MATH 33A – SUMMER 2022

Pablo S. Ocal

based on "Linear Algebra with Applications"
by Otto Bretscher.

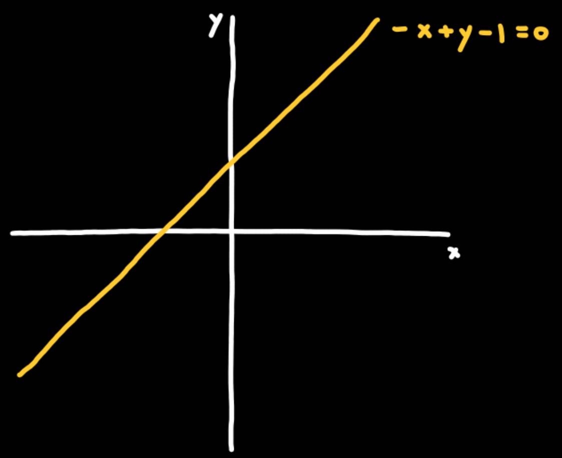
1. Introduction to Linear Algebra. (Chapter 1 and Chapter 2)

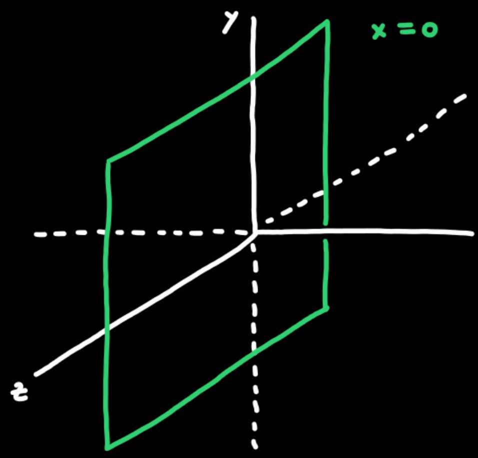
Linear algebra is the study of linear equations and linear transformations.

A linear equation has the form:

$$a_1 x_1 + a_2 x_2 + \cdots + a_n x_n + b = 0,$$

where $a_1, \dots, a_n$ are real numbers called coefficients, $x_1, \dots, x_n$ are variables, and b is a real number called the constant term. Geometrically, linear equations define lines, planes, and objects that we will call subspaces.



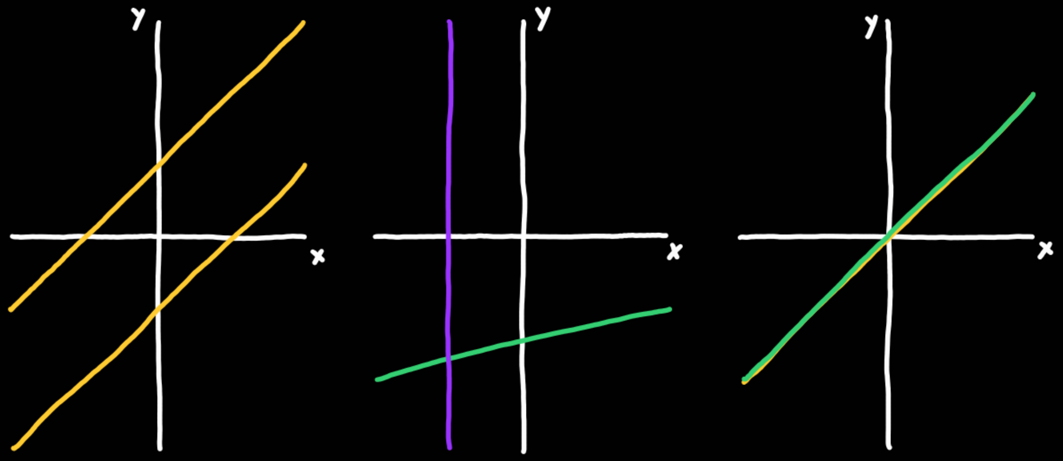


Systems of linear equations can have no solution, one solution, or infinitely many solutions.

$$a_{11} x_1 + a_{12} x_2 + \cdots + a_{1n} x_n + b_1 = 0$$
$$\vdots$$
$$a_{m1} x_1 + \cdots + a_{mn} x_n + b_m = 0$$

$$a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n + b_n = 0$$

This happens because the solutions are the intersection points of the geometric objects defined by the equations. There are either no intersections, one intersection, or infinitely many intersections.



parallel lines   one intersection   double line

To handle and solve systems of linear equations, we use <u>matrices</u>:

$$\begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nm} \end{bmatrix}$$ is an $n \times m$ matrix with entries a_{ij}.

A <u>matrix</u> is a rectangular array of numbers. If a matrix has n rows and m columns, we say that the <u>size</u> of the matrix is $n \times m$. We say that two matrices A and B are equal when their <u>entries</u> a_{ij} and b_{ij} are equal.

Some families of matrices receive special names:

(i) Square matrices.

(ii) Diagonal matrix.

(iii) Upper triangular matrix.

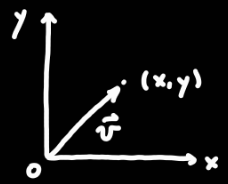
(iv) Lower diagonal matrix.

(v) Zero matrix.

A <u>vector</u> is a matrix with only one column. The entries of a vector are called its <u>components</u>.

The set of all column vectors with n components is denoted by $\mathbb{R}^n$. We will refer to $\mathbb{R}^n$ as a <u>vector space</u>.

The standard representation of a vector in the Cartesian coordinate plane is as an arrow from the origin: $\vec{v} = \begin{bmatrix} x \\ y \end{bmatrix}$ is represented as



. However, thinking about vectors conceptually as a list of numbers written in a column will be useful.

Given a system of n linear equations in m variables:

$$
\begin{aligned}
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m &= b_1 \\
&\vdots \\
a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nm}x_m &= b_n
\end{aligned}
$$

we store the information on an <u>augmented matrix</u>:

$$
\left[\begin{array}{cccc|c}
a_{11} & a_{12} & \cdots & a_{1m} & b_1 \\
\vdots & \vdots & & \vdots & \vdots \\
a_{n1} & a_{n2} & \cdots & a_{nm} & b_n
\end{array}\right]
$$

and simplify it using three <u>row operations</u>: (we will soon see that these correspond to multiplication by invertible matrices, specifically diagonal matrices and permutation matrices).

(1) Divide a row by a non-zero scalar.

(2) Subtract a multiple of a row from another row.

(3) Swap two rows.

<u>Example:</u>

The system of linear equations:

$$2x + 8y + 4z = 2$$
$$2x + 5y + z = 5$$
$$4x + 10y - z = 1$$

has augmented matrix

$$\left[\begin{array}{ccc|c} 2 & 8 & 4 & 2 \\ 2 & 5 & 1 & 5 \\ 4 & 10 & -1 & 1 \end{array}\right]$$

which can be simplified into:

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & 11 \\ 0 & 1 & 0 & -4 \\ 0 & 0 & 1 & 3 \end{array}\right]$$

giving the solution $x = 11$, $y = -4$, $z = 3$.

The simplified form is called <u>reduced row-echelon form</u>, and solves the system of linear equations.

A matrix is in reduced row-echelon form if it satisfies all the following conditions:

(i) If a row has non-zero entries, then the first non-zero entry is a 1.

This is called the <u>leading</u> 1, or <u>pivot</u>, of the row.

(ii) If a column contains a leading 1, then all the other entries in the column are 0.

(iii) If a row contains a leading 1, then each row above it contains a leading 1 further to the left.

If there are rows of zeroes, by (iii), they must appear at the bottom of the matrix.

<u>Example:</u> The zero matrix is in reduced row-echelon form.

<u>Example:</u> When reducing the augmented matrices of three systems we obtain:

(a) $\left[\begin{array}{ccc|c} 1 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right]$ (b) $\left[\begin{array}{ccc|c} 1 & 2 & 0 & 1 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{array}\right]$ (c) $\left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \end{array}\right]$

How many solutions are there in each case?

(a) No solutions. (b) Infinitely many solutions. (c) One solution.

A system of equations is called <u>consistent</u> if there is at least one solution, and <u>inconsistent</u> if there are no solutions.

<u>Theorem:</u> A linear system is inconsistent if and only if the reduced row-echelon form of its augmented matrix contains the row $[0 \cdots 0 \mid 1]$. If a linear system is consistent then:

(i) it has infinitely many solutions if there is at least one free variable, or

(ii) it has exactly one solution if all the variables are leading.

More useful information can be obtained from the reduced form of a matrix, like the rank.

The *rank* of a matrix A, denoted $\text{rank}(A)$, is the number of leading 1's in $\text{rref}(A)$.

Example: For $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$ we have $\text{rref}(A) = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$ so $\text{rank}(A) = 2$.

Theorem: Consider a system of n equations in m variables (so its coefficient matrix has size $n \times m$). Then: (why? justify this!)

(1) We have $\text{rank}(A) \leq n$ and $\text{rank}(A) \leq m$.

(2) If $\text{rank}(A) = n$, then the system is consistent.

(3) If $\text{rank}(A) = m$, then the system has at most one solution.

(4) If $\text{rank}(A) < m$, then the system has either zero or infinitely many solutions.

Example:

1. Suppose we have a system with fewer equations than variables. How many solutions could we have? Answer: no solutions or infinitely many, since $\text{rank}(A) \leq n < m$.

   That is, if a linear system has a unique solution, then there must be at least as many equations as variables.

2. Suppose we have a system with n equations and n variables. When do we have exactly one solution? Answer: if and only if the rank of the matrix is n.

Since matrices play such a big role in linear algebra, we have to get comfortable manipulating them. This includes addition of matrices, scalar multiples of matrices, and later multiplications.

<u>Addition:</u> The matrix $C = A + B$ has entries $c_{ij} = a_{ij} + b_{ij}$.

<u>Scalar multiplication:</u> The matrix $C = kA$ has entries $c = k a_{ij}$.

<u>Dot product:</u> The dot product of two vectors is a scalar: $\vec{x} \cdot \vec{y} = \sum_{i=1}^{n} x_i y_i$. (this is the precursor of matrix multiplication)

<u>Product of a matrix with a vector:</u>

$$A\vec{x} = \begin{bmatrix} — \vec{w}_1 — \\ \vdots \\ — \vec{w}_n — \end{bmatrix} \vec{x} = \begin{bmatrix} \vec{w}_1 \cdot \vec{x} \\ \vdots \\ \vec{w}_n \cdot \vec{x} \end{bmatrix}$$

$$A\vec{x} = \begin{bmatrix} | & & | \\ \vec{v}_1 & \cdots & \vec{v}_m \\ | & & | \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} = x_1 \vec{v}_1 + \cdots + x_m \vec{v}_m$$

A has size $n \times m$, $\vec{x}, \vec{w}_1, \ldots, \vec{w}_n$ are vectors in $\mathbb{R}^m$, $\vec{v}_1, \ldots, \vec{v}_m$ are vectors in $\mathbb{R}^n$.

<u>Algebraic rules:</u> $A(\vec{x} + \vec{y}) = A\vec{x} + A\vec{y}$ and $A(k\vec{x}) = kA\vec{x}$.

A vector $\vec{v}$ is a <u>linear combination</u> of the vectors $\vec{v}_1, \ldots, \vec{v}_m$ in $\mathbb{R}^n$ if there are scalars $a_1, \ldots, a_m$ such that $\vec{v} = a_1 \vec{v}_1 + \cdots + a_m \vec{v}_m$.

Given a linear system with augmented matrix $[A \mid \vec{b}]$, we can write it as an equality of matrices: $A\vec{x} = \vec{b}$ where $\vec{x}$ is the vector of variables.

<u>Example:</u>

The system of linear equations:

$$\begin{aligned} 2x + 8y + 4z &= 2 \\ 2x + 5y + z &= 5 \\ 4x + 10y - z &= 1 \end{aligned} \quad \text{is equivalent to the equation} \quad \begin{bmatrix} 2 & 8 & 4 \\ 2 & 5 & 1 \\ 4 & 10 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 2 \\ 5 \\ 1 \end{bmatrix}$$

Example: Are the following statements true or false?

1. There exists a 3×4 matrix of rank 4. False!

2. There exists a system of three linear equations with three unknowns that has exactly three solutions. False!

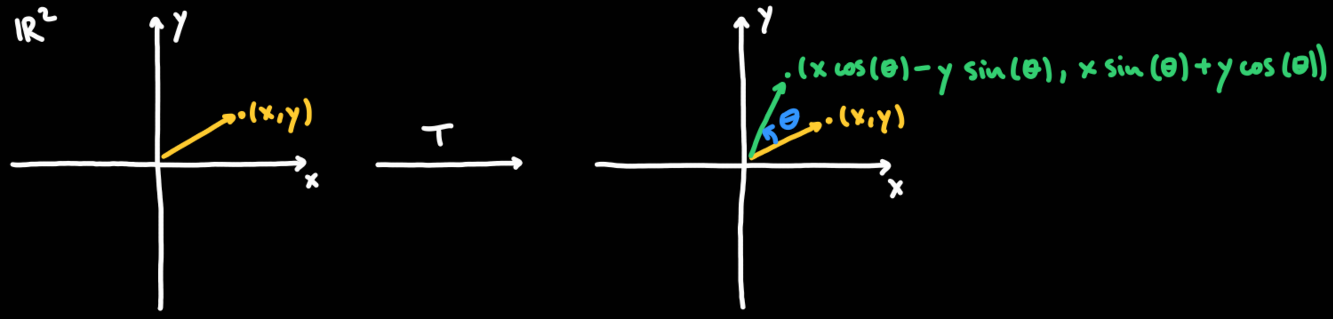
3. $\operatorname{rank} \begin{bmatrix} 2 & 2 & 2 \\ 2 & 2 & 2 \\ 2 & 2 & 2 \end{bmatrix} = 2.$ False!

4. If A is a 3×4 matrix of rank 3, then the system $A\vec{x} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ must have infinitely many solutions. True!

A function T from X to Y is an assignment of an unique element y of Y to each element x of X. We call X the domain of T and Y the range of T.

A linear transformation is a function T from $\mathbb{R}^m$ to $\mathbb{R}^n$ such that there exists an $n \times m$ matrix A with $T(\vec{x}) = A\vec{x}$ for all $\vec{x}$ in $\mathbb{R}^m$.

Example: Consider the function from $\mathbb{R}^2$ to $\mathbb{R}^2$ given by a rotation of angle θ.

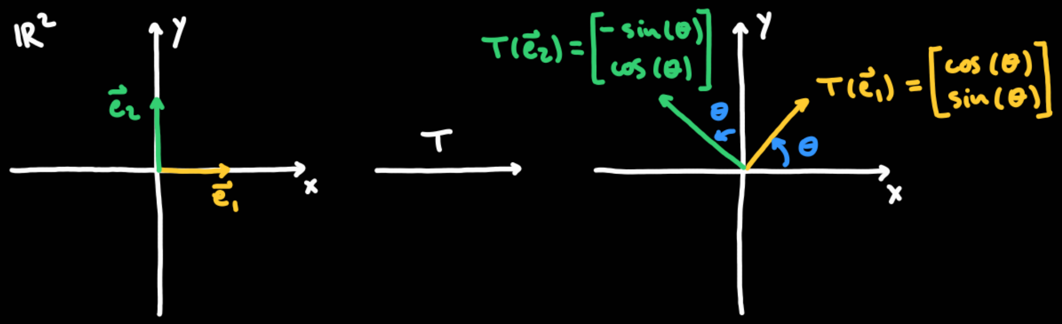


This rotation is a linear transformation because:

$$T(\vec{x}) = T\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x\cos(\theta) - y\sin(\theta) \\ x\sin(\theta) + y\cos(\theta) \end{bmatrix}.$$

<u>Theorem</u>: Let T be a linear transformation from $\mathbb{R}^m$ to $\mathbb{R}^n$. The columns of the matrix associated to T are $T(\vec{e}_1), \dots, T(\vec{e}_m)$ where $\vec{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \dots, \vec{e}_m = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$.

<u>Example</u>: Consider the function from $\mathbb{R}^2$ to $\mathbb{R}^2$ given by a rotation of angle θ.



So the matrix associated to T is $\begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$.

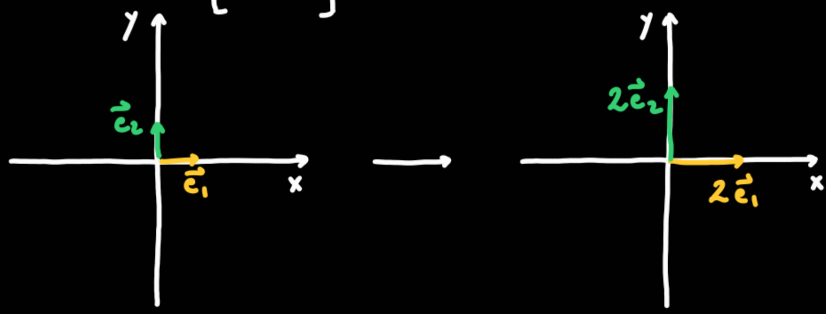
<u>Theorem</u>: A function T from $\mathbb{R}^m$ to $\mathbb{R}^n$ is a linear transformation if and only if:

(i) $T(\vec{v} + \vec{w}) = T(\vec{v}) + T(\vec{w})$ for all $\vec{v}, \vec{w}$ in $\mathbb{R}^m$, and

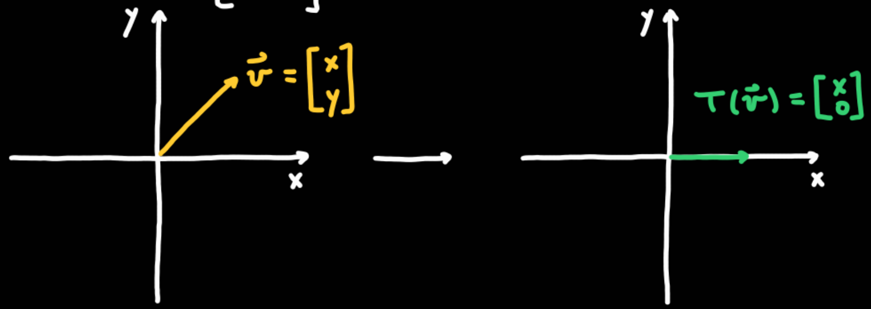
(ii) $T(\lambda\vec{v}) = \lambda T(\vec{v})$ for all $\vec{v}$ in $\mathbb{R}^m$ and λ in $\mathbb{R}$.

