My Pondering on Linear Algebra

Vectors

Orthogonality

Orthogonality is interesting as we need to define so many things which at first blush might not appear central to what follows, but we will make use of them to define orthonormal sets.

1. Given $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ we define the *dot product* to be

$$\mathbf{x} \cdot \mathbf{y} = \mathbf{x}_1\mathbf{y}_1 + \mathbf{x}_2\mathbf{y}_2 + \ldots + \mathbf{x}_n\mathbf{y}_n$$

   Note, $\mathbf{x} \cdot \mathbf{y} \in \mathbb{R}$ and **not an element of** $\mathbb{R}^n$; this means that $\mathbf{x} \cdot \mathbf{y} \cdot \mathbf{z}$ is not defined!
   Now, based on the definition, we can prove the following properties:

   Proposition. Let $\alpha \in \mathbb{R}$, and $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^n$.

   (I) $0 \leq \mathbf{x} \cdot \mathbf{x}$
   (II) $\mathbf{x} \cdot \mathbf{x} = 0$ if and only if $\mathbf{x} = \mathbf{0}$
   (III) $\mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x}$
   (IV) $(\alpha\mathbf{x}) \cdot \mathbf{y} = \alpha(\mathbf{x} \cdot \mathbf{y}) = \mathbf{x} \cdot (\alpha\mathbf{y})$
   (V) $\mathbf{x} \cdot (\mathbf{y} + \mathbf{z}) = \mathbf{x} \cdot \mathbf{y} + \mathbf{x} \cdot \mathbf{z}$ and $(\mathbf{x} + \mathbf{y}) \cdot \mathbf{z} = \mathbf{x} \cdot \mathbf{z} + \mathbf{y} \cdot \mathbf{z}$

2. Given $\mathbf{x} \in \mathbb{R}^n$, and knowing that $0 \leq \mathbf{x} \cdot \mathbf{x}$ we define its *norm*, or *length*, to be

$$\|\mathbf{x}\| = \sqrt{\mathbf{x} \cdot \mathbf{x}}$$

   Now, based on the definition, we can prove the following properties:

   Proposition. Let $\alpha \in \mathbb{R}$, and let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.

   (I) $0 \leq \|\mathbf{x}\|$
   (II) $\|\mathbf{x}\| = 0$ if and only if $\mathbf{x} = \mathbf{0}$
   (III) $\|\alpha\mathbf{x}\| = |\alpha|\|\mathbf{x}\|$
   (IV) $\|\mathbf{x} + \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + 2(\mathbf{x} \cdot \mathbf{y}) + \|\mathbf{y}\|^2$
   (V) [Cauchy-Schwarz] $|\mathbf{x} \cdot \mathbf{y}| \leq \|\mathbf{x}\|\|\mathbf{y}\|$
   (VI) [Triangle Inequality] $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$

   I encountered a proof for the Cauchy-Schwarz inequality, which I'd like to have it here.

   Proof. (V) For each $t \in \mathbb{R}$, we have

$$0 \leq \|\mathbf{x} + t\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + 2(\mathbf{x} \cdot (t\mathbf{y})) + \|t\mathbf{y}\|^2 = \|\mathbf{y}\|^2t^2 + (2(\mathbf{x} \cdot \mathbf{y}))t + \|\mathbf{x}\|^2$$

   This ensures that the quadratic cannot have distinct real roots, and so it must be that

$$[2(\mathbf{x} \cdot \mathbf{y})]^2 - 4\|\mathbf{y}\|^2\|\mathbf{x}\|^2 \leq 0$$

holds. Consequently,

$$4(\mathbf{x}\cdot\mathbf{y})^2 \le 4\|\mathbf{x}\|^2\|\mathbf{y}\|^2$$

and therefore

$$|\mathbf{x}\cdot\mathbf{y}| \le \|\mathbf{x}\|\|\mathbf{y}\|$$

□

3. Given $\mathbf{z}, \mathbf{w} \in \mathbb{C}^n$, we can form their dot product

$$\mathbf{z}\cdot\mathbf{w} = \mathbf{z}_1\mathbf{w}_1 + \mathbf{z}_2\mathbf{w}_2 + \ldots + \mathbf{z}_n\mathbf{w}_n$$

However, it is no longer the case that $\mathbf{z}\cdot\mathbf{z}$ is a non-negative real number; for example, if

$$\mathbf{z} = \begin{bmatrix} i \\ 0 \end{bmatrix}$$

then $\mathbf{z}\cdot\mathbf{z} = -1$. To make this consistent with reals, we define the *inner product* of $\mathbf{z}, \mathbf{w} \in \mathbb{C}^n$ to be

$$\langle \mathbf{z}, \mathbf{w}\rangle = \mathbf{z}_1\overline{\mathbf{w}}_1 + \ldots + \mathbf{z}_n\overline{\mathbf{w}}_n$$

With this definition, we can show the following properties: The following facts are true regarding this:

(I) $\langle \mathbf{z}, \mathbf{z}\rangle = |\mathbf{z}_1|^2 + \ldots + |\mathbf{z}_n|^2$; Reminder: $z\overline{z} = |z|^2$

(II) $0 \le \langle \mathbf{z}, \mathbf{z}\rangle$ and $\langle \mathbf{z}, \mathbf{z}\rangle = 0$ if and only if $\mathbf{z} = \mathbf{0}$

(III) $\langle \mathbf{z}, \mathbf{w}\rangle = \overline{\langle \mathbf{w}, \mathbf{z}\rangle}$: we lose the symmetry with inner products.

(IV) $\langle \alpha\mathbf{z}, \mathbf{w}\rangle = \alpha\langle \mathbf{z}, \mathbf{w}\rangle$ and $\langle \mathbf{z}, \alpha\mathbf{w}\rangle = \overline{\alpha}\langle \mathbf{z}, \mathbf{w}\rangle$

(V) $\langle \mathbf{z}+\mathbf{w}, \mathbf{u}\rangle = \langle \mathbf{z}, \mathbf{u}\rangle + \langle \mathbf{w}, \mathbf{u}\rangle$ and $\langle \mathbf{z}, \mathbf{w}+\mathbf{u}\rangle = \langle \mathbf{z}, \mathbf{w}\rangle + \langle \mathbf{z}, \mathbf{u}\rangle$

4. We say that $\mathbf{v}_1, \ldots, \mathbf{v}_m \in \mathbb{R}^n$ are *orthogonal* if

$$\mathbf{v}_j \cdot \mathbf{v}_k = 0$$

whenever $j \ne k$. For elements of $\mathbb{C}^n$, the dot product is replaced with the inner product. With this we can prove the Pythagorean theorem:

Proposition (Pythagorean Theorem)**.** Let $\mathbf{v}_1, \ldots, \mathbf{v}_m \in \mathbb{R}^n$ be orthogonal, and let $\alpha_1, \ldots, \alpha_m \in \mathbb{R}$. Then

$$\left\|\sum_{k=1}^m \alpha_k\mathbf{v}_k\right\|^2 = \sum_{k=1}^m (\alpha_k)^2\|\mathbf{v}_k\|^2$$

5. orthogonal sets of non-zero vectors are linearly independent.
One way to prove this is to use Pythagorean Theorem. Let $\mathbf{v}_1, \ldots, \mathbf{v}_m \in \mathbb{R}^n$ be orthogonal and such that $\mathbf{v}_k \ne \mathbf{0}$ for all $1 \le k \le m$. Then $\mathbf{v}_1, \ldots, \mathbf{v}_m$ are linearly independent.

