^{m\times n}$ and a scalar $\alpha\in\mathbb{R}$ we define $m\times n$ matrices $A+B$ and αA such that

$$\begin{aligned}(ij\text{ entry of }A+B)&=(ij\text{ entry of }A)+(ij\text{ entry of }B),\\(ij\text{ entry of }\alpha A)&=\alpha(ij\text{ entry of }A).\end{aligned}$$

It is easy to check that these operations make $\mathbb{R}^{m\times n}$ into a vector space over $\mathbb{R}$. Furthermore, there is a *standard basis* of matrices E_{ij} with $1\leq i\leq m$ and $1\leq j\leq n$, with the entry 1 in the ij position and all other entries equal to zero:

$$B_{ij}=\begin{array}{c c}& j\\ i & \begin{pmatrix} & \vdots & \\ \cdots & 1 & \\ & & \end{pmatrix}\end{array}$$

(When a matrix contains many zero entries we will simply leave them blank.) Since there mn such basis matrices it follows that

$$\dim\mathbb{R}^{m\times n}=mn.$$

In addition to the vector space structure, we have two additional operations on matrices. First we have *transposition* and *conjugate transposition*:

$$\begin{array}{ccc}\mathbb{R}^{m\times n}&\to&\mathbb{R}^{n\times m}\\A&\mapsto&A^T\end{array}$$

and

$$\begin{array}{ccc} \mathbb{C}^{m\times n} & \to & \mathbb{C}^{n\times m} \\ A & \mapsto & A^* \end{array}$$

Second we have the all-important operation of matrix multiplication:

$$\begin{array}{ccc} \mathbb{R}^{\ell\times m}\times\mathbb{R}^{m\times n} & \to & \mathbb{R}^{\ell\times n} \\ (A,B) & \mapsto & AB. \end{array}$$

Finally, we have two special classes of matrices. For any shape $m \times n$ we have a *zero matrix*:

$$O_{m\times n} = \begin{pmatrix} 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{pmatrix}.$$

[Note: I use the letter O for zero matrices.] And for any n we have a square *identity matrix*:

$$I_n = \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{pmatrix}.$$

This identity matrix corresponds to the *identity function* $\mathrm{id} : \mathbb{R}^n \to \mathbb{R}^n$, which sends each vector to itself. Indeed, for any linear function $T : \mathbb{R}^n \to \mathbb{R}^n$ recall that the ith column of the corresponding matrix $[T]$ is $T(\mathbf{e}_i)$. Since the ith column of $[\mathrm{id}]$ is $\mathrm{id}(\mathbf{e}_i) = \mathbf{e}_i$ we have $[\mathrm{id}] = I_n$.

Rules of Matrix Arithmetic. The operations of matrix arithmetic satisfy the following abstract rules. Here uppercase Roman letters represent matrices and lowercase Greek letters are scalars. Assume that the matrices have appropriate shape so the indicated matrix sums and products exist.

- **Vector Space Rules.**

$$\begin{aligned} A + B &= B + A, \\ A + (B + C) &= (A + B) + C, \\ A + O &= O + A = A, \\ 1A &= A, \\ 0A &= O, \\ \alpha(\beta A) &= (\alpha\beta)A, \\ (\alpha+\beta)A &= \alpha A + \beta A, \\ \alpha(A + B) &= \alpha A + \alpha B. \end{aligned}$$

- **Multiplication is not Commutative.** In general we have

$$AB \neq BA,$$

  even when both matrices are defined and have the same shape.

- **Multiplication is Bilinear.**

$$\begin{aligned} A(\beta B + \gamma C) &= \beta AB + \gamma AC, \\ (\alpha A + \beta B)C &= \alpha AC + \beta BC. \end{aligned}$$

- **Multiplication by O and I.**

$$\begin{aligned} AO &= O,\\ OA &= O,\\ AI &= A,\\ IA &= A. \end{aligned}$$

- **Properties of Transpose and Conjugate Transpose.**

$$\begin{aligned} (A^T)^T &= A, & (A^*)^* &= A,\\ (A+B)^T &= A^T + B^T, & (A+B)^* &= A^* + B^*\\ (\alpha A)^T &= \alpha A^T, & (\alpha A)^* &= \alpha^* A^*,\\ (AB)^T &= B^T A^T, & (AB)^* &= B^* A^*. \end{aligned}$$

Remark: If A is $\ell \times m$ and B is $m \times n$ then A^T is $m \times \ell$ and B^T is $n \times m$. The matrix $B^T A^T$ always exists and is equal to AB. In general, the matrix $A^T B^T$ does not exist.

In addition to arithmetic operations, we also need a way to measure the "size" of a matrix.