<u>Example</u>:

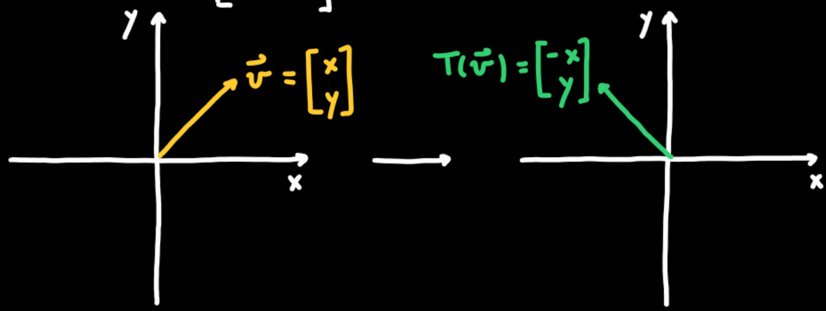
1. The matrix $\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$ is a dilation by 2 (or scaling).



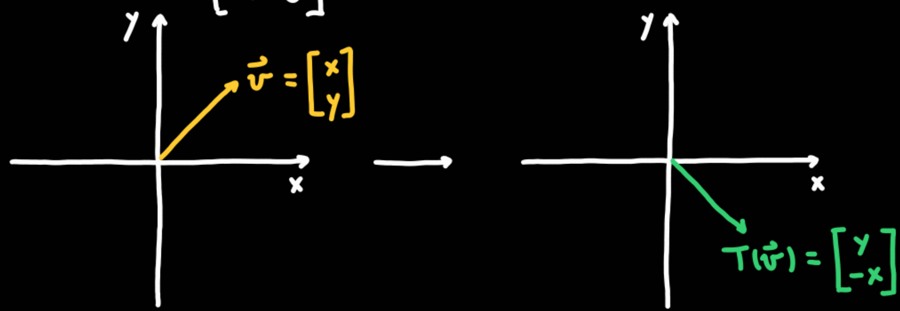
2. The matrix $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ is an orthogonal projection onto the horizontal axis:



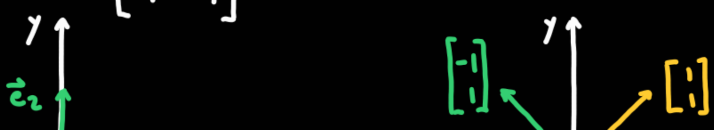
3. The matrix $\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$ is a reflection about the vertical axis:

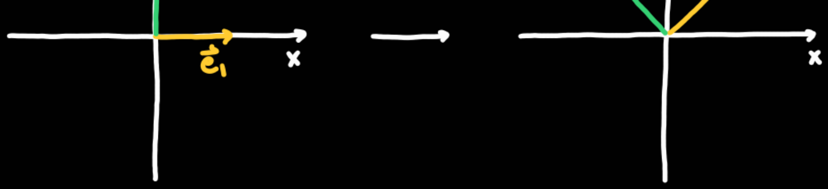


4. The matrix $\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ is a clockwise rotation of $\frac{\pi}{2}$.



5. The matrix $\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$ is a rotation of $\frac{\pi}{4}$ and a dilation of $\sqrt{2}$.





Scaling.

Is given by multiplying a vector $\vec{x}$ with a diagonal matrix $\begin{bmatrix} k & 0 \\ 0 & k \end{bmatrix}$, k in $\mathbb{R}$.

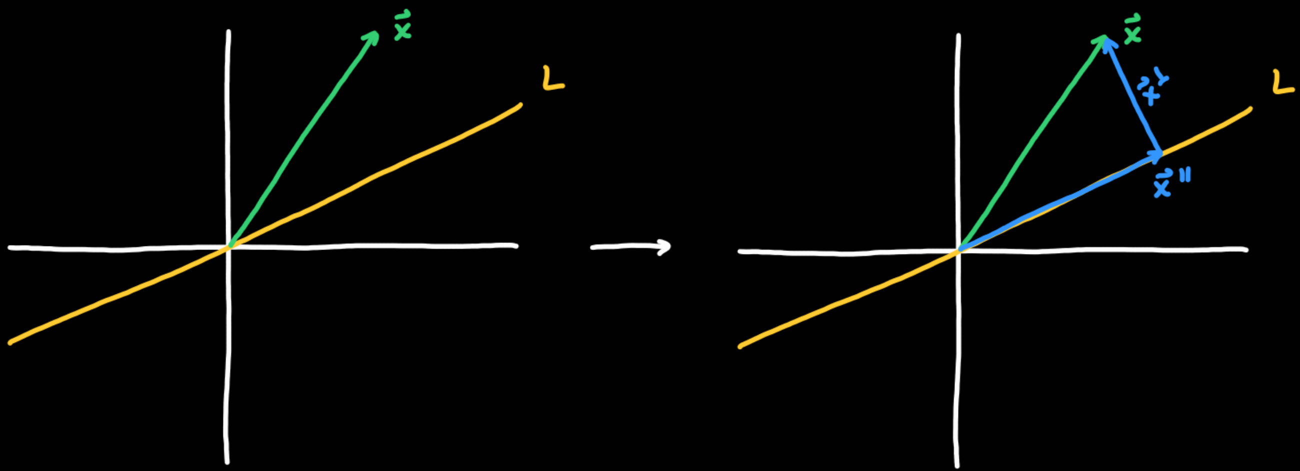
Orthogonal projections.

Given $\vec{x}$ in $\mathbb{R}^2$ and L a line through the origin, we can decompose:

$\vec{x} = \vec{x}^{\parallel} + \vec{x}^{\perp}$ with $\vec{x}^{\parallel}$ parallel to L and $\vec{x}^{\perp}$ perpendicular to L.

We call $\vec{x}^{\parallel}$ the <u>orthogonal projection</u> of $\vec{x}$ onto L, denoted $\text{proj}_L(\vec{x})$.

We have $\vec{x}^{\parallel} = \text{proj}_L(\vec{x}) = (\vec{x} \cdot \vec{u})\,\vec{u}$ where $\vec{u}$ is a unit vector parallel to L.

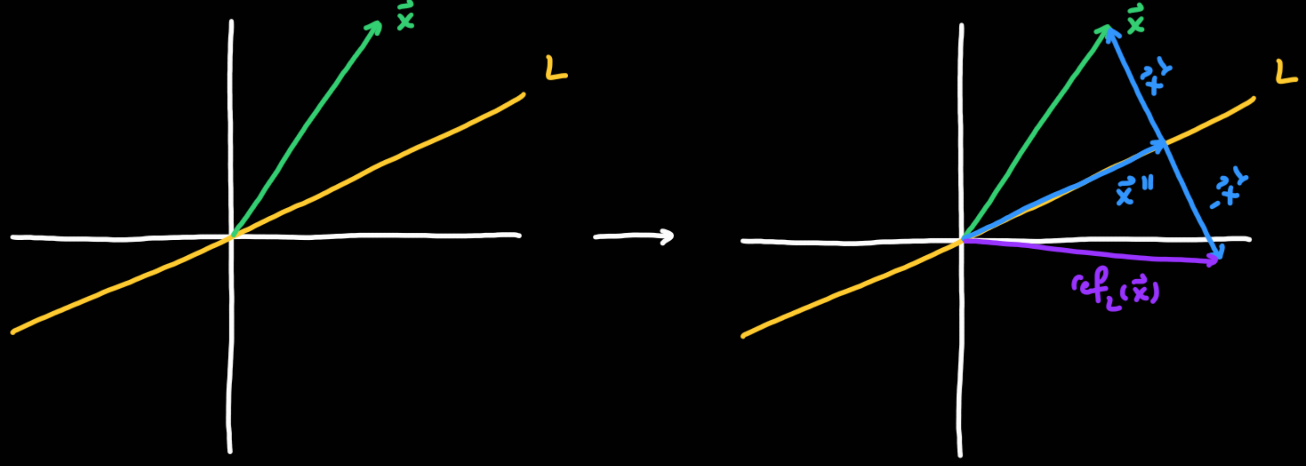


If $\vec{u}$ is non-zero and parallel to L, the associated matrix is $\frac{1}{u_1^2+u_2^2}\begin{bmatrix} u_1^2 & u_1u_2 \\ u_1u_2 & u_2^2 \end{bmatrix}$.

Reflections.

Given $\vec{x}$ in $\mathbb{R}^2$ and L a line through the origin, the <u>reflection</u> of $\vec{x}$ onto L is

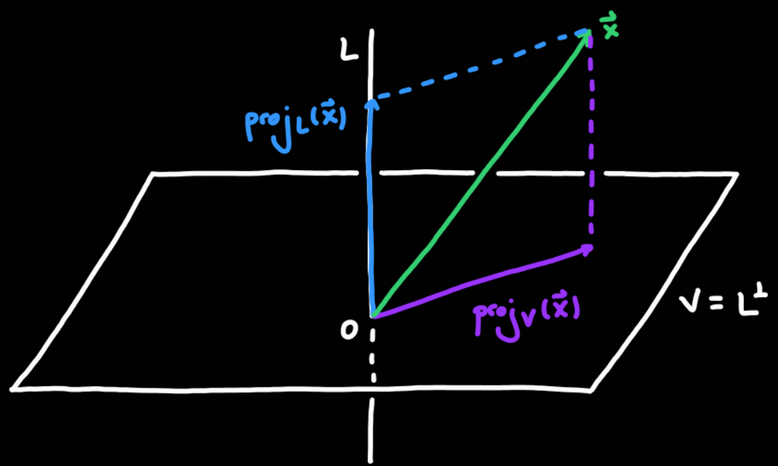
$\text{ref}_L(\vec{x}) = \vec{x}^{\parallel} - \vec{x}^{\perp}$.



If $\vec{u}$ is unitary and parallel to L, the associated matrix is $\begin{bmatrix} 2u_1^2-1 & 2u_1u_2 \\ 2u_1u_2 & 2u_2^2-1 \end{bmatrix}$.

A linear transformation is a reflection if and only if its associated matrix has the form $\begin{bmatrix} a & b \\ b & -a \end{bmatrix}$ with $a^2+b^2=1$.

We can do this same type of decompositions in higher dimensions! For $\mathbb{R}^3$, we have:



So given $\vec{x}$ in $\mathbb{R}^3$ and L a line through the origin, we can decompose $\vec{x} = \mathrm{proj}_L(\vec{x}) + \mathrm{proj}_V(\vec{x})$ where $\mathrm{proj}_L(\vec{x})$ is the orthogonal projection of $\vec{x}$ onto L and $\mathrm{proj}_V(\vec{x})$ is the projection of $\vec{x}$ onto V, the plane through the origin perpendicular to L. We have: (why? Read them on the picture!)

(i) $\mathrm{proj}_L(\vec{x}) = (\vec{x}\cdot\vec{u})\vec{u}$

(ii) $\text{proj}_V(\vec{x}) = \vec{x} - \text{proj}_L(\vec{x})$

(iii) $\text{ref}_L(\vec{x}) = \text{proj}_L(\vec{x}) - \text{proj}_V(\vec{x})$

(iv) $\text{ref}_V(\vec{x}) = \text{proj}_V(\vec{x}) - \text{proj}_L(\vec{x})$

<u>Example:</u>

Let V be the plane defined by $2x_1 + x_2 - 2x_3 = 0$ and $\vec{x} = \begin{bmatrix} 5 \\ 4 \\ -2 \end{bmatrix}$. A vector perpendicular to V is $\vec{v} = \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix}$, giving the unit vector $\vec{u} = \frac{\vec{v}}{\|\vec{v}\|} = \frac{1}{\sqrt{2^2+1^2+2^2}} \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix} = \begin{bmatrix} 2/3 \\ 1/3 \\ -2/3 \end{bmatrix}$, which is still perpendicular to V. Now:

(i) $\text{proj}_L(\vec{x}) = (\vec{x} \cdot \vec{u})\vec{u} = \left(5 \cdot \frac{2}{3} + 4 \cdot \frac{1}{3} + (-2) \cdot \frac{(-2)}{3}\right) \cdot \begin{bmatrix} 2/3 \\ 1/3 \\ -2/3 \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \\ -4 \end{bmatrix}$.

(ii) $\text{proj}_V(\vec{x}) = x - \text{proj}_L(\vec{x}) = \begin{bmatrix} 5 \\ 4 \\ -2 \end{bmatrix} - \begin{bmatrix} 4 \\ 2 \\ -4 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$.

(iii) $\text{ref}_L(\vec{x}) = \text{proj}_L(\vec{x}) - \text{proj}_V(\vec{x}) = \begin{bmatrix} 4 \\ 2 \\ -4 \end{bmatrix} - \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \\ -6 \end{bmatrix}$.

(iv) $\text{ref}_V(\vec{x}) = \text{proj}_V(\vec{x}) - \text{proj}_L(\vec{x}) = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} - \begin{bmatrix} 4 \\ 2 \\ -4 \end{bmatrix} = \begin{bmatrix} -3 \\ 0 \\ 6 \end{bmatrix}$.

<u>Rotations.</u>

A linear transformation is a rotation if and only if its associated matrix has the form $\begin{bmatrix} a & -b \\ b & a \end{bmatrix}$ with $a^2 + b^2 = 1$.

<u>Example:</u> To do a rotation combined with a scaling, first do a rotation, then do a

scaling. This is the same as first doing a scaling, and then a rotation.

How do we deal with consecutive linear transformations? If T is given by $\begin{bmatrix} 6 & 7 \\ 8 & 9 \end{bmatrix}$ and S is given by $\begin{bmatrix} 1 & 2 \\ 3 & 5 \end{bmatrix}$, and we would like to find $\vec{z} = T(S(\vec{x}))$, we do this in two steps. Call $\vec{y} = S(\vec{x})$, then $\vec{z} = T(\vec{x})$, and these two equations are:

$$\begin{aligned} y_1 &= x_1 + 2x_2 \\ y_2 &= 3x_1 + 5x_2 \end{aligned} \quad \text{and} \quad \begin{aligned} z_1 &= 6y_1 + 7y_2 \\ z_2 &= 8y_1 + 9y_2 \end{aligned} \quad \text{so} \quad \begin{aligned} z_1 &= 27x_1 + 47x_2 \\ z_2 &= 35x_1 + 61x_2 \end{aligned}.$$

This should mean that $\vec{z} = TS(\vec{x})$ is given by $\begin{bmatrix} 27 & 47 \\ 35 & 61 \end{bmatrix}$. This should be the product of the matrices T and S, namely $\begin{bmatrix} 6 & 7 \\ 8 & 9 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 3 & 5 \end{bmatrix}$.

<u>Matrix multiplication:</u>

Let B be an $n \times p$ matrix and A a $q \times m$ matrix. If (and only if) $p = q$ the <u>product</u> BA is the matrix of the linear transformation $T(\vec{x}) = B(A\vec{x})$, and it is an $n \times m$ matrix.

<u>Theorem:</u> Let B be an $n \times p$ matrix and A a $p \times m$ matrix. Then:

(i) $BA = B\begin{bmatrix} | & & | \\ \vec{v}_1 & \cdots & \vec{v}_m \\ | & & | \end{bmatrix} = \begin{bmatrix} | & & | \\ B\vec{v}_1 & \cdots & B\vec{v}_m \\ | & & | \end{bmatrix}$.

(ii) $C = BA = \begin{bmatrix} - & \vec{w}_1 & - \\ & \vdots & \\ - & \vec{w}_n & - \end{bmatrix}\begin{bmatrix} | & & | \\ \vec{v}_1 & \cdots & \vec{v}_m \\ | & & | \end{bmatrix}$ has entries $c_{ij} = \vec{w}_i \cdot \vec{v}_j = \sum_{k=1}^{p} b_{ik} a_{kj}$.

<u>Example:</u> Matrix multiplication is not commutative:

$$\begin{bmatrix} 1 & 2 \\ \end{bmatrix}\begin{bmatrix} 1 & 1 \\ \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ \end{bmatrix} \quad \text{but} \quad \begin{bmatrix} 1 & 1 \\ \end{bmatrix}\begin{bmatrix} 1 & 2 \\ \end{bmatrix} = \begin{bmatrix} 4 & 6 \\ \end{bmatrix}.$$

$$\begin{bmatrix} 3 & 4 \end{bmatrix}\begin{bmatrix} 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 3 & 7 \end{bmatrix} \qquad \begin{bmatrix} 0 & 1 \end{bmatrix}\begin{bmatrix} 3 & 4 \end{bmatrix} \quad \begin{bmatrix} 3 & 4 \end{bmatrix}$$

<u>Algebraic rules:</u>

(i) If A is an $n \times m$ matrix then: $AI_m = I_n A = A$.