Proof. Now, let $\alpha_1, \ldots, \alpha_m \in \mathbb{R}$ be such that

$$\mathbf{0} = \sum_{k=1}^m \alpha_k\mathbf{v}_k.$$

Then

$$0 = \left\| \sum_{k=1}^{m} \alpha_k \mathbf{v}_k \right\|^2 = \sum_{k=1}^{m} (\alpha_k)^2 \|\mathbf{v}_k\|^2.$$

As $\|\mathbf{v}_k\|^2 \neq 0$ holds for all $1 \leq k \leq m$, it must be that $\alpha_k = 0$ for all $1 \leq k \leq m$. $\square$

6. **Gram-Schmidt Process:** Now that we know orthogonal sets of non-zero vectors are linearly independent, it's important to note that the inverse is not true; it's possible for a set to be linearly independent meanwhile they fail to be orthogonal. However, any linearly independent set can be transformed into an orthogonal one using Gram-Schmidt Process. This process is done inductively. Suppose that we have constructed an orthogonal $\mathbf{v}_1, \ldots, \mathbf{v}_p \in \mathbb{R}^n$ such that $\mathbf{v}_k \neq 0$ for all $1 \leq k \leq p$ and $\text{span}(\mathbf{x}_1, \ldots, \mathbf{x}_p) = \text{span}(\mathbf{v}_1, \ldots, \mathbf{v}_p)$. Then, we'll define $\mathbf{v}_{p+1} \in \mathbb{R}^n$ using the following:

$$\mathbf{v}_{p+1} = \mathbf{x}_{p+1} - \sum_{k=1}^{p} \frac{\mathbf{x}_{p+1} \cdot \mathbf{v}_k}{\mathbf{v}_k \cdot \mathbf{v}_k} \mathbf{v}_k$$

7. We can upgrade the definition of orthogonal sets as follows: we say that $\mathbf{u}_1, \ldots, \mathbf{u}_m \subset \mathbb{R}^n$ are *orthonormal* if it is orthogonal and each $\mathbf{u}_k$ is a *unit vector*; i.e.,

$$\|\mathbf{u}_k\| = 1$$

This definition is really interesting as it gives us some nice properties about the span of such sets. For instance, based on this definition, $\forall \mathbf{x} \in \text{span}(\mathbf{u}_1, \ldots, \mathbf{u}_m)$ where $\mathbf{u}_1, \ldots, \mathbf{u}_m \in \mathbb{R}^n$ is orthonormal, we can show

$$\mathbf{x} = \sum_{k=1}^{m} (\mathbf{x} \cdot \mathbf{u}_k) \mathbf{u}_k$$

Also, we know how to make an orthogonal set out of linearly independent sets using the Gram-Schmidt Process. If we want this process to give us orthonormal sets, all we have to do is to normalize the set we want to make an orthogonal set from, and then apply the process.

Sample Questions

1. Perform the Gram-Schmidt process to

$$\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix}$$

Answer. Suppose the given vectors are x_1, x_2, x_3 from left to right. We proceed inductively by:

$$v_1 = x_1, \quad v_{p+1} = x_{p+1} - \sum_{k=1}^{p} \frac{x_{p+1} \cdot v_k}{v_k \cdot v_k} v_k$$

Hence

$$v_1 = x_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$$

$$\begin{aligned}
v_2 &= x_2 - \frac{x_2 \cdot v_1}{v_1 \cdot v_1} v_1 \\
&= \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} - \frac{\begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}}{\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \\
&= \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} - \begin{bmatrix} \frac{3}{2} \\ \frac{3}{2} \\ 0 \end{bmatrix} \\
&= \begin{bmatrix} -\frac{1}{2} \\ \frac{1}{2} \\ 0 \end{bmatrix}
\end{aligned}$$

$$\begin{aligned}
v_3 &= x_3 - \frac{x_3 \cdot v_1}{v_1 \cdot v_1} v_1 - \frac{x_2 \cdot v_2}{v_2 \cdot v_2} v_2 \\
&= \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} - \frac{\begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}}{\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} - \frac{\begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} -\frac{1}{2} \\ \frac{1}{2} \\ 0 \end{bmatrix}}{\begin{bmatrix} -\frac{1}{2} \\ \frac{1}{2} \\ 0 \end{bmatrix} \cdot \begin{bmatrix} -\frac{1}{2} \\ \frac{1}{2} \\ 0 \end{bmatrix}} \begin{bmatrix} -\frac{1}{2} \\ \frac{1}{2} \\ 0 \end{bmatrix} \\
&= \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} - \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} \\
&= \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}
\end{aligned}$$

2. Perform the Gram–Schmidt process on the following:

$$\begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 3 \\ 1 \end{bmatrix}$$

Answer. Suppose the given vectors are x_1, x_2, x_3 from left to right. We will proceed inductively using the following:

$$v_1 = x_1, \quad v_{p+1} = x_{p+1} - \sum_{k=1}^{p} \frac{x_{p+1} \cdot v_k}{v_k \cdot v_k} v_k$$

Hence

$$v_1 = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}$$

$$\begin{aligned}
v_2 &= x_2 - \frac{x_2 \cdot v_1}{v_1 \cdot v_1} v_1 \\
&= \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} - \frac{\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}}{\begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}} \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix} \\
&= \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} - 0 \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix} \\
&= \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}
\end{aligned}$$

$$\begin{aligned}
v_3 &= x_3 - \frac{x_3 \cdot v_1}{v_1 \cdot v_1} v_1 - \frac{x_3 v_2}{v_2 \cdot v_2} v_2 \\
&= \begin{bmatrix} 0 \\ 3 \\ 1 \end{bmatrix} - \frac{\begin{bmatrix} 0 \\ 3 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}}{\begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}} \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix} - \frac{\begin{bmatrix} 0 \\ 3 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}}{\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}} \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \\
&= \begin{bmatrix} 0 \\ 3 \\ 1 \end{bmatrix} - \begin{bmatrix} \frac{4}{3} \\ \frac{4}{3} \\ \frac{4}{3} \end{bmatrix} - \begin{bmatrix} -\frac{1}{2} \\ 0 \\ \frac{1}{2} \end{bmatrix} \\
&= \begin{bmatrix} -\frac{5}{6} \\ \frac{5}{3} \\ -\frac{5}{6} \end{bmatrix}
\end{aligned}$$

Matrices

1. In the community of Mathematics, matrices are written using bold symbols; i.e., $\mathbf{A} \in \mathbb{R}^{m \times n}$. This also includes column vectors.