Axioms of Matrix Norms. Let $\| - \|$ be a function that assigns to each matrix A a real number $\|A\|$. We call this a *matrix norm* when it satisfies the following axioms:

(a) $\|A\| \geq 0$ for all A, with $\|A\| = 0$ if and only if $A = O$.
(b) $\|\alpha A\| = |\alpha|\|A\|$
(c) $\|A + B\| \leq \|A\| + \|B\|$
(d) $\|AB\| \leq \|A\|\|B\|$.

Here are the two main examples.

The Frobenius Norm. We define this by analogy with the standard vector norm:

$$\|A\|_F := \begin{cases} \sqrt{\sum_{i,j} a_{ij}^2} & \text{over } \mathbb{R},\\ \sqrt{\sum_{i,j} |a_{ij}|^2} & \text{over } \mathbb{C}. \end{cases}$$

We observe that $\|\mathbf{v}\|_F = \|\mathbf{v}\|$ for all column vectors $\mathbf{v}$. The fact that $\| - \|_F$ satisfies (abc) follows from this vector case. You will prove that $\|AB\|_F \leq \|A\|_F \|B\|_F$ on the homework.

The L^2 Norm (Also Called the Operator Norm). The Frobenius norm only applies to matrices. The operator norm also applies to linear functions on infinite dimensional normed vector spaces:

$$\|A\|_2 := \max\{\|A\mathbf{u}\| : \text{over all unit vectors } \|\mathbf{u}\| = 1\}.$$

Since $\|A\mathbf{u}\| \geq 0$ for all $\mathbf{u}$ we have $\|A\|_2 \geq 0$. And if $\|A\|_2 = 0$ then we must have $\|A\mathbf{u}\| = 0$ (and hence $A\mathbf{u} = \mathbf{0}$) for all unit vectors $\mathbf{u}$. In particular, letting $\mathbf{u}$ range of the standard basis vectors we find that each column is A is a zero vector, hence $A = O$. This proves property (a). For property (b) we observe that $\|\alpha A\mathbf{u}\| = |\alpha|\|A\mathbf{u}\|$, hence the maximum value of $\|\alpha A\mathbf{u}\|$ is $|\alpha|$ times the maximum value of $\|A\mathbf{u}\|$. For part (c) we use the triangle inequality for vector norms to observe that[3]

$$\|(A + B)\mathbf{u}\| = \|A\mathbf{u} + B\mathbf{u}\| \leq \|A\mathbf{u}\| + \|B\mathbf{u}\| \text{ for all matrices } A, B \text{ and unit vectors } \mathbf{u}.$$

[3]Details: The maximum value of $\|(A + B)\mathbf{u}\|$ is $\leq$ the maximum value of $\|A\mathbf{u}\| + \|B\mathbf{u}\|$ which is $\leq$ the sum of the maximum values of $\|A\mathbf{u}\|$ and $\|B\mathbf{u}\|$.

To prove part (d) we first show that $\|A\mathbf{v}\|_2 \leq \|A\|_2\|\mathbf{v}\|$ for any nonzero vector $\mathbf{v}$. Indeed if $\mathbf{v}$ is nonzero then $\mathbf{v}/\|\mathbf{v}\|$ is a unit vector and hence

$$\|A\|_2 = \max\{\|A\mathbf{u}\| : \text{all unit vectors } \mathbf{u}\} \geq \|A(\mathbf{v}/\|\mathbf{v}\|)\| = \|A\mathbf{v}\|/\|\mathbf{v}\|.$$

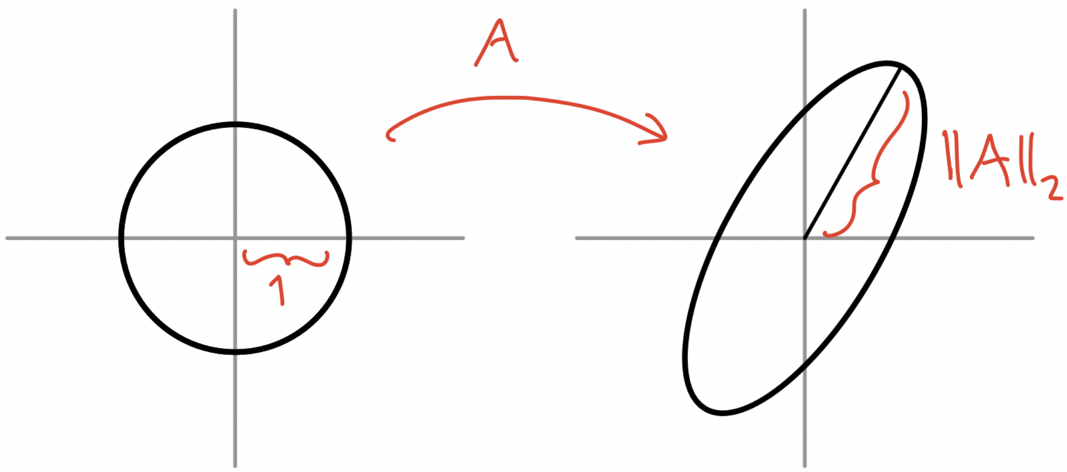
Finally, to show that $\|AB\|_2 \leq \|A\|_2\|B\|_2$, consider any unit vector $\mathbf{u}$. Note that $B\mathbf{u}$ is not necessarily a unit vector, but from the previous remark with $\mathbf{v} = B\mathbf{u}$ we still have[4]

$$\|(AB)\mathbf{u}\| = \|A(B\mathbf{u})\| \leq \|A\|_2\|B\mathbf{u}\| \leq \|A\|_2\|B\|_2.$$