(ii) Matrix multiplication is associative: $(AB)C = A(BC)$.

(iii) Matrix multiplication distributes over matrix addition:

$$(A+B)C = AC + BC \quad \text{and} \quad A(B+C) = AB + AC.$$

(iv) Multiplication by scalars can be factored out: $(kA)B = A(kB) = k(AB)$.

A function T from X to Y is called <u>invertible</u> if for each y in Y there is a unique x in X with $T(x) = y$. The <u>inverse</u> of T, denoted T^{-1}, is a function from Y to X given by $T^{-1}(y) = x$ (it assigns to each y the x such that $T(x) = y$).

A function T has inverse L if and only if:

$$T(L(y)) = y \text{ for all } y \text{ in } Y, \quad \text{and} \quad L(T(x)) = x \text{ for all } x \text{ in } X.$$

A square matrix A is said to be <u>invertible</u> if the linear transformation $\vec{y} = T(\vec{x}) = A\vec{x}$ is invertible. In this case, T^{-1} will also be a linear transformation, and we denote by A^{-1} its associated matrix: $\vec{x} = T^{-1}(\vec{y}) = A^{-1}\vec{y}$.

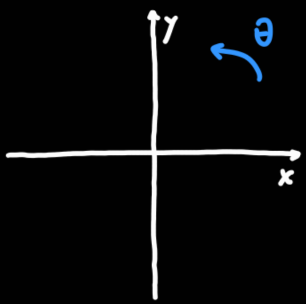
<u>Theorem:</u> Let A be an $n \times n$ matrix.

(i) A is invertible if and only if $\operatorname{rank}(A) = n$, if and only if $\operatorname{rref}(A) = I_n$.

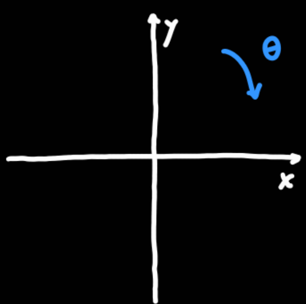
(ii) Let $\vec{b}$ be a vector in $\mathbb{R}^n$. If A is invertible then $A\vec{x} = \vec{b}$ has the unique solution $\vec{x} = A^{-1}\vec{b}$. If A is <u>not</u> invertible then $A\vec{x} = \vec{b}$ has zero or infinitely many solutions.

<u>Example:</u> Let A be an $n \times n$ matrix. The system $A\vec{x} = \vec{0}$ has $\vec{x} = \vec{0}$ as a solution. If A is invertible, this is the only solution. If A is not invertible, there are infinitely many solutions.

<u>Example:</u> Rotation by an angle θ counterclockwise is an invertible transformation with inverse rotation by an angle θ clockwise.



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



is given by $\begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} = A^{-1}!$

<u>Theorem:</u> To find the inverse of an $n \times n$ matrix A, compute $\operatorname{rref}([A \mid I_n])$.

(i) If $\operatorname{rref}([A \mid I_n]) = [I_n \mid B]$ then A is invertible with $A^{-1} = B$.

(ii) If $\mathrm{rref}([A \mid I_n]) \neq [I_n \mid B]$ then A is not invertible.

<u>Example:</u> The matrix $A = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 3 & 2 \\ 3 & 8 & 2 \end{bmatrix}$ is invertible because:

$$\mathrm{rref}\left(\left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 2 & 3 & 2 & 0 & 1 & 0 \\ 3 & 8 & 2 & 0 & 0 & 1 \end{array}\right]\right) = \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 10 & -6 & 1 \\ 0 & 1 & 0 & -2 & 1 & 0 \\ 0 & 0 & 1 & -7 & 5 & 1 \end{array}\right], \text{ so } A^{-1} = \begin{bmatrix} 10 & -6 & 1 \\ -2 & 1 & 0 \\ -7 & 5 & 1 \end{bmatrix}.$$

<u>Theorem:</u> If A, B are two invertible $n \times n$ matrices then:

(i) $AA^{-1} = I_n$ and $A^{-1}A = I_n$,

(ii) BA is invertible with inverse $(BA)^{-1} = A^{-1}B^{-1}$.

<u>Example:</u>

(i) $$\begin{bmatrix} 1 & 1 & 1 \\ 2 & 3 & 2 \\ 3 & 8 & 2 \end{bmatrix}\begin{bmatrix} 10 & -6 & 1 \\ -2 & 1 & 0 \\ -7 & 5 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 10 & -6 & 1 \\ -2 & 1 & 0 \\ -7 & 5 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 & 1 \\ 2 & 3 & 2 \\ 3 & 8 & 2 \end{bmatrix}.$$

(ii) $$\begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}\begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix}\begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}.$$

<u>Theorem:</u> If A, B are two $n \times n$ matrices such that $BA = I_n$ then:

(i) A and B are both invertible,

(ii) $A^{-1} = B$ and $B^{-1} = A$,

(iii) $AB = I_n$.

The 2×2 matrix $A = \begin{bmatrix} a & b \\ & \end{bmatrix}$ has <u>determinant</u> $\det(A) = ad - bc$. The matrix A is

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

invertible if and only if $\det(A) \neq 0$. If A is invertible then $A^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$.

Examples:

(i) For which values of k is the matrix $A = \begin{bmatrix} 1-k & 2 \\ 4 & 3-k \end{bmatrix}$ invertible?

The matrix has determinant $\det(A) = (k-5)(k+1)$, so if $k \neq 5, -1$ then

$\det(A) \neq 0$ and A is invertible.

(ii) Let A be a matrix representing the reflection about a line L passing through the

origin. Then $A = \begin{bmatrix} a & b \\ b & -a \end{bmatrix}$ with $a^2+b^2=1$, so $\det(A) = -a^2-b^2 = -1$ and the

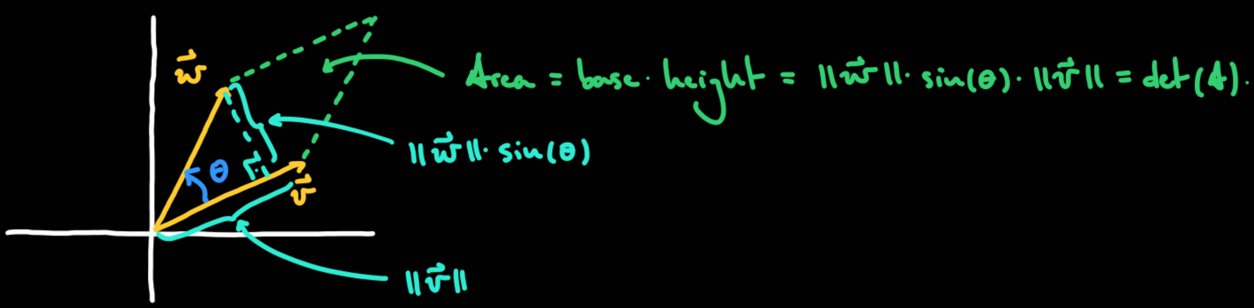
inverse is $A^{-1} = \frac{1}{-1}\begin{bmatrix} -a & -b \\ -b & a \end{bmatrix} = \begin{bmatrix} a & b \\ b & -a \end{bmatrix} = A$. That is, a reflection is its own

inverse.

For $A = [\vec{v}\ \vec{w}]$ with $\vec{v}, \vec{w}$ in $\mathbb{R}^2$ we have:

$\det(A) = \|\vec{v}\| \cdot \|\vec{w}\| \cdot \sin(\theta)$ with θ the angle between $\vec{v}$ and $\vec{w}$.

This determinant is the area of the parallelogram spanned by $\vec{v}$ and $\vec{w}$.



2. Subspaces of $\mathbb{R}^n$ (Chapter 3)

The <u>image</u> of a function T is the set of values $\text{im}(T)$ that the function takes.

$$\text{im}(T) = \{T(x) \mid x \in X\} = \{y \in Y \mid y = T(x) \text{ for some } x \in X\}.$$

<u>Example:</u> Find the image of the following linear transformations:

1. $T: \mathbb{R}^2 \to \mathbb{R}^2$ given by $T(\vec{x}) = \begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix} \vec{x}$.

$$T(\vec{x}) = \begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + x_2 \begin{bmatrix} 3 \\ 6 \end{bmatrix} = (x_1 + 3x_2)\begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

So the image of T is the scalar multiples of $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$:

$$\text{im}(T) = \left\{ k \begin{bmatrix} 1 \\ 2 \end{bmatrix} \,\middle|\, k \in \mathbb{R} \right\}.$$

2. $T: \mathbb{R}^3 \to \mathbb{R}^2$ given by $T(\vec{x}) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \vec{x}$.

This is an orthogonal projection onto the x_1-x_2-plane: $T\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ 0 \end{bmatrix}$, so the image consists of the x_1-x_2-plane:

$$\text{im}(T) = \left\{ \begin{bmatrix} x_1 \\ x_2 \\ 0 \end{bmatrix} \,\middle|\, x_1, x_2 \in \mathbb{R} \right\}.$$

Let $\vec{v}_1, \dots, \vec{v}_m$ be vectors in $\mathbb{R}^n$, the set of all linear combinations of $\vec{v}_1, \dots, \vec{v}_m$ is called their <u>span</u>:

$$\text{span}(\vec{v}_1, \dots, \vec{v}_m) = \{ c_1\vec{v}_1 + \dots + c_m\vec{v}_m \mid c_1, \dots, c_m \in \mathbb{R} \}.$$

Note that the image of the linear transformation $T(\vec{x}) = A\vec{x}$ is the span of the column vectors of A.

<u>Theorem:</u> Let $T: \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation, then: (why? What does T

linear imply ?)

(i) The zero vector $\vec{0} \in \mathbb{R}^n$ is in the image of T.

(ii) The image of T is closed under addition.

(iii) The image of T is closed under scalar multiplication.

<u>Example:</u> Consider $T: \mathbb{R}^2 \to \mathbb{R}^2$ given by $T(\vec{x}) = \begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix} \vec{x}$. Then:

(i) $T(\vec{0}) = \begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$.

(ii) $a \begin{bmatrix} 1 \\ 2 \end{bmatrix} + b \begin{bmatrix} 1 \\ 2 \end{bmatrix} = (a+b) \begin{bmatrix} 1 \\ 2 \end{bmatrix}$.

(iii) $k\left(a \begin{bmatrix} 1 \\ 2 \end{bmatrix}\right) = (ka) \begin{bmatrix} 1 \\ 2 \end{bmatrix}$.

The <u>kernel</u> of a function T is the set of values $\ker(T)$ that the function takes to $\vec{0}$.

$\ker(T) = \{ \vec{x} \in \mathbb{R}^m \mid T(\vec{x}) = \vec{0} \}$, the solutions of the equation $T(\vec{x}) = \vec{0}$.

<u>Example:</u> Find the kernel of the linear transformation $T: \mathbb{R}^3 \to \mathbb{R}^2$ given by $\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix}$.

We solve the equation $T(\vec{x}) = \vec{0}$, which is the linear system:

$$\begin{aligned} x_1 + x_2 + x_3 &= 0 \\ x_1 + 2x_3 + 3x_3 &= 0 \end{aligned}$$
, with augmented matrix $\left[\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 1 & 2 & 3 & 0 \end{array}\right]$.

This matrix has reduced form $\left[\begin{array}{ccc|c} 1 & 0 & -1 & 0 \\ 0 & 1 & 2 & 0 \end{array}\right]$, so by setting $x_1 = t$ the free variable

we have solutions $\vec{x} = \begin{bmatrix} t \\ -2t \\ t \end{bmatrix} = t \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$. Thus the kernel of T are scalar multiples of $\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$.

$\ker(T) = \left\{ k \begin{bmatrix} -2 \\ 1 \end{bmatrix} \mid k \in \mathbb{R} \right\}.$

<u>Example:</u> Let T be an invertible linear transformation. If $\vec{x} \in \ker(T)$, then $T(\vec{x}) = \vec{0}$

so $A\vec{x} = \vec{0}$ so $\vec{x} = A^{-1}A\vec{x} = A^{-1}\vec{0} = \vec{0}$. Hence $\ker(T) = \{\vec{0}\}$, often denoted $\ker(T) = 0$.

<u>Example:</u> Let A be an $n \times m$ matrix with $\ker(A) = 0$. Then, the system $A\vec{x} = \vec{0}$ has exactly one solution, and thus there are no free variables. In particular, all variables are leading, so $\text{rank}(A) = m$. (is the converse true? Why?)

<u>Theorem:</u> Let $T: \mathbb{R}^m \longrightarrow \mathbb{R}^n$ be a linear transformation, then: (why? What does T linear imply?)

(i) The zero vector $\vec{0} \in \mathbb{R}^m$ is in the kernel of T.

(ii) The kernel of T is closed under addition.

(iii) The kernel of T is closed under scalar multiplication.

<u>Theorem:</u> Let A be an $n \times m$ matrix.

(i) We have $\ker(A) = 0$ if and only if $\text{rank}(A) = m$.

(ii) If $\ker(A) = 0$ then $m \leq n$.

(iii) If $m > n$ then there are non-zero vectors in the kernel of A.

(iv) Let $n = m$, so A is a square matrix. We have $\ker(A) = 0$ if and only if A is invertible.

<u>Recall:</u> Let A be an $n \times n$ matrix. The following statements are equivalent:

(i) A is invertible.

(ii) The equation $A\vec{x} = \vec{b}$ has a unique solution for each $\vec{b} \in \mathbb{R}^n$.

(iii) $\operatorname{rref}(A) = I_n$.

(iv) $\operatorname{rank}(A) = n$.

(v) $\operatorname{im}(A) = \mathbb{R}^n$.

(vi) $\ker(A) = \{\vec{0}\}$.

A subset W of the vector space $\mathbb{R}^n$ is called a <u>linear subspace</u> of $\mathbb{R}^n$ if:

(i) The zero vector $\vec{0}$ is in W, and

(ii) W is closed under addition, and

(iii) W is closed under scalar multiplication.

In particular, W is closed under linear combinations.

<u>Example:</u> Let $T: \mathbb{R}^m \longrightarrow \mathbb{R}^n$ be a linear transformation. Then $\operatorname{im}(T)$ is a subspace of $\mathbb{R}^n$ and $\ker(T)$ is a subspace of $\mathbb{R}^m$.

<u>Example:</u>

1. The set $W = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \middle| \; x \geq 0, \; y \geq 0 \right\}$ is not a subspace of $\mathbb{R}^2$.

2. The only subspaces of $\mathbb{R}^2$ are:

(a) $\mathbb{R}^2$,

(b) $\{\vec{0}\}$,

(c) Any line L passing through the origin.

3. Let V be a plane in $\mathbb{R}^3$ given by the equation $x+2y+3z=0$.

(a) Find a matrix A such that $\ker(A) = V$.

We want a matrix that inputs a vector in $\mathbb{R}^3$ and whose kernel is a plane, that is, the equation $A\vec{x} = \vec{0}$ is the defining equation of V. Since V has only one defining equation, $A\vec{x} = \vec{0}$ has to already be one equation, so $\vec{0} = [0]$ is in $\mathbb{R}$. Thus A inputs a vector in $\mathbb{R}^3$ and outputs a vector in $\mathbb{R}$, so A is a 1×3 matrix. Now:

$$0 = A\vec{x} = \begin{bmatrix} a & b & c \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = ax + by + cz$$

has to be the equation $x + 2y + 3z = 0$ defining V. Thus $a=1$, $b=2$, $c=3$ and hence $A = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$.

(b) Find a matrix B such that $\operatorname{im}(B) = V$.

We know that the image of a matrix is the span of its columns, so if we describe V as the span of two vectors, the matrix B will have two columns, each column will be one of those vectors. To find two non parallel vectors, we set $z=0$ and obtain that $\begin{bmatrix}-2\\1\\0\end{bmatrix}$ is in V, and we set $y=0$ and obtain that $\begin{bmatrix}-3\\0\\1\end{bmatrix}$ is in V. Since their dot product is 6, which is not $5\sqrt{2}$, these vectors are not parallel, they are both in V, so they span V. Thus $B=\begin{bmatrix}-2 & -3\\1 & 0\\0 & 1\end{bmatrix}$.

<u>Example:</u> Find vectors in $\mathbb{R}^3$ that span the image of $A=\begin{bmatrix}1&2&1&2\\1&2&2&3\\1&2&3&4\end{bmatrix}$. Find the smallest number of vectors needed to span the image of A.