2. We will treat $\mathbb{R}^{n \times 1}$ and $\mathbb{R}^n$ as "the same." Both refer to column vectors.

3. Give $\mathbf{A} \in \mathbb{R}^{m \times n}$, the *jth row of* $\boldsymbol{A}$ is $\mathbf{A}_{j,*} \in \mathbb{R}^n$ and the *kth column of* $\boldsymbol{A}$ is $\mathbf{A}_{*,k} \in \mathbb{R}^m$. The interesting thing is how both $\mathbf{A}_{j,*}$ and $\mathbf{A}_{*,k}$ are both column vectors. Take the following example:

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

Then

$$\mathbf{A}_{1,*} = \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix} \quad \text{and} \quad \mathbf{A}_{*,3} = \begin{bmatrix} 3 \\ 0 \end{bmatrix}$$

It seems that we don't find row vectors wandering around the space by themselves unless we transpose column vectors. One reason for this might be the way linear algebra was originally formed, primarily to solve the following system of linear equations:

$$\begin{aligned} a_{1,1}x_1 + a_{1,2}x_2 + \ldots + a_{1,n}x_n &= b_1 \\ a_{2,1}x_1 + a_{2,2}x_2 + \ldots + a_{2,n}x_n &= b_2 \\ \vdots \qquad\qquad &\quad \vdots \\ a_{m,1}x_1 + a_{m,2}x_2 + \ldots + a_{m,n}x_n &= b_m \end{aligned}$$

which can now be viewed as

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

where

$$\mathbf{A} = \begin{bmatrix} a_{1,1} & \ldots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \ldots & a_{m,n} \end{bmatrix} \quad \mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$$

4. One interesting note is that:

$$(\mathbf{A}_{j,*})_k = (\mathbf{A}_{*,k})_j = \mathbf{A}_{j,k}$$

5. Four things are defined for matrices which will help us to prove some properties about them; Given $\alpha \in \mathbb{R}$ and $\mathbf{0}, \mathbf{A}, \mathbf{B} \in \mathbb{R}^{m \times n}$, then:

$$\begin{aligned} \mathbf{0}_{j,k} = 0 \quad &\text{and} \quad (-\mathbf{A})_{j,k} = -\mathbf{A}_{j,k} \\ (\alpha\mathbf{A})_{j,k} = \alpha\mathbf{A}_{j,k} \quad &\text{and} \quad (\mathbf{A}+\mathbf{B})_{j,k} = \mathbf{A}_{j,k} + \mathbf{B}_{j,k} \end{aligned}$$

With just these, we can prove the following properties:

(I) $\mathbf{A} + (\mathbf{B} + \mathbf{C}) = (\mathbf{A} + \mathbf{B}) + \mathbf{C}$

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

(III) $\mathbf{A} + \mathbf{0} = \mathbf{A}$

(IV) $\mathbf{A} + (-\mathbf{A}) = \mathbf{0}$

(V) $\alpha(\mathbf{A} + \mathbf{B}) = \alpha\mathbf{A} + \alpha\mathbf{B}$

(VI) $(\alpha + \beta)\mathbf{A} = \alpha\mathbf{A} + \beta\mathbf{A}$

(VII) $(\alpha\beta)\mathbf{A} = \alpha(\beta\mathbf{A})$

(VIII) $1\mathbf{A} = \mathbf{A}$

6. In linear algebra, the transpose of a matrix is an operator which flips a matrix over its diagonal; that is, it switches the row and column indices of the matrix A by producing another matrix:

$$(\mathbf{A}^\top)_{j,k} = \mathbf{A}_{k,j}$$

Again, with this simple definition, let $\alpha \in \mathbb{R}$, and let $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m \times n}$. we can prove the following properties:

(I) $(\mathbf{A}^\top)^\top = \mathbf{A}$

(II) $(\alpha\mathbf{A})^\top = \alpha\mathbf{A}^\top$

(III) $(\mathbf{A}+\mathbf{B})^\top = \mathbf{A}^\top + \mathbf{B}^\top$

Note that we have no idea what $\mathbf{AB}$ or $(\mathbf{AB})^\top$ is yet, since we haven't defined matrix multiplication.

7. Same as transpose, we have the conjugate of the matrix which is applying complex conjugation to each entry of the matrix:

$$\overline{\mathbf{A}} \in \mathbb{C}^{m\times n} \quad \text{by} \quad (\overline{\mathbf{A}})_{j,k} = \overline{\mathbf{A}_{j,k}}$$

Once again, with this simple definition, we can prove the following properties:

(I) $\overline{\alpha\mathbf{A}} = \overline{\alpha}\overline{\mathbf{A}}$

(II) $\overline{\mathbf{A}+\mathbf{B}} = \overline{\mathbf{A}} + \overline{\mathbf{B}}$

(III) $\overline{\overline{\mathbf{A}}} = \mathbf{A}$

7 & 6. Combining point 6 and 7, we will have the *adjoint* of the matrix A, $\mathbf{A}^* \in \mathbb{C}^{n\times m}$ which is the conjugate transpose of $\mathbf{A}$;

$$\mathbf{A}^* = (\overline{\mathbf{A}})^\top$$

Based on what we know about the properties of the conjugate and transpose of a matrix and the definition of the adjoint, we can prove the following properties:

(I) $\mathbf{A}^* = \overline{\mathbf{A}^\top}$

(II) Let $\mathbf{B} \in \mathbb{C}^{m\times n}$. Then $(\mathbf{A}+\mathbf{B})^* = \mathbf{A}^* + \mathbf{B}^*$

(III) Let $\alpha \in \mathbb{C}$. Then $(\alpha\mathbf{A})^* = \overline{\alpha}\mathbf{A}^*$

(IV) $(\mathbf{A}^*)^* = \mathbf{A}$

Since I like the proof of the last two properties, I'll write them to be here.

Proof. (III) Indeed,

$$(\alpha\mathbf{A})^* = (\overline{\alpha\mathbf{A}})^\top = (\overline{\alpha}\overline{\mathbf{A}})^\top = \overline{\alpha}(\overline{\mathbf{A}})^\top = \overline{\alpha}\mathbf{A}^*$$

(IV) Assume (I). Then,

$$(\mathbf{A}^*)^* = \overline{((\overline{\mathbf{A}})^\top)^\top} = \overline{(\overline{\mathbf{A}})} = \mathbf{A}$$

$\square$

8. Now we will jump into matrix multiplication. Matrix multiplication is a binary operation

$$\circ : \mathbb{R}^{m\times p} \times \mathbb{R}^{p\times n} \to \mathbb{R}^{m\times n}$$

such that

$$(\mathbf{AB})_{j,k} = \sum_{h=1}^{p} \mathbf{A}_{j,h}\mathbf{B}_{h,k}$$

Also, if we look at each element of $\mathbf{AB}$, it is the dot product of the two following vectors:

$$(\mathbf{AB})_{j,k} = \mathbf{A}_{j,*} \cdot \mathbf{B}_{*,k}$$

Which aligns with $\mathbf{A}_{j,*}$ and $\mathbf{B}_{*,k}$ being column vectors.

9. Before showing what properties hold for matrix multiplication, let's remind ourselves what doesn't!

   - $\mathbf{AB}$ and $\mathbf{BA}$ may not be equal (or even exist).
   - Non-zero matrices can multiply to zero:

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

   - Cancellation is not possible:

   $$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ \pi & \sqrt{2} \end{bmatrix}$$

10. The nice properties of matrix multiplications which can be proved by using the definition are the following:

    Proposition. Let $\mathbf{A}, \mathbf{D} \in \mathbb{R}^{m \times p}$, let $\mathbf{B}, \mathbf{C} \in \mathbb{R}^{p \times n}$, and let $\alpha \in \mathbb{R}$.

