Math Notes

Tyler Wilson

Contents

1 Linear Algebra

1.1 Vectors and Geometry

1.1.1 Vectors

Vectors arise from describing things that have both magnitude and direction (such as force or velocity). The coordinates of a vector $\vec{a}$ can be defined as $\langle a_1, a_2, a_3 \rangle$. This is the expression for a vector in three dimensions but a vector can be also be defined in an arbitrary number of dimensions.

The length of a vector can be determined using Pythagorean's theorem and is denoted by $\|\vec{a}\|$ for where $\vec{a}$ is a vector. The length is equal to,

$$\|\vec{a}\| = \sqrt{a_1^2 + a_2^2 + a_3^2 + \cdots + a_n^2}$$

Ex: The length of $\langle 1, 2, 3 \rangle$ is $\sqrt{1+4+9} = \sqrt{14}$

A vector can be multiplied by a scalar as such:

$$k\vec{a} = \langle ka_1, ka_2, ka_3, \ldots, ka_n \rangle$$

This stretches and contracts the vector. If you multiply by a negative, the vector will flip directions.
Ex: $2\langle 1, 1 \rangle = \langle 2, 2 \rangle$
Ex2: $-1\langle 2, 1 \rangle = \langle -2, -1 \rangle$

A unit vector is a vector whose magnitude is 1 and is considered to only have a directional component (sometimes called a direction vector)

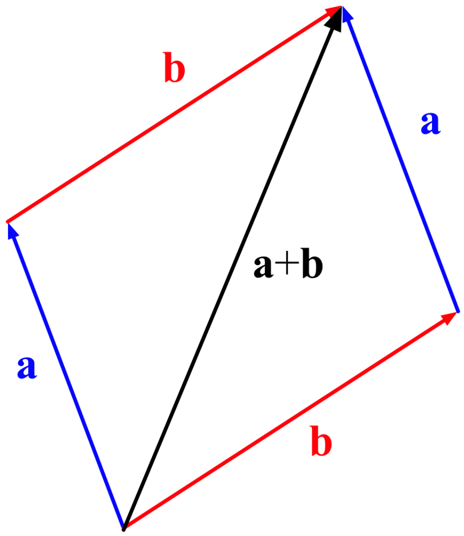
$$\hat{a} = \operatorname{dir} \vec{a} = \frac{\vec{a}}{\|\vec{a}\|}$$

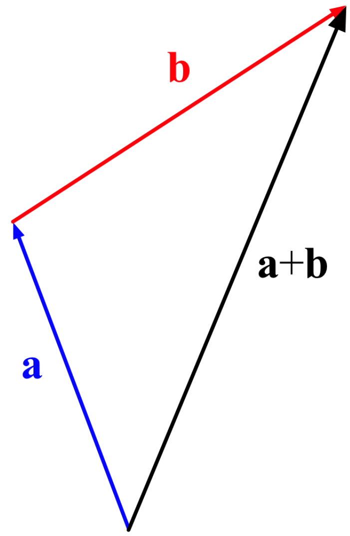
$$\text{Ex: } \operatorname{dir} \langle 2, 7 \rangle = \frac{\langle 2, 7 \rangle}{\sqrt{4+49}} = \left\langle \frac{2}{\sqrt{53}}, \frac{7}{\sqrt{53}} \right\rangle$$

Two vectors can be added by adding each element.

$$\vec{a} + \vec{b} = \langle a_1 + b_1, a_2 + b_2 \rangle$$

Ex: $\langle 2, 0 \rangle + \langle 1, 2 \rangle = \langle 3, 2 \rangle$ Geometrically, vector addition is the same as adding the two vectors tip to tail.

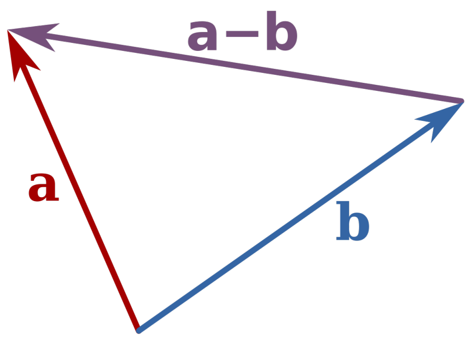




Similarly, vector subtraction is represented as

$$\vec{a} - \vec{b} = \langle a_1 - b_1,\, a_2 - b_2 \rangle$$

and is shown geometrically as the vector from $\vec{b}$ to $\vec{a}$



The dot product is one form of multiplication of vectors. It is defined as

$$\vec{a} \cdot \vec{b} = a_1 b_1 + a_2 b_2$$

*Notice that the output of a dot product is a scalar.
Ex: $\langle 2, 3 \rangle \cdot \langle 2, 1 \rangle = 2 \cdot 2 + 3 \cdot 1 = 4 + 3 = 7$ One useful identity of the dot product is

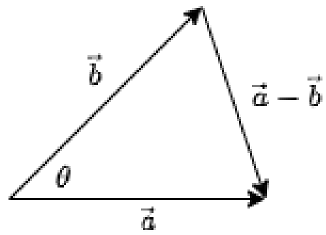
$$\vec{a} \cdot \vec{a} = \|\vec{a}\|^2$$

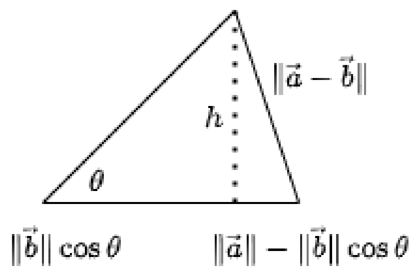
Another identity is

$$\vec{a} \cdot \vec{b} = \|\vec{a}\| \|\vec{b}\| \cos \theta$$

where θ is the angle between $\vec{a}$ and $\vec{b}$ Proof:

$$\|\vec{a} - \vec{b}\|^2 = (\vec{a} - \vec{b}) \cdot (\vec{a} - \vec{b}) = \vec{a} \cdot \vec{a} - \vec{a} \cdot \vec{b} - \vec{b} \cdot \vec{a} + \vec{b} \cdot \vec{b} = \|\vec{a}\|^2 - 2\vec{a} \cdot \vec{b} + \|\vec{b}\|^2$$





$$\|\vec{a}-\vec{b}\|^2 = h^2 + (\|\vec{a}\| - \|\vec{b}\|\cos\theta)^2$$
$$h = \|\vec{b}\|\sin\theta$$
$$\|\vec{a}-\vec{b}\|^2 = \|\vec{b}\|^2\sin^2\theta + \|\vec{a}\|^2 - 2\|\vec{a}\|\|\vec{b}\|\cos\theta + \|\vec{b}\|^2\cos^2\theta$$
$$= \|\vec{a}\|^2 + \|\vec{b}\|^2 - 2\|\vec{a}\|\|\vec{b}\|\cos\theta$$
$$\Rightarrow \|\vec{a}\|^2 - 2\vec{a}\cdot\vec{b} + \|\vec{b}\|^2 = \|\vec{a}\|^2 + \|\vec{b}\|^2 - 2\|\vec{a}\|\|\vec{b}\|\cos\theta$$
$$\Rightarrow \vec{a}\cdot\vec{b} = \|\vec{a}\|\|\vec{b}\|\cos\theta$$

Ex: Find the angle between $\langle 1,2\rangle$ and $\langle 1,1\rangle$

$$\cos\theta = \frac{\vec{a}\cdot\vec{b}}{\|\vec{a}\|\|\vec{b}\|} = \frac{\langle 1,2\rangle\cdot\langle 1,1\rangle}{\sqrt{5}\sqrt{2}} = \frac{3}{\sqrt{10}}$$
$$\theta = \arccos\left(\frac{3}{\sqrt{10}}\right)$$

Note that if $\vec{a}\cdot\vec{b}=0$ it implies that the angle between them is 90° and that $\vec{a}\perp\vec{b}$. In general, a vector orthogonal to $\vec{a}$ can be defined as $\vec{a}^\perp = \pm\langle a_2, -a_1\rangle$

Proof:

$$\vec{a}\cdot\vec{b} = a_1b_1 + a_2b_2 = 0$$
$$a_1b_1 = -a_2b_2$$
$$\Rightarrow b_1 = a_2,\ b_2 = -a_1$$
$$\Rightarrow \vec{b} = \vec{a}^\perp = \pm\langle a_2, -a_1\rangle$$

Projections:

The projection of a vector $\vec{a}$ in the direction of $\vec{b}$ is denoted by $\text{proj}_{\vec{b}}\,\vec{a}$.

The magnitude of the projection is $\|\vec{a}\|\cos\theta$ so we can define the projection vector as

$$\text{proj}_{\vec{b}}\,\vec{a} = (\vec{a}\cdot\hat{b})\hat{b} = \frac{\vec{a}\cdot\vec{b}}{\|\vec{b}\|^2}\vec{b}$$

Ex: Find the projection of $\langle 2,3\rangle$ onto $\langle 2,1\rangle$

$\vec{a} = \langle 2,3\rangle$, $\vec{b} = \langle 2,1\rangle$

$\text{proj}_{\vec{b}}\,\vec{a} = \dfrac{\langle 2,3\rangle \cdot \langle 2,1\rangle}{\|\langle 2,1\rangle\|^2}\langle 2,1\rangle = \dfrac{7}{5}\langle 2,1\rangle$

$= \left\langle \dfrac{14}{5}, \dfrac{7}{5} \right\rangle$

1.1.2 Determinants and Cross Products

The 2x2 determinant is defined as

$$\det \begin{bmatrix} a_1 & a_2 \\ b_1 & b_2 \end{bmatrix} = \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} = a_1b_2 - a_2b_1$$

Geometrically, it represents the signed area of the parallelogram spanned by vectors $\vec{a}$, $\vec{b}$.

$$A_{parallelogram} = \left| \det \begin{bmatrix} -\vec{a}- \\ -\vec{b}- \end{bmatrix} \right|$$

If the determinant is equal to zero it implies the area is zero and it means that $\vec{a}$ and $\vec{b}$ are along the same line and are considered *colinear.*

Determinants in $\mathbb{R}^3$

$$\begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} = a_1 \begin{vmatrix} b_2 & b_3 \\ c_2 & c_3 \end{vmatrix} - a_2 \begin{vmatrix} b_1 & b_3 \\ c_1 & c_3 \end{vmatrix} + a_3 \begin{vmatrix} b_1 & b_2 \\ c_1 & c_2 \end{vmatrix} = a_1b_2c_3 + a_2b_3c_1 + a_3b_1c_2 - a_1b_3c_2 - a_2b_1c_3 - a_3b_2c_1$$

Ex: $\det \begin{bmatrix} 1 & 2 & 0 \\ 2 & 7 & 1 \\ 5 & 3 & 3 \end{bmatrix}$

$= 1 \begin{vmatrix} 7 & 1 \\ 3 & 3 \end{vmatrix} - 2 \begin{vmatrix} 2 & 1 \\ 5 & 3 \end{vmatrix} + 0 \begin{vmatrix} 2 & 7 \\ 5 & 3 \end{vmatrix}$

$= (21 - 3) - 2(6 - 5)$

$= 16$

Geometrically, the 3x3 determinant represents the area of the parellelapiped spanned by $\vec{a}$, $\vec{b}$, $\vec{c}$

$$A_{parallelapiped} = \det \begin{bmatrix} -\vec{a}- \\ -\vec{b}- \\ -\vec{c}- \end{bmatrix} = \vec{a} \cdot (\vec{b} \times \vec{c})$$

Cross Product:
The cross product is defined as

$$\vec{a} \times \vec{b} = \langle a_2b_3 - a_3b_2,\ a_3b_1 - a_1b_3,\ a_1b_2 - a_2b_1 \rangle$$

It can be more easily interpreted as

$$\vec{a} \times \vec{b} = \begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix}$$

where $\hat{i} = \langle 1, 0, 0 \rangle$, $\hat{j} = \langle 0, 1, 0 \rangle$, and $\hat{k} = \langle 0, 0, 1 \rangle$
The result of $\vec{a} \times \vec{b}$ will be a vector orthogonal to both $\vec{a}$ and $\vec{b}$.
*Note that $\vec{a} \times \vec{b} = -\vec{b} \times \vec{a}$

$$\begin{aligned} \text{Ex: } & \langle 1, 1, 1 \rangle \times \langle 1, 2, 3 \rangle \\ &= \begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ 1 & 1 & 1 \\ 1 & 2 & 3 \end{vmatrix} \\ &= \hat{i} \begin{vmatrix} 1 & 1 \\ 2 & 3 \end{vmatrix} - \hat{j} \begin{vmatrix} 1 & 1 \\ 1 & 3 \end{vmatrix} + \begin{vmatrix} 1 & 1 \\ 1 & 2 \end{vmatrix} \\ &= \langle 1, -2, 1 \rangle \end{aligned}$$

A useful identity is that $\|\vec{a} \times \vec{b}\| = \|\vec{a}\| \|\vec{b}\| \sin \theta$. This also happens to be the area of a parallelogram in $\mathbb{R}^3$.

1.1.3 Lines and Planes

Parametric Representation of a line:
The parametric form of a plane can be thought of as a scaled direction vector plus a vector from the origin to some point on the line. If we let $\vec{x}$ represent our line then we have

$$L = \{\vec{x} : \vec{x} = s\vec{a} + \vec{p}, \; s \in \mathbb{R}\}$$

where $\vec{a}$ is the direction vector, s is a scaling variable, and $\vec{p}$ is some point on the line.
Ex: Find the parametric equation of a line that passes through $(1, 2)$ and $(3, 3)$

$$\begin{aligned} &\vec{a} = (3, 3) - (1, 2) = \langle 2, 1 \rangle \\ &\vec{p} = \langle 1, 2 \rangle \\ &\Rightarrow \vec{x} = \langle 2, 1 \rangle \, s + \langle 1, 2 \rangle \end{aligned}$$

Parametric Form of a Plane:
A plane is a 2D object, meaning that we will require 2 parameters (s and t) to describe it. The parametric form for a plane is very similar to that of a line in $\mathbb{R}^2$, just with the addition of another scaling variable, t, and another direction vector $\vec{b}$.

$$P = \left\{\vec{x} : \vec{x} = s\vec{a} + t\vec{b} + \vec{p}, \; s, t \in \mathbb{R}\right\}$$

Ex: The points $(1, 1, 1)$, $(2, 3, 7)$, and $(0, 2, 0)$ lie on a plane. Write the parametric form of the plane.

$$\vec{a} = (2, 3, 7) - (1, 1, 1) = \langle 1, 2, 6 \rangle$$

$\vec{b} = (0,2,0) - (1,1,1) = \langle -1,1,-1 \rangle$
$\vec{p} = \langle 1,1,1 \rangle$
$\Rightarrow \vec{x} = s\langle 1,2,6 \rangle + t\langle -1,1,-1 \rangle + \langle 1,1,1 \rangle$

Equation Form of a line in R^2:
Another way to express the equation of a line is

$$n_1x_1 + n_2x_2 = d$$

where n_1, n_2 are components of the normal vector and d is some constant that is found by plugging in a point on the line.
This equation is derived the following way:

let $\vec{n} = \vec{a}^{\perp}$ and $\vec{n} \perp \vec{a}$
$\vec{a} \cdot \vec{n} = 0$
$\vec{x} = \vec{a} + \vec{p} \Rightarrow \vec{a} = \vec{x} - \vec{p}$
$(\vec{x} - \vec{p}) \cdot \vec{n} = 0$
$\vec{x} \cdot \vec{n} - \vec{p} \cdot \vec{n} = 0$
$\vec{x} \cdot \vec{n} = \vec{p} \cdot \vec{n}$
let $\vec{p} \cdot \vec{n} = d$
$n_1x_1 + n_2x_2 = d$

*Note: a handy trick for calculating $\vec{n}$ in $\mathbb{R}^2$ is to use a similar method to the cross product:

$$\vec{n} = \pm \begin{vmatrix} \hat{i} & \hat{j} \\ a_1 & a_2 \end{vmatrix}$$

Ex: Write the equation of the line passing through $(1,2)$ and $(3,3)$.

$\vec{a} = (3,3) - (1,2) = \langle 2,1 \rangle$
$\vec{n} = \langle -1,2 \rangle$
$\vec{x} \cdot \vec{n} = d$
$-x_1 + 2x_2 = d$
plug in point $(1,2)$: $-1 + 4 = d = 3$
$-x_1 + 2x_2 = 3$

Ex2: Determine if the point $(5,5)$ is on the line in the example above.

$-5 + 2(5) = 5 \neq 3 \therefore (5,5)$ is not on the line.