We know that the image of A is spanned by the column vectors of A:

$$\vec{v}_1=\begin{bmatrix}1\\1\\1\end{bmatrix},\ \vec{v}_2=\begin{bmatrix}2\\2\\2\end{bmatrix},\ \vec{v}_3=\begin{bmatrix}1\\2\\3\end{bmatrix},\ \vec{v}_4=\begin{bmatrix}2\\3\\4\end{bmatrix},\ \operatorname{im}(A)=\operatorname{span}(\vec{v}_1,\vec{v}_2,\vec{v}_3,\vec{v}_4).$$

However, since $\vec{v}_2=2\cdot\vec{v}_1$ and $\vec{v}_4=\vec{v}_1+\vec{v}_3$, we have that $\vec{v}_2$ and $\vec{v}_4$ do not contribute to the span of $\vec{v}_1$ and $\vec{v}_3$. Namely, let $\vec{v}\in\operatorname{span}(\vec{v}_1,\vec{v}_2,\vec{v}_3,\vec{v}_4)$, then:

$$\vec{v}=c_1\vec{v}_1+c_2\vec{v}_2+c_3\vec{v}_3+c_4\vec{v}_4=(c_1+2c_2+c_4)\vec{v}_1+(c_3+c_4)\vec{v}_3$$

for c_1,c_2,c_3,c_4 real numbers, so $\vec{v}\in\operatorname{span}(\vec{v}_1,\vec{v}_3)$. Thus:

$$\operatorname{im}(A)=\operatorname{span}(\vec{v}_1,\vec{v}_2,\vec{v}_3,\vec{v}_4)=\operatorname{span}(\vec{v}_1,\vec{v}_3)$$

and since $\vec{v}_1\cdot\vec{v}_3=6$ they are not parallel, and two is the minimum number of vectors needed to span the image of A.

Let $\vec{v}_1, \dots, \vec{v}_m$ be vectors in $\mathbb{R}^n$. We say that a vector $\vec{v}_i$ is <u>redundant</u> if $\vec{v}_i$ is a linear combination of $\vec{v}_1, \dots, \vec{v}_{i-1}$. We say that $\vec{v}_1$ is <u>redundant</u> if $\vec{v}_1 = \vec{0}$. We say that the vectors $\vec{v}_1, \dots, \vec{v}_m$ are <u>linearly independent</u> if none of them is redundant. If at least one of them is redundant, we call them <u>linearly dependent</u>. We say that the vectors $\vec{v}_1, \dots, \vec{v}_m$ form a <u>basis</u> of a subspace V of $\mathbb{R}^n$ if they span V and are linearly independent.

<u>Example:</u> To construct a basis of the image of a matrix A we only need to list the column vectors of A and remove the redundant vectors from the list.

<u>Example:</u> We can identify linearly independence if enough entries are zero:

1. $\vec{v}_1 = \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}, \vec{v}_2 = \begin{bmatrix} 0 \\ 3 \\ 4 \end{bmatrix}, \vec{v}_3 = \begin{bmatrix} 9 \\ 7 \\ 3 \end{bmatrix}$ are linearly independent.
2. $\vec{v}_1 = \begin{bmatrix} 7 \\ 0 \\ 4 \\ 0 \\ 1 \\ 9 \\ 0 \end{bmatrix}, \vec{v}_2 = \begin{bmatrix} 6 \\ 0 \\ 7 \\ 1 \\ 4 \\ 8 \\ 0 \end{bmatrix}, \vec{v}_3 = \begin{bmatrix} 5 \\ 0 \\ 6 \\ 2 \\ 3 \\ 1 \\ 7 \end{bmatrix}, \vec{v}_4 = \begin{bmatrix} 4 \\ 5 \\ 3 \\ 3 \\ 2 \\ 2 \\ 4 \end{bmatrix}$ are linearly independent.

Let $\vec{v}_1, \dots, \vec{v}_m$ be vectors in $\mathbb{R}^n$. An equation of the form:

$$c_1\vec{v}_1 + \cdots + c_m\vec{v}_m = \vec{0}$$

is called a <u>linear relation</u> among $\vec{v}_1, \dots, \vec{v}_m$. If $c_1 = \cdots = c_m = 0$, the relation is called trivial. If at least one c_i is non zero, the relation is nontrivial.

Theorem: The vectors $\vec{v}_1, \ldots, \vec{v}_m$ in $\mathbb{R}^n$ are linearly dependent if and only if there is at least one nontrivial relation among them.

Example: Let A be an $n \times m$ matrix with linearly independent columns. Then the equation $A\vec{x} = \vec{0}$ with $A = \begin{bmatrix} | & & | \\ \vec{v}_1 & \cdots & \vec{v}_m \\ | & & | \end{bmatrix}$ gives a relation $x_1\vec{v}_1 + \cdots + x_m\vec{v}_m = \vec{0}$, so by linear independence this relation must be trivial, so $x_1 = \cdots = x_m = 0$, so $\ker(A) = 0$.

More generally, the vectors in the kernel of an $n \times m$ matrix A correspond to the linear relations among the column vectors of A, since writing $A\vec{x} = \vec{0}$ with $A = \begin{bmatrix} | & & | \\ \vec{v}_1 & \cdots & \vec{v}_m \\ | & & | \end{bmatrix}$ means that $x_1\vec{v}_1 + \cdots + x_m\vec{v}_m = \vec{0}$. In other words, the column vectors of A are linearly independent if and only if $\ker(A) = 0$, which happens if and only if $\operatorname{rank}(A) = m$.

Example: The columns of the matrix $\begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix}$ are linearly dependent since:

$$1 \cdot \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - 2 \cdot \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} + 1 \cdot \begin{bmatrix} 7 \\ 8 \\ 9 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \text{ and thus } \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \text{ is in } \ker(A).$$

Recall: Let $\vec{v}_1, \ldots, \vec{v}_m$ be vectors in $\mathbb{R}^n$. The following are equivalent:

(i) The vectors $\vec{v}_1, \ldots, \vec{v}_m$ are linearly independent.

(ii) None of the vectors $\vec{v}_1, \ldots, \vec{v}_m$ is redundant.

(iii) None of the vectors $\vec{v}_1, \ldots, \vec{v}_m$ is a linear combination of the others.

(iv) There is only one relation among the vectors $\vec{v}_1, \dots, \vec{v}_m$, the trivial relation.

(v) $\ker\left(\begin{bmatrix} | & & | \\ \vec{v}_1 & \cdots & \vec{v}_m \\ | & & | \end{bmatrix}\right) = 0.$

(vi) $\operatorname{rank}\left(\begin{bmatrix} | & & | \\ \vec{v}_1 & \cdots & \vec{v}_m \\ | & & | \end{bmatrix}\right) = m.$

<u>Theorem:</u> Let V be a subspace of $\mathbb{R}^n$, let $\vec{v}_1, \dots, \vec{v}_m$ be vectors in V. They form a basis of V if and only if every vector $\vec{v}$ in V can be expressed uniquely as a linear combination:

$$\vec{v} = c_1\vec{v}_1 + \cdots + c_m\vec{v}_m .$$

We call $c_1, \dots, c_m$ the <u>coordinates</u> of $\vec{v}$ with respect to the basis $\vec{v}_1, \dots, \vec{v}_m$.

<u>Theorem:</u>

(i) A spanning set is larger than or equal in size to a linearly independent set.

(ii) Any two basis of the same subspace have the same number of elements.

The number $\dim(V)$ of vectors in a basis of a subspace V is called the <u>dimension</u> of V.

<u>Theorem:</u> Let V be a subspace of $\mathbb{R}^n$ with $\dim(V) = m$.

(i) There are at most m linearly independent vectors in V.

(ii) We need at least m vectors to span V.

(iii) If m vectors in V are linearly independent, they form a basis of V.

(iv) If m vectors in V span V, they form a basis of V.

There are strong relations between the size of a matrix, its rank, and the dimensions of its kernel and image.

<u>Theorem:</u> Let A be a matrix, then $\dim(\text{im}(A)) = \text{rank}(A)$.

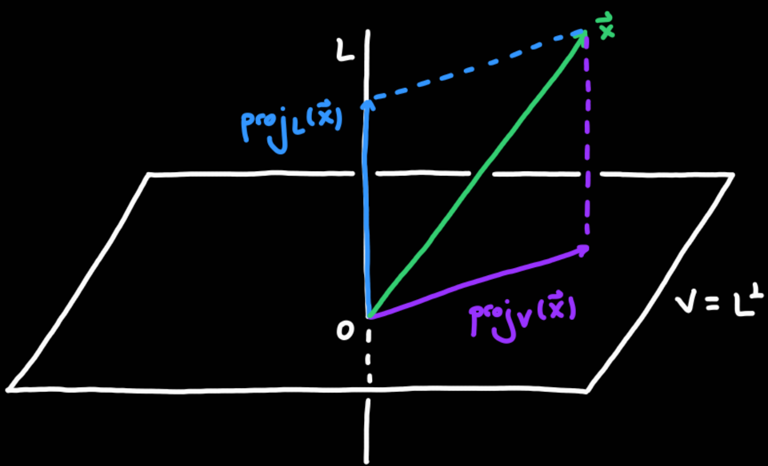
<u>Theorem:</u> (Rank-Nullity) Let A be an $n \times m$ matrix, then:

$$\dim(\ker(A)) + \dim(\text{im}(A)) = m.$$

We call the dimension of $\ker(A)$ the <u>nullity</u> of A. Then:

$$(\text{nullity of } A) + (\text{rank of } A) = m.$$

<u>Example:</u> Let T be the orthogonal projection onto a plane V in $\mathbb{R}^3$.



Here we have $T: \mathbb{R}^3 \longrightarrow \mathbb{R}^2$ so it is given by a 2×3 matrix so $m = 3$, also $\text{im}(T) = V$ so $\dim(\text{im}(T)) = 2$, and $\ker(T) = L$ so $\dim(\ker(T)) = 1$. Clearly $1 + 2 = 3$.

<u>Theorem:</u> Let $A = \begin{bmatrix} | & & | \\ \vec{v}_1 & \cdots & \vec{v}_m \\ | & & | \end{bmatrix}$ be an $n \times m$ matrix with $\text{rref}(A) = \begin{bmatrix} | & & | \\ \vec{w}_1 & \cdots & \vec{w}_m \\ | & & | \end{bmatrix}$.

(i) To construct a basis of the image of A, pick the column vectors of A that correspond to the columns of $\mathrm{rref}(A)$ containing the leading 1's.

(ii) The column vectors of A that correspond to the columns of $\mathrm{rref}(A)$ that do not contain leading 1's are redundant, and can be used to find a basis of $\ker(A)$.

(iii) Suppose that column i of $\mathrm{rref}(A)$ does not contain a leading 1, let $w_{j_1}, \dots, w_{j_r}$ be the entries of $\vec{w}_i$ that have a leading 1 to its left, say in columns $c_1, \dots, c_r$ respectively. Then we have the relation: $\vec{v}_i = w_{j_1}\vec{v}_{c_1} + \dots + w_{j_r}\vec{v}_{c_r}$.

In particular, this last relation means that the vector with entry 1 in position i and entries $-w_{j_1}, \dots, -w_{j_r}$ in positions $c_1, \dots, c_r$ is in the kernel of A, and all these vectors form a basis of $\ker(A)$.

<u>Example:</u> Find bases of the image and kernel of $A = \begin{bmatrix} 1 & 2 & 0 & 1 & 2 \\ 1 & 2 & 0 & 2 & 3 \\ 1 & 2 & 0 & 3 & 4 \\ 1 & 2 & 0 & 4 & 5 \end{bmatrix}$.

We have $\mathrm{rref}(A) = \begin{bmatrix} 1 & 2 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$, so seeing $A = \begin{bmatrix} | & & | \\ \vec{v}_1 & \cdots & \vec{v}_m \\ | & & | \end{bmatrix}$ we have $\vec{v}_1$ and $\vec{v}_4$ as basis of the image:

$$\mathrm{im}(A) = \mathrm{span}\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} \right\}.$$

Moreover, since columns 2, 3, 5 do not have leading 1's, vectors $\vec{v}_2, \vec{v}_3, \vec{v}_5$ are redundant.

We can read from $\mathrm{rref}(A)$ that:

$$\vec{v}_2 = 2\vec{v}_1 \quad \text{so} \quad -2\vec{v}_1 + \vec{v}_2 = \vec{0} \quad \text{so} \quad \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \text{ is in } \ker(A).$$

$$\vec{v}_3 = 0 \cdot \vec{v}_1 = \vec{0} \quad \text{so} \quad \vec{v}_3 = \vec{0} \quad \text{so} \quad \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \text{ is in } \ker(A).$$

$$\vec{v}_5 = \vec{v}_1 + \vec{v}_4 \quad \text{so} \quad -\vec{v}_1 - \vec{v}_4 + \vec{v}_5 = \vec{0} \quad \text{so} \quad \begin{bmatrix} -1 \\ 0 \\ 0 \\ -1 \\ 1 \end{bmatrix} \text{ is in } \ker(A).$$

In fact, these three vectors form a basis of $\ker(A)$:

$$\ker(A) = \mathrm{span}\left\{ \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 0 \\ -1 \\ 1 \end{bmatrix} \right\}.$$

So far, the only basis of $\mathbb{R}^n$ that we have seen is the canonical one, namely:

$$\mathbb{R}^n = \mathrm{span}\left\{ \vec{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \dots, \vec{e}_n = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \right\}.$$

However, there are many more.

<u>Theorem:</u> The vectors $\vec{v}_1, \dots, \vec{v}_n$ of $\mathbb{R}^n$ form a basis of $\mathbb{R}^n$ if and only if the matrix

$$\begin{bmatrix} | & & | \\ \vec{v}_1 & \cdots & \vec{v}_n \\ | & & | \end{bmatrix} \text{ is invertible.}$$

<u>Example:</u> For which values of the constant k do the vectors $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ k \\ k^2 \end{bmatrix}$ form a basis of $\mathbb{R}^3$? It suffices to examine the matrix $\begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & k \\ 1 & 1 & k^2 \end{bmatrix}$, and to determine when it is invertible. Since $\mathrm{rref}\left(\begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & k \\ 1 & 1 & k^2 \end{bmatrix}\right) = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & (1-k)/2 \\ 0 & 0 & k^2-1 \end{bmatrix}$, the only thing that we need for this reduced matrix to be further reducible to I_3 is that $k^2 - 1 \neq 0$, that is $k \neq 1, -1$.

Coordinates in a subspace of $\mathbb{R}^n$.

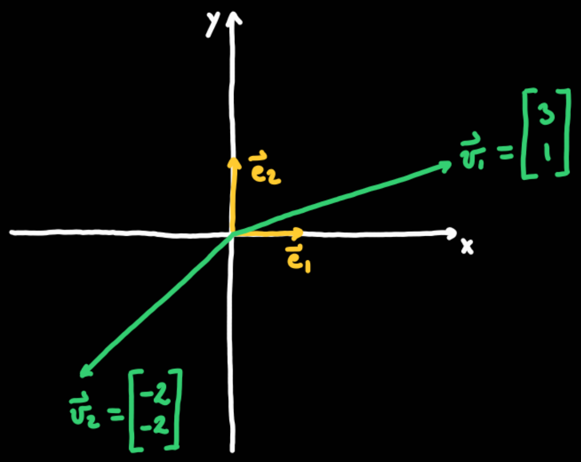
Let $\mathcal{B} = \{\vec{v}_1, \dots, \vec{v}_m\}$ be a basis of a subspace V of $\mathbb{R}^n$. Any $\vec{x}$ in V can be written as $\vec{x} = c_1\vec{v}_1 + \cdots + c_m\vec{v}_m$ for some real scalars $c_1, \dots, c_m$, called the <u>$\mathcal{B}$-coordinates</u> of $\vec{x}$. The vector $\begin{bmatrix} c_1 \\ \vdots \\ c_m \end{bmatrix}$ is called the <u>$\mathcal{B}$-coordinate vector</u> of $\vec{x}$, denoted $[\vec{x}]_{\mathcal{B}}$.

<u>Example:</u> Note that:

$$\vec{x} = c_1\vec{v}_1 + \cdots + c_m\vec{v}_m = \begin{bmatrix} | & & | \\ \vec{v}_1 & \cdots & \vec{v}_m \\ | & & | \end{bmatrix}\begin{bmatrix} c_1 \\ \vdots \\ c_m \end{bmatrix} = S[\vec{x}]_{\mathcal{B}}.$$

The matrix S inputs vectors with $\mathcal{B}$-coordinates, and outputs vectors in the standard basis $\vec{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \dots, \vec{e}_m = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$. (later, we will call this a change of basis matrix)

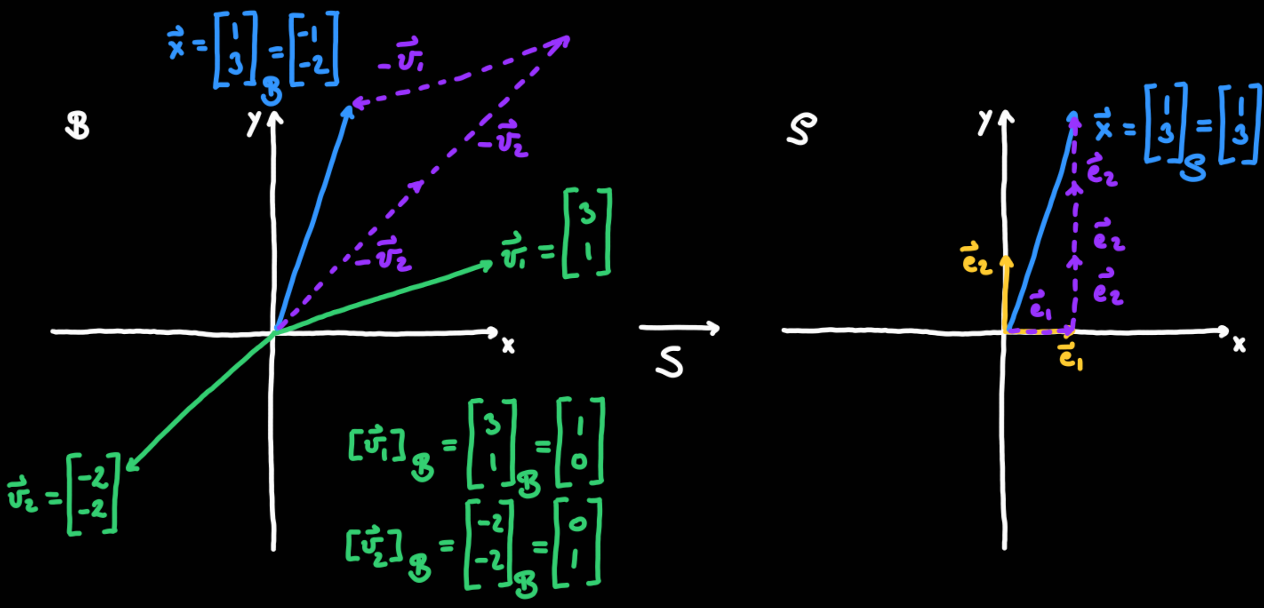
<u>Example:</u> The plane $\mathbb{R}^2$ has a basis $\mathcal{B} = \left\{\begin{bmatrix} 3 \\ 1 \end{bmatrix}, \begin{bmatrix} -2 \\ -2 \end{bmatrix}\right\}$.