    (I) $\mathbf{A}(\mathbf{B}+\mathbf{C}) = \mathbf{AB} + \mathbf{AC}$ and $(\mathbf{A}+\mathbf{D})\mathbf{B} = \mathbf{AB} + \mathbf{DB}$

    (II) $\alpha(\mathbf{AB}) = (\alpha\mathbf{A})\mathbf{B}$ and $\alpha(\mathbf{AB}) = \mathbf{A}(\alpha\mathbf{B})$

    (III) $\mathbf{A0} = \mathbf{0}$ and $\mathbf{0A} = \mathbf{0}$

    (IV) $-(\mathbf{AB}) = (-\mathbf{A})\mathbf{B}$ and $-(\mathbf{AB}) = \mathbf{A}(-\mathbf{B})$

    (V) $\mathbf{AB} = (-\mathbf{A})(-\mathbf{B})$

    (VI) $\mathbf{AI}_p = \mathbf{A}$ and $\mathbf{I}_m\mathbf{A} = \mathbf{A}$

    (VII) $(\mathbf{AB})^\top = \mathbf{B}^\top\mathbf{A}^\top$

    (VIII) $(\mathbf{AB})^* = \mathbf{B}^*\mathbf{A}^*$

    Again, I like the proof of the last two properties, so I'll write them to be here.

    Proof. (VII) Indeed,

    $$((\mathbf{AB})^\top)_{j,k} = (\mathbf{AB})_{k,j} = \sum_{h=1}^{p} \mathbf{A}_{k,h}\mathbf{B}_{h,j} = \sum_{h=1}^{p} (\mathbf{A}^\top)_{h,k}(\mathbf{B}^\top)_{j,h} = \sum_{h=1}^{p} (\mathbf{B}^\top)_{j,h}(\mathbf{A}^\top)_{h,k} = (\mathbf{B}^\top\mathbf{A}^\top)_{j,k}$$

(VIII)

$$
\begin{aligned}
((\mathbf{AB})^*)_{j,k} &= (\overline{(\mathbf{AB})^\top})_{j,k} \\
&= \overline{(\mathbf{AB})^\top_{j,k}} \\
&= \overline{(\mathbf{B}^\top\mathbf{A}^\top)_{j,k}} \\
&= \overline{\sum_{h=1}^{p} \mathbf{B}^\top_{j,h} \cdot \mathbf{A}^\top_{h,k}} \\
&= \sum_{h=1}^{p} \overline{\mathbf{B}^\top_{j,h} \cdot \mathbf{A}^\top_{h,k}} \\
&= \sum_{h=1}^{p} \overline{\mathbf{B}^\top_{j,h}} \cdot \overline{\mathbf{A}^\top_{h,k}} \\
&= \sum_{h=1}^{p} (\overline{\mathbf{B}^\top})_{j,h} \cdot (\overline{\mathbf{A}^\top})_{h,k} \\
&= \sum_{h=1}^{p} (\mathbf{B}^*)_{j,h} \cdot (\mathbf{A}^*)_{h,k} \\
&= \mathbf{B}^*\mathbf{A}^*
\end{aligned}
$$

□

11. We now know that any system of linear equations

$$
\begin{aligned}
a_{1,1}x_1 + a_{1,2}x_2 + \ldots + a_{1,n}x_n &= b_1 \\
a_{2,1}x_1 + a_{2,2}x_2 + \ldots + a_{2,n}x_n &= b_2 \\
\vdots \qquad & \qquad \vdots \\
a_{m,1}x_1 + a_{m,2}x_2 + \ldots + a_{m,n}x_n &= b_m
\end{aligned}
$$

can written as the matrix equation

$$\mathbf{Ax} = \mathbf{b}$$

where

$$
\mathbf{A} = \begin{bmatrix} a_{1,1} & \ldots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \ldots & a_{m,n} \end{bmatrix} \quad \mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}
$$

Since matrix multiplication is associative, we now know that if we can find a matrix $\mathbf{B}$ with $\mathbf{BA} = \mathbf{I}$ then we can solve for $\mathbf{x}$ as follows:

$$\mathbf{x} = \mathbf{Ix} = (\mathbf{BA})\mathbf{x} = \mathbf{B}(\mathbf{Ax}) = \mathbf{Bb}$$

For the inverse of $\mathbf{A}$ to exist, first and foremost, $\mathbf{A}$ has to be a square matrix. Now, let $\mathbf{A} \in \mathbb{R}^{n\times n}$. Then $\mathbf{A}$ is invertible if and only if any of the following conditions hold:

(a) The reduced row echelon form of A is I.
(b) The columns of A are linearly independent.
(c) The rows of A are linearly independent.
(d) $\mathcal{N}(A) = 0$.

12. Now, I'll prove the following famous formula which we all have encountered at least once in our life:
Let

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

Then $ad - bc \neq 0$ if and only if $\mathbf{A}$ is invertible. In which case,

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

Proof. Let

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

And

$$\mathbf{A}^{-1} = \begin{bmatrix} e & f \\ g & h \end{bmatrix}$$

Hence

$$\begin{bmatrix} e & f \\ g & h \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} ae + cf & be + df \\ ag + ch & bg + dh \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Our approach is to systematically solve each equation twice, isolating different variables from e to h in each instance. By equating the resulting expressions, we derive explicit formulas for e to h in terms of a to d.

$$ae + cf = 1$$

Solving for e and f:

$$e = \frac{1 - cf}{a}, \quad f = \frac{1 - ae}{c}$$

Equating:

$$1 - ae = -be \quad \Rightarrow \quad d - ade = -bce$$

$$e = \frac{d}{ad - bc}, \quad f = \frac{-b}{ad - bc}$$

$$be + df = 0$$

Solving for e and d:

$$e = \frac{-df}{b}, \quad d = \frac{ade - bce}{d}$$

Equating:

$$d = e(ad - bc)$$

$$ag + ch = 0$$

Solving for g and h:

$$g = \frac{-ch}{a}, \quad h = \frac{-ag}{c}$$

Equating:

$$a - adh = -bch$$

$$h = \frac{a}{ad - bc}$$

$$bg + dh = 1$$

Solving for g and h:

$$g = \frac{1 - dh}{b}, \quad h = \frac{1 - bg}{d}$$

Equating:

$$c - bcg = -adg$$

$$g = \frac{c}{ad - bc}$$

Thus, the inverse of A is:

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

$\square$

13. As usual, the following properties hold for an invertible matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$:

(I) $\mathbf{A}^{-1}$ is invertible and $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$.

(II) Let $\mathbf{B}$ be invertible. Then $\mathbf{AB}$ is invertible and $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$.

(III) $\mathbf{A}^\top$ is invertible and $(\mathbf{A}^\top)^{-1} = (\mathbf{A}^{-1})^\top$

(IV) Let $\alpha \in \mathbb{R} \setminus \{0\}$. Then $\alpha\mathbf{A}$ is invertible and $(\alpha\mathbf{A})^{-1} = \frac{1}{\alpha}\mathbf{A}^{-1}$.