It follows that

$$\|AB\|_2 = \max\{\|AB\mathbf{u}\| : \text{all unit vectors } \mathbf{u}\} \leq \|A\|_2\|B\|_2.$$

Here is a picture of the L^2 norm of a 2×2 matrix:



The matrix A sends the unit circle to an ellipse. The operator norm $\|A\|_2$ is the longest axis of the ellipse. More generally, the longest axis is called the first *singular value* σ_1 and the smaller axis is the second singular value σ_2. We will discuss the SVD (singular value decomposition) in a later section.

The Frobenius norm is harder to visualize.

4. Inverse Matrices

We have seen how to multiply matrices, but can we also divide? If we can then this will be extremely useful for solving matrix equations. For example, suppose we have an equation

$$AX = B,$$

where A and B are given matrices and X is an unknown matrix. If we can find a matrix C such that $CA = I$ then multiplying both sides on the left by C gives

$$\begin{aligned} AX &= B \\ C(AX) &= CB \\ (CA)X &= CB \\ IX &= CB \\ X &= CB. \end{aligned}$$

[4]If $B\mathbf{u} = \mathbf{0}$ then we have $\|AB\mathbf{u}\| = 0$ and there is nothing to show.

Definition of Inverse Matrices. Let A be an $m \times n$ matrix. Any $n \times m$ matrix B satisfying

$$AB = I_m$$

is called a *right inverse* of A. Any $n \times m$ matrix C satisfying

$$CA = I_n$$

is called a *left inverse* of A. Left and right inverses, if they exist, need not be unique. However, suppose that A has **both** a right inverse B **and** a left inverse C. Then we must have

$$B = I_n B = (CA)B = C(AB) = CI_m = C.$$

In this case $B = C$ is the unique *two-sided inverse* of A, and we write

$$A^{-1} = B = C.$$

When A has a two-sided inverse we say that A is *invertible*. It will follow from the Fundamental Theorem below that **an invertible matrix must be square** (i.e., have $m = n$) but this theorem is surprisingly difficult to prove.

For example, consider the following non-square matrix:

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{pmatrix}$$

If B is a right inverse of A then it must have two columns $\mathbf{b}_1, \mathbf{b}_2 \in \mathbb{R}^3$ and it must satisfy the block matrix equation

$$\begin{aligned} I_2 &= AB \\ \left(\begin{array}{c|c} 1 & 0 \\ 0 & 1 \end{array}\right) &= A \left(\begin{array}{c|c} \mathbf{b}_1 & \mathbf{b}_2 \end{array}\right) \\ \left(\begin{array}{c|c} \mathbf{e}_1 & \mathbf{e}_2 \end{array}\right) &= \left(\begin{array}{c|c} A\mathbf{b}_1 & A\mathbf{b}_2 \end{array}\right) \end{aligned}$$

From Linear Algebra I you know how to solve the systems $A\mathbf{b}_1 = \mathbf{e}_1$ and $A\mathbf{b}_2 = \mathbf{e}_2$ to obtain all possible column vectors $\mathbf{b}_1, \mathbf{b}_2$. In turns out $\mathbf{b}_1 = (2+s, -1-2s, s)$ and $\mathbf{b}_2 = (-1+t, 1-2t, t)$ for any parameters s, t. Thus we obtain a two-dimensional family of right inverses:[5]

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 2+s & -1+t \\ -1-2s & 1-2t \\ s & t \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

This already tells us that A has **no left inverse**, since if it did then any two right inverses would be equal. Indeed, let B, B' be any two right inverses of A and suppose that A has a left inverse C. Then we get

$$\begin{aligned} I_2 &= I_2 \\ AB &= AB' \\ C(AB) &= C(AB') \\ (CA)B &= (CA)B' \\ I_3 B &= I_3 B' \\ B &= B'. \end{aligned}$$

Since our matrix A has many different right inverses, no left inverse can exist.

[5]Note: The family of right inverses of A is not a vector subspace of $\mathbb{R}^{3\times 2}$ because it does not contain the zero matrix. However, it is an *affine subspace* of $\mathbb{R}^{3\times 2}$, i.e., a translation of a linear subspace.

I mentioned above that any invertible matrix (i.e., any matrix with a two-sided inverse) must be square. It is also true that any left inverse of a given square matrix must also be a right inverse, and vice versa. I will state these theorems now, but the proofs are surprisingly subtle and are postponed until the next section.