The matrix $S = \begin{bmatrix} 3 & -2 \\ 1 & -2 \end{bmatrix}$ takes $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ to $\vec{v}_1$ and $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ to $\vec{v}_2$. It is a transformation

$$S: \mathbb{R}^2_{\mathcal{B}} \longrightarrow \mathbb{R}^2_{\mathcal{S}}.$$

The standard basis of $\mathbb{R}^2$ is $\mathcal{S} = \left\{\vec{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \vec{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right\}$. When we write a vector $\vec{x}$ as $\vec{x} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$, we are implicitly understanding $\begin{bmatrix} 1 \\ 3 \end{bmatrix}_{\mathcal{S}} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$ in the target of the linear transformation S. When we write $\vec{x}$ as $\vec{x} = -1 \cdot \begin{bmatrix} 3 \\ 1 \end{bmatrix} - 2 \cdot \begin{bmatrix} -2 \\ -2 \end{bmatrix}$, so $\begin{bmatrix} 1 \\ 3 \end{bmatrix}_{\mathcal{B}} = \begin{bmatrix} -1 \\ -2 \end{bmatrix}$, we are now in the source of the linear transformation S.



The matrix S is sending a vector in the basis $\mathcal{B}$ to a vector in the basis $\mathcal{S}$.

When we are given a vector in the basis $\mathcal{S}$ and we are asked to find its coordinates in the basis $\mathcal{B}$, we are being asked to solve the system $\vec{x} = S[\vec{x}]_{\mathcal{B}}$, where $\vec{x}$ and S are known, and $[\vec{x}]_{\mathcal{B}}$ is unknown. Since S has for columns the vectors in the basis $\mathcal{B}$, the matrix S has all columns linearly independent. If S is a square matrix, S will then be invertible, and $[\vec{x}]_{\mathcal{B}} = S^{-1}\vec{x}$.

Let $\vec{x} = \begin{bmatrix} 3 \\ -3 \end{bmatrix}$, to find $[\vec{x}]_{\mathcal{B}}$ we solve:

$\begin{bmatrix} 3 \\ -3 \end{bmatrix} = \begin{bmatrix} 3 & -2 \\ 1 & -2 \end{bmatrix}\begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$, with augmented matrix $\left[\begin{array}{cc|c} 3 & -2 & 3 \\ 1 & -2 & -3 \end{array}\right]$, which reduces to $\left[\begin{array}{cc|c} 1 & 0 & 3 \\ 0 & 1 & 3 \end{array}\right]$, and thus $c_1 = 3$ and $c_2 = 3$, so $[\vec{x}]_{\mathcal{B}} = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$, namely $\begin{bmatrix} 3 \\ -3 \end{bmatrix}_{\mathcal{B}} = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$.

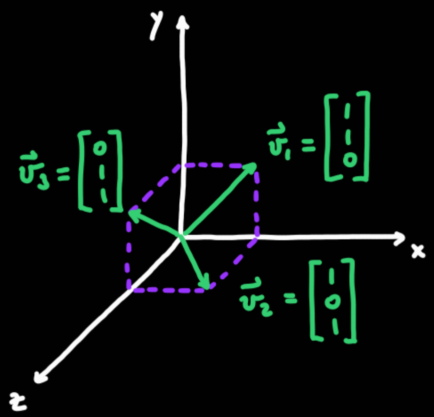
Alternatively, we compute:

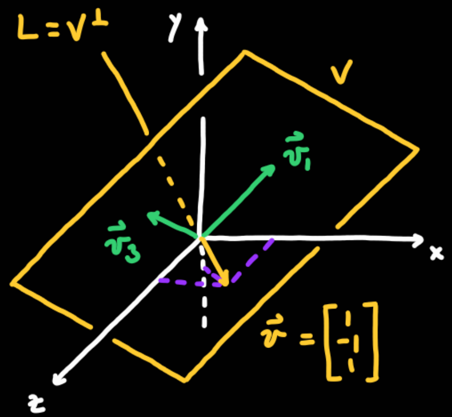
$$[\vec{x}]_{\mathcal{B}} = \begin{bmatrix} 3 & -2 \\ 1 & -2 \end{bmatrix}^{-1}\begin{bmatrix} 3 \\ -3 \end{bmatrix} = \frac{1}{-4}\begin{bmatrix} -2 & 2 \\ -1 & 3 \end{bmatrix}\begin{bmatrix} 3 \\ -3 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \end{bmatrix}.$$

Theorem: Let $T: \mathbb{R}^n \longrightarrow \mathbb{R}^n$ be a linear transformation, $\mathfrak{B} = \{\vec{v}_1, \dots, \vec{v}_n\}$ a basis of $\mathbb{R}^n$.

Then there exists a unique $n \times n$ matrix B transforming $[\vec{x}]_{\mathfrak{B}}$ into $[T(\vec{x})]_{\mathfrak{B}}$, namely $[T(\vec{x})]_{\mathfrak{B}} = B[\vec{x}]_{\mathfrak{B}}$. Moreover: $B = \begin{bmatrix} | & & | \\ [T(\vec{v}_1)]_{\mathfrak{B}} & \cdots & [T(\vec{v}_n)]_{\mathfrak{B}} \\ | & & | \end{bmatrix}$.

Example: Let $\mathfrak{B} = \left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \right\}$ be a basis of $\mathbb{R}^3$. Consider $T: \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ the linear transformation that projects any vector orthogonally onto the plane V spanned by $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$.





To find the defining equation of V, we compute $\vec{v}$ the cross product of $\vec{v}_1$ and $\vec{v}_2$, obtaining a vector perpendicular to V, so V is given by $v_1 x + v_2 y + v_3 z = 0$. In our particular case, $\vec{v} = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$ so V is given by $x - y + z = 0$.

To find the matrix associated to T in the standard basis $\mathcal{S} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$, we know that $A = \begin{bmatrix} | & | & | \\ T(\vec{e}_1) & T(\vec{e}_2) & T(\vec{e}_3) \\ | & | & | \end{bmatrix}$. Recall that:

$$T(\vec{x}) = \mathrm{proj}_V(\vec{x}) = \vec{x} - \mathrm{proj}_L(\vec{x}) = \vec{x} - (\vec{x} \cdot \vec{u}) \cdot \vec{u}$$

with $\vec{u} = \dfrac{\vec{v}}{\|\vec{v}\|} = \dfrac{1}{\sqrt{3}}\begin{bmatrix}1\\-1\\1\end{bmatrix}$, so now:

$$T(\vec{e}_1) = \vec{e}_1 - \frac{1}{\sqrt{3}} \cdot \frac{1}{\sqrt{3}}\begin{bmatrix}1\\-1\\1\end{bmatrix} = \begin{bmatrix}2/3\\1/3\\-1/3\end{bmatrix},$$

$$T(\vec{e}_2) = \vec{e}_2 + \frac{1}{\sqrt{3}} \cdot \frac{1}{\sqrt{3}}\begin{bmatrix}1\\-1\\1\end{bmatrix} = \begin{bmatrix}1/3\\2/3\\1/3\end{bmatrix},$$

$$T(\vec{e}_3) = \vec{e}_3 - \frac{1}{\sqrt{3}} \cdot \frac{1}{\sqrt{3}}\begin{bmatrix}1\\-1\\1\end{bmatrix} = \begin{bmatrix}-1/3\\1/3\\2/3\end{bmatrix},$$

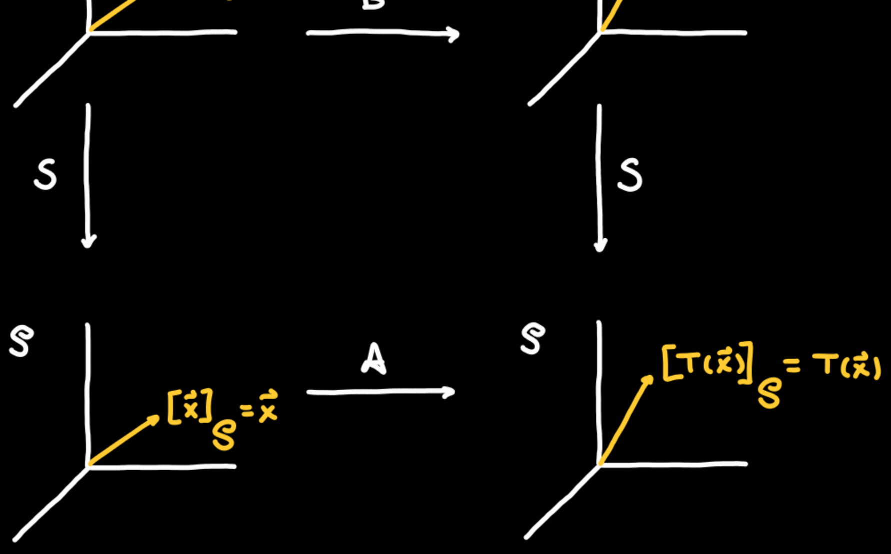
and thus $A = \begin{bmatrix}2/3 & 1/3 & -1/3\\1/3 & 2/3 & 1/3\\-1/3 & 1/3 & 2/3\end{bmatrix}$.

All the above work was done over the standard basis S. If we work with the basis $\mathcal{B}$, the linear transformation $T: \mathbb{R}^3 \to \mathbb{R}^3$ is projecting onto the basis vectors $\vec{v}_1$ and $\vec{v}_3$. Note that:

$\vec{v}_2 = \frac{1}{3}\vec{v}_1 + \frac{1}{3}\vec{v}_3 + \frac{2}{3}\vec{v}$, where we computed $\text{proj}_{L}(\vec{v}_2) = \frac{2}{3}\vec{v}$ as the component of $\vec{v}_2$ that gets sent to zero by T.

Thus if we write $\vec{x} = c_1\vec{v}_1 + c_2\vec{v}_2 + c_3\vec{v}_3 = \left(c_1 + \frac{c_2}{3}\right)\vec{v}_1 + \left(c_3 + \frac{c_2}{3}\right)\vec{v}_3 + \frac{2c_2}{3}\vec{v}$, we have $T(\vec{x}) = \left(c_1 + \frac{c_2}{3}\right)\vec{v}_1 + \left(c_3 + \frac{c_2}{3}\right)\vec{v}_3$. As before, we know that working over $\mathcal{B}$, the matrix associated to T should be $B = \begin{bmatrix} | & | & | \\ [T(\vec{v}_1)]_{\mathcal{B}} & [T(\vec{v}_2)]_{\mathcal{B}} & [T(\vec{v}_3)]_{\mathcal{B}} \\ | & | & | \end{bmatrix}$,

namely $B = \begin{bmatrix}1 & 1/3 & 0\\0 & 0 & 0\\0 & 1/3 & 1\end{bmatrix}$. The following is happening:



Namely :

$$T(\vec{x}) = AS[\vec{x}]_B \text{ and } T(\vec{x}) = SB[\vec{x}]_B \text{ for all } \vec{x}, \text{ so } AS = SB$$

where $S = \begin{bmatrix} | & | & | \\ \vec{v}_1 & \vec{v}_2 & \vec{v}_3 \\ | & | & | \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$ is the matrix who inputs vectors in the basis $\mathcal{B}$ and outputs vectors in the basis $\mathcal{S}$. In other words :

$$\begin{tikzcd} \mathbb{R}^3, \mathcal{S} \arrow[r, "A"] \arrow[d, "S^{-1}"'] & \mathbb{R}^3, \mathcal{S} \\ \mathbb{R}^3, \mathcal{B} \arrow[r, "B"] & \mathbb{R}^3, \mathcal{B} \arrow[u, "S"'] \end{tikzcd}$$

to compute T we can work in any basis, each basis will have a matrix that is associated to T, and we can move back and forth between the different basis using the change of basis matrix.

We can check that :

$$SBS^{-1} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1/3 & 0 \\ 0 & 0 & 0 \\ 0 & 1/3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}^{-1} =$$

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

<u>Theorem</u>: Let $T: \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation, $\mathcal{B} = \{\vec{v}_1, \dots, \vec{v}_n\}$ a basis of $\mathbb{R}^n$, A the matrix associated to T in the standard basis, B the matrix associated to T in the basis $\mathcal{B}$, and $S = \begin{bmatrix} | & & | \\ \vec{v}_1 & \cdots & \vec{v}_n \\ | & & | \end{bmatrix}$. Then $AS = SB$.

Let A, B be two $n \times n$ matrices. We say that A is <u>similar</u> to B if there exists an invertible matrix S such that $AS = SB$.

<u>Example</u>: Let A be similar to B, then A^T is similar to B^T:

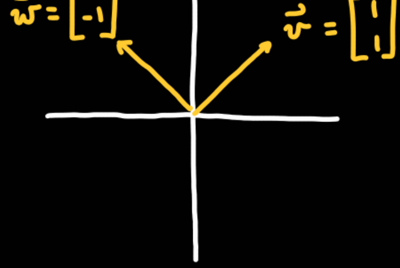
$$A^T = (SBS^{-1})^T = (S^{-1})^T B^T S^T = (S^T)^{-1} B^T S^T = R^{-1} B^T R \text{ with } R = S^T.$$

3. Orthogonality. (Chapter 5)

Two vectors $\vec{v}, \vec{w}$ in $\mathbb{R}^n$ are <u>perpendicular</u> (or <u>orthogonal</u>) if $\vec{v} \cdot \vec{w} = 0$. The <u>length</u> (or <u>magnitude</u>) of a vector $\vec{v}$ is $\|\vec{v}\| = \sqrt{\vec{v} \cdot \vec{v}}$. A vector $\vec{u}$ is called a <u>unit</u> vector if it has length one. A vector $\vec{x}$ in $\mathbb{R}^n$ is said to be <u>orthogonal</u> to a subspace V of $\mathbb{R}^n$ if it is orthogonal to all the vectors in V: $\vec{x} \cdot \vec{v} = 0$ for all $\vec{v}$ in V.

<u>Remark</u>: A vector $\vec{x}$ is orthogonal to V if and only if $\vec{x}$ is orthogonal to a basis of V.

<u>Example</u>: The vectors $\vec{v} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\vec{w} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ are orthogonal: $\vec{v} \cdot \vec{w} = 1 \cdot 1 - 1 \cdot 1 = 0$.



Both vectors have length $\sqrt{2}$:

$$\|\vec{v}\| = \sqrt{\vec{v}\cdot\vec{v}} = \sqrt{2}\ , \quad \|\vec{w}\| = \sqrt{\vec{w}\cdot\vec{w}} = \sqrt{2}.$$

<u>Example</u>: Given a plane $ax+by+cz=0$ in $\mathbb{R}^3$, the vector $\vec{x} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$ is orthogonal to every vector in the plane. To see this, let $\vec{v} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$ be a vector in this plane, then:

$$\vec{x}\cdot\vec{v} = \begin{bmatrix} a & b & c \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = ax+by+cz=0.$$

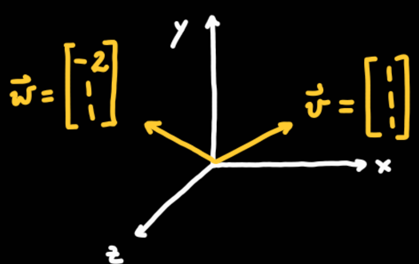
The vectors $\vec{v}_1, \ldots, \vec{v}_m$ in $\mathbb{R}^n$ are called <u>orthonormal</u> if they are all unit vectors and orthogonal to one another: $\vec{v}_i\cdot\vec{v}_j = 0$ for $i \neq j$ and $v_i \cdot v_i = 1$.

<u>Theorem</u>:

(i) Orthonormal vectors are linearly independent.

(ii) If $\vec{v}_1, \ldots, \vec{v}_n$ in $\mathbb{R}^n$ are orthonormal, then $\mathcal{B} = \{\vec{v}_1, \ldots, \vec{v}_n\}$ is a basis of $\mathbb{R}^n$.

<u>Example</u>: We can make any two orthogonal vectors into an orthonormal basis of a plane.



Since $\vec{v}\cdot\vec{w} = (-2)\cdot 1 + 1\cdot 1 + 1\cdot 1 = 0$, they are perpendicular.