Now that we know enough about matrix multiplications, I'd like to show the following:

$$\frac{\partial RSS(\mathbf{w})}{\partial \mathbf{w}} = -2\mathbf{X}^T\mathbf{y} + 2\mathbf{X}^T\mathbf{X}\mathbf{w}$$

Which represents the gradient of the Residual Sum of Squares (RSS) function with respect to the weight vector w in a linear regression problem. The RSS function measures the discrepancy between the observed values y and the predicted values $\mathbf{w}^\top\mathbf{X}$, where $\mathbf{X}$ is the feature matrix.

$$RSS(\mathbf{w}) = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

Equivalently:

$$\begin{aligned}
RSS(\mathbf{w}) &= (\mathbf{y} - \mathbf{X}\mathbf{w})^\top(\mathbf{y} - \mathbf{X}\mathbf{w}) \\
&= (\mathbf{y}^\top - (\mathbf{X}\mathbf{w})^\top)(\mathbf{y} - \mathbf{X}\mathbf{w}) \\
&= (\mathbf{y}^\top - \mathbf{w}^\top\mathbf{X}^\top)(\mathbf{y} - \mathbf{X}\mathbf{w}) \\
&= \mathbf{y}^\top\mathbf{y} - \mathbf{w}^\top\mathbf{X}^\top\mathbf{y} - \mathbf{y}^\top\mathbf{X}\mathbf{w} + \mathbf{w}^\top\mathbf{X}^\top\mathbf{X}\mathbf{w} \\
&= \mathbf{y}^\top\mathbf{y} - 2\mathbf{w}^\top\mathbf{X}^\top\mathbf{y} + \mathbf{w}^\top\mathbf{X}^\top\mathbf{X}\mathbf{w}
\end{aligned}$$

Now

$$\begin{aligned}
\frac{\partial RSS(\mathbf{w})}{\partial \mathbf{w}} &= \frac{\partial(\mathbf{y}^\top\mathbf{y} - 2\mathbf{w}^\top\mathbf{X}^\top\mathbf{y} + \mathbf{w}^\top\mathbf{X}^\top\mathbf{X}\mathbf{w})}{\partial \mathbf{w}} \\
&= 0 - 2\mathbf{X}^\top\mathbf{y} + 2\mathbf{X}^\top\mathbf{X}\mathbf{w}
\end{aligned}$$

Consequently, to find the global optimum of $RSS(\mathbf{w})$ we have to solve the following equation:

$$\begin{aligned}
\mathbf{0} &= -2\mathbf{X}^\top\mathbf{y} + 2\mathbf{X}^\top\mathbf{X}\mathbf{w} \\
&= -\mathbf{X}^\top\mathbf{y} + \mathbf{X}^\top\mathbf{X}\mathbf{w} \\
&\Rightarrow \mathbf{X}^\top\mathbf{X}\mathbf{w} = \mathbf{X}^\top\mathbf{y} \\
&\Rightarrow \mathbf{w} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y} \\
&\Rightarrow \mathbf{w} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y} \\
&\Rightarrow \mathbf{w} = \mathbf{X}^{-1}(\mathbf{X}^\top)^{-1}\mathbf{X}^\top\mathbf{y}
\end{aligned}$$

Sample Questions

Linear Transformations

Sample Questions

1. Let $T\colon V \to W$ be a linear transform, and let $\dim V = \dim W$
   (a) If T is one-to-one, then prove that T is a linear isomorphism.
   (b) If T is onto, then prove that T is a linear isomorphism.

Answer. In both parts, we would like to show that T is both one-to-one and onto leading to being a linear isomorphism.

(a) (a) Since T is one-to-one, then for a basis $B \subset V$, $T[B]$ is a basis for $T[V]$, hence $\dim T[V] = \dim V$ which results in $\dim T[V] = \dim W$. Therefore, T is onto.

(b) (b) Fix a basis for W such as $w_1, w_2, \cdots w_n$. Since T is onto, then exist $v_1, v_2 \cdots v_n \in V$ such that

$$T(v_i) = w_i$$

Furthermore, as $\dim V = \dim W$, we have $v_1, \cdots, v_n$ as a basis for V.
Now, let $x \in \ker(T)$. Then $x = \alpha_1 v_1 + \cdots + \alpha_n v_n$ holds for some $\alpha_1, \cdots, \alpha_n \in \mathbb{R}$. Then,

$$0_W = T(x) = T(\alpha_1 v_1 + \cdots + \alpha_n v_n) = \alpha_1 T(v_1) + \cdots + \alpha_n T(v_n) = \alpha_1 w_1 + \cdots + \alpha_n w_n.$$

Since $w_1, \cdots, w_n$ are linearly independent, it follows that $\alpha_1, \cdots, \alpha_n = 0$, which means:

$$x = \alpha_1 v_1 + \cdots + \alpha_n v_n = 0_V,$$

and

$$\ker(T) = \{0_V\}.$$

Hence, T is one-to-one.

2. (a) Prove that there is no onto linear transformation $T\colon \mathbb{R}^2 \to \mathbb{R}^3$.
   (b) Create a one-to-one linear transformation $T\colon \mathbb{R}^2 \to \mathbb{R}^3$
   (c) Prove that there is no one-to-one linear transformation $T\colon \mathbb{R}^3 \to \mathbb{R}^2$
   (d) Create an onto linear transformation $T\colon \mathbb{R}^3 \to \mathbb{R}^2$

Answer. Let $\mathbf{A}$ be the matrix associated with this linear transformation. Then $\mathbf{A} \in \mathbb{R}^{3\times 2}$

(a) Based on rank-nullity therm, we know:

$$\dim T[v] = \dim \operatorname{col}(\mathbf{A}), \quad \dim \operatorname{Ker}(T[v]) = \dim \backslash(\mathbf{A})$$

Hence

$$\dim T[v] + \dim \operatorname{Ker}(T[v]) = 2$$

However, since T is onto, we have $\dim T[v] = \dim W = 3$, which results in $\dim \operatorname{Ker}(T[v]) = -1$.

(b) Consider the following linear mapping

$$T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x \\ y \\ 0 \end{bmatrix}$$

Then

$$T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad \text{and} \quad T\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

And

$$T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

Now, suppose $v = \begin{bmatrix} x \\ y \end{bmatrix} \in \ker(T)$, then

$$T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x \\ y \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Which results in $x = 0, y = 0$, and $\mathbf{v} = \mathbf{0}$. Hence the kernel is trivial, and T in onto.

(c) Let $\mathbf{A}$ be the matrix associated with this linear transformation. Then $\mathbf{A} \in \mathbb{R}^{2\times 3}$, which means

$$\text{Rank}(\mathbf{A}) \leq 2$$

. Based on rank-nullity therm, we know:

$$\dim T[v] = \dim \text{col}(\mathbf{A}), \quad \dim \text{Ker}(T[v]) = \dim \backslash(\mathbf{A})$$

And

$$\text{Rank}(\mathbf{A}) + \dim \text{Ker}(T[v]) = 3$$

Which means $\dim \text{Ker}(T[v]) \geq 1$ and kernel in not trivial, and T is one-to-one.