Two Subtle Theorems.

- Any invertible matrix must be square.
- For any square matrices A and B of the same size, we have

$$AB = I \quad \Longleftrightarrow \quad BA = I.$$

To be concrete, consider the matrices

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad \text{and} \quad A' = \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix}.$$

The matrix equation $AA' = I$ is equivalent to the following four equations:

$$\begin{cases} aa' + bc' & = 1, \\ ab' + bd' & = 0, \\ ca' + dc' & = 0, \\ cb' + dd' & = 1. \end{cases}$$

And the matrix equation $A'A = I$ is equivalent to the system

$$\begin{cases} a'a + b'c & = 1, \\ a'b + b'd & = 0, \\ c'a + d'c & = 0, \\ c'b + d'd & = 1. \end{cases}$$

The second theorem above tells us that these two systems of equations have the same solutions for the eight unknowns $a, b, c, d, a', b', c', d'$. It is tempting to look for a direct algebraic proof of this but you won't be able to find one because this is the wrong approach. The correct approach requires us to consider the dimensions of certain vector spaces associated to the matrices. See the Fundamental Theorem in the next section.

For now we will prove some easy and purely symbolic properties of inverse matrices.

Algebraic Properties of Inverse Matrices.

(a) Suppose that A^{-1} exists. Then $(A^*)^{-1}$ exists and is equal to $(A^{-1})^*$.
(b) Suppose that A^{-1}, B^{-1} and AB exist. Then $(AB)^{-1}$ exists and is equal to $B^{-1}A^{-1}$.

Proof. (a): We only need to show that $A^*(A^{-1})^* = I$ and $(A^{-1})^*A^* = I$. For the first identity we have[6]

$$A^*(A^{-1})^* = (A^{-1}A)^* = I^* = I.$$

The other direction is similar. (b): We only need to show that $(AB)(B^{-1}A^{-1}) = I$ and $(B^{-1}A^{-1})AB = I$. This follows easily from the associativity of matrix multiplication:

$$(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AIA^{-1} = AA^{-1} = I.$$

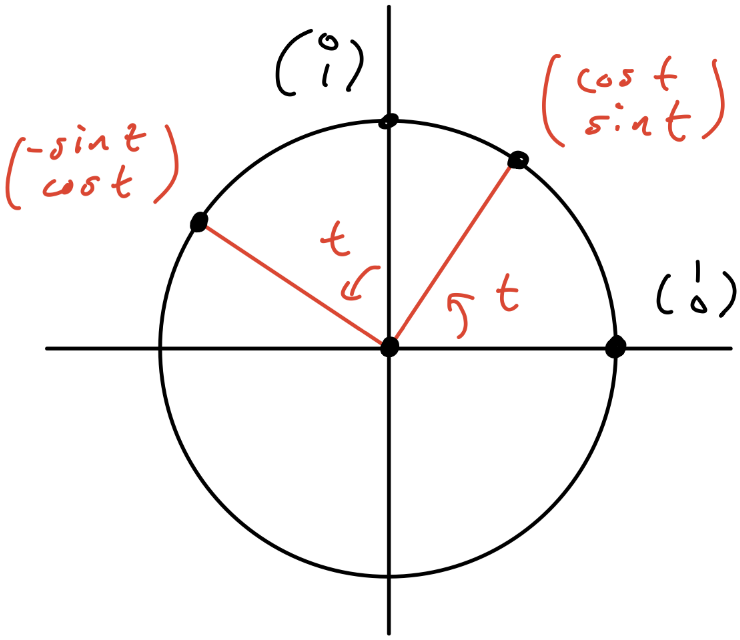
The other direction is similar. $\square$

[6] Recall that $A^*B^* = (BA)^*$.

5. Examples

It is high time for some examples.

Rotations. Consider the function $R_t : \mathbb{R}^2 \to \mathbb{R}^2$ that rotates each point by angle t, counterclockwise around the origin. This function is linear because it sends the origin to itself and it sends parallelograms to parallelograms. To determine the corresponding matrix we only need to rotate the standard basis vectors:



Since no confusion will result, I will use the notation R_t for the function and for the corresponding matrix. Thus we have

$$R_t = \left(R_t \begin{pmatrix}1\\0\end{pmatrix} \quad R_t \begin{pmatrix}0\\1\end{pmatrix} \right) = \begin{pmatrix}\cos t & -\sin t\\ \sin t & \cos t\end{pmatrix}.$$

Once we have the matrix we can use this to rotate a general point:

$$R_t \begin{pmatrix}x\\y\end{pmatrix} = \begin{pmatrix}\cos t & -\sin t\\ \sin t & \cos t\end{pmatrix} \begin{pmatrix}x\\y\end{pmatrix} = \begin{pmatrix}x\cos t - y\sin t\\ x\sin t + y\cos t\end{pmatrix}.$$

