MATH 317 - Theory of Linear Algebra **Spring 2025**

My Pondering on Linear Algebra

Vector Operations

Orthogonality

Orthogonality is interesting as we need define so many things which at first blush might not appear central to what follows, but we will make a 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

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

$\square$

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 orthogonal set out of linearly independent sets using 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 ortogonal set from, and then apply the process.

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)

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

$\square$

11. We now know that any system of linear equations

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

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:

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

□

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:

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

Representation

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

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