The plane they define is perpendicular to $\vec{u} = \begin{bmatrix} 0 \\ 3 \\ -3 \end{bmatrix}$, so $\vec{v}$ and $\vec{w}$ are both in the plane $3y-3z=0$. By normalizing $\vec{v}$ and $\vec{w}$:

$$\vec{v}' = \frac{\vec{v}}{\|\vec{v}\|} = \frac{1}{\sqrt{3}}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}\ , \quad \vec{w}' = \frac{1}{\sqrt{6}}\begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix},$$

we obtain two orthonormal vectors. They are both in the plane $3y-3z=0$, they

are linearly independent because:

$$\begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{-2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} \end{bmatrix}$$ has rank 2, and thus $\vec{v}'$ and $\vec{w}'$ form an orthonormal basis of $3y-3z=0$.

In the same way that we could project onto a line and onto a plane, we can project onto any subspace V of $\mathbb{R}^n$. Let $\vec{x}$ be a vector in $\mathbb{R}^n$, we can write $\vec{x} = \vec{x}^{\parallel} + \vec{x}^{\perp}$ where $\vec{x}^{\parallel}$ is in V and $\vec{x}^{\perp}$ is orthogonal to V. If $\vec{u}_1, \dots, \vec{u}_m$ is an orthonormal basis of V, then: $\vec{x}^{\parallel} = \text{proj}_V(\vec{x}) = (\vec{x}\cdot\vec{u}_1)\vec{u}_1 + \dots + (\vec{x}\cdot\vec{u}_m)\vec{u}_m$.

<u>Example:</u> Consider the plane V with orthonormal basis $\vec{u}_1 = \frac{1}{\sqrt{3}}\begin{bmatrix}1\\1\\1\end{bmatrix}$, $\vec{u}_2 = \frac{1}{\sqrt{6}}\begin{bmatrix}-2\\1\\1\end{bmatrix}$, and

$\vec{x} = \begin{bmatrix}4\\5\\6\end{bmatrix}$. Then:

$$\vec{x}^{\parallel} = \left(\frac{1}{\sqrt{3}}\begin{bmatrix}1 & 1 & 1\end{bmatrix}\begin{bmatrix}4\\5\\6\end{bmatrix}\right)\frac{1}{\sqrt{3}}\begin{bmatrix}1\\1\\1\end{bmatrix} + \left(\frac{1}{\sqrt{6}}\begin{bmatrix}-2 & 1 & 1\end{bmatrix}\begin{bmatrix}4\\5\\6\end{bmatrix}\right)\frac{1}{\sqrt{6}}\begin{bmatrix}-2\\1\\1\end{bmatrix} = \begin{bmatrix}5\\5\\5\end{bmatrix} + \begin{bmatrix}-1\\1/2\\1/2\end{bmatrix} = \begin{bmatrix}4\\11/2\\11/2\end{bmatrix}.$$

Now $\vec{x}^{\parallel}$ is in V since V is given by the equation $3y-3z=0$ and $3\cdot\frac{11}{2} - 3\cdot\frac{11}{2} = 0$.

Moreover:

$$\vec{x}^{\perp} = \vec{x} - \vec{x}^{\parallel} = \begin{bmatrix}4\\5\\6\end{bmatrix} - \begin{bmatrix}4\\11/2\\11/2\end{bmatrix} = \begin{bmatrix}0\\-\frac{1}{2}\\\frac{1}{2}\end{bmatrix}$$ is a multiple of $\begin{bmatrix}0\\-3\\3\end{bmatrix}$, which is perpendicular to V, so indeed $\vec{x}^{\perp}$ is orthogonal to V. We can check:

$$\vec{x} = \begin{bmatrix}4\\5\\6\end{bmatrix} = \begin{bmatrix}4\\11/2\\11/2\end{bmatrix} + \begin{bmatrix}0\\-\frac{1}{2}\\\frac{1}{2}\end{bmatrix} = \vec{x}^{\parallel} + \vec{x}^{\perp}.$$

If $\vec{u}_1, \ldots, \vec{u}_n$ is an orthonormal basis of $\mathbb{R}^n$, then:

$$\vec{x} = (\vec{x}\cdot\vec{u}_1)\vec{u}_1 + \cdots + (\vec{x}\cdot\vec{u}_n)\vec{u}_n .$$

Namely, we can project $\vec{x}$ onto the lines determined by the basis vectors, add the projections, and we recover $\vec{x}$.

Let V be a subspace of $\mathbb{R}^n$, the <u>orthogonal complement</u> $V^\perp$ of V is the set of all vectors in $\mathbb{R}^n$ that are orthogonal to all vectors in V:

$$V^\perp = \{\vec{x} \in \mathbb{R}^n \mid \vec{x}\cdot\vec{v} = 0 \text{ for all } \vec{v} \in V\}.$$

If $T: \mathbb{R}^n \longrightarrow \mathbb{R}^n$ is the orthogonal projection onto V, any vector $\vec{x}$ that is orthogonal to V will be sent to zero: $T(\vec{x}) = 0$. In fact, $V^\perp$ is the kernel of this projection.

<u>Example:</u> Consider $T: \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ the linear transformation that projects any vector orthogonally onto the plane V spanned by $\begin{bmatrix}1\\1\\0\end{bmatrix}$ and $\begin{bmatrix}0\\1\\1\end{bmatrix}$. We already computed the associated matrix of this transformation as $A = \begin{bmatrix}2/3 & 1/3 & -1/3\\ 1/3 & 2/3 & 1/3\\ -1/3 & 1/3 & 2/3\end{bmatrix}$. Now:

$A\vec{x} = \vec{0}$ has augmented matrix $\left[\begin{array}{ccc|c}2/3 & 1/3 & -1/3 & 0\\ 1/3 & 2/3 & 1/3 & 0\\ -1/3 & 1/3 & 2/3 & 0\end{array}\right]$ which reduces to

$\left[\begin{array}{ccc|c}1 & 0 & -1 & 0\\ 0 & 1 & 1 & 0\\ 0 & 0 & 0 & 0\end{array}\right]$ so by setting $x_3 = t$ we have $\vec{x} = \begin{bmatrix}t\\-t\\t\end{bmatrix} = t\cdot\begin{bmatrix}1\\-1\\1\end{bmatrix}$.

Thus $\ker(T) = \operatorname{span}\left(\begin{bmatrix}1\\-1\\1\end{bmatrix}\right)$ which is exactly the line perpendicular to V.

<u>Theorem:</u> Let V be a subspace of $\mathbb{R}^n$. Then:

(i) The orthogonal complement $V^\perp$ is a subspace of $\mathbb{R}^n$.

(ii) The intersection of V and $V^\perp$ is exactly the zero vector.

(iii) $\dim(V) + \dim(V^\perp) = n$.

(iv) $(V^\perp)^\perp = V$.

<u>Theorem:</u> Let $\vec{x}, \vec{y}$ be in $\mathbb{R}^n$, let θ be the angle between them, let V be a subspace. Then:

(i) $\|\vec{x} + \vec{y}\|^2 = \|\vec{x}\|^2 + \|\vec{y}\|^2$ if and only if $\vec{x}$ and $\vec{y}$ are orthogonal.

(ii) $\|\mathrm{proj}_V(\vec{x})\| \leq \|\vec{x}\|$, with equality if and only if $\vec{x}$ is in V.

(iii) $|\vec{x} \cdot \vec{y}| \leq \|\vec{x}\| \cdot \|\vec{y}\|$, with equality if and only if $\vec{x}$ and $\vec{y}$ are parallel.

(iv) $\cos(\theta) = \dfrac{\vec{x} \cdot \vec{y}}{\|\vec{x}\| \cdot \|\vec{y}\|}$.

<u>Example:</u> Let $\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \right\}$ be a basis of $\mathbb{R}^3$. This is not an orthonormal basis for many reasons:

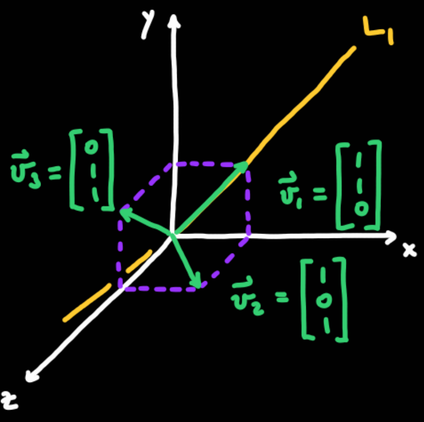
(i) The vectors have length $\sqrt{2}$.

(ii) The dot product of any two vectors is 1, so the angle between any two vectors satisfies $\cos(\theta) = \frac{1}{\sqrt{2} \cdot \sqrt{2}} = \frac{1}{2}$, so $\theta = \frac{\pi}{3}$, which is not $\frac{\pi}{2}$.

Having orthonormal basis is extremely useful. Given any basis of a subspace of $\mathbb{R}^n$, there is a methodical way of finding an orthonormal basis.

Example: Let $\mathcal{B} = \left\{ \begin{bmatrix}1\\1\\0\end{bmatrix}, \begin{bmatrix}1\\0\\1\end{bmatrix}, \begin{bmatrix}0\\1\\1\end{bmatrix} \right\}$ be a basis of $\mathbb{R}^3$. We begin by normalizing the first vector:

$$\vec{u}_1 = \frac{\vec{v}_1}{\|\vec{v}_1\|} = \frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\\0\end{bmatrix},$$ who also defines a line in $\mathbb{R}^3$:



Then, we decompose the second vector in the components parallel and perpendicular to $\vec{u}_1$:

$$\vec{v}_2 = \vec{v}_2^{\parallel} + \vec{v}_2^{\perp}$$

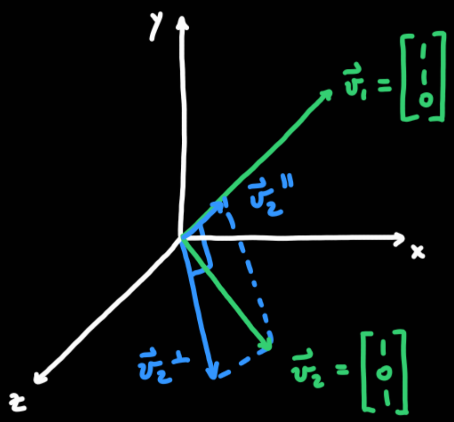
$$\vec{v}_2^{\parallel} = \text{proj}_{L_1}(\vec{v}_2) = (\vec{v}_2 \cdot \vec{u}_1)\,\vec{u}_1 = \left([1\ 0\ 1] \cdot \begin{bmatrix}1/\sqrt{2}\\1/\sqrt{2}\\0\end{bmatrix} \right) \frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\\0\end{bmatrix} = \frac{1}{2}\begin{bmatrix}1\\1\\0\end{bmatrix}$$

$$\vec{v}_2^{\perp} = \vec{v}_2 - \vec{v}_2^{\parallel} = \begin{bmatrix}1\\0\\1\end{bmatrix} - \frac{1}{2}\begin{bmatrix}1\\1\\0\end{bmatrix} = \begin{bmatrix}1/2\\-1/2\\1\end{bmatrix}$$

and we normalize the perpendicular component:

$$\vec{u}_2 = \frac{\vec{v}_2^{\perp}}{\|\vec{v}_2^{\perp}\|} = \frac{2}{\sqrt{6}}\begin{bmatrix}1/2\\-1/2\\1\end{bmatrix} = \frac{1}{\sqrt{6}}\begin{bmatrix}1\\-1\\2\end{bmatrix}$$

Now $\vec{u}_1$ and $\vec{u}_2$ span a plane V in $\mathbb{R}^3$, it is the same plane spanned by $\vec{v}_1$ and $\vec{v}_2$.



To finish, we decompose the third vector in the components inside V and orthogonal to V. The vector $\vec{v} = \begin{bmatrix}1\\-1\\-1\end{bmatrix}$ is normal to V, so:

$$\vec{v}_3 = \vec{v}_3^{\parallel} + \vec{v}_3^{\perp}$$

$$\vec{v}_3^{\perp} = \text{proj}_{V^\perp}(\vec{v}_3) = \left(\vec{v}_3 \cdot \frac{\vec{v}}{\|\vec{v}\|}\right)\frac{\vec{v}}{\|\vec{v}\|} = \left(\begin{bmatrix} 0 & 1 & 1 \end{bmatrix} \cdot \frac{1}{\sqrt{3}}\begin{bmatrix} \\ -1 \\ -1 \end{bmatrix}\right)\frac{1}{\sqrt{3}}\begin{bmatrix} \\ -1 \\ -1 \end{bmatrix} = \frac{2}{3}\begin{bmatrix} \\ 1 \\ 1 \end{bmatrix}$$

$$\vec{v}_3^{\parallel} = \vec{v}_3 - \vec{v}_3^{\perp} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} - \frac{2}{3}\begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}$$

and we normalize the perpendicular component:

$$\vec{u}_3 = \frac{\vec{v}_3^{\perp}}{\|\vec{v}_3^{\perp}\|} = \frac{1}{\sqrt{3}} \cdot \frac{3}{2} \cdot \frac{2}{3}\begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix} = \frac{1}{\sqrt{3}}\begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$$

Now $\mathcal{R} = \{\vec{u}_1, \vec{u}_2, \vec{u}_3\} = \left\{ \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \frac{1}{\sqrt{6}}\begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}, \frac{1}{\sqrt{3}}\begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix} \right\}$ is an orthonormal basis.

This is known as the <u>Gram-Schmidt process</u>. Let $\vec{v}_1, \dots, \vec{v}_m$ be a basis of a subspace V of $\mathbb{R}^n$. For each $j = 2, \dots, m$, decompose $\vec{v}_j$ into its components parallel and perpendicular to the span of the preceding vectors $\vec{v}_1, \dots, \vec{v}_{j-1}$: $\vec{v}_j = \vec{v}_j^{\parallel} + \vec{v}_j^{\perp}$. Then:

$$\vec{u}_1 = \frac{\vec{v}_1}{\|\vec{v}_1\|}, \quad \vec{u}_2 = \frac{\vec{v}_2^{\perp}}{\|\vec{v}_2^{\perp}\|}, \quad \dots, \quad \vec{u}_m = \frac{\vec{v}_m^{\perp}}{\|\vec{v}_m^{\perp}\|}$$

is an orthonormal basis of V.

This process is a change of basis from $\{\vec{v}_1, \dots, \vec{v}_m\}$ to $\{\vec{u}_1, \dots, \vec{u}_m\}$. This is encoded in the <u>QR factorization</u> of a matrix M. Let M be an $n \times m$ matrix with linearly independent columns $\vec{v}_1, \dots, \vec{v}_m$, then there exists an $n \times m$ matrix Q whose columns $\vec{u}_1, \dots, \vec{u}_m$ are orthonormal, and an upper triangular matrix R with positive diagonal entries, such that $M = QR$.

In fact, $r_{11} = \|\vec{v}_1\|$ and $r_{jj} = \|\vec{v}_j^{\perp}\|$ for $j = 2, \dots, m$, and $r_{ij} = \vec{u}_i \cdot \vec{v}_j$ for $i < j$

In fact, $r_{11} = \|\vec{v}_1\|$ and $r_{jj} = \|\vec{v}_j^{\perp}\|$ for $j = 2, \dots, m$, and $r_{ij} = \vec{u}_i \cdot \vec{v}_j$ for $i < j$.

We can compute Q and R column by column:

1st column of R:

$$r_{11} = \|\vec{v}_1\|.$$

1st column of Q:

$$\vec{u}_1 = \frac{\vec{v}_1}{r_{11}}$$

2nd column of R:

$$r_{12} = \vec{u}_1 \cdot \vec{v}_2, \quad \vec{v}_2^{\perp} = \vec{v}_2 - r_{12}\vec{u}_1$$

$$r_{22} = \|\vec{v}_2^{\perp}\|$$

2nd column of Q:

$$\vec{u}_2 = \frac{\vec{v}_2^{\perp}}{r_{22}}$$

3rd column of R:

$$r_{13} = \vec{u}_1 \cdot \vec{v}_3,$$

$$r_{23} = \vec{u}_2 \cdot \vec{v}_3, \quad \vec{v}_3^{\perp} = \vec{v}_3 - r_{13}\vec{u}_1 - r_{23}\vec{u}_2$$

$$r_{33} = \|\vec{v}_3^{\perp}\|$$

3rd column of R:

$$\vec{u}_3 = \frac{\vec{v}_3^{\perp}}{r_{33}}$$

and so on, the i-th column of R and Q would be:

$$r_{1i} = \vec{u}_1 \cdot \vec{v}_i,$$

$$r_{2i} = \vec{u}_2 \cdot \vec{v}_i,$$

$$\vdots$$

$$r_{i-1\,i} = \vec{u}_{i-1} \cdot \vec{v}_i, \quad \vec{v}_i^{\perp} = \vec{v}_i - r_{1i}\vec{u}_1 - \cdots - r_{i-1\,i}\vec{u}_{i-1}$$

$$\vec{u}_i = \frac{\vec{v}_i^{\perp}}{r_{ii}}$$

$r_{ii} = \|\vec{v}_i^{\perp}\|$

<u>Example:</u> Find the QR factorization of $M = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$. We have already computed

$$\vec{u}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad \vec{u}_2 = \frac{1}{\sqrt{6}}\begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}, \quad \vec{u}_3 = \frac{1}{\sqrt{3}}\begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}, \text{ so } Q = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{6} & -1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3} \\ 0 & 2/\sqrt{6} & 1/\sqrt{3} \end{bmatrix}.$$

Now:

$$r_{11} = \vec{u}_1 \cdot \vec{v}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \sqrt{2}, \quad r_{12} = \vec{u}_1 \cdot \vec{v}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \frac{1}{\sqrt{2}},$$

$$r_{13} = \vec{u}_1 \cdot \vec{v}_3 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = \frac{1}{\sqrt{2}}, \quad r_{22} = \vec{u}_2 \cdot \vec{v}_2 = \frac{1}{\sqrt{6}}\begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \frac{\sqrt{3}}{\sqrt{2}},$$

$$r_{23} = \vec{u}_2 \cdot \vec{v}_3 = \frac{1}{\sqrt{6}}\begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = \frac{1}{\sqrt{6}}, \quad r_{33} = \vec{u}_3 \cdot \vec{v}_3 = \frac{1}{\sqrt{3}}\begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = \frac{2}{\sqrt{3}}$$

and thus $R = \begin{bmatrix} \sqrt{2} & 1/\sqrt{2} & 1/\sqrt{2} \\ 0 & \sqrt{3}/\sqrt{2} & 1/\sqrt{6} \\ 0 & 0 & 2/\sqrt{3} \end{bmatrix}$. We can check:

$$QR = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{6} & -1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3} \\ 0 & 2/\sqrt{6} & 1/\sqrt{3} \end{bmatrix}\begin{bmatrix} \sqrt{2} & 1/\sqrt{2} & 1/\sqrt{2} \\ 0 & \sqrt{3}/\sqrt{2} & 1/\sqrt{6} \\ 0 & 0 & 2/\sqrt{3} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} = M.$$

A linear transformation $T: \mathbb{R}^n \to \mathbb{R}^n$ is said to be <u>orthogonal</u> if it preserves the length of the vectors: $\|T(\vec{x})\| = \|\vec{x}\|$ for all $\vec{x}$ in $\mathbb{R}^n$.