(d) Consider the following linear mapping

$$T\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} x \\ y \end{bmatrix}$$

Then given $w \in W$

$$\begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

We can consider $v = \begin{bmatrix} w_1 \\ w_2 \\ 0 \end{bmatrix}$, which leads to

$$T\left(\begin{bmatrix} w_1 \\ w_2 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

Determinants

1. **Definition**: The determinant is a function that takes a square matrix as an input and produces a scalar as an output. Hence, every square matrix has a value associated with called determinant, which will reveal properties such as its null space.

2. Representation: for a matrix $\mathbf{A}$, the determinant is shown with:

$$\det(\mathbf{A}) = \mid \mathbf{A} \mid$$

3. Determinant most importantly tells us about the singularity of a matrix. Matrix $\mathbf{A}$ is singular if its determinant is zero. Equivalently, is invertible, if the determinate is non-zero. Hence, the determinant is a test for invertibility.

4. The big reason that we need determinants is that they help us to calculate eigen values.

5. **Properties**:

   (a) $\det(\mathbf{I}) = 1$

   (b) exchange a row = reverse the sign of the determinate. Therefore, having n exchanges leads to:

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

   This means you cannot produce the same matrix with seven row exchanges, or the matrix that you get with ten row exchanges isn't the same as the one you get with sevens.

   (c) scaling a row by t scales the determinate:

   $$\det(\mathbf{A'}) = t \det(\mathbf{A})$$

   hence

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

   This is like volume! Suppose we have a box, and we double all the sides, then the volume is multiplied by 8.

   (d) two identical rows imply that $\det = 0$, which means the matrix is not invertible. When two rows are equal, the rank of the matrix is less than n, indicating that there are fewer than n linearly independent rows. As a result, the matrix is singular.

   (e) subtract $l \times row_i$ from row_k. Then the determinate doesn't change.

   (f) this property can expand to square matrices with higher dimensions which is:

   $$\det\left(\begin{bmatrix} a+a' & b+b' \\ c & d \end{bmatrix}\right) = \det\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) + \det\left(\begin{bmatrix} a' & b' \\ c & d \end{bmatrix}\right)$$

   (g) scaling a row and adding it to another $\Rightarrow$ the determinant stays the same. Now, suppose we scale the first row by a scalar k and add it to the second row:

   $$\begin{aligned} \det\left(\begin{bmatrix} a & b \\ c+ka & d+kb \end{bmatrix}\right) &= \det\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) + \det\left(\begin{bmatrix} a & b \\ ka & kb \end{bmatrix}\right) \\ &= \det(\mathbf{A}) + k \det\left(\begin{bmatrix} a & b \\ a & b \end{bmatrix}\right) \\ &= \det(\mathbf{A}) + 0 \\ &= \det(\mathbf{A}) \end{aligned}$$

   (h) if a matrix $\mathbf{A}$ has a row of zeros $\Rightarrow \det(\mathbf{A}) = 0$.

   (i) $\det(\mathbf{AB}) = \det(\mathbf{A}) \det(\mathbf{B})$

   (j) $\det(\mathbf{A}^\top) = \det(\mathbf{A})$. This says a lot! all the properties that evolved around row operations are also true with respect to column operations since doing the column operations is equivalent to doing the row operation on $\mathbf{A}^\top$ which is determinant is equal to the $\mathbf{A}$ determinant.

Suppose that A is a square matrix of size n. The following are equivalent.

i. A is nonsingular.

ii. A row-reduces to the identity matrix.

iii. The null space of A contains only the zero vector, $\mathcal{N}(A) = \{\mathbf{0}\}$.

iv. The linear system $\mathcal{LS}(A, \mathbf{b})$ has a unique solution for every possible choice of $\mathbf{b}$.

v. The columns of A are a linearly independent set.

vi. A is invertible.

vii. The column space of A is $\mathbb{C}^n$, $C(A) = \mathbb{C}^n$.

viii. The columns of A are a basis for $\mathbb{C}^n$.

ix. The rank of A is n, $r(A) = n$.

x. The nullity of A is zero, $n(A) = 0$.

xi. The determinant of A is nonzero, $\det(A) \neq 0$.

Sample Questions

1. Suppose we have the following triangular matrix:

$$\mathbf{A} = \begin{bmatrix} d_1 & * & * & \cdots & * \\ 0 & d_2 & * & \cdots & * \\ 0 & 0 & d_3 & \cdots & * \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & d_n \end{bmatrix}$$

show that:

$$\det(\mathbf{A}) = d_1 d_2 \cdots d_n$$

Answer. Do the following operation to make the non-diagonal entries zero. Since all of these operations involve scaling a row, and adding it to another, the determinate won't change.

for $j = n$ **to** 2 **step** -1 **do**
    for $i = 1$ **to** $j - 1$ **do**
        $\text{Row}_i \leftarrow \text{Row}_i - \dfrac{a_{i,j}}{a_{j,j}} \cdot \text{Row}_j$
    end for
end for

Hence, the matrix that we'll end up with is:

$$\mathbf{A} = \begin{bmatrix} d_1 & 0 & 0 & \cdots & 0 \\ 0 & d_2 & 0 & \cdots & 0 \\ 0 & 0 & d_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & d_n \end{bmatrix}$$

and $\det(\mathbf{A}) = d_1 d_2 d_3 \cdots d_n \det(\mathbf{I}) = d_1 d_2 d_3 \cdots d_n$

2. What's the $\det\left(\mathbf{A}^{-1}\right)$?
Answer. We know:

$$1 = \det(\mathbf{I}) = \det\left(\mathbf{A}^{-1}\mathbf{A}\right) = \det\left(\mathbf{A}^{-1}\right)\det(\mathbf{A})$$

Hence

$$\det\left(\mathbf{A}^{-1}\right) = \frac{1}{\det(\mathbf{A})}$$

3. Doing the computation **by hand**, find the determinant of the matrix

$$\mathbf{A} = \begin{bmatrix} -2 & 3 & -2 \\ -4 & -2 & 1 \\ 2 & 4 & 2 \end{bmatrix}$$

Answer.

$$\begin{aligned} \det(\mathbf{A}) &= (-1)^{1+1}(-2)\det\left(\begin{bmatrix} -2 & 1 \\ 4 & 2 \end{bmatrix}\right) + (-1)^{1+2}(3)\det\left(\begin{bmatrix} -4 & 1 \\ 2 & 2 \end{bmatrix}\right) + (-1)^{1+3}(-2)\det\left(\begin{bmatrix} -4 & -2 \\ 2 & 4 \end{bmatrix}\right) \\ &= 16 + 30 + 24 \\ &= 70 \end{aligned}$$

4. (a) Let a be a non-zero real number. Verify that the matrix

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

satisfies $\mathbf{A}^2 = -\mathbf{I}$

(b) Prove that there is no 3×3 real matrix $\mathbf{A}$ such that $\mathbf{A}^2 = -\mathbf{I}$

Answer.

(a)

$$\begin{aligned} \mathbf{AA} &= \begin{bmatrix} 0 & a \\ -1/a & 0 \end{bmatrix}\begin{bmatrix} 0 & a \\ -1/a & 0 \end{bmatrix} \\ &= \begin{bmatrix} (-1/a)\cdot a & 0 \\ 0 & a\cdot(-1/a) \end{bmatrix} \\ &= \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \end{aligned}$$

satisfies $\mathbf{A}^2 = -\mathbf{I}$

(b) Assume the contrary. Then:

$$\det\left(\mathbf{A}^2\right) = \det(\mathbf{A})\det(\mathbf{A}) = \det(\mathbf{A})^2 = \det(-\mathbf{I}) = (-1)^3\det(\mathbf{I}) = -1$$

which cannot bet true, since $\mathbf{A}$ is a real matrix.