Equation Form of a Plane in $\mathbb{R}^3$:
The equation form of a plane follows the exact same pattern but contains one more term.

$$n_1x_1 + n_2x_2 + n_3x_3 = d$$

The only part that is slightly different is finding the normal vector $\vec{n}$. In this case, you must take the cross product of two direction vectors to find the normal (as it will be orthogonal to both and,

therefore, the plane). $\vec{n} = \pm\vec{a} \times \vec{b}$
Ex: Find the equation form of the plane that contains $(1, 1, 1)$, $(2, 3, 7)$, and $(0, 2, 0)$

$$\vec{a} = (2, 3, 7) - (1, 1, 1) = \langle 1, 2, 6\rangle$$
$$\vec{b} = (0, 2, 0) - (1, 1, 1) = \langle -1, 1, -1\rangle$$
$$\vec{n}\begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ 1 & 2 & 6 \\ -1 & 1 & -1 \end{vmatrix} = \langle -8, -5, 3\rangle$$
$$-8x_1 - 5x_2 + 3x_3 = d$$
$$\text{plug in } (0, 2, 0): \ -5(2) = d = -10$$
$$-8x_1 - 5x_2 + 3x_3 = -10$$

Equation Form of a Line in $\mathbb{R}^3$:
This consists of two equations for x. for a point to be on the line, it must satisfy both equations.

$$L : \begin{cases} n_1x_1 + n_2x_2 = d_1 \\ m_1x_1 + m_2x_2 = d_2 \end{cases}$$

where $\vec{n}$ and $\vec{m}$ are both 2D normal vectors. A handy trick is to set $\vec{n}$ and $\vec{m}$ to be

$$\vec{n} = \det \begin{bmatrix} \hat{i} & \hat{j} \\ a_1 & a_2 \end{bmatrix}, \vec{m} = \det \begin{bmatrix} \hat{j} & \hat{k} \\ a_2 & a_3 \end{bmatrix}$$

Ex: Find the equation form of the line passing through $(0, 1, 5)$ and $(2, 2, 2)$

$$\vec{a} = (2, 2, 2) - (0, 1, 5) = \langle 2, 1, -3\rangle$$
$$\vec{n} = \det \begin{bmatrix} \hat{i} & \hat{j} \\ 2 & 1 \end{bmatrix} = \langle 1, -2, 0\rangle$$
$$\vec{m} = \det \begin{bmatrix} \hat{j} & \hat{k} \\ 1 & -3 \end{bmatrix} = \langle 0, -3, -1\rangle$$
$$L : \begin{cases} x_1 - 2x_2 = d_1 \\ -3x_2 - x_3 = d_2 \end{cases}$$
$$\text{plug in } (0, 1, 5)$$
$$\begin{cases} -2(1) = d_1 = -2 \\ -3(1) - 5 = d_2 = -8 \end{cases}$$
$$L : \begin{cases} x_1 - 2x_2 = -2 \\ -3x_2 - x_3 = -8 \end{cases}$$

1.1.4 Distances in Space

Note: This section uses some math in it that is not covered until later sections
Intersection of Lines in $\mathbb{R}^2$:
To solve, put both lines in equation form and solve the augmented matrix

$$\left[cc\middle| \begin{matrix} n_1x_1 & n_2x_2 & d_! \\ m_1x_1 & m_2x_2 & d_2 \end{matrix}\right]$$

Alternatively, this can also be solved using Pre-Calculus methods. The calculations work out similarly in either case.

Intersection of Planes in $\mathbb{R}^3$:
The solution of two intersecting planes will generally be a line with direction vector $\vec{a} = \pm\vec{n} \times \vec{m}$
This line can be found by solving the following matrix in terms of a free variable

$$\left[\begin{array}{c|cccc} ccc & n_1x_1 & n_2x_2 & n_3x_3 & d_1 \\ & m_1x_1 & m_2x_2 & m_3x_3 & d_2 \end{array}\right]$$

Ex: Find the intersection of the planes $x_1 + x_2 + x_3 = 2$ and $x_1 + 2x_2 + 3x_3 = -1$

$$\left[\begin{array}{c|cccc} ccc & 1 & 1 & 1 & 2 \\ & 1 & 2 & 3 & -1 \end{array}\right] \rightarrow \left[\begin{array}{c|cccc} ccc & 1 & 1 & 1 & 2 \\ & 0 & 1 & 2 & -3 \end{array}\right] \rightarrow \left[\begin{array}{c|cccc} ccc & 1 & 0 & -1 & 5 \\ & 0 & 1 & 2 & -3 \end{array}\right]$$

$x_3 = t \Rightarrow x_1 = 5 + t,\ x_2 = -3 - 2t$

$\vec{x} = \langle 1, -2, 1 \rangle\, t + \langle 5, -3, 0 \rangle$

Similarly, we could take $\vec{a} = \vec{n} \times \vec{m} = \langle 1, 1, 1 \rangle \times 1, \vec{2}, 3 = \langle 1, -2, 1 \rangle$ to get the direction vector and then find some point that exists on both planes, giving the overall equation.

Intersection between a line and plane:

1. Set up with $L : \vec{x} = t\vec{a} + \vec{p}$ and $P : n_1x_1 + n_2x_2 + n_3x_3 = d$
2. Plug the line equation into the plane equation in place of $\vec{x}$
3. Solve for t
4. Plug t back into the line equation to calculate $\vec{x}$

*This same method works for finding the intersection of 2 lines.
Ex: Find the intersection of $x_1 + y_1 + 2x_3 = 5$ and the line passing through $(1, 0, 0)$ and $(1, 2, 3)$

$\vec{a} = \langle 0, 2, 3 \rangle,\ \vec{p} = \langle 1, 0, 0 \rangle$

$\vec{x} = \langle 0, 2, 3 \rangle\, t + \langle 1, 0, 0 \rangle = \langle 1, 2t, 3t \rangle$

$(1) + (2t) + 2(3t) = 5$

$8t = 4 \Rightarrow t = \dfrac{1}{2}$

$$\vec{x} = \left\langle 1, 2\left(\frac{1}{2}\right), 3\left(\frac{1}{2}\right) \right\rangle = (1, 1, 1.5)$$

Distance of an object from a hyperplane:
Def: A hyperplane is defined as an object in Euclidean space that separates it in two halves. For example, a hyperplane in $\mathbb{R}^2$ is a line and a hyperplane in $\mathbb{R}^3$ is a plane.

1. Let $\vec{p}$ be some point that lies on the hyperplane and let $\vec{q}$ be some point on the object you are trying to find the distance from.

2. Compute a vector $\vec{pq}$ (order doesn't matter)

3. Find the normal vector to the hyperplane, $\vec{n}$

4. The distance will be $|\operatorname{proj}_{\vec{n}}(\vec{pq})|$

General Method of Finding Distances:

1. Set up both objects in parametric form

2. Set the distance, d to be $d = \|\vec{x}_2 - \vec{x}_1\|$

3. Set $\nabla d = 0$ and solve the system of equations for t_1, t_2 and so on. (note that the denominator of the square root derivative can be ignored)

4. Plug in the values for $t_1, t_2, \ldots$ to get a value for $\vec{x}_1$ and $\vec{x}_2$ to calculate d.

1.2 Linear Systems

1.2.1 Linear Combinations

A combination of vectors $s_1\vec{a}_1$ and $s_2\vec{a}_2$, $s_1, s_2 \in \mathbb{R}$ is called a *linear combination* of $\{\vec{a}_1, \vec{a}_2\}$
The set of all linear combinations of a set is called the *span* of the set of vectors.
Ex: $\operatorname{span}\{\langle 1,0,0\rangle, \langle 0,1,0\rangle, \langle 0,0,1\rangle\} = \mathbb{R}^3$
Ex2: $\operatorname{span}\{\langle 1,0,0\rangle, \langle 0,1,0\rangle\} = \mathbb{R}^2$
Ex3: If $\vec{a} = \langle 2,3\rangle$ and $\vec{b} = \langle 1,2\rangle$, is $(1,1)$ in the span of $\left\{\vec{a}, \vec{b}\right\}$? If so, find the linear combination.

Require s and t such that $s\vec{a} + t\vec{b} = \langle 1,1\rangle$
$s\langle 2,3\rangle + t\langle 1,2\rangle = \langle 1,1\rangle$
$\langle 2s+t, 3s+2t\rangle = \langle 1,1\rangle$
$$\begin{cases} 2s+t=1 \\ 3s+2t=1 \end{cases} \Rightarrow s=1,\ t=-1$$
$\rightarrow$ any point, (c_1, c_2) can be written as a linear combination of $\vec{a}$ and $\vec{b}$
$\therefore \operatorname{span}\left\{\vec{a}, \vec{b}\right\} = \mathbb{R}^2$

A collection of vectors is called *linearly dependent* if some nontrivial (not all 0) combination of equal zero. Otherwise, it is called *linearly independent*
$\vec{a}$, $\vec{b}$, $\vec{b}$ are linearly dependent if
$$\det\begin{bmatrix} -\vec{a}- \\ -\vec{b}- \\ -\vec{b}- \end{bmatrix} = 0$$
or if
$$\begin{bmatrix} | & | & | \\ \vec{a} & \vec{b} & \vec{b} \\ | & | & | \end{bmatrix} \begin{bmatrix} | \\ \vec{0} \\ | \end{bmatrix}$$

has rank $< n$. i.e. no unique solution
Ex: Is $\langle 1,0,1\rangle$, $\langle 1,1,0\rangle$, $\langle 0,1,1\rangle$ linearly independent?

$$\det \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} = \begin{vmatrix} 1 & 0 \\ 1 & 1 \end{vmatrix} - 0 \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} + \begin{vmatrix} 1 & 1 \\ 0 & 1 \end{vmatrix} = 1 - 0 + 1 = 2 \neq 0$$

$\therefore$ linearly independent

A collection of n linearly dependent vectors in $\mathbb{R}^n$ dimensional space is called a *basis*. So if $\vec{a}$, $\vec{b}$, $\vec{b}$ are linearly independent they form a basis and any vector, $\vec{x}$, can be formed from a linear combination of $\vec{a}$, $\vec{b}$, $\vec{b}$.

$$\begin{bmatrix} | & | & | \\ \vec{a} & \vec{b} & \vec{b} \\ | & | & | \end{bmatrix} \begin{bmatrix} | \\ \vec{k} \\ | \end{bmatrix} = \begin{bmatrix} | \\ \vec{x} \\ | \end{bmatrix}$$

1.2.2 Gaussian Elimination

Say we have a system of 3 equations and 3 unknowns. We can use an augmented matrix to solve this system.

$$\text{Ex: } \begin{cases} x_2 + x_3 = 1 \\ x_1 + x_2 = 2 \\ x_1 + x_2 + x_3 = 3 \end{cases} = \left[\begin{array}{ccc|c} 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 2 \\ 1 & 1 & 1 & 3 \end{array}\right]$$

Once in augmented matrix form on the right, we can perform the following elementary row operations:

- Multiply a row by a non-zero number
- Add a multiple of one row to another row
- Interchange two rows

Using these operations we can perform Gaussian elimination to put the matrix into echelon form and come to our solution.
Gaussian Elimination Method:

1. kill off first column so that there is only one nonzero number remaining
2. Use a row with a 0 first entry to kill off the second column
3. Continue the process until you are able to solve for a variable
4. Use the variable you solved for to work backwards and solve for the rest

$$\text{Ex: } \left[\begin{array}{ccc|c} 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 2 \\ 1 & 1 & 1 & 3 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 1 & 0 & 2 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 3 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 1 & 0 & 2 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{array}\right]$$

$x_3 = 1$

$x_2 = 1 - x_3 = 1 - 1 = 0$

$x_1 = 2 - x_2 = 2$

$\vec{x} = \langle 2, 0, 1\rangle$

1.2.3 Solution Spaces

Not all systems will have a unique solution. Some may have no solutions and some may have infinite solutions

$$\text{Ex2: } \left[\begin{array}{ccc|c} 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right]$$

if you look at the last row, it states that $0 = 1$. This implies that there is no solution to the system

$$\text{Ex3: } \left[\begin{array}{ccc|c} 1 & 2 & 3 & 0 \\ 1 & 1 & 1 & 1 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 2 & 3 & 0 \\ 0 & -1 & -2 & 1 \end{array}\right]$$

In this example, we have no equation to determine x_3 so we call it a free variable and assign it a variable value t instead of a number.

$$\begin{aligned}
&x_3 = t \\
&-x_2 - 2x_3 = 1 \Rightarrow x_2 = -1 - 2x_3 = -1 - 2t \\
&x_1 + 2x_2 + 3x_3 = 0 \Rightarrow x_1 = -2x_2 - 3x_3 = -2(-1 - 2t) - 3t = 2 + t \\
&\vec{x} = \langle 2 + t, -1 - 2t, t \rangle
\end{aligned}$$

Notice, the solution space is also the equation of a line.
A handy way to imagine solution spaces is through lines and planes. In a 3x3 matrix, each row represents a plane and the solution is the intersection of the three planes. Similarly, if two rows are linearly dependent, we will effectively only have two planes and the solution space will be the intersection of these two planes which is a line.
One way to determine the number of solutions a system has is using rank.
The *rank* of a matrix is the number of nonzero rows of the matrix in echelon form.
r represents the rank of the matrix
r_A represents the rank of the augmented matrix (inclusive of the rightmost column)
If we consider a matrix to be of size $m \times n$ then m is the number of rows and n is the number of columns. (n is also the number of variables).

- If $\text{rank}(A) < \text{rank}([A|b])$, there are no solutions (irrational argument)
- If $\text{rank}(A) = \text{rank}([A|b]) = n$, there will be a unique solution
- If $\text{rank}(A) = \text{rank}([A|b]) < n$, there will be infinite solutions

A special type of augmented matrix comes from homogeneous systems. These have 0 for all the constant terms and the matrix will either have one solution ($\vec{x} = 0$) or infinite solutions.

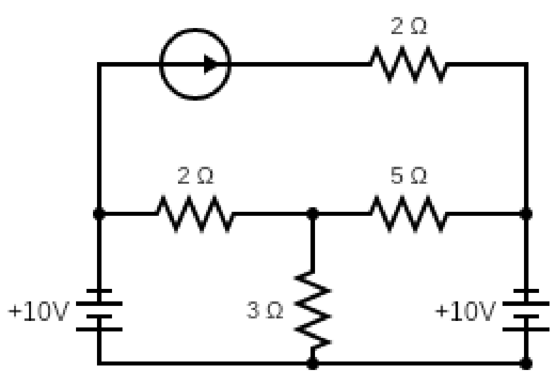
$$\left[\begin{array}{ccc|c} a_{11} & a_{12} & a_{13} & 0 \\ a_{21} & a_{22} & a_{23} & 0 \\ a_{31} & a_{32} & a_{33} & 0 \end{array}\right]$$

1.2.4 Resistor Networks

Mesh Current Method:

1. Determine variables: the current in every loop (I_1, I_2, $\dots I_n$) and the voltage drop across the current sources (E_1, E_2 $\dots E_n$)
2. Set up equations so that the sum of the voltage in each loop is zero and equations that relate the specific currents of the current source
3. Set up in a matrix and solve for the unknowns

*The current must be travelling in the same direction in each loop (i.e. arbitrarily set all clockwise)
*We consider the current specific to the loop we are considering to be positive
Ex: If the current source is $4A$, find the voltage drop across all resistors and the current source



Loop 1: $-E = 2I_1 = 5(I_1 - I_3) + 2(I_1 - I_2) = 0$
Loop 2: $10 + 2(I_2 - I_1) + 3(I_2 - I_3) = 0$
Loop 3: $5(I_3 - I_1) - 10 + 3(I_3 - I_2) = 0$
Current Source: $I_1 = 4$

$$\begin{cases} 2I_2 + 5I_3 + E = 36 \\ 5I_2 - 3I_3 = -2 \\ -3I_2 + 8I_3 = 30 \end{cases}$$

$$\left[\begin{array}{ccc|c} 2 & 5 & 1 & 36 \\ 5 & -3 & 0 & -2 \\ -3 & 8 & 0 & 30 \end{array}\right] \rightsquigarrow \begin{cases} I_1 = 4A \\ I_2 \approx 2.3871A \\ I_3 \approx 4.6452A \\ E = 8V \end{cases}$$

$V_{R_2} = 2|I_1 - I_2| \approx 3.2V$
$V_{R_5} = 5|I_1 - I_3| \approx 3.2V$
$V_{R_3} = 3|I_2 - I_3| \approx 6.8V$

1.2.5 Polynomial Interpolation

If we define $\phi_0(t) = 1,\ \phi_1(t) = t, \dots,\ \phi_d(t) = t^d$ then a polynomial of degree d is of the form

$$p(t) = c_0\phi_0(t) + c_1\phi_1(t) + \cdots + c_d\phi_d(t)$$

We can say that the polynomials $\phi_k(t)$ form a basis of a d-degree polynomial, $\mathbb{P}_d$ and the set $\{\phi_0(t), \dots, \phi_d(t)\}$ is called the *monomial basis* of $\mathbb{P}_d$.
We wish to find a polynomial, $p(t) = \left\{c_0 + c_1 t + \cdots + c_d t^d :\ c_0, c_1, \dots, c_d \in \mathbb{R}\right\}$, that interpolates the points $(t_0, y_0), (t_1, y_1), \dots, (t_n, y_n)$.
Each data point will give an equation:

$$\begin{cases} c_0 + c_1 t_0 + c_2 t_0^2 + \cdots + c_d t_0^d = y_0 \\ c_0 + c_1 t_1 + c_2 t_0^2 + \cdots + c_d t_1^d = y_1 \\ \qquad\qquad\vdots \\ c_0 + c_1 t_d + c_2 t_d^2 + \cdots + c_d t_d^d = y_d \end{cases}$$

which gives $d+1$ equations for $d+1$ unknowns.
Writing this in the form $A\vec{c} = \vec{y}$, we get

$$\begin{bmatrix} 1 & t_0 & \cdots & t_0^d \\ 1 & t_1 & \cdots & t_1^d \\ \vdots & \vdots & \ddots & \vdots \\ 1 & t_d & \cdots & t_d^d \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_d \end{bmatrix} = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_d \end{bmatrix}$$

where the matrix A is called the *Vandermode matrix*.
The Vandermode matrix has the identity that

$$\det(A) = \prod_{0 \le i < j \le d} (t_j - t_i)$$

Ex: For a 3x3 Vandermode matrix,

$$\begin{vmatrix} 1 & t_0 & t_0^2 \\ 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \end{vmatrix} = (t_1 - t_0)(t_2 - t_0)(t_2 - t_1)$$

Ex: Find a polynomial that passes through the points $(-1, 2), (0, 1), (1, 4)$

$$A = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix},\ \vec{y} = \begin{bmatrix} 2 \\ 1 \\ 4 \end{bmatrix}$$

$$A\vec{c} = \vec{y} = \left[\begin{array}{ccc|c} 1 & -1 & 1 & 2 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 4 \end{array}\right] \rightsquigarrow \left[\begin{array}{ccc|c} 1 & -1 & 1 & 2 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 2 & 4 \end{array}\right]$$

$$c_3 = 2 \Rightarrow c_2 = -1 + c_3 = 1 \Rightarrow c_1 = 2 + c_2 - c_3 = 1$$

$$\vec{c} = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} \Rightarrow p(t) = 1 + t + 2t^2$$

A downside to polynomial interpolation is that the condition number of the Vandermode matrix gets very large with an increase in the amount of points.