It would be much harder to solve this problem without the theory of matrices. Next we consider the composition of two rotations. Thinking in terms of functions, it is clear that $R_sR_t = R_{s+t} = R_tR_s$, since rotating first by one angle and then by the other angle is the same as rotating once by the sum of the two angles. On the other hand, since matrix multiplication is the same as functional composition, we obtain the following matrix identity, which is equivalent to the angle sum trigonometric identities:

$$\begin{pmatrix}\cos s & -\sin s\\ \sin s & \cos s\end{pmatrix} \begin{pmatrix}\cos t & -\sin t\\ \sin t & \cos t\end{pmatrix} = \begin{pmatrix}\cos(s+t) & -\sin(s+t)\\ \sin(s+t) & \cos(s+t)\end{pmatrix}.$$

Note that rotation clockwise by angle t is the same as rotation counterclockwise by angle $-t$. Thus the functions R_t and R_{-t} are inverses:

$$R_tR_{-t} = R_{-t}R_t = R_0 = I.$$

Note that rotation by angle zero is just the identity function. It is interesting to observe that

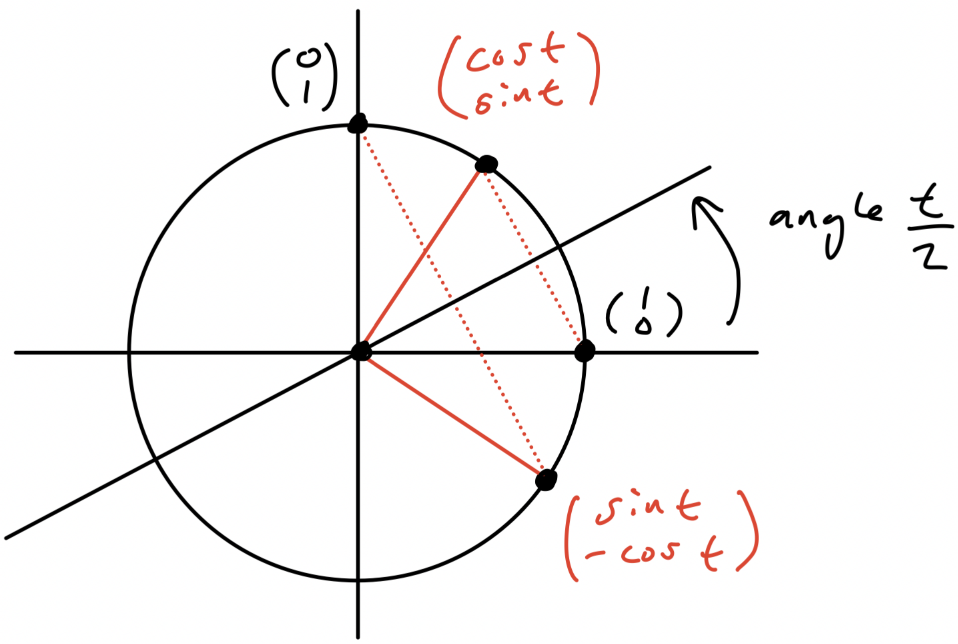
$$\begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix}^{-1} = (R_t)^{-1} = R_{-t} = \begin{pmatrix} \cos(-t) & -\sin(-t) \\ \sin(-t) & \cos(-t) \end{pmatrix} = \begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix} = (R_t)^T.$$

We will see below the matrices satisfying $A^{-1} = A^T$ are called *orthogonal matrices.* Finally, let me remark that the determinant of a rotation matrix is always 1:

$$\det(R_t) = \det \begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix} = \cos^2 t + \sin^2 t = 1.$$

We will discuss the general theory of determinants later.

Reflections. Let $F_t : \mathbb{R}^2 \to \mathbb{R}^2$ be the function that reflects each point across the line that makes angle $t/2$ from the positive x-axis. Again, this is a linear function because it sends the origin to itself and sends parallelograms to parallelograms. To determine the corresponding matrix we reflect the standard basis vectors:



Thus we obtain the matrix

$$F_t = \left(F_t \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad F_t \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right) = \begin{pmatrix} \cos t & \sin t \\ \sin t & -\cos t \end{pmatrix}.$$

The composition of two reflections in two different lines turns out to be a rotation:

$$\begin{aligned} F_s F_t &= \begin{pmatrix} \cos s & \sin s \\ \sin s & -\cos s \end{pmatrix} \begin{pmatrix} \cos t & \sin t \\ \sin t & -\cos t \end{pmatrix} \\ &= \begin{pmatrix} \cos s \cos t + \sin s \sin t & \cos s \sin t - \sin s \cos t \\ \sin s \cos t - \cos s \sin t & \sin s \sin t + \cos s \cos t \end{pmatrix} \\ &= \begin{pmatrix} \cos(s-t) & -\sin(s-t) \\ \sin(s-t) & \cos(s-t) \end{pmatrix} \\ &= R_{s-t}. \end{aligned}$$