(a) Consider matrix $\mathbf{A}$ be the following:

$$\mathbf{A} = \begin{bmatrix} -3 & -5 & -7 \\ 2 & 4 & 3 \\ 0 & 1 & -1 \end{bmatrix}$$

what's the $\det \mathbf{A}$, $\operatorname{adj}(\mathbf{A})$, and $\mathbf{A}^{-1}$.

Answer.

$$\begin{aligned} \det(\mathbf{A}) &= (-1)^{1+1}(-3)\det\left(\begin{bmatrix} 4 & 3 \\ 1 & -1 \end{bmatrix}\right) + (-1)^{1+2}(-5)\det\left(\begin{bmatrix} 2 & 3 \\ 0 & -1 \end{bmatrix}\right) + (-1)^{1+3}(-7)\det\left(\begin{bmatrix} 2 & 4 \\ 0 & 1 \end{bmatrix}\right) \\ &= (-3)(-7) + 5(-2) + (-7)(2) \\ &= -3 \end{aligned}$$

$$\operatorname{adj}(\mathbf{A}) =$$

(b) Doing the computations by hand, find the determinant of the matrix A.

$$A = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 2 & -1 & -1 & 1 \\ 2 & 5 & 3 & 0 \\ 1 & -1 & 0 & 1 \end{bmatrix}$$

Answer.

$$\begin{aligned} \det\left(\begin{bmatrix} 1 & 0 & 1 & 1 \\ 2 & -1 & -1 & 1 \\ 2 & 5 & 3 & 0 \\ 1 & -1 & 0 & 1 \end{bmatrix}\right) &= -\det\left(\begin{bmatrix} 1 & 0 & 1 & 1 \\ 2 & -1 & -1 & 1 \\ 1 & -1 & 0 & 1 \\ 2 & 5 & 3 & 0 \end{bmatrix}\right) = \det\left(\begin{bmatrix} 2 & -1 & -1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & -1 & 0 & 1 \\ 2 & 5 & 3 & 0 \end{bmatrix}\right) \\ &= \det\left(\begin{bmatrix} 3 & -1 & 0 & 2 \\ 1 & 0 & 1 & 1 \\ 1 & -1 & 0 & 1 \\ 2 & 5 & 3 & 0 \end{bmatrix}\right) = \det\left(\begin{bmatrix} 2 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & -1 & 0 & 1 \\ 2 & 5 & 3 & 0 \end{bmatrix}\right) \\ &= \det\left(\begin{bmatrix} 2 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & -1 & -1 & 0 \\ 2 & 5 & 3 & 0 \end{bmatrix}\right) = \det\left(\begin{bmatrix} 1 & 0 & -1 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & -1 & -1 & 0 \\ 2 & 5 & 3 & 0 \end{bmatrix}\right) \\ &= \det\left(\begin{bmatrix} 1 & 0 & -1 & 0 \\ 2 & 0 & 0 & 1 \\ 0 & -1 & -1 & 0 \\ 2 & 5 & 3 & 0 \end{bmatrix}\right) = \det\left(\begin{bmatrix} 1 & 0 & -1 & 0 \\ 2 & 0 & 0 & 1 \\ 0 & -1 & -1 & 0 \\ 0 & 5 & 5 & 0 \end{bmatrix}\right) \\ &= \det\left(\begin{bmatrix} 1 & 0 & -1 & 0 \\ 2 & 0 & 0 & 1 \\ 0 & -1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}\right) = 0 \end{aligned}$$

Eigenvalues, Eigenvectors, and Eigen-spaces

1. The act of multiplying a vector $\mathbf{x}$ by a matrix $\mathbf{A}$, resulting in another vector $\mathbf{Ax}$, can be viewed from a functional perspective. For some vectors, this seemingly complex computation is no more difficult than scalar multiplication. Sometimes, the question is whether, for a given matrix $\mathbf{A}$, such vectors exist—and if so, which ones.

2. Definition: Suppose that $\mathbf{A}$ is a square matrix of size n, $\mathbf{x} \neq 0$ is a vector in $\mathbb{C}^n$, and λ is a scalar in $\mathbb{C}$. Then we say that $\mathbf{x}$ is an **eigenvector** of $\mathbf{A}$ with **eigenvalue** λ if

$$\mathbf{Ax} = \lambda\mathbf{x}.$$

3. every matrix has at least one eigenvalue, and an eigen-vector to go with it.

4. Given as linear operator $T : V \to V$ and an eigenvalue λ we define

$$E_\lambda = \{v \in V : T(v) = \lambda v\}$$

the eigen space associated with λ.

5. E_λ is a subspace and non-trivial, since it has to have at least on non-zero $u \in E_\lambda$.

6. E_λ is a null-space for a matrix

$$E_\lambda = \mathcal{N}(A - \lambda I_n)$$

A matrix $\mathbf{D} \in \mathbb{R}^{n\times n}$ is said to be diagonal if $\mathbf{D}_{*,k} = d_k e_k$ for some $d_k \in \mathbb{R}$.

We say that $\mathbf{A} \in \mathbb{R}^{n\times n}$ is diagonalizable if it is similar to a diagonal matrix $\mathbf{D}$, meaning that there is an invertible matrix $\mathbf{P} \in \mathbb{R}^{n\times n}$ with

$$\mathbf{A} = \mathbf{P}^{-1}\mathbf{DP}$$

Sample Questions

1. let $\mathbf{A}$ be an invertible $n \times n$ matrix, and let λ be an eigenvalue.

   (a) Prove that $\lambda \neq 0$

   (b) Prove that $1/\lambda$ is an eigenvalue for $\mathbf{A}^{-1}$

   (c) Let E_λ be the eigenspace for λ and let $E_{1/\lambda}$ be the eigenspace for $1/\lambda$. Prove that $E_\lambda = E_{1/\lambda}$.

   (d) Show that $\mathbf{A}$ and $\mathbf{A}^\top$ have the same eigen-values.

   Answer.

   (a) Suppose $\lambda = 0$. Then, since E_λ is non trivial, then exists at least one non-zero $v \in E_\lambda$ which results in:

$$\mathbf{Ax} = \lambda\mathbf{x} = 0$$

   Which results in the $\mathcal{N}(\mathbf{A})$ being non-trivial, and $\mathbf{A}$ being non-invertible.