1.2.6 Cubic Spline Interpolation

Cubic spline interpolation is the method of interpolating data points using a piecewise cubic function. The rational is that a cubic function allows you to match the function, its derivative, and its 2nd derivative, giving it the appearance of a very smooth interpolation of the data.
If we have $N+1$ data points, $(t_0, y_0), (t_1, y_1), \ldots, (t_N, y_N)$, it will be defined by N cubic polynomials defined by the general equation

$$p_k(t) = a_k(t - t_{k-1})^3 + b_k(t - t_{k-1})^2 + c_k(t - t_{k-1}) + d_k,\ \ t \in [t_{k-1}, t_k]$$

This leaves us with $4N$ unknowns we have to solve for.
We can get our unknowns using the following constraints:

- Left endpoints: $p_k(t_{k-1}) = y_{k-1}$ gives N equations
- Right endpoints: $p_k(t_k) = y_k$ gives N equations
- Continuity of $p'(t)$: $p'_k(t_k) = p'_{k+1}(t_k)$ gives $N-1$ equations
- Continuity of $p''(t)$: $p''_k(t) = p''_{k+1}(t_k)$ gives $N-1$ equations

We can get our last 2 equations using arbitrary boundary conditions.
We define the natural cubic spline to have the property that

- $p''_1(t_0) = p''_N(t_N) = 0$

We express the cubic spline with its coefficient matrix:

$$C = \begin{bmatrix} a_1 & a_2 & \cdots & a_N \\ b_1 & b_2 & \cdots & b_N \\ c_1 & c_2 & \cdots & c_N \\ d_1 & d_2 & \cdots & d_N \end{bmatrix}$$

We can solve for these coefficients using a behemoth of a linear system:

$$\begin{bmatrix} A(L_1) & B & & & \\ & A(L_2) & B & & \\ & & \ddots & \ddots & \\ & & & A(L_{N-1}) & B \\ T & & & & V \end{bmatrix} \begin{bmatrix} a_1 \\ b_1 \\ c_1 \\ \vdots \\ a_N \\ b_N \\ c_N \end{bmatrix} = \begin{bmatrix} y_1 - y_0 \\ 0 \\ 0 \\ \vdots \\ y_N - y_{N-1} \\ 0 \\ 0 \end{bmatrix}$$

where $L_k = t_k - t_{k-1}$ and

$$A(L) = \begin{bmatrix} L^3 & L^2 & L \\ 3L^2 & 2L & 1 \\ 6L & 2 & 0 \end{bmatrix} \qquad B = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & -2 & 0 \end{bmatrix}$$

$$T = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad V = \begin{bmatrix} L_N^3 & L_N^2 & L_N \\ 0 & 0 & 0 \\ 6L_N & 2 & 0 \end{bmatrix}$$

The last 2 rows of the augmented matrix correspond to the natural cubic spline condition and would be different for a different cubic spline model.

1.3 Matrices

1.3.1 Matrix Operations

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \\ a_{m1} & a_{m2} & & a_{mn} \end{bmatrix} = \left[a_{ij}\right]_{m\times n}$$

*m represents the number of rows and n represents the number of columns. ij represents the current subscripts.
For a matrix A to multiply a matrix B, the number of columns of A must match the number of rows of B. Ex: $A_{4\times 2}B_{2\times 3}$ is allowed because the bottom middle numbers are both 2.
The method for matrix multiplication, AB, is to dot the rows of A with the columns of B

$$\begin{aligned} \text{Ex: } & \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 5 \\ 2 & 3 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 2+2 & 0+3 & 2\cdot 5+1 \\ 1+2\cdot 2 & 0+2\cdot 3 & 5+2 \end{bmatrix} \\ &= \begin{bmatrix} 4 & 3 & 11 \\ 5 & 6 & 7 \end{bmatrix} \end{aligned}$$

*note that in general, $AB \neq BA$
A matrix of particular interest is the identity matrix. Multiplying by this matrix is the same as multiplying by 1.

$$I = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \\ 0 & 0 & & 1 \end{bmatrix}$$

Transpose of a Matrix:
The transpose is an operation that makes the rows become columns and the columns become rows and is expressed as A^T.

$$\text{Ex: if } A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \text{ then } A^T = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix}$$

If $A = A^T$ then we consider A to be a symmetric matrix.
Properties of the transpose:

$$\begin{aligned} &\left(A^T\right)^T = a \\ &(A+B)^T = A^T + B^T \\ &(sA)^T = sA^T \\ &(AB)^T = B^T A^T \end{aligned}$$

1.3.2 Linear Transformations

We can express vectors as matrices: A $m \times 1$ matrix is a column vector and a $1 \times n$ matrix is a row vector.
In the case of column vectors, we can multiply them by a matrix and get out another vector.

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

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

We can think of this matrix A as some function applied to $\vec{x}$. $\vec{f}(\vec{x}) = A\vec{x} = \vec{b}$
This is represented as

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

and has the properties that

$$T(\vec{x} + \vec{y}) = T(\vec{x}) + T(\vec{y}) \text{ and } T(s\vec{x}) = sT(\vec{x})$$

We can take advantage of these identities and express the transformation matrix in the form

$$A = \begin{bmatrix} | & | & & | \\ T(\vec{e}_1) & T(\vec{e}_2) & \cdots & T(\vec{e}_n) \\ | & | & & | \end{bmatrix}$$

where $\vec{e}_1 = \hat{\imath}$, $\vec{e}_2 = \hat{\jmath}$, ...

Ex: $T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ and $T\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$. Find $T\left(\begin{bmatrix} 3 \\ 2 \end{bmatrix}\right)$

$$T(\vec{e}_1) = T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

$$T(\vec{e}_2) = T\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) - T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 2 \\ 1 \end{bmatrix} - \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 0 \\ -2 \end{bmatrix}$$

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

$$T\left(\begin{bmatrix} 3 \\ 2 \end{bmatrix}\right) = \begin{bmatrix} 2 & 0 \\ 3 & -2 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 6 \\ 5 \end{bmatrix}$$

If we have a composition of transformations, $S(\vec{x}) = B\vec{x}$ and $T(\vec{x}) = A\vec{x}$, then $S(T(\vec{x})) = BA\vec{x}$.
2D Rotation Matrix:

$$\mathrm{Rot}_\theta(\vec{x}) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

Ex: $\mathrm{Rot}_{\pi/2} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \rightarrow \mathrm{Rot}_{\pi/2} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} -a_2 \\ a_1 \end{bmatrix} = \vec{a}^\perp$

2D Projection Matrix:

$$\mathrm{proj}_{\hat{a}}(\vec{x}) = \begin{bmatrix} a_1^2 & a_1a_2 \\ a_1a_2 & a_2^2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

where $\hat{a}$ is the unit vector of the line that the vector $\vec{x}$ is projected onto. Proof:

$$\text{proj}_{\hat{a}}(\vec{x}) = (\vec{x}\cdot\hat{a})\hat{a} = (x_1a_1+x_2a_2)\begin{bmatrix}a_1\\a_2\end{bmatrix} = \begin{bmatrix}(x_1a_1+x_2a_2)a_1\\(x_1a_1+x_2a_2)a_2\end{bmatrix} = \begin{bmatrix}x_1a_1^2+x_2a_1a_2\\x_1a_1a_2+x_2a_2^2\end{bmatrix}$$
$$= \begin{bmatrix}a_1^2 & a_1a_2\\a_1a_2 & a_2^2\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}$$

We can also express the line of projection as the angle of a line.
If $\hat{a} = \langle\cos\theta,\sin\theta\rangle$, $\theta\in[0,2\pi)$ we can then denote $\text{proj}_{\hat{a}} = \text{proj}_\theta$ where

$$\text{proj}_\theta = \begin{bmatrix}\cos^2\theta & \cos\theta\sin\theta\\\cos\theta\sin\theta & \sin^2\theta\end{bmatrix}$$

Ex: Find the matrix that represents the projection onto the line $y=x$.

$$\vec{a} = \langle 1,1\rangle \Rightarrow \|\vec{a}\| = \sqrt{2}$$
$$\hat{a} = \left\langle\frac{1}{\sqrt{2}},\frac{1}{\sqrt{2}}\right\rangle$$
$$\text{proj}_{\hat{a}} = \begin{bmatrix}\frac{1}{2} & \frac{1}{2}\\\frac{1}{2} & \frac{1}{2}\end{bmatrix}$$

2D Reflection Matrix:
Ref_θ is the reflection across the line making an angle θ with the x-axis.

$$\text{Ref}_\theta(\vec{x}) = 2\,\text{proj}_\theta(\vec{x}) - \vec{x} = \begin{bmatrix}\cos(2\theta) & \sin(2\theta)\\\sin(2\theta) & -\cos(2\theta)\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}$$

1.3.3 Inverse of a Matrix

Consider a linear system $ax=b$. Its solution will be $x=a^{-1}b$. The same sort of thing can be done with matrices and we can take the inverse of a matrix.
One way to do this is using a super-augmented matrix.

$$[A|I] \to [I|A]$$

Also, it is important to note that A^{-1} only exists if $\det A \neq 0$.

Ex: Find the inverse of $\begin{bmatrix}1&1&1\\1&2&3\\1&4&9\end{bmatrix}$

$$\left[\begin{array}{ccc|ccc}1&1&1&1&0&0\\1&2&3&0&1&0\\1&4&9&0&0&1\end{array}\right] \to \left[\begin{array}{ccc|ccc}1&1&1&1&0&0\\0&1&2&-1&1&0\\0&3&8&-1&0&1\end{array}\right] \to \left[\begin{array}{ccc|ccc}1&1&1&1&0&0\\0&1&2&-1&1&0\\0&0&2&2&-3&1\end{array}\right]$$
$$\to \left[\begin{array}{ccc|ccc}1&1&1&1&0&0\\0&1&2&-1&1&0\\0&0&1&1&-\frac{3}{2}&\frac{1}{2}\end{array}\right] \to \left[\begin{array}{ccc|ccc}1&1&0&0&\frac{3}{2}&-\frac{1}{2}\\0&1&0&-3&4&-1\\0&0&1&1&-\frac{3}{2}&\frac{1}{2}\end{array}\right] \to \left[\begin{array}{ccc|ccc}1&0&0&3&-\frac{5}{2}&\frac{1}{2}\\0&1&0&-3&4&1\\0&0&1&1&-\frac{3}{2}&\frac{1}{2}\end{array}\right]$$
$$A^{-1} = \begin{bmatrix}3&-\frac{5}{2}&\frac{1}{2}\\-3&4&1\\1&-\frac{3}{2}&\frac{1}{2}\end{bmatrix}$$

Another way to find the inverse is to use the method

$$A^{-1} = \frac{\text{adj}(A)}{\det(A)}$$

This method is more complex but can be quicker in some cases. The adjoint of a matrix is calculated by finding the cofactors of each entry and then taking the transpose of the cofactor matrix.

$$\text{adj}(A) = C^T = \begin{bmatrix} C_{11} & C_{21} & C_{31} \\ C_{12} & C_{22} & C_{32} \\ C_{13} & C_{23} & C_{33} \end{bmatrix}$$

where

$$C_{ij} = (-1)^{i+j} M_{ij}$$

M_{ij} is the minor which is defined to be the determinant of the entries not in the row i and column j.

Ex: The minor of cell 32 in a 3x3 matrix, $\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$ is $\det \begin{bmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{bmatrix}$

The cofactors are calculated similarly to how determinants are calculated.
This method leads to a handy formula for 2x2 matrices

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

$$C = \begin{bmatrix} a_{22} & -a_{21} \\ -a_{12} & a_{11} \end{bmatrix}$$

$$\text{adj}(A) = \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}$$

$$A^{-1} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}$$

Larger matrices don't have a nice formula but still use the same method.

Ex: $A = \begin{bmatrix} 1 & 3 & 1 \\ 0 & 3 & 1 \\ 4 & 2 & 0 \end{bmatrix}$

$$C_{11} = \begin{vmatrix} 3 & 1 \\ 2 & 0 \end{vmatrix} = 2$$

$$C_{12} = -\begin{vmatrix} 0 & 1 \\ 4 & 0 \end{vmatrix} = 4$$

$$C_{13} = \begin{vmatrix} 0 & 3 \\ 4 & 2 \end{vmatrix} = -12$$

$$C_{21} = -\begin{vmatrix} 3 & 1 \\ 2 & 0 \end{vmatrix} = 2$$

$$C_{22} = \begin{vmatrix} 1 & 1 \\ 4 & 0 \end{vmatrix} = -4$$

$$C_{23} = -\begin{vmatrix} 1 & 3 \\ 4 & 2 \end{vmatrix} = 10$$

$$C_{31} = \begin{vmatrix} 3 & 1 \\ 3 & 1 \end{vmatrix} = 0$$

$$C_{32} = -\begin{vmatrix} 1 & 1 \\ 0 & 1 \end{vmatrix} = -1$$

$$C_{33} = \begin{vmatrix} 1 & 3 \\ 0 & 3 \end{vmatrix} = 3$$

$$C = \begin{bmatrix} -2 & 4 & -12 \\ 2 & -4 & 10 \\ 0 & -1 & 3 \end{bmatrix}$$

$$\operatorname{adj}(A) = C^T = \begin{bmatrix} -2 & 2 & 0 \\ 4 & -4 & -1 \\ -12 & 10 & 3 \end{bmatrix}$$

$$\det(A) = \begin{vmatrix} 3 & 1 \\ 2 & 0 \end{vmatrix} - 3\begin{vmatrix} 0 & 1 \\ 4 & 0 \end{vmatrix} + \begin{vmatrix} 0 & 3 \\ 4 & 2 \end{vmatrix} = -2 + 12 - 12 = -2$$

$$A^{-1} = \frac{\operatorname{adj}(A)}{\det(A)} = \begin{bmatrix} 1 & -1 & 0 \\ -2 & 2 & \frac{1}{2} \\ 6 & -5 & -\frac{3}{2} \end{bmatrix}$$

One use for matrix inverses is to give an alternative way to solve systems of equations.

Ex: $\begin{cases} x + 3y = 5 \\ 2x + 4y = 6 \end{cases}$

$$AX = B$$

$$A^{-1}AX = A^{-1}B$$

$$X = A^{-1}B$$

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

$$X = -\frac{1}{2}\begin{bmatrix} 4 & -3 \\ -2 & 1 \end{bmatrix}\begin{bmatrix} 5 \\ 6 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$$

Inverse Properties:

$$AA^{-1} = A^{-1}A = I$$

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

$$(kA)^{-1} = \frac{1}{k}A^{-1}$$

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

$$\left(A^{-1}\right)^T = \left(A^T\right)^{-1}$$

Ex: Simplify $(AB^TB)^{-1}(BB^TBA^T)^T(A^TB^{-1})^T$

$(B^TB)^{-1}A^{-1}(BA^T)^T(BB^T)^T(B^{-1})^TA$

$B^{-1}(B^T)^{-1}A^{-1}AB^TBB^T(B^{-1})^TA$

$B^{-1}(B^T)^{-1}B^TBA$

$B^{-1}BA$

A

1.3.4 Determinants

While we know how to calculate determinants already, there are some tricks that make calculating them easier, especially larger determinants.
If A is an upper or lower triangular matrix then $\det(A)$ is the product of the diagonals.
Also, $\det(A) = \det(A^T)$
We can also perform row operations on determinants to simplify them as much as possible.