This would be more difficult to see geometrically. In particular, we find that reflection matrices do not commute in general:

$$F_sF_t = R_{s-t} \neq R_{t-s} = F_tF_s \quad \text{unless angles } s-t \text{ and } t-s \text{ are equal.}$$

Taking $s = t$ shows that the composition of a reflection with itself is the identity matrix:

$$F_t^2 = F_tF_t = R_{t-t} = R_0 = I.$$

In other words, reflecting in the same line twice is the same thing as doing nothing. This implies that each reflection matrix F_t is equal to its own inverse:

$$(F_t)^{-1} = F_t.$$

It also happens that $(F_t)^T = F_t$, so F_t is another example of an orthogonal matrix. Finally, let me remark that the determinant of any reflection matrix is -1:

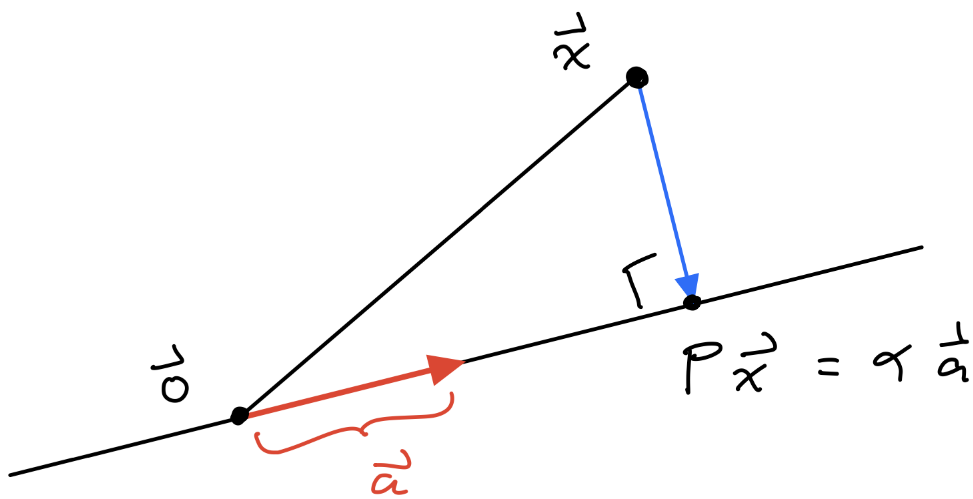
$$\det(F_t) = \det\begin{pmatrix} \cos t & \sin t \\ \sin t & -\cos t \end{pmatrix} = -\cos^2 t - \sin^2 t = -1.$$

Projections. Consider the following matrix:

$$P_t = \begin{pmatrix} \cos^2 t & \cos t \sin t \\ \cos t \sin t & \sin^2 t \end{pmatrix}.$$

As with any 2×2 matrix, this defines a linear function $\mathbb{R}^2 \to \mathbb{R}^2$. What is the geometric description of this function? It is convenient to solve this problem in greater generality.

Suppose that we want to project[7] a point $\mathbf{x} \in \mathbb{R}^n$ onto the line in $\mathbb{R}^n$ generated by a vector $\mathbf{a}$:



Since projection is a linear function there will be some $n \times n$ matrix P that achieves this projection. We know exactly two things about this situation:

(1) Since the projection $P\mathbf{x}$ is on the line generated by $\mathbf{a}$ we must have $P\mathbf{x} = \alpha\mathbf{x}$ for some scalar α. This scalar will change depending on the point $\mathbf{x}$.
(2) Since the projection is orthogonal we know that the blue vector $P\mathbf{x} - \mathbf{x}$ is orthogonal to the red vector $\mathbf{a}$.

[7]Here we are talking about *orthogonal projection*, i.e., projection at right angles. Later we will talk about more general kinds of projection.

Putting these two facts together gives[8]

$$
\begin{aligned}
\mathbf{a}^T(P\mathbf{x} - \mathbf{x}) &= 0 && (2)\\
\mathbf{a}^T(\alpha\mathbf{a} - \mathbf{x}) &= 0 && (1)\\
\alpha\mathbf{a}^T\mathbf{a} - \mathbf{a}^T\mathbf{x} &= 0\\
\alpha &= \mathbf{a}^T\mathbf{x}/\mathbf{a}^T\mathbf{a}\\
\alpha &= \mathbf{a}^T\mathbf{x}/\|\mathbf{a}\|^2.
\end{aligned}
$$

Hence the projection of $\mathbf{x}$ is given by

$$P\mathbf{x} = \alpha\mathbf{a} = \underbrace{\left(\frac{\mathbf{a}^T\mathbf{x}}{\|\mathbf{a}\|^2}\right)}_{\text{scalar}} \underbrace{\mathbf{a}}_{\text{vector}}.$$