(b)

$$\begin{aligned}
\mathbf{A}^{-1}\mathbf{x} &= \mathbf{A}^{-1}1 \cdot \mathbf{x} \\
&= \mathbf{A}^{-1}(\frac{1}{\lambda} \cdot \lambda) \cdot \mathbf{x} \\
&= \mathbf{A}^{-1}\frac{1}{\lambda} \cdot (\lambda \cdot \mathbf{x}) \\
&= \mathbf{A}^{-1}\frac{1}{\lambda}\mathbf{A}\mathbf{x} \\
&= \frac{1}{\lambda}\mathbf{A}^{-1}\mathbf{A}\mathbf{x} \\
&= \frac{1}{\lambda}\mathbf{I}\mathbf{x} \\
&= \frac{1}{\lambda}\mathbf{x}
\end{aligned}$$

(c) Let E_λ be the eigenspace for λ and let $E_{1/\lambda}$ be the eigenspace for $1/\lambda$. Prove that $E_\lambda = E_{1/\lambda}$. We have shown for $\mathbf{x} \in E_\lambda$, $\mathbf{x} \in E_{1/\lambda}$. Hence $E_\lambda \subset E_{1/\lambda}$. The other way can be shown similarly. Hence:

$$E_\lambda = E_{1/\lambda}$$

(d) Show that $\mathbf{A}$ and $\mathbf{A}^\top$ have the same eigen-values.

$$\begin{aligned}
P_A(x) &= \det(A - tI_n) \\
&= \det\Big((A - tI_n)^\top\Big) \\
&= \det\Big(A^\top - (tI_n)^\top\Big) \\
&= \det\Big(A^\top - tI_n^\top\Big) \\
&= \det\Big(A^\top - tI_n\Big)
\end{aligned}$$

Therefore, $\mathbf{A}$ and $\mathbf{A}^\top$ have the same characteristic polynomials, hence their eigenvalues are the same.

2. Find the eigenvalues and corresponding eigenspaces for the matrix

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

Answer. The characteristic polynomial of the given matrix is:

$$t^2 - 2t + (1 - 1) = t^2 - 2t = t(t - 2)$$

Hence the eigen values are $t = 0, 2$.

E_2:

$$(\mathbf{A} - 2I_2)v = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} v_1 - v_2 \\ -v_1 + v_2 \end{bmatrix}$$

$\begin{bmatrix} v_1 - v_2 \\ -v_1 + v_2 \end{bmatrix}$ has to be equal to $\mathbf{0}$ which results in $v_1 = v_2$. Hence:

$$E_2 = \left\{ t \begin{bmatrix} 1 \\ 1 \end{bmatrix} : t \in \mathbb{R} \right\}$$

E_0:

$$(\mathbf{A} - 0I_2)v = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} v_1 + v_2 \\ -v_1 + v_2 \end{bmatrix}$$

$\begin{bmatrix} v_1 + v_2 \\ v_1 + v_2 \end{bmatrix}$ has to be equal to $\mathbf{0}$ which results in $v_1 = -v_2$. Hence:

$$E_2 = \left\{ t \begin{bmatrix} 1 \\ -1 \end{bmatrix} : t \in \mathbb{R} \right\}$$

3. Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be such that $\mathbf{A}^2 = \mathbf{A}$. Prove that the only possible eigenvalues of $\mathbf{A}$ are $\lambda = 0$ or $\lambda = 1$
   Answer.
   $$\lambda\mathbf{x} = \mathbf{A}\mathbf{x} = \mathbf{A}^2\mathbf{x} = \mathbf{A}\mathbf{A}\mathbf{x} = \mathbf{A}\lambda\mathbf{x} = \lambda\mathbf{A}\mathbf{x} = \lambda\lambda\mathbf{x} = \lambda^2\mathbf{x}$$
   Hence
   $$\lambda\mathbf{x} = \lambda^2\mathbf{x}$$
   And since $\mathbf{x}$ is non-zero, then $\lambda \in \{0, 1\}$.

4. For the inverse of the previous question to be true, we need an additional example which is $\mathbf{A}$ has to be diagonalizable.
   Let $\mathbf{A}$ be a diagonalizable matrix such that its eigenvalues are 0 or 1. Prove that $\mathbf{A}^2 = \mathbf{A}$.
   Answer. By hypothesis, we know $\exists \mathbf{P}, \mathbf{D} \in \mathbb{R}^{n \times n}$ that
   $$\mathbf{A} = \mathbf{P}^{-1}\mathbf{D}\mathbf{P}$$
   Note that $\mathbf{D}_{*,k} = d_k e_k$ must hold where d_k is the eigen value. Hence $d_k \in \{0, 1\}$. With that, we have:
   $$(\mathbf{D})^2_{*,k} = \mathbf{D}(\mathbf{D})_{*,k} = \mathbf{D}(d_k e_k) = d_k(\mathbf{D}e_k) = d_k(\mathbf{D})_{*,k} = d_k(d_k e_k) = (d_k)^2 e_k = \mathbf{D}_{*,k}$$
   Which mean $\mathbf{D}^2 = \mathbf{D}$, and:
   $$\mathbf{A}^2 = (\mathbf{P}^{-1}\mathbf{D}\mathbf{P})(\mathbf{P}^{-1}\mathbf{D}\mathbf{P}) = \mathbf{P}^{-1}\mathbf{D}\mathbf{P}\mathbf{P}^{-1}\mathbf{D}\mathbf{P} = \mathbf{P}^{-1}\mathbf{D}\mathbf{I}\mathbf{D}\mathbf{P} = \mathbf{P}^{-1}(\mathbf{D})^2\mathbf{P} = \mathbf{P}^{-1}\mathbf{D}\mathbf{P} = \mathbf{A}$$

Representation for Exams

1. Null-Spaces
   Consider the following matrices:
   $$A = \begin{bmatrix} 1 & 0 \\ 2 & 6 \\ 1 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 3 & 0 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{bmatrix}$$
   Here are their RREF's:
   $$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Here are their null spaces:

$$N(A) = \{\mathbf{0}\}, \quad N(B) = \left\{ \begin{bmatrix} -z \\ 0 \\ z \end{bmatrix} : z \in \mathbb{R} \right\} \quad N(C) = \{\mathbf{0}\}$$

2. Sets of solutions to System of Linear Equations

$$\begin{aligned} 4x + 5y + 6z &= 1 \\ x + 2y + 3z &= 1 \\ 7x + 8y + 9z &= 1 \end{aligned}$$

Augmented matrix form:

$$\left[\begin{array}{ccc|c} 4 & 5 & 6 & 1 \\ 1 & 2 & 3 & 1 \\ 7 & 8 & 9 & 1 \end{array}\right] \Longrightarrow \left[\begin{array}{ccc|c} 1 & 0 & -1 & -1 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

Go back to equations:

$$\begin{aligned} x - z &= -1 \\ y + 2z &= 1 \end{aligned}$$

Note:

$$\begin{aligned} x &= -1 + z \\ y &= 1 - 2z \end{aligned}$$

Parametric form for solution:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} -1 + z \\ 1 - 2z \\ z \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} + z \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$$

3. Simplified Form of Spanning Sets

$$\begin{aligned} \text{span}\left(\begin{bmatrix} 1 \\ 4 \\ 7 \end{bmatrix}, \begin{bmatrix} 2 \\ 5 \\ 8 \end{bmatrix} \right) &= |row\rangle\langle row| \left(\begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \end{bmatrix} \right) \\ &= |row\rangle\langle row| \left(\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \end{bmatrix} \right) \\ &= \text{span}\left(\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} \right) \\ &= \left\{ \alpha \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} + \beta \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} : \alpha, \beta \in \mathbb{R} \right\} \\ &= \left\{ \begin{bmatrix} \alpha \\ \beta \\ 2\beta - \alpha \end{bmatrix} : \alpha, \beta \in \mathbb{R} \right\} \end{aligned}$$