- Swap Rows: $\det(B) = -\det(A)$
- Multiply a Row by a Constant, k: $\det(B) = k\det(A)$
- Add a Multiple of One Row to Another: $\det(B) = \det(A)$

Ex: $\det\begin{bmatrix} 1 & 2 & -2 & -7 \\ 1 & 2 & -1 & -5 \\ 0 & 3 & 0 & -3 \\ -1 & 4 & 1 & 1 \end{bmatrix}$

$$\begin{vmatrix} 1 & 2 & -2 & -7 \\ 1 & 2 & -1 & -5 \\ 0 & 3 & 0 & -3 \\ -1 & 4 & 1 & 1 \end{vmatrix} = \begin{vmatrix} 1 & 2 & -2 & -7 \\ 0 & 0 & 1 & 2 \\ 0 & 3 & 0 & -3 \\ -1 & 4 & 1 & 1 \end{vmatrix} = \begin{vmatrix} 1 & 2 & -2 & -7 \\ 0 & 0 & 1 & 2 \\ 0 & 3 & 0 & -3 \\ 0 & 6 & -1 & -6 \end{vmatrix} = \begin{vmatrix} 1 & 2 & -2 & -7 \\ 0 & 0 & 1 & 2 \\ 0 & 3 & 0 & -3 \\ 0 & 0 & -1 & 0 \end{vmatrix}$$

$$= -\begin{vmatrix} 1 & 2 & -2 & -7 \\ 0 & 3 & 0 & -3 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & -1 & 0 \end{vmatrix} = -\begin{vmatrix} 1 & 2 & -2 & -7 \\ 0 & 3 & 0 & -3 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 2 \end{vmatrix} = -(1)(3)(1)(2) = -6$$

Some properties of determinants are as follows:

$\det(A)\det(A^T)$

$\det(A^{-1}) = \dfrac{1}{\det(A)}$

$\det(A^x) = (\det(A))^x$

$\det(AB) = \det(A)\det(B)$

$\det(kA) = k^n\det(A)$ where n is the matrix size

1.3.5 LU Decomposition

We can express a matrix as the multiplication between a lower and upper triangular matrix, L and U respectively.

$$A = LU$$

Such that L is a matrix representation of row operations and U is the matrix A in reduced echelon form.
A lower triangular matrix is a matrix with zeros above the diagonal:

$$\begin{bmatrix} * & & \\ * & * & \\ * & * & * \\ * & * & * \end{bmatrix}$$

An upper triangular matrix is a matrix with zeros below the diagonal:

$$\begin{bmatrix} * & * & * \\ & * & * \\ & & * \\ & & \end{bmatrix}$$

A unit triangular matrix is a square triangular matrix with ones on the diagonal
We can express the elementary row operations (as performed in Gaussian elimination) can be expressed as matrix multiplications:

- Interchange rows i and j:
  Ex: Switch rows 2 and 4

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

  Corresponds to the identity matrix but with $a_{i,i} = a_{j,j} = 0$ and $a_{i,j} = a_{j,i} = 1$

- Multiply row i by a scalar k:
  Ex: multiply row 3 by k

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

  Corresponds to the identity matrix but with $a_{i,i} = k$

- Add c times row j to row i:
  Ex: Add 5 times row 1 to row 4

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

Corresponds to the identity matrix but with $a_{i,j} = c$
To find the inverse, we merely flip the sign

$$E^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -5 & 0 & 0 & 1 \end{bmatrix}$$

We can multiply these matrices to a matrix A to get the desired result.
If we can write a matrix in row echelon form using only row operations, it will have an LU decomposition and we can express L as the product of these matrix row operations.

$$E_3E_2E_1A = U \Rightarrow A = (E_3E_2E_1)^{-1}U = E_1^{-1}E_2^{-1}E_3^{-1}U = LU$$

Ex: Find the LU decomposition

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

$R_2 \to -3R_1 + R_2$

$$E_1A = \begin{bmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 2 \\ 3 & 1 & 3 \\ -1 & 3 & 5 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 2 \\ 0 & -2 & -3 \\ -1 & 3 & 5 \end{bmatrix}$$

$R_3 \to R_1 + R_3$

$$E_2(E_1A) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 2 \\ 0 & -2 & -3 \\ -1 & 3 & 5 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 2 \\ 0 & -2 & -3 \\ 0 & 4 & 7 \end{bmatrix}$$

$R_3 \to 2R_2 + R_3$

$$E_3(E_2E_1A) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 2 \\ 0 & -2 & -3 \\ 0 & 4 & 7 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 2 \\ 0 & -2 & -3 \\ 0 & 0 & 1 \end{bmatrix} = U$$

$$L = E_1^{-1}E_2^{-1}E_3^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ -1 & -2 & 1 \end{bmatrix}$$

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

In order to receive a lower triangular matrix we require that $i < j$ for where we add i times row j to row i.
We are not always able to find an LU decomposition. If we interchange row, we can come up with a PLU decomposition where P is a permutation matrix.
If we want to solve a linear system, $A\vec{x} = \vec{b}$, using LU decomposition, we would use the following method:

$$A\vec{x} = LU\vec{x} = \vec{b}$$

Set $U\vec{x} = \vec{y}$

Solve $L\vec{y} = \vec{b}$ for $\vec{y}$

Solve $U\vec{x} = \vec{y}$ for $\vec{x}$

Some properties of LU decomposition:

- $\text{rank}(A) = \text{rank}(U)$
- $\det(A) = \det(U)$ where A is a square matrix

1.3.6 Error Analysis

The norm of a matrix (operator norm) is given by

$$\|A\| = \max_{\vec{x}\neq 0}\left\{\frac{\|A\vec{x}\|}{\|\vec{x}\|}\right\} = \max\{\|A\hat{x}\|\}$$

If D is a diagonal matrix with entries d_i then $\|D\| = \max_i\{|d_i|\}$

For an inverse matrix, the norm is given by

$$\|A^{-1}\| = \frac{1}{\min\{\|A\hat{x}\|\}}$$

and similarly, if D is diagonal, $\|D^{-1}\| = \frac{1}{\min_i\{|d_i|\}}$

The norm of a matrix is the maximum stretch of a unit vector by that matrix and the norm of the inverse is 1 over the maximum compression of a unit vector by the matrix.

The condition number $\text{cond}(A)$ tells us how sensitive the solution $\vec{x}$ is to the changes in $\vec{b}$ and is defined to be

$$\text{cond}(A) = \|A\|\|A^{-1}\| = \frac{\max\{\|A\hat{x}\|\}}{\min\{\|A\hat{x}\|\}}$$

By convention, if $\det(A) = 0$ we define $\text{cond}(A) = \infty$.

Suppose a small error, $\Delta\vec{b}$, produces a change in the solution, $\Delta\vec{x}$. We want to know how large that error will be

$$\frac{\|\Delta\vec{x}\|}{\|\vec{x}\|} \leq \text{cond}(A)\frac{\|\Delta\vec{b}\|}{\|\vec{b}\|}$$

Proof:

$A\vec{x} = \vec{b}$ and $A\Delta\vec{x} = \Delta\vec{b}$

$\|A\vec{x}\| = \|\vec{b}\|$

$\Delta x = A^{-1}\vec{b} \Rightarrow \|\Delta x\| = \|A^{-1}\vec{b}\|$

$\|A^{-1}\Delta\vec{b}\|\|A\vec{x}\| = \|\Delta x\|\|\vec{b}\|$

$\|A^{-1}\|\|\Delta\vec{b}\|\|A\|\|\vec{x}\| \geq \|\Delta\vec{x}\|\|\vec{b}\|$

$\dfrac{\|\Delta\vec{x}\|}{\|\vec{x}\|} \leq \|A\|\|A^{-1}\|\dfrac{\|\Delta\vec{b}\|}{\|\vec{b}\|}$

$\dfrac{\|\Delta\vec{x}\|}{\|\vec{x}\|} \leq \text{cond}(A)\dfrac{\|\Delta\vec{b}\|}{\|\vec{b}\|}$

This implies that if $\text{cond}(A)$ is large then small changes in $\vec{b}$ may result in very large changes in $\vec{x}$.

1.4 Orthogonality

1.4.1 Subspaces

Definitions:
A subset $U \in \mathbb{R}^n$ is a subspace if:

1. U contains the zero vector, $\vec{0}$
2. $\vec{u}_1 + \vec{u}_2 \in U$ for all $\vec{u}_1, \vec{u}_2 \in U$
3. $c\vec{u} \in U$ for all $c \in \mathbb{R},\ \vec{u} \in U$

The linear combination of vectors $\vec{u}_1, \dots, \vec{u}_m \in \mathbb{R}^n$ is given by $c_1\vec{u}_1 + \cdots + c_m\vec{u}_m$
The span of a set of vectors is the set of all linear combinations of them.
A set of vectors $\{\vec{u}_1, \dots, \vec{u}_m\} \subset \mathbb{R}^n$ forms a linearly independent set if the vectors satisfy the property $c_1\vec{u}_1 + \cdots c_m\vec{u}_m = \vec{0}$ if and only if $c_1 = \cdots c_m = 0$
We can determine if the vectors are linearly independent by setting up a matrix where the columns are the vectors.

$$A = \begin{bmatrix} | & & | \\ \vec{u}_1 & \cdots & \vec{u}_m \\ | & & | \end{bmatrix}$$

If the only solution is the trivial solution, $\vec{x} = \vec{0}$ then the vectors are linearly independent.
If $U \subseteq \mathbb{R}^n$ is a subspace then a set of vectors $\{\vec{u}_1, \dots, \vec{u}_m\}$ forms a basis of U if:

- $\{\vec{u}_1, \dots, \vec{u}_m\}$ is a linearly independent set
- $\operatorname{span}\{\vec{u}_1, \dots, \vec{u}_m\} = U$

There are infinitely many different bases of a subspace U but each basis of U has the same number of vectors.
The dimension of U is the number of vectors in a basis of U. We write this as $\dim(U)$
Ex: Find a basis and dimension of U

$$U = \operatorname{span}\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ -2 \\ 0 \\ -1 \end{bmatrix} \right\}$$

check if $\vec{u}_1, \vec{u}_2, \vec{u}_3, \vec{u}_4$ are linearly independent

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

If we choose a smaller set of $\{\vec{u}_1, \vec{u}_2, \vec{u}_3\}$ then we get

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & -2 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix} \to \text{unique solution to } A\vec{c} = \vec{0}$$

$\therefore \{\vec{u}_1, \vec{u}_2, \vec{u}_3\}$ is a basis of U, $\dim(U) = 3$

Nullspace: Let A be an $m \times n$ matrix. The nullspace of A is $\mathcal{N}(A) = \left\{\vec{x} \in \mathbb{R}^n : \ A\vec{x} = \vec{0}\right\}$
What this means is that the nullspace is all the values that get mapped to $\vec{0}$ under the linear transformation of A.
$\mathcal{N}(A)$ is a subspace of $\mathbb{R}^n$. This is because the nullspace is determined based on the input vectors which are elements in $\mathbb{R}^n$
Ex: Find a basis of $\mathcal{N}(A)$

$$A = \begin{bmatrix} 1 & 4 & -1 & -2 \\ -2 & -6 & 2 & 8 \\ 1 & 5 & -1 & 0 \end{bmatrix}$$

$$\rightsquigarrow \begin{bmatrix} 1 & 4 & -1 & -2 \\ 0 & 2 & 0 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$x_3 = s, \ x_4 = t$$

$$2x_2 + 4t = 0 \Rightarrow x_2 = -2t$$

$$x_1 + 4x_2 - s - 2t = 0 \Rightarrow x_1 = s + 10t$$

$$\vec{x} = s \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} 10 \\ -2 \\ 0 \\ 1 \end{bmatrix}$$

$$N(A) = \text{span}\left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 10 \\ -2 \\ 0 \\ 1 \end{bmatrix} \right\}, \ \dim(N(A)) = 2$$

Range:
Let A be an $m \times n$ matrix. The range of A is $\mathcal{R}(A) = \{\vec{y} \in \mathbb{R}^m : \ A\vec{x} = \vec{y}, \ \vec{x} \in \mathbb{R}^n\}$
This means that $\mathcal{R}(A)$ is the span of all vectors that can result from the linear transformation of A.
Note: $\mathcal{R}(A)$ is also called the image of A of the column space of A.
$\mathcal{R}(A)$ is a subspace of $\mathbb{R}^m$. This is because the range is determined by the outputs of $A\vec{x}$ which are elements in $\mathbb{R}^m$.
We will also have that

$$\dim(\mathcal{R}(A)) = \text{rank}(A)$$

$$A\vec{x} = \begin{bmatrix} | & | & & | \\ \vec{a}_1 & \vec{a}_2 & \cdots & \vec{a}_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1\vec{a}_1 + x_2\vec{a}_2 + \cdots + x_n\vec{a}_n$$

This means that $\mathcal{R}(A) = \text{span}\{\vec{a}_1, \vec{a}_2, \dots, \vec{a}_n\}$
Ex: find a basis of $\mathcal{R}(A)$

$$A = \begin{bmatrix} 1 & 4 & -1 & -2 \\ -2 & -6 & 2 & 8 \\ 1 & 5 & -1 & 0 \end{bmatrix}$$

$$\rightsquigarrow \begin{bmatrix} 1 & 4 & -1 & -2 \\ 0 & 2 & 0 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

take $\vec{a}_1$ and $\vec{a}_2$

$$\begin{bmatrix} 1 & 4 \\ -2 & -6 \\ 1 & 5 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 4 \\ 0 & 2 \\ 0 & 0 \end{bmatrix}$$

unique solution so $\vec{a}_1$ and $\vec{a}_2$ are linearly independent

$$\Rightarrow R(A) = \text{span}\left\{ \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}, \begin{bmatrix} 4 \\ -6 \\ 5 \end{bmatrix} \right\}$$

Given an $m \times n$ matrix, the Rank-Nullity theorem is given by

$$\dim(\mathcal{R}(A)) + \dim(\mathcal{N}(A)) = n$$

We also have that for an LU decomposition of A with column vectors of L being $\vec{l}_1, \dots, \vec{l}_n$ and $r = \text{rank}(A)$ then

$$\mathcal{R}(A) = \text{span}\left\{\vec{l}_1, \dots, \vec{l}_r\right\}$$

Related to LU decomposition, we will have the following identities;

- $\mathcal{R}(A) = \mathcal{R}(L)$
- $\mathcal{N}(A) = \mathcal{N}(U)$
- $\mathcal{R}(A^T) = \mathcal{R}(U^T)$
- $\mathcal{N}(A^T) = \mathcal{N}(L^T)$

(note that L must only be the matrix containing the first r columns for these properties to hold)

1.4.2 Inner Product and Orthogonality

The inner product of two vectors, $\vec{x},\ \vec{y} \in \mathbb{R}^n$ is defined to be

$$\langle \vec{x}, \vec{y} \rangle = \sum_{k=1}^{n} x_k y_k = x_1 y_1 + x_2 y_2 + \cdots x_n y_n$$

The inner product is also called the dot product, written as $\vec{x} \cdot \vec{y}$
For complex vectors, $\vec{x},\ \vec{y} \in \mathbb{C}^n$, we conjugate the 2nd vector in the inner product.

$$\langle \vec{x}, \vec{y} \rangle = \sum_{k=1}^{n} x_k \bar{y}_k = x_1 \bar{y}_1 + \cdots x_n \bar{y}_n$$

Some properties of the dot inner product

- $\langle a\vec{x} + b\vec{y}, \vec{z}\rangle = a\langle \vec{x}, \vec{z}\rangle + b\langle \vec{y}, \vec{z}\rangle$
- $\langle \vec{x}, \vec{y}\rangle = \langle \vec{y}, \vec{x}\rangle$
- $\langle \vec{x}, \vec{x}\rangle = \|\vec{x}\|^2 \geq 0$
- $\langle \vec{x}, \vec{y}\rangle = \|\vec{x}\|\|\vec{y}\|\cos\theta$ where θ is the angle between vectors
- $\langle \vec{x}, \vec{y}\rangle = \vec{x}^T\vec{y}$

A less intuitive property is that
$$\langle A\vec{x}, \vec{y}\rangle = \left\langle \vec{x}, A^T\vec{y}\right\rangle$$
Proof:
$$\langle A\vec{x}, \vec{y}\rangle = (A\vec{x})^T\vec{y} = \vec{x}^T A^T\vec{y} = \left\langle \vec{x}, A^T\vec{y}\right\rangle$$
Two vectors $\vec{x},\ \vec{y} \in \mathbb{R}^n$ are *orthogonal* if $\langle \vec{x}, \vec{y}\rangle = 0$
We can define an *orthogonal set* if we have $\{\vec{x}_1, \ldots, \vec{x}_m\} \subseteq \mathbb{R}^n$ such that $\langle \vec{x}_i, \vec{x}_j\rangle = 0$ for all $i \neq j$
If each of these vectors in the set has magnitude 1 (set of unit vectors) then the set is considered to be *orthonormal*.
Naturally, if a set is orthogonal, then they are linearly independent.
We can say that two sets are orthogonal if $\langle \vec{x}, \vec{y}\rangle = 0\ \forall \vec{x} \in S_1 \wedge \vec{y} \in S_2$. We write this as $S_1 \perp S_2$.

This idea of the inner product and orthogonality can be extended to functions as well. The vectors we're used to dealing with exist in what's called Euclidean space which is a finite dimensional vector space. A function will have infinite inputs and so cannot be expressed in Euclidean space. What we can do instead is create a new vector space that has an inifnite dimensional product space. This is called the *Hilbert space*. Using this, we will have an infinite sum representing the infinte input values that are contained in this vector space. This implies that the definition of the inner product will involve integration instead of differentiation. And so the inner product between two real functions can be defined as
$$\boxed{\langle f, g\rangle = \int_{-\infty}^{\infty} f(x)g(x)dx}$$
For functions that are 2π periodic, we can define the inner product over a different domain
$$\langle f, g\rangle = \int_{0}^{2\pi} f(x)g(x)dx$$
Two functions are said to be orthogonal if their inner product gives 0.
Orthogonal functions are particularly useful when aiming to create a set of basis functions. In particular, trigonometric functions have some nice orthogonality properties:
For where $m, n \in \mathbb{N}$
$$\int_{-L}^{L} \sin\left(\frac{n\pi x}{L}\right)\sin\left(\frac{m\pi x}{L}\right)dx = \begin{cases} 0 & \text{if } m \neq n \\ L & \text{if } m = n \end{cases}$$
$$\int_{-L}^{L} \cos\left(\frac{n\pi x}{L}\right)\cos\left(\frac{m\pi x}{L}\right)dx = \begin{cases} 0 & \text{if } m \neq n \\ L & \text{if } m = n \neq 0 \\ 2L & \text{if } m = n = 0 \end{cases}$$

$$\int_{-L}^{L} \sin\left(\frac{n\pi x}{L}\right)\cos\left(\frac{m\pi x}{L}\right)dx = 0$$

Derivation:

$$I = \int_{-L}^{L} \sin\left(\frac{n\pi x}{L}\right)\sin\left(\frac{m\pi x}{L}\right)dx$$

case $m \neq n$:

$$I = 2\int_{0}^{L} \sin\left(\frac{n\pi x}{L}\right)\sin\left(\frac{m\pi x}{L}\right)dx$$

$$\cos(A-B) = \cos A\cos B + \sin A\sin B$$

$$\cos(A+B) = \cos A\cos B - \sin A\sin B$$

$$\sin A\sin B = \frac{1}{2}\left(\cos(A-B) - \cos(A+B)\right)$$

$$I = \frac{2}{2}\int_{0}^{L}\left(\cos\left(\frac{n-m}{L}\pi x\right) - \cos\left(\frac{n+m}{L}\pi x\right)\right)dx$$

$$I = \left[\frac{L}{(n-m)\pi}\sin\left(\frac{n-m}{L}\pi x\right)\right]_0^L - \left[\frac{L}{(n+m)\pi}\sin\left(\frac{n+m}{L}\pi x\right)\right]_0^L = 0$$

case $m = n$:

$$I = 2\int_{0}^{L}\sin^2\left(\frac{n\pi x}{L}\right)dx$$

$$\cos 2A = 1 - \sin^2 A \Rightarrow \sin^2 A = \frac{1-\cos 2A}{2}$$

$$I = \frac{2}{2}\int_{0}^{L}\left(1 - \cos\left(\frac{2n\pi x}{L}\right)\right)dx$$

$$I = x\Big|_0^L - \frac{L}{2n\pi}\sin\left(\frac{2n\pi x}{L}\right)\Big|_0^L = L$$

Note that this result makes use of the fact that $\sin(kx) = 0$ for where $k \in \mathbb{Z}$, hence why some terms appear to drop out. This also means that these specific orthogonality calculations only hold for integer values of m and n.