<u>Example:</u> Rotations and reflections are orthogonal linear transformations.

<u>Theorem:</u> Let $T: \mathbb{R}^n \to \mathbb{R}^n$ be an orthogonal transformation, let $\vec{x}$ and $\vec{y}$ be orthogonal vectors in $\mathbb{R}^n$. Then $T(\vec{x})$ and $T(\vec{y})$ are orthogonal.

<u>Theorem:</u>

(i) A linear transformation $T: \mathbb{R}^n \to \mathbb{R}^n$ is orthogonal if and only if $\{T(\vec{e}_1), \dots, T(\vec{e}_n)\}$

is an orthonormal basis of $\mathbb{R}^n$.

(ii) An $n \times n$ matrix A is orthogonal if and only if its columns form an orthonormal basis of $\mathbb{R}^n$.

(iii) Let A and B be $n \times n$ orthogonal matrices, then AB is orthogonal.

(iv) Let A be an $n \times n$ orthogonal matrix, then A^{-1} is orthogonal.

<u>Example:</u> The following matrices are orthogonal:

$$A = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \quad A^{-1} = A,$$

$$B = \frac{1}{3} \begin{bmatrix} 2 & -2 & 1 \\ 1 & 2 & 2 \\ 2 & 1 & -2 \end{bmatrix}, \quad B^{-1} = \frac{1}{3} \begin{bmatrix} 2 & 1 & 2 \\ -2 & 2 & 1 \\ 1 & 2 & -2 \end{bmatrix}.$$

And for $M = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$, in its QR decomposition we have that Q is orthogonal:

$$Q = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{6} & -1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3} \\ 0 & 2/\sqrt{6} & 1/\sqrt{3} \end{bmatrix}, \quad Q^{-1} = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 1/\sqrt{6} & -1/\sqrt{6} & 2/\sqrt{6} \\ -1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \end{bmatrix}.$$

Let A be an $m \times n$ matrix. The <u>transpose</u> of A, denoted A^T, is the $n \times m$ matrix whose ij-th entry is the ji-th entry of A. In other words, the columns of A^T are the rows of A, and the rows of A^T are the columns of A.

<u>Example:</u> Let $A = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$, then $A^T = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$.

A square matrix A is said to be <u>symmetric</u> if $A^T = A$, and <u>skew symmetric</u> if $A^T = -A$.

Example: The matrix $A = \begin{bmatrix} 1 & 4 & 5 \\ 4 & 2 & 6 \\ 5 & 6 & 3 \end{bmatrix}$ is symmetric because $A^T = \begin{bmatrix} 1 & 4 & 5 \\ 4 & 2 & 6 \\ 5 & 6 & 3 \end{bmatrix} = A$.

The matrix $B = \begin{bmatrix} 0 & -2 & -3 \\ 2 & 0 & 4 \\ 3 & -4 & 0 \end{bmatrix}$ is skew symmetric because $B^T = \begin{bmatrix} 0 & 2 & 3 \\ -2 & 0 & -4 \\ -3 & 4 & 0 \end{bmatrix} = -B$.

Theorem: Let A be an $n \times n$ matrix, it is orthogonal if and only if $A^T A = I_n$, that is, if and only if $A^{-1} = A^T$.

Example: Recall the above orthogonal matrices and their inverses:

$$A = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \quad A^{-1} = A = A^T,$$

$$B = \frac{1}{3} \begin{bmatrix} 2 & -2 & 1 \\ 1 & 2 & 2 \\ 2 & 1 & -2 \end{bmatrix}, \quad B^{-1} = \frac{1}{3} \begin{bmatrix} 2 & 1 & 2 \\ -2 & 2 & 1 \\ 1 & 2 & -2 \end{bmatrix} = B^T.$$

$$Q = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{6} & -1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3} \\ 0 & 2/\sqrt{6} & 1/\sqrt{3} \end{bmatrix}, \quad Q^{-1} = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 1/\sqrt{6} & -1/\sqrt{6} & 2/\sqrt{6} \\ -1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \end{bmatrix} = Q^T.$$

Theorem: Let A, B be matrices of the appropriate sizes, invertible if necessary, and k a real scalar. Then:

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

(ii) $(kA)^T = kA^T$,

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

(iv) $\operatorname{rank}(A^T) = \operatorname{rank}(A)$,

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

The matrix of an orthogonal projection:

Let V be a subspace of $\mathbb{R}^n$ with orthonormal basis $\vec{u}_1, \dots, \vec{u}_m$. The matrix P of the orthogonal projection onto V is: (careful with the order!)

$$P = QQ^T = \begin{bmatrix} | & & | \\ \vec{u}_1 & \cdots & \vec{u}_m \\ | & & | \end{bmatrix} \begin{bmatrix} - & \vec{u}_1 & - \\ & \vdots & \\ - & \vec{u}_m & - \end{bmatrix}$$

Note that $P^T = (QQ^T)^T = (Q^T)^T Q^T = QQ^T = P$, so P is symmetric.

Example: Consider $T: \mathbb{R}^3 \to \mathbb{R}^3$ the linear transformation that projects any vector orthogonally onto the plane V spanned by $\vec{v} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ and $\vec{w} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$. Since these vectors are linearly independent, we can use the Gram-Schmidt process to obtain an orthonormal basis of V:

$$\vec{u}_1 = \frac{\vec{v}}{\|\vec{v}\|} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad \text{so } \vec{w}^\perp = \vec{w} - (\vec{w} \cdot \vec{u}_1)\vec{u}_1 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} - \left(\frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \right) \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix}$$

$$\vec{u}_2 = \frac{\vec{w}^\perp}{\|\vec{w}^\perp\|} = \frac{2}{\sqrt{6}} \cdot \frac{1}{2} \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix} = \frac{1}{\sqrt{6}} \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix}$$

and now the matrix of the orthogonal projection onto V is:

$$P = \begin{bmatrix} | & | \\ \vec{u}_1 & \vec{u}_2 \\ | & | \end{bmatrix} \begin{bmatrix} - & \vec{u}_1 & - \\ - & \vec{u}_2 & - \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{6} \\ 1/\sqrt{2} & 1/\sqrt{6} \\ 0 & 2/\sqrt{6} \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ -1/\sqrt{6} & 1/\sqrt{6} & 2/\sqrt{6} \end{bmatrix} = \begin{bmatrix} 2/3 & 1/3 & -1/3 \\ 1/3 & 2/3 & 1/3 \\ -1/3 & 1/3 & 2/3 \end{bmatrix}$$

which is exactly what we found before.

We will be using these ideas when finding the least squares solution of a linear system. We also need to better understand transposes. Let A be an $n \times m$ matrix, then A^T is an $m \times n$ matrix, so they define two linear transformations:

$A : \mathbb{R}^m \longrightarrow \mathbb{R}^n$ and $A^T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$, giving $A^TA : \mathbb{R}^m \longrightarrow \mathbb{R}^m$.

<u>Theorem:</u> Let A be an $n\times m$ matrix, then $(\mathrm{im}(A))^\perp = \ker(A^T)$ as subspaces of $\mathbb{R}^n$. Also:

(i) $\ker(A) = \ker(A^TA)$.

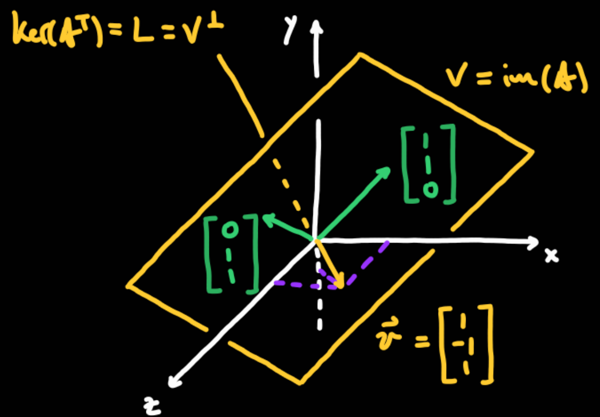
(ii) If $\ker(A) = \{\vec{0}\}$ then A^TA is invertible.

In particular, the orthogonal complement of $\mathrm{im}(A)$ is $\ker(A^T)$.

<u>Example:</u> Consider $T : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ the linear transformation that projects any vector orthogonally onto the plane V spanned by $\begin{bmatrix}1\\1\\0\end{bmatrix}$ and $\begin{bmatrix}0\\1\\1\end{bmatrix}$. It is given by the matrix $A = \begin{bmatrix}2/3 & 1/3 & -1/3\\ 1/3 & 2/3 & 1/3 \\ -1/3 & 1/3 & 2/3\end{bmatrix}$.

Now:

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



and the kernel of A is precisely the line perpendicular to V. We can also compute the kernel of A^T directly:

$\vec{x} \in \ker(A^T)$ if and only if $A^T\vec{x} = \vec{0}$, so we are solving:

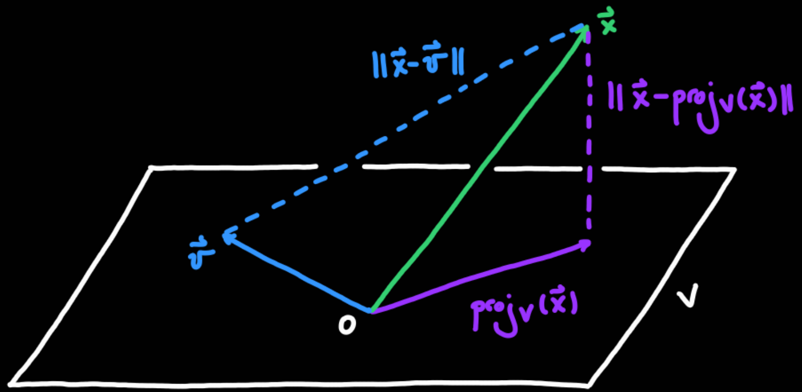
$$\begin{bmatrix}2/3 & 1/3 & -1/3\\ 1/3 & 2/3 & 1/3 \\ -1/3 & 1/3 & 2/3\end{bmatrix}\begin{bmatrix}x\\y\\z\end{bmatrix} = \begin{bmatrix}0\\0\\0\end{bmatrix} \text{ with augmented matrix } \left[\begin{array}{ccc|c}2/3 & 1/3 & -1/3 & 0\\ 1/3 & 2/3 & 1/3 & 0\\ -1/3 & 1/3 & 2/3 & 0\end{array}\right],$$

reducing to $\left[\begin{array}{ccc|c}1 & 0 & -1 & 0\\ 0 & 1 & 1 & 0\end{array}\right]$, so setting $z = t$ we have $\vec{x} = \begin{bmatrix}t\\-t\end{bmatrix} = t\cdot\begin{bmatrix}1\\-1\end{bmatrix} = t\cdot\vec{v}$

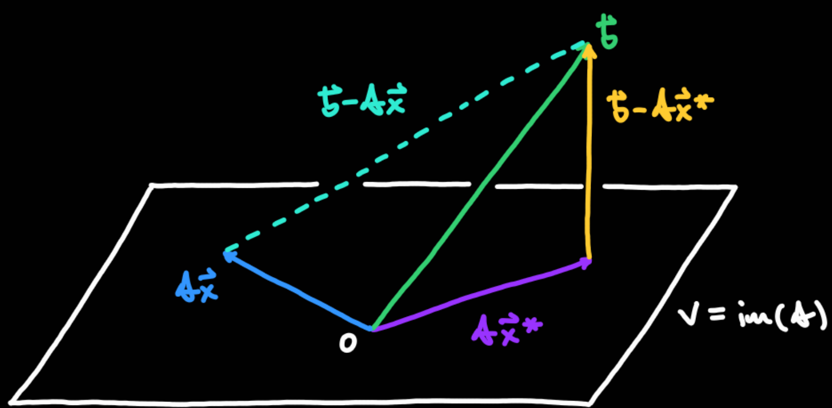
and thus $\ker(A^T) = \operatorname{span}(\vec{v}) = L = V^{\perp} = (\operatorname{im}(A))^{\perp}$.

When computing the orthogonal projection of a vector $\vec{x}$ in $\mathbb{R}^n$ onto a subspace V, we are in fact finding the vector $\operatorname{proj}_V(\vec{x})$ in V that is closest to $\vec{x}$. That is, $\operatorname{proj}_V(\vec{x})$ is the vector $\vec{v}$ in V such that the distance between $\vec{x}$ and $\vec{v}$ is the smallest possible:

$$\|\vec{x} - \operatorname{proj}_V(\vec{x})\| < \|\vec{x} - \vec{v}\| \quad \text{for all } \vec{v} \text{ in } V \text{ different from } \operatorname{proj}_V(\vec{x}).$$



Let A be an $n \times m$ matrix, $\vec{b}$ a vector in $\mathbb{R}^n$, and $A\vec{x} = \vec{b}$ a linear system. A vector $\vec{x}^*$ in $\mathbb{R}^m$ is called a <u>least-squares solution</u> of the system if $\|\vec{b} - A\vec{x}^*\| \leq \|\vec{b} - A\vec{x}\|$ for all $\vec{x} \in \mathbb{R}^m$.



The least-squares solutions of $A\vec{x} = \vec{b}$ are the exact solutions of $A^T A\vec{x} = A^T\vec{b}$, which is always a consistent system which we call the normal equation of $A\vec{x}=\vec{b}$.

Theorem: Let A be an $n\times m$ matrix, $\vec{b}$ a vector in $\mathbb{R}^n$. If $\ker(A)=\{\vec{0}\}$ then the system $A\vec{x}=\vec{b}$ has a unique least-squares solution $\vec{x}^*=(A^TA)^{-1}A^T\vec{b}$.

Warning: The matrices A and A^T are not necessarily square, so A^{-1} and $(A^T)^{-1}$ do not make sense in general. Thus $(A^TA)^{-1}=A^{-1}(A^T)^{-1}$ does not make sense in general.

Example: Consider $T:\mathbb{R}^3\longrightarrow\mathbb{R}^3$ the linear transformation that projects any vector orthogonally onto the plane V spanned by $\begin{bmatrix}1\\1\\0\end{bmatrix}$ and $\begin{bmatrix}0\\1\\1\end{bmatrix}$. It is given by the matrix $P=\begin{bmatrix}2/3 & 1/3 & -1/3\\ 1/3 & 2/3 & 1/3\\ -1/3 & 1/3 & 2/3\end{bmatrix}$. Find the least-square solution $\vec{x}^*$ of $A\vec{x}=\vec{b}$ where $\vec{b}=\begin{bmatrix}1\\0\\1\end{bmatrix}$ and $A=\begin{bmatrix}1&0\\1&1\\0&1\end{bmatrix}$. We have:

$$\vec{x}^*=(A^TA)^{-1}A^T\vec{b}=\left(\begin{bmatrix}1&1&0\\0&1&1\end{bmatrix}\begin{bmatrix}1&0\\1&1\\0&1\end{bmatrix}\right)^{-1}\begin{bmatrix}1&1&0\\0&1&1\end{bmatrix}\begin{bmatrix}1\\0\\1\end{bmatrix}=\begin{bmatrix}2&1\\1&2\end{bmatrix}^{-1}\begin{bmatrix}1\\1\end{bmatrix}=$$

$$=\frac{1}{3}\begin{bmatrix}2&-1\\-1&2\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix}=\frac{1}{3}\begin{bmatrix}1\\1\end{bmatrix}.$$

Since $A=\begin{bmatrix}1&0\\1&1\\0&1\end{bmatrix}$, then $\mathrm{im}(A)=\mathrm{span}\left(\begin{bmatrix}1\\1\\0\end{bmatrix},\begin{bmatrix}0\\1\\1\end{bmatrix}\right)$, say $\vec{v}_1=\begin{bmatrix}1\\1\\0\end{bmatrix}$ and $\vec{v}_3=\begin{bmatrix}0\\1\\1\end{bmatrix}$.