To find a formula for the $n \times n$ projection matrix P we simply rearrange using the fact that matrix multiplication is associative:[9]

$$
\begin{aligned}
P\mathbf{x} &= \left(\frac{\mathbf{a}^T\mathbf{x}}{\|\mathbf{a}\|^2}\right)\mathbf{a}\\
&= \mathbf{a}\left(\frac{\mathbf{a}^T\mathbf{x}}{\|\mathbf{a}\|^2}\right) && \text{scalars commute with matrices}\\
&= \frac{1}{\|\mathbf{a}\|^2}\mathbf{a}(\mathbf{a}^T\mathbf{x})\\
&= \underbrace{\frac{1}{\|\mathbf{a}\|^2}(\mathbf{a}\mathbf{a}^T)}_{n\times n \text{ matrix}} \underbrace{\mathbf{x}}_{\text{vector}}.
\end{aligned}
$$

Since this identity holds for any vector $\mathbf{x}$[10] we conclude that the projection matrix is given by

$$P = \frac{1}{\|\mathbf{a}\|^2}\mathbf{a}\mathbf{a}^T.$$

If $\mathbf{a} = \mathbf{u}$ is a unit vector then the formula is particularly simple:

$$P = \mathbf{u}\mathbf{u}^T = \text{the projection onto the line in } \mathbb{R}^n \text{ spanned by unit vector } \mathbf{u}.$$

Now we go back to two dimensions. Consider the line in $\mathbb{R}^2$ that makes angle t counterclockwise from the positive x-axis. This line is generated by the unit vector $\mathbf{u} = (\cos t, \sin t)$. Hence the matrix that projects onto this line is

$$P_t = \mathbf{u}\mathbf{u}^T = \begin{pmatrix}\cos t\\ \sin t\end{pmatrix}\begin{pmatrix}\cos t & \sin t\end{pmatrix} = \begin{pmatrix}\cos^2 t & \cos t\sin t\\ \cos t\sin t & \sin^2 t\end{pmatrix}.$$

The image of a general point $\mathbf{x} = (x, y)$ under this projection is

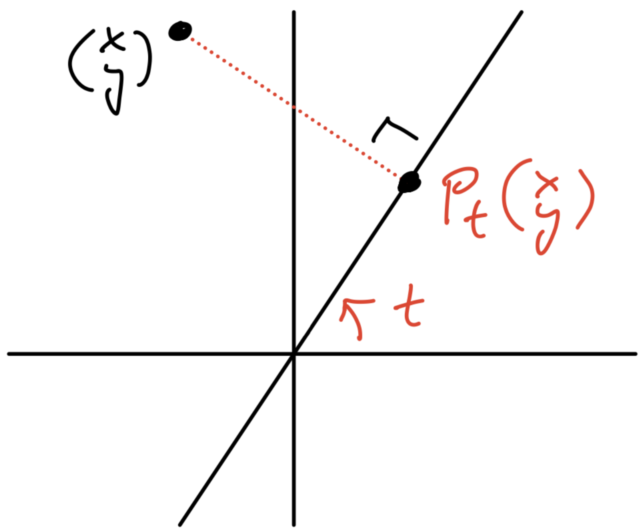
$$P_t\begin{pmatrix}x\\ y\end{pmatrix} = \begin{pmatrix}\cos^2 t & \cos t\sin t\\ \cos t\sin t & \sin^2 t\end{pmatrix}\begin{pmatrix}x\\ y\end{pmatrix} = \begin{pmatrix}x\cos^2 t + y\cos t\sin t\\ x\cos t\sin t + y\sin^2 t\end{pmatrix}.$$

Here is a picture:

[8] I will express this using inner products because the ideas generalize beyond Euclidean space.

[9] The associativity of matrix multiplication is behind many clever proofs like this.

[10] Let A, B be two $n \times n$ matrices such that $A\mathbf{x} = B\mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$. If $\mathbf{x} = \mathbf{e}_j$ then the identity $A\mathbf{e}_j = B\mathbf{e}_j$ tells us that the jth columns of A and B are the same. Since this holds for any j we conclude that A and B are the same matrix.



Note that this projection is **not invertible**. To see this, let's consider the point $(-\sin t, \cos t)$. This point gets projected to the origin:

$$\begin{aligned} P_t \begin{pmatrix} -\sin t \\ \cos t \end{pmatrix} &= \begin{pmatrix} \cos^2 t & \cos t \sin t \\ \cos t \sin t & \sin^2 t \end{pmatrix} \begin{pmatrix} -\sin t \\ \cos t \end{pmatrix} \\ &= \begin{pmatrix} -\cos^2 t \sin t + \cos^2 t \sin t \\ -\cos t \sin^2 t + \cos t \sin^2 t \end{pmatrix} \\ &= \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \end{aligned}$$