1.4.3 Orthogonal Complement

Def: If $U \in \mathbb{R}^n$ is a subspace, then the *orthogonal complement* of U is the set $U^\perp = \{\vec{x} \in \mathbb{R}^n : \langle \vec{x}, \vec{u}\rangle = 0\ \forall \vec{u} \in U\}$
What this means is that the orthogonal complement of U is the subspace that contains all of the vectors that are not contained in U within $\mathbb{R}^n$.
This means that if $U \in \mathbb{R}^n$

$$\dim(U) + \dim(U^\perp) = n$$

To find $U^\perp$, we use a basis of U. Let $\{\vec{u}_1, \ldots, \vec{u}_m\}$ be a basis of U. A vector $\vec{x}$ is an element of $U^\perp$ if the inner product with $\vec{x}$ and every basis vector is 0

$$\vec{x} \in U^\perp \Rightarrow \begin{cases} \langle \vec{x}, \vec{u}_1\rangle = 0 \\ \vdots \\ \langle \vec{x}, \vec{u}_m\rangle = 0 \end{cases} = \begin{cases} \vec{u}_1^T\vec{x} = 0 \\ \vdots \\ \vec{u}_m^T\vec{x} = 0 \end{cases} = \begin{bmatrix} \vec{u}_1^T \\ \vdots \\ \vec{u}_m^T \end{bmatrix}\vec{x} = 0$$

If we define

$$A = \begin{bmatrix} \vec{u}_1 & \cdots & \vec{u}_m \end{bmatrix}$$

Then the transpose is

$$A^T = \begin{bmatrix} \vec{u}_1^T \\ \vdots \\ \vec{u}_m^T \end{bmatrix}$$

And so we get that $\vec{x} \in U^\perp$ iff $A^T\vec{x} = 0$. This implies that $U^\perp = \mathcal{N}(A^T)$. If you notice that U in this case is the same as the range of A, we can write a more general formula

$$\mathcal{R}(A)^\perp = \mathcal{N}(A^T)$$

We can rearrange this formula to get 3 more identities:

$$\mathcal{N}(A) = \mathcal{R}(A^T)^\perp$$

$$\mathcal{N}(A)^\perp = \mathcal{R}(A^T)$$

$$\mathcal{R}(A) = \mathcal{N}(A^T)^\perp$$

Note that the dimensions of these fundamental subspaces will be as follows for an $m \times n$ matrix A.

$$\dim(\mathcal{R}(A)) + \dim(\mathcal{R}(A)^\perp) = n$$

$$\dim(\mathcal{N}(A)) + \dim(\mathcal{N}(A)^\perp) = m$$

(the nullspace corresponds to the dimension of the input of A and the range corresponds to the dimension of the output of A)
Ex: Find a basis for $U^\perp$ where

$$U = \text{span}\left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} -2 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \\ 4 \end{bmatrix} \right\}$$

$$U^\perp = \mathcal{N}(A^T) \text{ where } A^T = \begin{bmatrix} 1 & 2 & 3 \\ -2 & 1 & -1 \\ 3 & 1 & 4 \end{bmatrix}$$

Find all $\vec{x}$ such that $A^T\vec{x} = 0$

$$\left[\begin{array}{ccc|c} 1 & 2 & 3 & 0 \\ -2 & 1 & -1 & 0 \\ 3 & 1 & 4 & 0 \end{array}\right] \to \begin{bmatrix} 1 & 2 & 3 \\ 0 & 5 & 5 \\ 0 & 0 & 0 \end{bmatrix} \Rightarrow \vec{x} = t \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}$$

$$U^\perp = \text{span}\left\{ \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix} \right\}$$

Ex2: Find the dimensions of $\mathcal{N}(A), \mathcal{R}(A), \mathcal{N}(A^T), \mathcal{R}(A^T)$

$$
\begin{aligned}
&A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & -2 & 1 \end{bmatrix} \begin{bmatrix} 2 & 3 & 1 & 4 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & 0 & 2 \end{bmatrix} \\
&\dim(A) = \text{rank}(U) = 3 = \dim(\mathcal{R}(A)) \\
&\mathcal{R}(A) \in \mathbb{R}^3 \\
&\mathcal{N}(A^T) = \mathcal{R}(A)^\perp \\
&\dim(\mathcal{R}(A)^\perp) + \dim(\mathcal{R}(A)) = 3 \Rightarrow \dim(\mathcal{R}(A)^\perp) = \dim(\mathcal{N}(A^T)) = 0 \\
&\dim(\mathcal{N}(A)) + \dim(\mathcal{R}(A)) = 4 \Rightarrow \dim(\mathcal{N}(A)) = 1 \\
&\mathcal{N}(A)^\perp = \mathcal{R}(A^T) \\
&\mathcal{N}(A)^\perp = \mathcal{R}(A^T) \\
&\dim(\mathcal{N}(A)) + \dim(\mathcal{N}(A)^\perp) = 4 \Rightarrow \dim(\mathcal{N}(A)^\perp) = 3 \\
&\mathcal{N}(A)^\perp = \mathcal{R}(A^T) \Rightarrow \dim(\mathcal{R}(A^T)) = 3
\end{aligned}
$$

A handy identity is that $\text{rank}(A) = \text{rank}(A^T)$
Proof:

$$
\begin{aligned}
&\text{rank}(A) = \dim(\mathcal{R}(A)) = n - \dim(\mathcal{N}(A)) \\
&\mathcal{N}(A) \in \mathbb{R}^n \\
&\dim(\mathcal{N}(A)) + \dim(\mathcal{N}(A)^\perp) = n \\
&\dim(\mathcal{N}(A)) = n - \dim(\mathcal{N}(A)^\perp) \\
&\mathcal{N}(A)^\perp = \mathcal{R}(A^T) \\
&\dim(\mathcal{N}(A)) = n - \dim(\mathcal{R}(A^T)) \\
&\text{rank}(A) = n - \big(n - \dim(\mathcal{R}(A^T))\big) \\
&\text{rank}(A) = \text{rank}(A^T) \ \square
\end{aligned}
$$

1.4.4 Orthogonal Projection

The projection of a vector $\vec{x} \in \mathbb{R}^n$ onto another vector $\vec{u} \in \mathbb{R}^n$ is given by

$$
\text{proj}_{\vec{u}}(\vec{x}) = \frac{\langle \vec{x}, \vec{u} \rangle}{\langle \vec{u}, \vec{u} \rangle} \vec{u}
$$

We can write $\text{proj}_{\vec{u}}(\vec{x})$ as a matrix multiplication:

$$
\begin{aligned}
&\text{proj}_{\vec{u}}(\vec{x}) = \frac{\langle \vec{x}, \vec{u} \rangle}{\langle \vec{u}, \vec{u} \rangle} \vec{u} = \frac{1}{\langle \vec{u}, \vec{u} \rangle} \vec{u} \langle \vec{x}, \vec{u} \rangle = \frac{1}{\|\vec{u}\|^2} \vec{u}\vec{u}^T \vec{x} = P\vec{x} \\
&P = \frac{1}{\|\vec{u}\|^2} \vec{u}\vec{u}^T = \frac{1}{\|\vec{u}\|^2} \begin{bmatrix} u_1 u_1 & \cdots & u_1 u_n \\ \vdots & \ddots & \vdots \\ u_n u_1 & \cdots & u_n u_n \end{bmatrix}
\end{aligned}
$$

Properties of the projection matrix

- $P^T = P$ (symmetric matrix)
- $P^2 = P$ (eigenpotent matrix)
- $\text{rank}(P) = 1$

An *orthogonal basis* of U is a basis of U that consists of orthogonal vectors. If these vectors all have a magnitude of 1, then it is called an *orthonormal basis.*
To find an orthogonal basis, we can use the Gram-Schmidt algorithm. The algorithm produces the following vectors from the basis vectors $\{\vec{u}_1, \ldots, \vec{u}_m\}$

$$\begin{aligned}
&\vec{v}_1 = \vec{u}_1 \\
&\vec{v}_2 = \vec{u}_2 - \text{proj}_{\vec{v}_1}(\vec{u}_2) \\
&\vec{v}_3 = \vec{u}_3 - \text{proj}_{\vec{v}_1} - \text{proj}_{\vec{v}_2}(\vec{u}_2) \\
&\vdots \\
&\vec{v}_m = \vec{u}_m - \text{proj}_{\vec{v}_1}(\vec{u}_m) - \cdots - \text{proj}_{\vec{v}_{m-1}}(\vec{u}_m)
\end{aligned}$$

The vectors $\{\vec{v}_1, \ldots, \vec{v}_m\}$ will form an orthogonal basis of U.
We can also normalize the vectors, such that $\vec{w}_i = \frac{\vec{v}_i}{\|\vec{v}_i\|}$ so that $\{\vec{w}_1, \ldots, \vec{w}_m\}$ forms an orthonormal basis of U.
Ex: Construct an orthonormal basis of

$$U = \text{span}\left\{\begin{bmatrix}1\\0\\1\\0\end{bmatrix}, \begin{bmatrix}2\\1\\0\\-2\end{bmatrix}, \begin{bmatrix}3\\5\\-1\\1\end{bmatrix}\right\} \subseteq \mathbb{R}^4$$

$$\vec{v}_1 = \vec{u}_1 = \begin{bmatrix}1\\0\\1\\0\end{bmatrix}$$

$$\vec{v}_2 = \vec{u}_2 - \text{proj}_{\vec{v}_1}(\vec{u}_2) = \begin{bmatrix}2\\1\\0\\-2\end{bmatrix} - \frac{\langle \vec{u}_2, \vec{v}_1\rangle}{\langle \vec{v}_1, \vec{v}_1\rangle}\vec{v}_1 = \begin{bmatrix}2\\1\\0\\-2\end{bmatrix} - \frac{2}{2}\begin{bmatrix}1\\0\\1\\0\end{bmatrix} = \begin{bmatrix}1\\1\\-1\\-2\end{bmatrix}$$

$$\vec{v}_3 = \vec{u}_3 - \frac{\langle \vec{u}_3, \vec{v}_2\rangle}{\langle \vec{v}_2, \vec{v}_2\rangle}\vec{v}_2 - \frac{\langle \vec{u}_3, \vec{v}_1\rangle}{\langle \vec{v}_1, \vec{v}_1\rangle}\vec{v}_1 = \begin{bmatrix}3\\5\\-1\\1\end{bmatrix} - \frac{7}{7}\begin{bmatrix}1\\1\\-1\\-2\end{bmatrix} - \frac{2}{2}\begin{bmatrix}1\\0\\1\\0\end{bmatrix} = \begin{bmatrix}1\\4\\-1\\3\end{bmatrix}$$

$\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ forms an orthogonal basis
normalize to get

$$\left\{\frac{1}{\sqrt{2}}\begin{bmatrix}1\\0\\1\\0\end{bmatrix}, \frac{1}{\sqrt{7}}\begin{bmatrix}1\\1\\-1\\-2\end{bmatrix}, \frac{1}{3\sqrt{3}}\begin{bmatrix}1\\4\\-1\\3\end{bmatrix}\right\}$$

If $\{\vec{u}_1, \ldots, \vec{u}_m\}$ forms an orthonormal basis of $U \subseteq \mathbb{R}^n$ then the orthogonal projection of $\vec{x} \in \mathbb{R}^n$ onto the subspace U is given by the sum of the projections onto the basis vectors

$$\text{proj}_U(\vec{x}) = \sum_{i=1}^{m} \text{proj}_{\vec{u}_i}(\vec{x}) = \sum_{i=1}^{m} \frac{\langle \vec{x}, \vec{u}_i \rangle}{\langle \vec{u}_i, \vec{u}_i \rangle} \vec{u}_i$$

The projection matrix is given by $\text{proj}_U(\vec{x}) = P\vec{x}$ where

$$P = \sum_{i=1}^{m} \frac{1}{\|\vec{u}_i\|^2} \vec{u}_i \vec{u}_i^T$$

This projection matrix will have the properties

- $P^T = P$
- $P^2 = P$
- $\text{rank}(P) = \dim(U)$

We can also write P as $Q_1 Q_1^T$.
If $U \subseteq \mathbb{R}^n$ is a subspace then any $\vec{x} \in \mathbb{R}^n$ can be written as $\vec{x} = \vec{u} + \vec{v}$ for some $\vec{u} \in U$ and $\vec{v} \in U^\perp$. This means that we can write any vector as the sum of orthogonal projections

$$\vec{x} = \text{proj}_U(\vec{x}) + \text{proj}_{U^\perp}(\vec{x})$$

This allows us to express the projection onto the orthogonal complement of U as

$$P^\perp = P - I$$

Ex: Find the projection matrix that projects onto U for where

$$U = \text{span}\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \\ 1 \end{bmatrix} \right\}$$

find an orthonormal basis of U

$$\vec{v}_1 = \vec{u}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\vec{v}_2 = \vec{u}_2 - \frac{\langle \vec{u}_2, \vec{v}_1 \rangle}{\langle \vec{v}_1, \vec{v}_1 \rangle} \vec{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 0 \\ 1 \end{bmatrix} - \frac{4}{4} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ -1 \\ 0 \end{bmatrix}$$

$$\vec{v}_3 = \vec{u}_3 - \frac{\langle \vec{u}_3, \vec{v}_1 \rangle}{\langle \vec{v}_1, \vec{v}_1 \rangle} \vec{v}_1 - \frac{\langle \vec{u}_3, \vec{v}_2 \rangle}{\langle \vec{v}_2, \vec{v}_2 \rangle} \vec{v}_2 = \begin{bmatrix} -1 \\ 0 \\ 1 \\ 1 \end{bmatrix} - \frac{1}{4} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} - \frac{-1}{2} \begin{bmatrix} 0 \\ 1 \\ -1 \\ 0 \end{bmatrix} = \begin{bmatrix} -\frac{5}{4} \\ \frac{1}{4} \\ \frac{1}{4} \\ \frac{3}{4} \end{bmatrix}$$

$$\Rightarrow \left\{ \begin{bmatrix}1\\1\\1\\1\end{bmatrix}, \begin{bmatrix}0\\1\\-1\\0\end{bmatrix}, \begin{bmatrix}-5\\1\\1\\3\end{bmatrix} \right\} \text{ is an orthogonal basis}$$

$$P = \frac{1}{\|\vec{v}_1\|^2}\vec{v}_1\vec{v}_1^T + \frac{1}{\|\vec{v}_2\|^2}\vec{v}_2\vec{v}_2^T + \frac{1}{\|\vec{v}_3\|^2}\vec{v}_3\vec{v}_3^T$$

$$P = \frac{1}{4}\begin{bmatrix}1&1&1&1\\1&1&1&1\\1&1&1&1\\1&1&1&1\end{bmatrix} + \frac{1}{2}\begin{bmatrix}0&0&0&0\\0&1&-1&0\\0&-1&1&0\\0&0&0&0\end{bmatrix} + \frac{1}{36}\begin{bmatrix}25&-5&-5&-15\\-5&1&1&3\\-5&1&1&3\\-15&3&3&9\end{bmatrix}$$

$$P = \frac{1}{36}\begin{bmatrix}34&4&4-6&\\4&28&-8&12\\4&-8&28&12\\-6&12&12&18\end{bmatrix}$$

This was quite computationally intensive. It may be easier in some cases to instead find $P^\perp$ and then use $P = I - P_\perp$ to get P.
We can say that $U = \mathcal{R}(A)$ where

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

$$U^\perp = \mathcal{R}(A)^\perp = \mathcal{N}(A^T)$$

$$A^T = \begin{bmatrix}1&1&1&\\1&2&0&1\\-1&0&1&1\end{bmatrix} \to \begin{bmatrix}1&1&1&1\\0&1&-1&0\\0&1&-2&2\end{bmatrix} \to \begin{bmatrix}1&1&1&1\\0&1&-1&0\\0&0&3&2\end{bmatrix}$$

$$x_4 = t \Rightarrow x_3 = -\frac{2}{3}t \Rightarrow x_2 = -\frac{2}{3}t \Rightarrow x_1 = \frac{1}{3}t$$

$$\vec{x} = \frac{t}{3}\begin{bmatrix}1\\-2\\-2\\3\end{bmatrix} \Rightarrow U^\perp = \text{span}\begin{bmatrix}1\\-2\\-2\\3\end{bmatrix}$$

$$P_\perp = \frac{1}{\|\vec{v}_1\|^2}\vec{v}_1\vec{v}_1^T = \frac{1}{18}\begin{bmatrix}1&-2&-2&3\\-2&4&4&-6\\2&4&4&-6\\3&-6&-6&9\end{bmatrix}$$

$$P = I - P_\perp = \frac{1}{18}\begin{bmatrix}17&2&2&-3\\2&14&-4&6\\2&-4&14&6\\-3&6&6&9\end{bmatrix}$$

We can find the shortest distance between a vector and a subspace using projections.
The vector distance $\|\vec{x} - \text{proj}_U(\vec{x})\| = \|\text{proj}_{U^\perp}(\vec{x})\|$ will be the shortest distance between the vector

$\vec{x}$ and the subspace U.
Ex: Find the shortest distance between $\vec{x}$ and U for

$$\vec{x} = \begin{bmatrix} 1 \\ 1 \\ -2 \\ 1 \end{bmatrix} \text{ and } U = \text{span}\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \\ 1 \end{bmatrix} \right\}$$

from the previous example, we found that

$$U^{\perp} = \text{span}\left\{ \begin{bmatrix} 1 \\ -2 \\ -2 \\ 3 \end{bmatrix} \right\}$$

$$\| \operatorname{proj}_{U^{\perp}}(\vec{x})\| = \left\| \frac{\langle \vec{x}, \vec{w} \rangle}{\langle \vec{w}, \vec{w} \rangle} \vec{w} \right\| = \left| \frac{\langle \vec{x}, \vec{w} \rangle}{\langle \vec{w}, \vec{w} \rangle} \right| \|\vec{w}\| = \left| \frac{6}{18} \right| \sqrt{18} = \sqrt{2}$$

1.4.5 QR Decomposition

The idea behind a QR decomposition is to form a decomposition that contains orthonormal bases of both $\mathcal{R}(A)$ and $\mathcal{R}(A)^{\perp}$.
A matrix Q is orthogonal if

- $Q^T Q = I$
- $Q Q^T = I$

This can more simply be written as

$$Q^T = Q^{-1}$$

An orthogonal matrix is square and invertible.
A matrix is orthogonal if the columns are orthonormal ($Q^T Q = I$) and the rows are orthonormal ($QQ^T = I$).
Note that the norm of an orthonormal matrix is 1:

$$\|Q\| = 1$$

We will also have the condition number of an orthogonal matrix is 1.
Some examples of orthogonal matrices include rotations and reflections.

$$\text{Rot}_{\theta} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

$$\text{Ref} = 2P - I$$

We can write the QR decomposition of A as $A = QR$, where Q is an orthogonal matrix and R is an upper triangular matrix.
We can do this a few different ways. One way is using Gram-Schmidt.
If we let the column vectors of A be $\{\vec{a}_1, \ldots, \vec{a}_n\}$ and we let $\{\vec{w}_1, \ldots, \vec{w}_n\}$ be an orthonormal basis of $\mathcal{R}(A)$ then we can express each $\vec{a}_k$ as the sum of its projections onto $\mathcal{R}(A)$.

$$\vec{a}_1 = \langle \vec{a}, \vec{w}_1 \rangle \vec{w}_1$$

$$\vec{a}_2 = \langle \vec{a}_2, \vec{w}_1 \rangle \vec{w}_1 + \langle \vec{a}_2, \vec{w}_2 \rangle \vec{w}_2$$

$$\vdots$$

$$\vec{a}_n = \langle \vec{a}_n, \vec{w}_1 \rangle \vec{w}_1 + \cdots + \langle \vec{a}_n, \vec{w}_n \rangle \vec{w}_n$$

(note that we don't need to normalize because the norm of $\vec{w}_i$ is 1)
We can write this as matrix multiplication;

$$A = \underbrace{\begin{bmatrix} \vec{w}_1 & \cdots & \vec{w}_n \end{bmatrix}}_{Q_{1_{m\times n}}} \underbrace{\begin{bmatrix} \langle \vec{w}_1, \vec{a}_1 \rangle & \langle \vec{w}_1, \vec{a}_2 \rangle & \cdots & \langle \vec{w}_1, \vec{a}_n \rangle \\ 0 & \langle \vec{w}_2, \vec{a}_2 \rangle & \cdots & \langle \vec{w}_2, \vec{a}_n \rangle \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \langle \vec{w}_n, \vec{a}_n \rangle \end{bmatrix}}_{R_{1_{n\times m}}}$$

$A = Q_1 R_1$ is called the *thin QR decomposition* (or reduced QR) of A.
Ex: Find the QR decomposition of

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

$$\vec{v}_1 = \vec{a}_1 = \begin{bmatrix} 1 \\ 0 \\ -1 \\ 0 \end{bmatrix}$$

$$\vec{v}_2 = \vec{a}_2 - \frac{\langle \vec{a}_2, \vec{v}_1 \rangle}{\langle \vec{v}_1, \vec{v}_1 \rangle} \vec{v}_1 = \begin{bmatrix} 3 \\ 2 \\ -1 \\ -2 \end{bmatrix} - \frac{4}{2} \begin{bmatrix} 1 \\ 0 \\ -1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 1 \\ -2 \end{bmatrix}$$

$$\vec{v}_3 = \vec{a}_3 - \frac{\langle \vec{a}_3, \vec{v}_1 \rangle}{\langle \vec{v}_1, \vec{v}_1 \rangle} \vec{v}_1 - \frac{\langle \vec{a}_3, \vec{v}_2 \rangle}{\langle \vec{v}_2, \vec{v}_2 \rangle} \vec{v}_2 = \begin{bmatrix} 1 \\ 3 \\ 1 \\ -1 \end{bmatrix} - 0\vec{v}_1 - \frac{10}{10} \begin{bmatrix} 1 \\ 2 \\ 1 \\ -2 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}$$

normalize to get

$$\vec{w}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ -1 \\ 0 \end{bmatrix}, \ \vec{w}_2 = \frac{1}{\sqrt{10}} \begin{bmatrix} 1 \\ 2 \\ 1 \\ -2 \end{bmatrix}, \ \vec{w}_3 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}$$

$$Q_1 = \begin{bmatrix} \vec{w}_1 & \vec{w}_2 & \vec{w}_3 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{10}} & 0 \\ 0 & \frac{2}{\sqrt{10}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{10}} & 0 \\ 0 & -\frac{2}{\sqrt{10}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

$$R_1 = \begin{bmatrix} \langle \vec{w}_1, \vec{a}_1 \rangle & \langle \vec{w}_1, \vec{a}_2 \rangle & \langle \vec{w}_3, \vec{a}_3 \rangle \\ 0 & \langle \vec{w}_2, \vec{a}_2 \rangle & \langle \vec{w}_2, \vec{a}_3 \rangle \\ 0 & 0 & \langle \vec{w}_3, \vec{a}_3 \rangle \end{bmatrix} = \begin{bmatrix} \sqrt{2} & 2\sqrt{2} & 0 \\ 0 & \sqrt{10} & \sqrt{10} \\ 0 & 0 & \sqrt{2} \end{bmatrix}$$

For the full QR decomposition, we continue by extending $\{\vec{w}_1, \ldots, \vec{w}_n\}$ to an orthonormal basis of $\mathbb{R}^m$.

$$\underbrace{\vec{w}_1, \ldots, \vec{w}_n}_{\text{orthonormal basis of } \mathcal{R}(A)}, \quad \underbrace{\vec{w}_{n+1}, \ldots, \vec{w}_m}_{\text{orthonormal basis of } \mathcal{R}(A)^\perp}$$

Then we get that

$$A = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} R_1 \\ 0 \end{bmatrix}$$

For the full QR, we extend $\{\vec{w}_1, \vec{w}_2, \vec{w}_3\}$ to an orthonormal basis of $\mathbb{R}^n$

$$\mathcal{R}(A)^\perp = \mathcal{N}(A^T)$$

$$A^T = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 3 & 2 & -1 & -2 \\ 1 & 3 & 1 & -1 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 1 & -1 \\ 0 & 0 & -1 & 2 \end{bmatrix} \Rightarrow \vec{x} = t \begin{bmatrix} 2 \\ -1 \\ 2 \\ 1 \end{bmatrix}$$

$$\vec{w}_n = \frac{1}{\sqrt{10}} \begin{bmatrix} 2 \\ -1 \\ 2 \\ 1 \end{bmatrix}$$

$$Q = \begin{bmatrix} \vec{w}_1 & \vec{w}_2 & \vec{w}_3 & \vec{w}_n \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{10}} & 0 & \frac{2}{\sqrt{10}} \\ 0 & \frac{2}{\sqrt{10}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{10}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{10}} & 0 & \frac{2}{\sqrt{10}} \\ 0 & -\frac{2}{\sqrt{10}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{10}} \end{bmatrix}$$

$$R = \begin{bmatrix} \sqrt{2} & 2\sqrt{2} & 0 \\ 0 & \sqrt{10} & \sqrt{10} \\ 0 & 0 & \sqrt{2} \\ 0 & 0 & 0 \end{bmatrix}$$

1.4.6 Least Squares Approximation

Say we want to fit a large collection of data points to some function (often we use a linear function), we can use the least squares approximation.
We want to find the vector $\vec{x}$ that minimizes the least squares error, $\|A\vec{x} - \vec{b}\|$. where $\vec{b}$ represents the y-data, A represents the x-data for each coefficient, and $\vec{x}$ contains the coefficients.
For any $\vec{x} \in \mathbb{R}^n$ we have that $A\vec{x} \in \mathcal{R}(A)$. The projection theorem states that the vector in $\mathcal{R}(A)$ nearest to $\vec{b}$ is the orthogonal projection of $\vec{b}$ onto $\mathcal{R}(A)$.

$$A\vec{x} = \text{proj}_{\mathcal{R}(A)}(\vec{b})$$
$$\vec{b} - A\vec{x} \in \mathcal{R}(A)^\perp = \mathcal{N}(A^T)$$
$$\Rightarrow A^T(\vec{b} - A\vec{x}) = \vec{0}$$
$$A^T A\vec{x} = A^T\vec{b}$$

Thus, the least squares approximation to $A\vec{x} = \vec{b}$ will be the given by the vector $\vec{x}$ that solves

$$A^T A\vec{x} = A^T\vec{b}$$

- A is $m \times n$ with $\text{rank}(A) = n$
- A^TA is $n \times n$ with $\text{rank}(A^TA) = n$

A is full rank which implies A^TA is a square matrix with full rank. This means that A^TA is invertible, implying that there is a unique solution for $\vec{x}$.
We can also solve the least squares approximation using QR decomposition.
We can say $Q^T\vec{b} = \begin{bmatrix} \vec{c}_1 \\ \vec{c}_2 \end{bmatrix}$ and that $Q_1\vec{b} = \vec{c}_1$.

$$\begin{aligned}
&\|A\vec{x} - \vec{b}\|^2 = \|QR\vec{x} - QQ^T\|^2 = \|Q(R\vec{x} - Q^T\vec{b})\|^2 = \|R\vec{x} - Q^T\vec{b}\|^2 \\
&= \left\| \begin{bmatrix} R_1\vec{x} \\ \vec{0} \end{bmatrix} - \begin{bmatrix} \vec{c}_1 \\ \vec{c}_2 \end{bmatrix} \right\|^2 = \left\| \begin{bmatrix} R_1\vec{x} - \vec{c}_1 \\ \vec{0} \end{bmatrix} + \begin{bmatrix} \vec{0} \\ -\vec{c}_2 \end{bmatrix} \right\|^2 = \left\| \begin{bmatrix} R_1\vec{x} - \vec{c}_1 \\ \vec{0} \end{bmatrix} \right\|^2 + \left\| \begin{bmatrix} \vec{0} \\ -\vec{c}_2 \end{bmatrix} \right\|^2 \\
&= \|R_1\vec{x} - \vec{c}\|^2 + \|\vec{c}_2\|^2
\end{aligned}$$

This means that $\|A\vec{x} - \vec{b}\|$ is minimal when

$$R_1\vec{x} = \vec{c}_1$$

Ex: Find the least squares approximation to $A\vec{x} = \vec{b}$ where

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}, \ \vec{b} = \begin{bmatrix} 0 \\ 3 \\ 1 \end{bmatrix}$$

Using normal equations: $A^TA\vec{x} = A^T\vec{b}$

$$A^TA = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix}$$

$$A^T\vec{b} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 0 \\ 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 9 \end{bmatrix}$$

$$\left[\begin{array}{cc|c} 3 & 6 & 4 \\ 6 & 14 & 9 \end{array}\right] \to \left[\begin{array}{cc|c} 3 & 6 & 4 \\ 0 & 2 & 1 \end{array}\right] \Rightarrow x_2 = \frac{1}{2} \Rightarrow x_1 = \frac{1}{3} \Rightarrow \vec{x} = \begin{bmatrix} 1/3 \\ 1/2 \end{bmatrix}$$

Using QR decomposition $R_1\vec{x} = Q_1^T\vec{b}$

$$\vec{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\vec{v}_2 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - \frac{6}{3}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

$$A = Q_1R_1 = \begin{bmatrix} \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & 0 \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \sqrt{3} & 2\sqrt{3} \\ 0 & \sqrt{2} \end{bmatrix}$$

$$\vec{c}_1 = Q_1^T\vec{b} = \begin{bmatrix} \frac{4}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

$$R_1\vec{x} = Q_1^T\vec{b}$$

$$\left[\begin{array}{cc|c} \sqrt{3} & 2\sqrt{3} & \frac{4}{\sqrt{3}} \\ 0 & \sqrt{2} & \frac{1}{\sqrt{2}} \end{array}\right] \rightarrow \left[\begin{array}{cc|c} 3 & 6 & 4 \\ 0 & 2 & 1 \end{array}\right] \Rightarrow \vec{x} = \begin{bmatrix} 1/3 \\ 1/2 \end{bmatrix}$$

1.5 Eigen-analysis

1.5.1 Eigenvalues and Eigenvectors

If we think of a matrix A as a linear transformation, an eigenvector is a nonzero vector that's direction is unchanged by the transformation. This can be expressed as

$$A\vec{v} = \lambda\vec{v}$$

or more generally as,

$$A^n\vec{v} = \lambda^n\vec{v}$$

The value λ is called the *eigenvalue* that corresponds to the given eigenvector.
To solve for these eigenvalues, we can do the following,

$$\begin{aligned} &A\vec{v} = \lambda\vec{v} \\ &A\vec{v} = \lambda I\vec{v} \\ &A\vec{v} - \lambda I\vec{v} = 0 \\ &(A - \lambda I)\vec{v} = 0 \\ &\Rightarrow \det(A - \lambda I) = 0 \end{aligned}$$

To then find the eigenvectors, we can solve the homogeneous system for each λ_i formed by $(A - \lambda_i I)\vec{v} = 0$
This will give a solution space of a line for $\vec{v}_i$. Because we only care about the direction of the eigenvectors, we can take the simplest eigenvector.

Ex: Compute eigenvectors for the matrix $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$

$$\det(A - \lambda I) = \det \begin{bmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{bmatrix} = (2-\lambda)(2-\lambda) - 1 = \lambda^2 - 4\lambda + 3 = (\lambda - 3)(\lambda - 1) = 0$$

$$\lambda_1 = 1,\ \lambda_2 = 3$$

$$\lambda_1 :\ A - \lambda_1 I = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \Rightarrow x_1 + x_2 = 0 \Rightarrow x_1 = -x_2 \Rightarrow \begin{cases} x_1 = -t \\ x_2 = t \end{cases}$$

$$\vec{v}_1 = [1, -1]^T$$

$$\lambda_2 :\ A - \lambda_2 I = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \rightarrow \begin{bmatrix} -1 & 1 \\ 0 & 0 \end{bmatrix} \Rightarrow -x_1 + x_2 = 0 \Rightarrow x_1 = x_2 \Rightarrow \begin{cases} x_1 = t \\ x_2 = t \end{cases}$$

$$\vec{v}_2 = [1, 1]^T$$

The eigenvectors of a matrix also form a basis, meaning that we can express any vector in that space as a linear combination of eigenvectors. This makes eigen-analysis very helpful for computing large powers of matrices.