But the origin gets projected to itself: $P_t\mathbf{0} = \mathbf{0}$. If P_t had an inverse matrix $(P_t)^{-1}$ then this would imply that

$$\begin{aligned} P_t \begin{pmatrix} -\sin t \\ \cos t \end{pmatrix} &= P_t \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\ (P_t)^{-1} P_t \begin{pmatrix} -\sin t \\ \cos t \end{pmatrix} &= (P_t)^{-1} P_t \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\ \begin{pmatrix} -\sin t \\ \cos t \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \end{aligned}$$

Contradiction.[11] More generally, let $P = \mathbf{u}\mathbf{u}^T$ be the matrix that projects onto the line in $\mathbb{R}^n$ generated by some unit vector $\mathbf{u}$ and let $\mathbf{v} \in \mathbb{R}^n$ be any vector that is perpendicular to $\mathbf{u}$, so that $\mathbf{u}^T\mathbf{v} = 0$. Then we have

$$P\mathbf{v} = (\mathbf{u}\mathbf{u}^T)\mathbf{v} = \mathbf{u}(\mathbf{u}^T\mathbf{v}) = \mathbf{u}(0) = \mathbf{0}.$$

This shows that the projection onto a line in $\mathbb{R}^n$ is never invertible. Finally, let me note that the matrix P_t has determinant zero:

$$\det P_t = \det \begin{pmatrix} \cos^2 t & \cos t \sin t \\ \cos t \sin t & \sin^2 t \end{pmatrix} = \cos^2 t \sin^2 t - \cos^2 t \sin^2 t = 0.$$

[11] More generally, a linear function that is not injective cannot have a left inverse.

Later we will see that a square matrix A is invertible if and only if $\det A \neq 0$.

The Group of Orthogonal Matrices. I mentioned above that a square matrix A satisfying $A^{-1} = A^T$ is called an *orthogonal matrix*. We denote the set of all such matrices by

$$\mathrm{O}_n(\mathbb{R}) = \{A \in \mathbb{R}^{n\times n} : A^TA = I \text{ and } AA^T = I\}.$$

Sometimes the set $\mathrm{O}_n(\mathbb{R})$ is called the *orthogonal group*, because it satisfies the three group axioms from abstract algebra:

- The identity matrix is in $\mathrm{O}_n(\mathbb{R})$. Indeed, we have $I^T = I$ and $II = I$, so that $I^TI = II = I$ and $II^T = II = I$.
- If A is in $\mathrm{O}_n(\mathbb{R})$ then A is invertible and A^{-1} is also in $\mathrm{O}_n(\mathbb{R})$. Indeed, the conditions $A^TA = I$ and $AA^T = I$ just tell us that A is invertible with $A^{-1} = A^T$. But then we also have
$$(A^{-1})^{-1} = A = (A^T)^T = (A^{-1})^T,$$
which tells us that A^{-1} is in $\mathrm{O}_n(\mathbb{R})$.
- If A and B are in $\mathrm{O}_n(\mathbb{R})$ (i.e., if $A^{-1} = A^T$ and $B^{-1} = B^T$) then so is their product AB. Indeed, we have
$$(AB)^{-1} = B^{-1}A^{-1} = B^TA^T = (AB)^T.$$

Remark: Particle physicists are particularly interested in matrix groups but they prefer the complex version of orthogonal matrices, which are called *unitary matrices*:

$$\mathrm{U}_n(\mathbb{C}) = \{A \in \mathbb{C}^{n\times n} : A^*A = I \text{ and } AA^* = I\}.$$

It is worth mentioning a geometric interpretation of orthogonal matrices:[12]

$$A^TA = I \quad \Longleftrightarrow \quad \text{the columns of } A \text{ are orthonormal.}$$

Indeed, suppose that $A \in \mathbb{R}^{n\times n}$ has column vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^n$, so that A^T has row vectors $\mathbf{a}_1^T, \ldots, \mathbf{a}_n^T$. Then the i,j entry of the matrix A^TA is the dot product of $\mathbf{a}_i$ and $\mathbf{a}_j$:

$$\begin{aligned}(i,j \text{ entry of } A^TA) &= (i\text{th row of } A^T)(j\text{th col of } A)\\ &= \mathbf{a}_i^T\mathbf{a}_j\\ &= \mathbf{a}_i \bullet \mathbf{a}_j.\end{aligned}$$

On the other hand, the i,j entry of the identity matrix is the Kronecker delta δ_{ij}. Hence we have $A^TA = I$ if and only if

$$\mathbf{a}_i \bullet \mathbf{a}_j = \delta_{ij},$$

i.e., if and only if the column vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^n$ are orthonormal. This is one reason for the term "orthogonal matrix". On the homework you used this fact to prove that every 2×2 orthogonal matrix is either a rotation or a rotation.

The Fundamental Theorem, which we will prove in the next section, tells us that the equations $A^TA = I$ and $AA^T = I$ are equivalent when A is square, which means that the columns of a square matrix are orthonormal if and only if the rows are orthonormal. I find this mysterious.

[12] The same result holds for unitary matrices, with respect to the Hermitian inner product.