Using the matrix from the above example, compute $A^{10}\begin{bmatrix}2\\3\end{bmatrix}$

$\begin{bmatrix}2\\3\end{bmatrix} = c_1\vec{v}_1 + c_2\vec{v}_2 = c_1\begin{bmatrix}1\\-1\end{bmatrix} + c_2\begin{bmatrix}1\\1\end{bmatrix}$

$\begin{bmatrix}1 & 1\\-1 & 1\end{bmatrix}\begin{bmatrix}c_1\\c_2\end{bmatrix} = \begin{bmatrix}2\\3\end{bmatrix}$

$\left[\begin{array}{cc|c}1 & 1 & 2\\-1 & 1 & 3\end{array}\right] \to \left[\begin{array}{cc|c}1 & 1 & 2\\0 & 2 & 5\end{array}\right] \to \left[\begin{array}{cc|c}1 & 1 & 2\\0 & 1 & \frac{5}{2}\end{array}\right] \to \left[\begin{array}{cc|c}1 & 0 & -\frac{1}{2}\\0 & 1 & \frac{5}{2}\end{array}\right]$

$\begin{bmatrix}c_1\\c_2\end{bmatrix} = \begin{bmatrix}-\frac{1}{2}\\\frac{5}{2}\end{bmatrix}$

$\begin{bmatrix}2\\3\end{bmatrix} = -\dfrac{1}{2}\vec{v}_1 + \dfrac{5}{2}\vec{v}_2$

$A^{10}\begin{bmatrix}2\\3\end{bmatrix} = A^{10}\left(-\dfrac{1}{2}\vec{v}_1 + \dfrac{5}{2}\vec{v}_2\right) = A^{10}\left(-\dfrac{1}{2}\vec{v}_1\right) + A^{10}\left(\dfrac{5}{2}\vec{v}_2\right)$

recall $A^n\vec{v} = \lambda^n\vec{v}$

$A^{10}\begin{bmatrix}2\\3\end{bmatrix} = -\dfrac{1}{2}(1^{10}\vec{v}_1) + \dfrac{5}{2}(3^{10}\vec{v}_2) = \begin{bmatrix}-\frac{1}{2} + \frac{5}{2}\cdot 3^{10}\\\frac{1}{2} + \frac{5}{2}\cdot 3^{10}\end{bmatrix} = \begin{bmatrix}147622\\147623\end{bmatrix}$

A handy method for finding/checking eigenvalues is to use the trace and determinant of the matrix. (Trace is the sum of the diagonal elements)

$$\operatorname{tr}(A) = \sum_{i=1}^{n}\lambda_i$$

$$\det(A) = \prod_{i=1}^{n}\lambda_i$$

1.5.2 Complex Eigen-Analysis

If A is a real matrix with complex eigenvalues then the complex eigenvalues and eigenvectors will always come in conjugate pairs.
This will also lead to the coefficients being conjugate pairs and allows us to express these conjugate pairs as,

$$2\Re(c_i\lambda_i^n\vec{v}_i)$$

We can do this because the imaginary components in a real matrix will always cancel.

Ex: for the matrix $A = \begin{bmatrix}0 & 0 & 3\\1 & 0 & -1\\0 & 1 & 3\end{bmatrix}$ compute $A^{13}\begin{bmatrix}1\\0\\0\end{bmatrix}$

$$\det(A-\lambda I)=\det\begin{bmatrix}-\lambda & 0 & 3\\ 1 & -\lambda & -1\\ 0 & 1 & 3-\lambda\end{bmatrix}=-\lambda\begin{vmatrix}-\lambda & -1\\ 1 & 3-\lambda\end{vmatrix}+3\begin{vmatrix}1 & -\lambda\\ 0 & 1\end{vmatrix}$$

$$=-\lambda(-\lambda(3-\lambda)+1)+3=-\lambda^3+3\lambda^2-\lambda+3=0$$

$$\Rightarrow(\lambda-3)(-\lambda^2-1)=0\Rightarrow\begin{cases}\lambda_1=3\\ \lambda_2=i\\ \lambda_3=-i\end{cases}$$

$$\lambda_1:\ A-3I=\begin{bmatrix}-3 & 0 & 3\\ 1 & -3 & -1\\ 0 & 1 & 0\end{bmatrix}\to\begin{bmatrix}1 & 0 & -1\\ 0 & -3 & 0\\ 0 & 1 & 0\end{bmatrix}\to\begin{bmatrix}1 & 0 & -1\\ 0 & 1 & 0\\ 0 & 0 & 0\end{bmatrix}$$

$$\Rightarrow\vec{v}_1=\begin{bmatrix}1 & 0 & 1\end{bmatrix}^T$$

$$\lambda_2:\ A-iI=\begin{bmatrix}-i & 0 & 3\\ 1 & -i & -1\\ 0 & 1 & 3-i\end{bmatrix}\to\begin{bmatrix}1 & -i & -1\\ 0 & 1 & 3-i\\ 0 & 0 & 0\end{bmatrix}$$

$$\Rightarrow\vec{v}_2=\begin{bmatrix}-3i & -3+i & 1\end{bmatrix}^T$$

$$\therefore\ \vec{v}_3=\begin{bmatrix}3i & -3-i & 1\end{bmatrix}^T$$

$$\begin{bmatrix}1\\0\\0\end{bmatrix}=c_1\vec{v}_1+c_2\vec{v}_2+c_3\vec{v}_3=c_1\begin{bmatrix}1\\0\\1\end{bmatrix}+c_2\begin{bmatrix}-3i\\-3+i\\1\end{bmatrix}+c_3\begin{bmatrix}3i\\-3-i\\1\end{bmatrix}$$

$$\begin{bmatrix}1 & -3i & 3i\\ 0 & -3+i & -3-i\\ 1 & 1 & 1\end{bmatrix}\begin{bmatrix}c_1\\c_2\\c_3\end{bmatrix}=\begin{bmatrix}1\\0\\0\end{bmatrix}$$

$$\left[\begin{array}{ccc|c}1 & -3i & 3i & 1\\ 0 & -3+i & -3-i & 0\\ 1 & 1 & 1 & 0\end{array}\right]\rightsquigarrow\vec{c}=\begin{bmatrix}1\\ -\frac{1}{20}+\frac{3}{20}i\\ \frac{1}{20}-\frac{3}{20}i\end{bmatrix}$$

$$A^{13}\begin{bmatrix}1\\0\\0\end{bmatrix}=c_1\lambda_1^{13}\vec{v}_1+c_2\lambda_2^{13}\vec{v}_2+c_3\lambda_3^{13}\vec{v}_3=c_1\lambda_1^{13}\vec{v}_1+2\Re(c_2\lambda_2^{13}\vec{v}_2)$$

$$\Re(c_2\lambda_2^{13}\vec{v}_2)=\Re\left(\left(-\frac{1}{20}+\frac{3}{20}i\right)\begin{bmatrix}-3i\\-3+i\\1\end{bmatrix}\right)=\begin{bmatrix}-\frac{3}{20}\\ \frac{1}{2}\\ -\frac{3}{20}\end{bmatrix}$$

$$A^{13}\begin{bmatrix}1\\0\\0\end{bmatrix}=3^{13}+2\begin{bmatrix}-\frac{3}{20}\\ \frac{1}{2}\\ -\frac{3}{20}\end{bmatrix}=\begin{bmatrix}3^{13}-\frac{3}{10}\\ 1\\ 3^{13}-\frac{3}{10}\end{bmatrix}$$

If A has repeated eigenvalues, it may or may not have a basis of eigenvectors. However, if A is symmetric ($A=A^T$), it always has a basis of eigenvectors, even if eigenvalues are repeated, and all eigenvalues are real. This introduces the idea of eigenspaces.

1.5.3 Eigenspaces

The eigenvalues of a matrix A are the roots of the characteristic polynomial $c_A(\lambda)$. The eigenvalues are given by

$$\det(A - \lambda I) = 0$$

If λ is an eigenvalue of A then the *eigenspace* for λ is

$$E_\lambda = \{\vec{v} \in \mathbb{R}^n : \; A\vec{v} = \lambda\vec{v}\} = \mathcal{N}(A - \lambda I)$$

If we write the characteristic polynomial as

$$c_A(x) = (x - \lambda_1)^{\alpha_1}(x - \lambda_2)^{\alpha_2} \cdots (x - \lambda_k)^{\alpha_k}$$

then the value α_i is called the *algebraic multiplicity* of λ_i.
We can also define the *geometric multiplicity* of λ_i to be d_i which is the dimension of the eigenspace:

$$d_i = \dim(E_{\lambda_i}) = \dim(\mathcal{N}(A - \lambda_i I))$$

We will always have that the geometric multiplicity is less than or equal to the algebraic multiplicity

$$d_i \le \alpha_i$$

The eigenvalue is called *defective* if $d_i < \alpha_i$
Ex:

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

$$\det(A - \lambda I) = \begin{vmatrix} 5-\lambda & 1 \\ -1 & 3-\lambda \end{vmatrix} = (5-x)(3-x) + 1 = x^2 - 8x + 16 = (x-4)^2 \Rightarrow \lambda = 4$$

$$\alpha_1 = 2$$

$$E_\lambda = \mathcal{N}(A - 4I) = \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix} \to \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \Rightarrow \vec{v} = t\begin{bmatrix} -1 \\ 1 \end{bmatrix} \Rightarrow E_\lambda = \text{span}\left\{\begin{bmatrix} -1 \\ 1 \end{bmatrix}\right\}$$

$$d_1 = 1$$

$d_1 < \alpha_1 \Rightarrow A$ is a defective matrix

Properties:

- An $n \times n$ matrix always has n eigenvalues when counted with multiplicities. $\sum_{i=1}^{k} \alpha_i = n$
- Every eigenvalue has at least one eigenvector, so $1 \le \dim(E_{\lambda_i}) \le \alpha_i$

1.5.4 Diagonalization

If A is an $n \times n$ matrix, there exists an invertible matrix P and a diagonal matrix D such that

$$A = PDP^{-1}$$

If A is diagonalizable, then the columns of P are eigenvectors of A and the entries of D are corresponding eigenvalues.

$$AP = PD = \begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_n \end{bmatrix} \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

Ex:

$$A = \begin{bmatrix} 1 & 3 \\ 2 & 2 \end{bmatrix}, \quad \begin{matrix} \lambda_1 = 4 & E_{\lambda_1} = \text{span}\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\} \\ \lambda_2 = -1 & E_{\lambda_2} = \text{span}\left\{ \begin{bmatrix} -3 \\ 2 \end{bmatrix} \right\} \end{matrix}$$

$$AP = PD = \begin{bmatrix} 1 & -3 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & -1 \end{bmatrix}$$

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

Note that the diagonalization is not unique - we can switch the order of our eigenvalues/eigenvectors as long as we're consistent.
Also note that a matrix is only diagonalizable if A has n linearly independent eigenvectors. This is identical to saying that the algebraic and geometric multiplicities are equal for all eigenvalues of A.
Properties of real symmetric matrices:

- A has only real eigenvalues
- For $\lambda_1 \neq \lambda_2$, $E_{\lambda_1} \perp E_{\lambda_2}$
- Diagonalization is possible for symmetric matrices
- AA^T and A^TA are symmetric matrices

$$AA^T = PD_1P^T$$
$$A^TA = QD_2Q^T$$

1.5.5 Transition Matrices

Transition matrices, also known as random walks or probability matrices, are matrices that have every column sum to 1.
Take a 3x3 probability matrix. It will have 3 states, 1, 2, and 3 and there will be some probability that an object from one state moves to another state. The probability it goes from state i to state j is represented as P_{ij}. We can represent these probabilities in a matrix as shown

$$P = \begin{bmatrix} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{bmatrix}$$

We let the vector $\vec{x}^{(n)}$ be the probability that an object is in each state at the time n. So $x_1^{(0)}$ would be the probability it is in state 1 at time 0. Using $\vec{x}^{(0)}$ as the starting probability, we can find $\vec{x}^{(n)}$ through

$$\vec{x}^{(n)} = P^n \vec{x}^{(0)}$$

If $\lim_{n\to\infty} P^n \vec{x}_0 = \vec{p}$ for every $\vec{x}_0$, we say the transition matrix has an *equilibrium probability vector*, $\vec{p}$. A matrix will only contain $\vec{p}$ if it has an eigenvalue $\lambda = 1$. In this case, $\vec{p}$ can be found by scaling the eigenvector associated with $\lambda = 1$ such that the sum of its components is 1.

Ex: Find $\vec{p}$ for the random walk $P = \begin{bmatrix} \frac{1}{2} & \frac{2}{3} \\ \frac{1}{2} & \frac{1}{3} \end{bmatrix}$

$$\text{tr}(P) = \frac{5}{6} = \lambda_1 + \lambda_2$$
$$\det(P) = -\frac{1}{6} = \lambda_1 \lambda_2$$
$$\Rightarrow \lambda_1 = 1,\ \lambda_2 = -\frac{1}{6}$$
$$\lambda_1 = 1: \begin{bmatrix} -\frac{1}{2} & \frac{2}{3} \\ \frac{1}{2} & -\frac{2}{3} \end{bmatrix} \to \begin{bmatrix} -\frac{1}{2} & \frac{2}{3} \\ 0 & 0 \end{bmatrix} \to \begin{bmatrix} -3 & 4 \\ 0 & 0 \end{bmatrix} \to \vec{v}_1 = \begin{bmatrix} 4, 3 \end{bmatrix}^T$$
$$\sum \vec{v}_1 = 7 \Rightarrow \vec{p} = \frac{1}{7}\vec{v}_1$$
$$\therefore \vec{p} = \begin{bmatrix} \frac{4}{7} \\ \frac{3}{7} \end{bmatrix}$$

Trends of Transition Matrices:

- $|\lambda| \leq 1$ for all transition matrices
- Every transition matrix with all nonzero entries will contain $\lambda = 1$
- The term with $\lambda = 1$ will always tend to $\vec{p}$

1.5.6 Singular Value Decomposition

We may not always be able to get the diagonalization of a matrix, however, we can always get the singular value decomposition of a matrix.

$$A = P\Sigma Q^T$$

where

$$\Sigma = \begin{bmatrix} \sigma_1 & & & 0 \\ & \ddots & & \vdots \\ & & \sigma_r & 0 \\ 0 & \cdots & 0 & 0 \end{bmatrix}$$

Note that Σ is a diagonal matrix that is zero-padded to be of size $m \times n$ (Σ will always be the same size as A).
The values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r \geq 0$ are the *singular values* of A.

The decomposition comes from $AA^T = PD_1P^T$ and $A^TA = QD_2Q^T$.
The set of eigenvalues for AA^T and A^TA are equal and contains only positive eigenvalues.
So we can write the singular values in terms of these eigenvalues of A^TA or AA^T as

$$\sigma_k = \sqrt{\lambda_k}$$

We can construct the matrix Q as

$$Q = \begin{bmatrix} \vec{q}_1 & \cdots & \vec{q}_n \end{bmatrix}_{n\times n}$$

where $\vec{q}_1, \ldots, \vec{q}_r$ are the orthonormal eigenvectors of A^TA and $\vec{q}_{r+1}, \ldots, \vec{q}_n$ are found as an orthonormal basis of the nullspace of A^TA
We can construct the matrix P as

$$P = \begin{bmatrix} \vec{p}_1 & \cdots & \vec{p}_m \end{bmatrix}_{m\times m}$$

where $\vec{p}_1, \ldots, \vec{p}_r$ are the orthonormal eigenvectors of AA^T and $\vec{p}_{r+1}, \ldots, \vec{p}_n$ are found as an orthonormal basis of the nullspace of AA^T. We can solve for either P or Q and then quickly get the one from the other using these relationships;

$$\vec{p}_k = \frac{1}{\sigma_k} A\vec{q}_k$$

$$\vec{q}_k = \frac{1}{\sigma_k} A^T\vec{p}_k$$

Note that P and Q must be orthonormal. The reason we are able to find the vectors without using Gram-Schmidt is because the eigenvectors of a symmetric matrix will be orthogonal already.
Ex: Find the singular value decomposition of

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

We can choose to compute A^TA or AA^T. Let's choose A^TA so we can work with a 2x2 matrix.

$$A^TA = \begin{bmatrix} 1 & -2 & 0 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -2 & 2 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 5 & -4 \\ -4 & 5 \end{bmatrix}$$

Now we find the eigenvalues and construct Σ

$$\begin{vmatrix} 5-\lambda & -4 \\ -4 & 5-\lambda \end{vmatrix} = 0$$

$$(5-\lambda)^2 - 16 = \lambda^2 - 10\lambda + 9 = 0 \Rightarrow (\lambda-1)(\lambda-9) = 0$$

$$\lambda = 1, 9$$

$$\sigma_1 = \sqrt{9} = 3, \quad \sigma_2 = \sqrt{1} = 1$$

$$\Sigma = \begin{bmatrix} 3 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

Now we can construct Q.
First we find eigenvectors of A^TA

$$\lambda = 1: \; A^TA - I = \begin{bmatrix} 4 & -4 \\ -4 & 4 \end{bmatrix} \Rightarrow \vec{v}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\Rightarrow \vec{q}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\lambda = 9: \; A^TA - 9I = \begin{bmatrix} -4 & -4 \\ -4 & -4 \end{bmatrix} \Rightarrow \vec{v} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

$$\Rightarrow \vec{q}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

$$Q = \begin{bmatrix} \vec{q}_1 & \vec{q}_2 \end{bmatrix} = \begin{bmatrix} -1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$

Now we construct P

$$\vec{p}_1 = \frac{1}{\sigma_1} A\vec{q}_1 = \frac{1}{3} \begin{bmatrix} 1 & 0 \\ -2 & 2 \\ 0 & 1 \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \frac{1}{3\sqrt{2}} \begin{bmatrix} -1 \\ 4 \\ 1 \end{bmatrix}$$

$$\vec{p}_2 = \frac{1}{\sigma_2} A\vec{q}_2 = \frac{1}{1} \begin{bmatrix} 1 & 0 \\ -2 & 2 \\ 0 & 1 \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

We require one more column vector for P (needs to be a square matrix). This will be a basis of the nullspace of AA^T. Because we know that P is an orthonormal matrix, we can also get it from the orthogonal complement of $\vec{p}_1$ and $\vec{p}_2$

$$\vec{p}_3 \in \operatorname{span}\{\vec{p}_1, \vec{p}_2\}^{\perp}$$

$$\begin{bmatrix} -1 & 4 & 1 \\ 1 & 0 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \to \vec{v} = \begin{bmatrix} -1 \\ -1/2 \\ 1 \end{bmatrix}$$

$$\vec{p}_3 = \frac{1}{3} \begin{bmatrix} -2 \\ -1 \\ 2 \end{bmatrix}$$

$$P = \begin{bmatrix} -\frac{1}{3\sqrt{2}} & \frac{1}{\sqrt{2}} & -\frac{2}{3} \\ \frac{4}{3\sqrt{2}} & 0 & -\frac{1}{3} \\ \frac{1}{3\sqrt{2}} & \frac{1}{\sqrt{2}} & \frac{2}{3} \end{bmatrix}$$

A nice property of the SVD is that the magnitude of A is the largest singular value.

$$\|A\| = \sigma_1$$

Proof:

$$\|A\| = \max \|A\hat{x}\| = \max \|P\Sigma Q^T \hat{x}\| = \max \|\Sigma Q^T \hat{x}\|$$

$\hat{y} = Q^T \hat{x}$
$\|A\| = \max \|\Sigma \hat{y}\| = \|\Sigma\| = \sigma_1$

The inverse of Σ is given to be

$$\Sigma^{-1} = \begin{bmatrix} \frac{1}{\sigma_1} & & \\ & \ddots & \\ & & \frac{1}{\sigma_n} \end{bmatrix}$$

By a similar proof, we can get that the magnitude of A^{-1} is

$$\|A^{-1}\| = \frac{1}{\sigma_n}$$

And so the condition number can be expressed as

$$\text{cond}(A) = \frac{\sigma_1}{\sigma_n}$$

Another property is that the number of singular values is equal to the rank of the matrix. This comes from the fact that the P and Q matrices are invertible and, therefore, full rank.
The four fundamental subspaces can also be gotten from the SVD decomposition

- $\mathcal{R}(A) = \text{span}\{\vec{p}_1, \ldots, \vec{p}_r\}$
- $\mathcal{R}(A)^{\perp} = \text{span}\{\vec{p}_{r+1}, \ldots, \vec{p}_m\}$
- $\mathcal{N}(A)^{\perp} = \{\vec{q}_1, \ldots, \vec{q}_r\}$
- $\mathcal{N}(A) = \text{span}\{\vec{q}_{r+1}, \ldots, \vec{q}_n\}$

One nice use for the SVD decomposition is that we can compress data stored in a matrix by taking only the largest singular values. For example, if our matrix represents the data that displays an image, the largest singular values capture the most "essence" of the image while the smaller singular values caputure fine details.
Normally we have

$$A = P\Sigma Q^T = \sum_{k=1}^{r} \sigma_k \vec{p}_k \vec{q}_k^T$$

The compressed form of the matrix can be written as

$$A \approx \sum_{k=1}^{s} \sigma_k \vec{p}_k \vec{q}_k^T, \qquad 1 \leq s \leq r$$

1.5.7 Principal Component Analysis

Principal component analysis (PCA) looks at a matrix of data and will create a basis of vectors that capture the most information from the data. The fist basis vector (or weight vector) will be projected in the direction that has the largest standard deviation in the data. The second weight vector will be perpendicular to the first one and projected in the direction of the next most standard deviation and so on.

For our computations, we can assume that the data will be normalized: $\sum_{k=1}^{n} \vec{x}_k = \vec{0}$

We wish to find the vector, $\vec{w}_1$ which captures the most variance in the data. This will be the unit vector $\vec{w}_1$ which maximizes

$$\sum_{k=1}^{n} \| \operatorname{proj}_{\vec{w}_1}(\vec{x}_k)\| = \sum_{k=1}^{n} \langle \vec{x}_k, \vec{w}_1 \rangle^2 = \|X\vec{w}_1\|$$

where X is called the data matrix

$$X = \begin{bmatrix} \vec{x}_1^T \\ \vdots \\ \vec{x}_n^T \end{bmatrix}$$

$\vec{w}_1$ is called the first weight vector. More generally, given the weight vectors $\vec{w}_1, \ldots, \vec{w}_{k-1}$, the kth weight vector $\vec{w}_k$ is the weight vector which maximizes $\|X_k\vec{w}_k\|^2$ where X_k is the projection of the data matrix onto span $\{\vec{w}_1, \ldots, \vec{w}_k\}^{\perp}$

$$X_k = X - \sum_{i=1}^{k-1} X\vec{w}_i\vec{w}_i^T$$

We can say that the weight vectors are right singular vectors of X which means that $\vec{w}_k = \vec{q}_k$ in the SVD decomposition.

For general k, we can show that

$$X_k = P \begin{bmatrix} 0 & & & & & \\ & \ddots & & & & \\ & & 0 & & & \\ & & & \sigma_k & & \\ & & & & \ddots & \\ & & & & & \sigma_p \end{bmatrix} Q^T$$

So the largest remaining singular value is σ_k and $\|X_k\vec{w}_k\|^2$ is maximum when $\vec{w}_k = \vec{q}_k$. Using this, we are able to find the weight vectors for a given data matrix.

Ex: Find weight vectors for the data $(-2, 2), (-1, 1), (0, -1), (1, 2), (2, 0)$

$$X = \begin{bmatrix} -2 & -2 \\ -1 & 1 \\ 0 & -1 \\ 1 & 2 \\ 2 & 0 \end{bmatrix}$$

$$X^T X = \begin{bmatrix} 10 & 5 \\ 5 & 10 \end{bmatrix}$$

$$\rightsquigarrow \lambda = 15, 5$$

$$\lambda = 15 : \begin{bmatrix} -5 & 5 \\ 5 & -5 \end{bmatrix} \Rightarrow \vec{v} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \Rightarrow \vec{w}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\lambda = 5 : \begin{bmatrix} 5 & 5 \\ 5 & 5 \end{bmatrix} \Rightarrow \vec{v} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \Rightarrow \vec{w}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

1.5.8 Pseudoinverse

If A is invertible then we can solve $A\vec{x} = \vec{b}$ as $\vec{x} = A^{-1}\vec{b}$. For a general A that isn't necessarily invertible, we can approximate the solution to $A\vec{x} = \vec{b}$ using the pseudoinverse $\vec{x} = A^{+}\vec{b}$.
If we consider the SVD decomposition of A, $A = P\Sigma Q^T$ the the SVD of A^{-1} will be $A^{-1} = Q\Sigma^{-1}P^T$.
The pseudoinverse of A can be written as $A^{+} = Q\Sigma^{+}P^T$ where

$$\Sigma^{+} = \begin{bmatrix} \frac{1}{\sigma_1} & & & 0 \\ & \ddots & & \vdots \\ & & \frac{1}{\sigma_r} & 0 \\ 0 & \cdots & 0 & 0 \end{bmatrix}_{n \times m}$$

If A is invertible then $A^{+} = A^{-1}$
The pseudoinverse has the property that

- $AA^{+}A = A$ and $A^{+}AA^{+} = A^{+}$

The pseudoinverse also gives a very simple solution to the least squares approximation.
$\vec{x} = A^{+}\vec{b}$ solves the least squares problem $A\vec{x} \approx \vec{b}$
The pseudoinverse is computed by using the SVD decomposition of A and written as

$$A^{+} = Q\Sigma^{+}P^T$$

where Σ^{+} is the same size as A^{-1}

1.5.9 Computing Eigenvalues

For large matrices, computing eigenvalues can be challenging as it involves solving an n degree polynomial. There are many numerical methods, however, that we can use to approximate eigenvalues.

Power method:
An eigenvalue is dominant if λ has an algebraic multiplicity of 1 and $|\lambda| > |\lambda_k|$ for all other eigenvalues.
We can pick out this dominant eigenvalue through iterative matrix multiplication.
We can choose a random vector $\vec{x}_0$ and repeatedly multiply it by A and the result will tend toward the dominant eigenvector.

$$\vec{x}_{k+1} = A\vec{x}_k$$
$$\Rightarrow \vec{x}_k = A^k\vec{x}_0$$

$$\lim_{k\to\infty} \frac{A^k\vec{x}_0}{\lambda_1^k} = c_1\vec{v}_1$$

Rayleigh quotient:
Another computational method is given by the formula below:

$$\lim_{k\to\infty} \frac{\vec{x}_k^T A\vec{x}_k}{\vec{x}_k^T\vec{x}_k} = \lambda_1$$

1.6 Complex Vector Spaces

1.6.1 Complex Operations

If we have vectors that contain complex values, they will exist in $\mathbb{C}^n$. This will change some of our definitions as we will see, however, we would like to keep some of our same properties. One such property is that the square of the length of a vector should be the inner product of that vector with itself.
If we take a complex vector and take the inner product with itself, using the traditional definition, there is a chance that we get a complex value

$$\begin{bmatrix} i+1 & 1 \end{bmatrix} \begin{bmatrix} i+1 \\ 1 \end{bmatrix} = 1 + 2i$$

This doesn't fit with what we would expect as the length of our vector and so this definition doesn't hold. We can, however, take advantage of the property $z\overline{z} = |z|^2$. So what we can do is redefine the inner product for complex vectors:

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

Ex:

$$\left\langle \begin{bmatrix} i \\ 1 \end{bmatrix}, \begin{bmatrix} i \\ 1 \end{bmatrix} \right\rangle = \begin{bmatrix} i & 1 \end{bmatrix} \begin{bmatrix} -i \\ 1 \end{bmatrix} = 2$$

Some properties of the complex inner product are as follows:

- $\langle c\vec{v}, \vec{w} \rangle = c \langle \vec{v}, \vec{w} \rangle$
- $\langle \vec{v}, c\vec{w} \rangle = \overline{c} \langle \vec{v}, \vec{w} \rangle$
- $\langle \vec{v}, \vec{w} \rangle = \overline{\langle \vec{w}, \vec{v} \rangle}$
- $\|\vec{v}\|^2 = \langle \vec{v}, \vec{v} \rangle \geq 0$

This implies that we will also have to introduce conjugates with some of our other definitions as well:
The conjugate transpose of a matrix is defined to be $A^* = \overline{A}^T$
A property of the inner product will be that

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

A matrix is said to be *hermitian* if

$$\boxed{A = A^*}$$

This is analogous to symmetric matrices.
Some properties are

- $\langle A\vec{x}, \vec{y} \rangle = \langle \vec{x}, A\vec{y} \rangle$
- A has only real eigenvalues and A is diagonalizable

- The diagonal entries are real numbers

A matrix is said to be *unitary* if

$$\boxed{A^{-1} = A^*}$$

This is analogous to orthogonal matrices.
It will have the properties

- $\langle A\vec{x}, A\vec{y}\rangle = \langle \vec{x}, \vec{y}\rangle$
- $\|A\vec{x}\| = \|\vec{x}\|$

Note that this conjugation also extends to the inner product definition of complex functions:

$$\langle f, g\rangle = \int_{-\infty}^{\infty} f(x)\overline{g(x)}dx$$

1.6.2 Roots of Unity

A complex number $\omega \in \mathbb{C}$ is an Nth root of unity if $\omega^N = 1$.
Ex: $N = 2$

$$\omega^2 = 1 \Rightarrow \omega = \pm 1$$

so 1 and -1 are 2nd roots of unity
Ex2: $N = 3$

$$\omega^3 = 1 \Rightarrow (\omega - 1)(\omega^2 + \omega + 1) = 0$$

$$\omega = 1,\ \omega = \frac{-1 \pm i\sqrt{3}}{2}$$

Ex3: $N = 4$

$$\omega^4 = 1 \Rightarrow \omega^2 = \pm 1$$

$$\omega = \{1, -1, i, -i\}$$

For general N we can express the roots of unity as

$$\omega_N^N = 1 = e^{2\pi i} \Rightarrow \omega_N = e^{i\frac{2\pi}{N}}$$

And so the roots of unity will be

$$\omega_N = \left\{1, e^{i\frac{2\pi}{N}}, e^{i\frac{2\pi(2)}{N}}, \ldots, e^{i\frac{2\pi(N-1)}{N}}\right\}$$

Some properties of the roots of unity are

- $\omega_N^{N-1} = \overline{\omega_N}$

$$\omega_N^{N-1} = e^{i\frac{2\pi(N-1)}{N}} = e^{i2\pi}e^{-i\frac{2\pi}{N}} = e^{-i\frac{2\pi}{N}} = \overline{\omega_N}$$

- $\overline{\omega_N} = \omega_N^{-1}$

$$\omega_N\overline{\omega_N} = e^{i\frac{2\pi}{N}}e^{-i\frac{2\pi}{N}} = 1$$

- $\sum_{m=0}^{N-1} \omega_N^m = 0$

1.6.3 Discrete Fourier Transform

The standard basis of $\mathbb{C}^N$ is $\{\vec{c}_0, \vec{c}_1, \ldots, \vec{c_{N-1}}\}$ where

$$\vec{c}_0 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \ \vec{c}_1 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \ldots$$

We can define the Fourier basis of $\mathbb{C}^N$ to be $\left\{\vec{f}_0, \vec{f}_1, \ldots, \vec{f}_{N-1}\right\}$ where

$$\vec{f}_k = \begin{bmatrix} 1 \\ \omega_N^k \\ \omega_N^{2k} \\ \vdots \\ \omega_N^{(N-1)k} \end{bmatrix}$$

Ex: The Fourier basis of $\mathbb{C}^4$ is given by

$$\vec{f}_0 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \ \vec{f}_1 = \begin{bmatrix} 1 \\ i \\ -1 \\ -i \end{bmatrix}, \ \vec{f}_2 = \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix}, \ \vec{f}_3 = \begin{bmatrix} 1 \\ -i \\ -1 \\ i \end{bmatrix}$$

The Fourier basis forms an orthogonal basis of $\mathbb{C}^N$ with

$$\left\langle \vec{f}_k, \vec{f}_m \right\rangle = \begin{cases} 0 & k \neq m \\ N & k = m \end{cases}$$

Note that this is not an orthonormal basis as $\|\vec{f}_k\| = \sqrt{N}$
By the properties of the roots of unity we can get that

$$\overline{\vec{f}}_k = \vec{f}_{N-k}$$

which states that the Fourier basis vectors have conjugate symmetry about $N/2$.

The discrete Fourier transform (DTF) of $\vec{x}$ is the vector of coefficients of $\vec{x}$ with respect to the Fourier basis. Essentially, it takes a vector and projects it onto the Fourier basis. The resulting vector tells us what linear combination of Fourier basis vectors made up the original vector $\vec{x}$.
We can define the Fourier matrix as

$$F_N = \begin{bmatrix} \overline{\vec{f}}_0^T \\ \vdots \\ \overline{\vec{f}}_{N-1}^T \end{bmatrix} = \begin{bmatrix} \vec{f}_0^T \\ \vec{f}_{N-1}^T \\ \vdots \\ \vec{f}_1^T \end{bmatrix}$$

The discrete Fourier transform of $\vec{x}$ is given by

$$\mathrm{DFT}(\vec{x}) = F_N\vec{x}$$

Ex: Compute $\mathrm{DFT}(\vec{x})$ for $\vec{x} = \begin{bmatrix}1 & 2 & 0 & 1\end{bmatrix}^T$

$$N = 4$$

$$\vec{f_0} = \begin{bmatrix}1\\1\\1\\1\end{bmatrix},\ \vec{f_1} = \begin{bmatrix}1\\i\\-1\\-i\end{bmatrix},\ \vec{f_2} = \begin{bmatrix}1\\-1\\1\\-1\end{bmatrix},\ \vec{f_3} = \begin{bmatrix}1\\-i\\-1\\i\end{bmatrix}$$

$$F_4 = \begin{bmatrix}1 & 1 & 1 & 1\\ 1 & -i & -1 & i\\ 1 & -1 & 1 & -1\\ 1 & i & -1 & -i\end{bmatrix}$$

$$\mathrm{DFT}(\vec{x}) = F_4\vec{x} = \begin{bmatrix}4\\1-i\\-2\\1+i\end{bmatrix}$$

The Fourier matrix has the property

$$F_N F_N^T = NI$$

and

$$F_N^{-1} = \frac{1}{N}\overline{F}_N^T$$

This implies that $\frac{1}{\sqrt{N}}F_N$ is a unitary matrix.

A sinusoid can be expressed as a vector of the form $\vec{x} = A\cos(2\pi k\vec{t} + \phi)$ where A is the amplitude, k is the frequency, and ϕ is the phase shift.

$$\vec{x} = A\cos(2\pi k\vec{t} + \phi) = \begin{bmatrix}A\cos(\phi)\\ A\cos\left(2\pi k\frac{1}{N} + \phi\right)\\ A\cos\left(2\pi k\frac{2}{N} + \phi\right)\\ \vdots\\ A\cos\left(2\pi k\frac{N-1}{N} + \phi\right)\end{bmatrix}$$

Ex:

$$\vec{x} = 2\cos\left(4\pi\vec{t} + \frac{\pi}{2}\right) = \begin{bmatrix}0\\-2\\0\\2\\0\\-2\\0\\2\end{bmatrix}$$

We can easily express sinusoids using the Fourier basis. For a sinusoid $\vec{x} = A\cos(2\pi k\vec{t} + \phi)$ with $0 < k < N$, we can write

$$\mathrm{DFT}(\vec{x}) = \frac{AN}{2}e^{i\phi}\vec{e}_k + \frac{AN}{2}e^{-i\phi}\vec{e}_{N-k}$$

For $k = 0$, $\mathrm{DFT}(\vec{x}) = AN\vec{e}_0$
We can use the DFT to write the input signal as a sum of sinusoids:

- Look at the entry at index k ($0 \leq k \leq \frac{N}{2}$)
- If the number is nonzero, include a sinusoid with frequency k.
- The modulus divided by $\frac{N}{2}$ gives the amplitude of that sinusoid.
- The argument gives the phase shift.
- Entries at index 0 and $\frac{N}{2}$ (if N is even) are a bit special. Both are always real numbers, and the amplitude is given by the modulus divided by N instead
- If positive amplitude then $\phi = 0$ and if negative amplitude, $\phi = \pi$