The vectors $\vec{v}_1$ and $\vec{v}_3$ behave like the basis of $\mathrm{im}(A)$. Geometrically, $\vec{x}^*$ corresponds to the vector in $V=\mathrm{im}(A)$ that is closest to $\vec{b}=\begin{bmatrix}1\\0\\1\end{bmatrix}$. To compute the vector in V, we multiply:

$$A\vec{x}^*=\begin{bmatrix}1&0\\1&1\\0&1\end{bmatrix}\frac{1}{3}\begin{bmatrix}1\\1\end{bmatrix}=\frac{1}{3}\begin{bmatrix}1\\2\\1\end{bmatrix}=\frac{1}{3}\cdot\vec{v}_1+\frac{1}{3}\cdot\vec{v}_3,$$ and note

$$P\vec{b} = \begin{bmatrix} 2/3 & 1/3 & -1/3 \\ 1/3 & 2/3 & 1/3 \\ -1/3 & 1/3 & 2/3 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1/3 \\ 2/3 \\ 1/3 \end{bmatrix} = \frac{1}{3}\cdot\vec{v}_1 + \frac{1}{3}\cdot\vec{v}_3 .$$

Finding the least-square solution is indeed the same thing as projecting onto the plane.

Theorem: Let V be a subspace of $\mathbb{R}^n$ with basis $\vec{v}_1, \dots, \vec{v}_m$. The matrix P of the orthogonal projection onto V is: (there is no orthonormality required)

$$P = A(A^TA)^{-1}A^T = \begin{bmatrix} | & & | \\ \vec{v}_1 & \cdots & \vec{v}_m \\ | & & | \end{bmatrix}\left(\begin{bmatrix} - & \vec{v}_1 & - \\ & \vdots & \\ - & \vec{v}_m & - \end{bmatrix}\begin{bmatrix} | & & | \\ \vec{v}_1 & \cdots & \vec{v}_m \\ | & & | \end{bmatrix}\right)^{-1}\begin{bmatrix} - & \vec{v}_1 & - \\ & \vdots & \\ - & \vec{v}_m & - \end{bmatrix}$$

where $A = \begin{bmatrix} | & & | \\ \vec{v}_1 & \cdots & \vec{v}_m \\ | & & | \end{bmatrix}$.

Example: Consider $T: \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ the linear transformation that projects any vector orthogonally onto the plane V spanned by $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$. It is given by:

$$P = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix}\left(\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix}\right)^{-1}\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix}\frac{1}{3}\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 2/3 & 1/3 & -1/3 \\ 1/3 & 2/3 & 1/3 \\ -1/3 & 1/3 & 2/3 \end{bmatrix}$$

which is indeed the expected result.

4. Determinants. (Chapter 6)

Denote by $M_n(\mathbb{R})$ the set of all $n\times n$ matrices, a determinant is a function $\det : M_n(\mathbb{R}) \longrightarrow \mathbb{R}$ that inputs a square matrix and outputs a real number and that satisfies the following:

(i) It is linear with respect to each column:

$$\det\begin{bmatrix} | & & | & & | \\ \vec{c}_1, & \dots, & \vec{c}_i + \vec{c}_i', & \dots, & \vec{c}_n \\ | & & | & & | \end{bmatrix} = \det\begin{bmatrix} | & & | & & | \\ \vec{c}_1, & \dots, & \vec{c}_i, & \dots, & \vec{c}_n \\ | & & | & & | \end{bmatrix} + \det\begin{bmatrix} | & & | & & | \\ \vec{c}_1, & \dots, & \vec{c}_i', & \dots, & \vec{c}_n \\ | & & | & & | \end{bmatrix}$$

and :

$$\det \begin{bmatrix} | & & | & & | \\ \vec{c}_1, & \dots, & k\,\vec{c}_i, & \dots, & \vec{c}_n \\ | & & | & & | \end{bmatrix} = k \cdot \det \begin{bmatrix} | & & | & & | \\ \vec{c}_1, & \dots, & \vec{c}_i, & \dots, & \vec{c}_n \\ | & & | & & | \end{bmatrix}$$

(ii) It is alternating in the columns :

$$\det \begin{bmatrix} | & & | & & | & & | \\ \vec{c}_1, & \dots, & \vec{c}_i, & \dots, & \vec{c}_i, & \dots, & \vec{c}_n \\ | & & | & & | & & | \end{bmatrix} = 0$$

(iii) The determinant of the identity matrix is 1:

$$\det \begin{bmatrix} | & & | \\ \vec{e}_1, & \dots, & \vec{e}_n \\ | & & | \end{bmatrix} = 1$$

Such a function exists, and it is unique! There are many ways of computing determinants, but they will all give the same result.

<u>Example:</u> For $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ then:

$\det(A) = ad - bc$, and A is invertible if and only if $\det(A) \neq 0$.

<u>Example:</u> For $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$ then :

$$\det(A) = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33}.$$

<u>Example:</u> For an upper triangular matrix $A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1\,n-1} & a_{1n} \\ 0 & a_{22} & \cdots & a_{2\,n-1} & a_{2n} \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & a_{n-1\,n-1} & a_{n-1\,n} \\ 0 & 0 & \cdots & 0 & a_{nn} \end{bmatrix}$ then :

$\det(A) = a_{11} \cdot a_{22} \cdots a_{n-1\,n-1} \cdot a_{nn}$,

the determinant of A is the product of its diagonal entries. Similarly, if B is a lower triangular matrix, the determinant of B is the product of its diagonal entries.

Although conceptually this definition of determinant is extremely powerful (it is known as the "universal property" of the determinant), it is hard to compute determinants using it.

<u>Theorem</u>: Let A be an invertible $n\times n$ matrix. If when computing rref(A) we swap rows s times and we divide rows by the scalars $k_1, \dots, k_r$, then:

$$\det(A) = (-1)^s \cdot k_1 \cdots k_r.$$

<u>Example</u>: Let $A = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \\ 2 & 1 & 2 \end{bmatrix}$, now:

$$\begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \\ 2 & 1 & 2 \end{bmatrix} \xrightarrow[R_3-2R_1]{R_2-3R_1} \begin{bmatrix} 1 & 2 & 3 \\ 0 & -4 & -8 \\ 0 & -3 & -4 \end{bmatrix} \xrightarrow{\frac{1}{-4}\cdot R_2} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & -3 & -4 \end{bmatrix} \xrightarrow[R_3+3R_2]{R_1-2R_2} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 2 \end{bmatrix}$$

$$\xrightarrow{\frac{1}{2}\cdot R_3} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow[R_2-2R_3]{R_1+R_3} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

where we have never swapped rows and we have divided by -4 and 2. Thus $\det(A) = -8$.

We can check that:

$$\det(A) = 1\cdot 2\cdot 2 + 2\cdot 1\cdot 2 + 3\cdot 3\cdot 1 - 3\cdot 2\cdot 2 - 1\cdot 1\cdot 1 - 2\cdot 3\cdot 2 = -8.$$

We can also compute the determinant in a recursive way.

Let A be an $n\times n$ matrix, the $(n-1)\times(n-1)$ matrix A_{ij} obtained by omitting the i-th row and the j-th column of A is called a <u>submatrix</u> of A. The determinant of A_{ij} is called a <u>minor</u> of A.

Theorem: Let A be an $n \times n$ matrix, then:

$$\det(A) = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \cdot \det(A_{ij})$$

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \cdot \det(A_{ij})$$

Where the first equality is an expansion down the j-th column, and the second equality is an expansion along the i-th row. We can iterate this process to compute $\det(A_{ij})$.

Example: Let $A = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \\ 2 & 1 & 2 \end{bmatrix}$. Expanding down the second column:

$$\det(A) = (-1)^{1+2} \cdot 2 \cdot \det\begin{bmatrix} 3 & 1 \\ 2 & 2 \end{bmatrix} + (-1)^{2+2} \cdot 2 \cdot \det\begin{bmatrix} 1 & 3 \\ 2 & 2 \end{bmatrix} + (-1)^{3+2} \cdot 1 \cdot \det\begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix} =$$

$$= -2 \cdot 4 + 2 \cdot (-4) - 1 \cdot (-8) = -8$$

Expanding along the third row:

$$\det(A) = (-1)^{3+1} \cdot 2 \cdot \det\begin{bmatrix} 2 & 3 \\ 2 & 1 \end{bmatrix} + (-1)^{3+2} \cdot 1 \cdot \det\begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix} + (-1)^{3+3} \cdot 2 \cdot \det\begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix} =$$

$$= 2 \cdot (-4) - 1 \cdot (-8) + 2 \cdot (-4) = -8.$$

Remark: The determinant is symmetric with respect to rows and columns, because both expansions above are the same. Thus:

(i) It is linear with respect to each row:

$$\det\begin{bmatrix} — \vec{r}_1 — \\ \vdots \\ — \vec{r}_i + \vec{r}_i{}' — \\ \vdots \\ — \vec{r}_n — \end{bmatrix} = \det\begin{bmatrix} — \vec{r}_1 — \\ \vdots \\ — \vec{r}_i — \\ \vdots \\ — \vec{r}_n — \end{bmatrix} + \det\begin{bmatrix} — \vec{r}_1 — \\ \vdots \\ — \vec{r}_i{}' — \\ \vdots \\ — \vec{r}_n — \end{bmatrix}$$

and:

$$\det\begin{bmatrix} — \vec{r}_1 — \\ \vdots \\ — k\vec{r}_i — \\ \vdots \\ — \vec{r}_n — \end{bmatrix} = k \cdot \det\begin{bmatrix} — \vec{r}_1 — \\ \vdots \\ — \vec{r}_i — \\ \vdots \\ — \vec{r}_n — \end{bmatrix}$$

(ii) It is alternating in the rows:

$$\det\begin{bmatrix} — \vec{r}_1 — \\ \vdots \\ — \vec{r}_i — \\ \vdots \\ — \vec{r}_i — \\ \vdots \\ — \vec{r}_n — \end{bmatrix} = 0$$

(iii) The determinant of the identity matrix is 1:

$$\det\begin{bmatrix} — \vec{e}_1 — \\ \vdots \\ — \vec{e}_n — \end{bmatrix} = 1$$

In particular, for all $n \times n$ matrices A we have $\det(A^T) = \det(A)$.

<u>Example:</u> let $A = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \\ 2 & 1 & 2 \end{bmatrix}$ so $A^T = \begin{bmatrix} 1 & 3 & 2 \\ 2 & 2 & 1 \\ 3 & 1 & 2 \end{bmatrix}$. Expanding down the third column:

$$\det(A^T) = (-1)^{1+3} \cdot 2 \cdot \det\begin{bmatrix} 2 & 2 \\ 3 & 1 \end{bmatrix} + (-1)^{2+3} \cdot 1 \cdot \det\begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix} + (-1)^{3+3} \cdot 2 \cdot \det\begin{bmatrix} 1 & 3 \\ 2 & 2 \end{bmatrix} =$$

$$= 2 \cdot (-4) - 1 \cdot (-8) + 2 \cdot (-4) = -8$$

and we already knew that $\det(A) = -8$.

It is useful to remember how elementary row operations affect the determinant:

(i) If B is obtained from A by dividing a row of A by k then:

$$\det(B) = \frac{1}{k} \cdot \det(A)$$

(ii) If B is obtained from A by swapping two rows then:

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

(iii) If B is obtained from A by adding a multiple of a row of A to another row of A then:

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

Finally, as the name indicates, the determinant detects when a matrix is invertible.

__Theorem:__ A square matrix is invertible if and only if it has non-zero determinant.

__Theorem:__ Let A be an invertible square matrix, then $\det(A^{-1}) = \frac{1}{\det(A)} = \det(A)^{-1}$.

__Example:__ Let $A = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \\ 2 & 1 & 2 \end{bmatrix}$, recalling the operations we used to reduce it, we find:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow[R_3-2R_1]{R_2-3R_1} \begin{bmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix} \xrightarrow{\frac{1}{-4}\cdot R_2} \begin{bmatrix} 1 & 0 & 0 \\ 3/4 & -1/4 & 0 \\ -2 & 0 & 1 \end{bmatrix} \xrightarrow[R_3+3R_2]{R_1-2R_2} \begin{bmatrix} -1/2 & 1/2 & 0 \\ 3/4 & -1/4 & 0 \\ 1/4 & -3/4 & 1 \end{bmatrix}$$

$$\xrightarrow{\frac{1}{2}\cdot R_3} \begin{bmatrix} -1/2 & 1/2 & 0 \\ 3/4 & -1/4 & 0 \\ 1/8 & -3/8 & 1/2 \end{bmatrix} \xrightarrow[R_2-2R_3]{R_1+R_3} \begin{bmatrix} -3/8 & 1/8 & 1/2 \\ 1/2 & 1/2 & -1 \\ 1/8 & -3/8 & 1/2 \end{bmatrix} \quad \text{so} \quad A^{-1} = \begin{bmatrix} -3/8 & 1/8 & 1/2 \\ 1/2 & 1/2 & -1 \\ 1/8 & -3/8 & 1/2 \end{bmatrix}.$$

To reduce this inverse to the identity, we only have to reverse all the previous arrows.

Doing so does not swap any rows, and we multiply by 2 and -4. Hence:

$$\det(A^{-1}) = \frac{1}{2}\cdot\frac{1}{-4} = \frac{1}{-8} = \frac{1}{\det(A)}.$$

__Remark:__ The determinant is linear on the columns and rows of the matrices, but not on the matrices themselves: $\det(A+B) \neq \det(A) + \det(B)$ in general. For example, if $A = I_n$ and $B = -I_n$ for n even integer then $\det(A) = 1 = \det(B)$

so $\det(A) + \det(B) = 2$ but $\det(A+B) = \det \begin{bmatrix} 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{bmatrix} = 0$.

Also $\det(kA) \neq k\det(A)$ in general, for example if $A = I_n$ and $k \neq 1, 0$ then

$k \cdot \det(A) = k$ but $\det(kA) = k^n$. In fact,

<u>Theorem</u>: Let A, B be two $n \times n$ matrices, let k be a real number. Then:

(i) $\det(kA) = k^n \cdot \det(A)$

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

(iii) $\det(A^m) = \det(A)^m$

(iv) If A and B are similar, then $\det(A) = \det(B)$.

We can use these properties of the determinant to great effect.

<u>Theorem</u>: The determinant of an orthogonal matrix is ± 1. (why?)

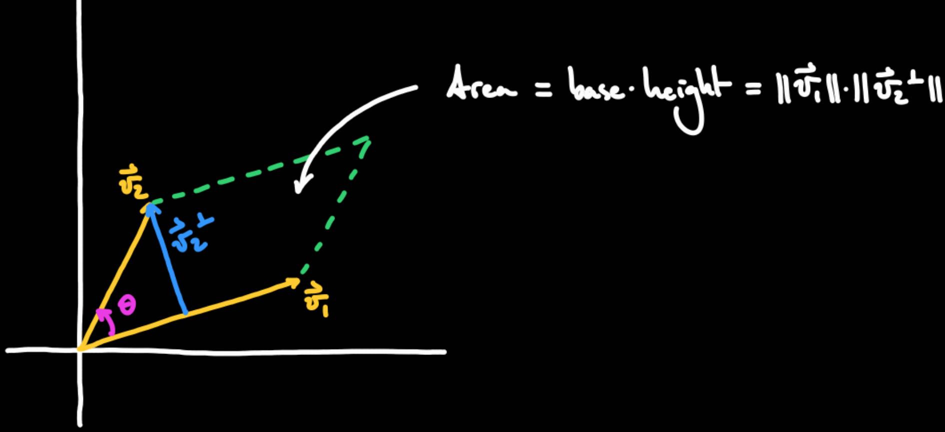
<u>Geometric interpretation of the determinant:</u>

Recall that if $A = \begin{bmatrix} | & | \\ \vec{v}_1 & \vec{v}_2 \\ | & | \end{bmatrix}$ is a 2×2 matrix, then $\det(A) = \|\vec{v}_1\| \cdot \|\vec{v}_2\| \cdot \sin(\theta)$, where

θ is the angle between $\vec{v}_1$ and $\vec{v}_2$. Note that $\|\vec{v}_2\| \cdot |\sin(\theta)|$ is exactly the length of

$\vec{v}_2^{\perp}$, the component of $\vec{v}_2$ perpendicular to $\vec{v}_1$. Thus $|\det(A)| = \|\vec{v}_1\| \cdot \|\vec{v}_2^{\perp}\|$ is the

area of the parallelogram spanned by $\vec{v}_1$ and $\vec{v}_2$.



Using the Gram-Schmidt process and the QR decomposition we can write any invertible matrix $A = \begin{bmatrix} | & & | \\ \vec{v}_1 & \cdots & \vec{v}_n \\ | & & | \end{bmatrix}$ as $A = QR$ with Q orthogonal and R upper triangular with diagonal entries $r_{11} = \|\vec{v}_1\|$, $r_{22} = \|\vec{v}_2^{\perp}\|, \ldots, r_{nn} = \|\vec{v}_n^{\perp}\|$, so:

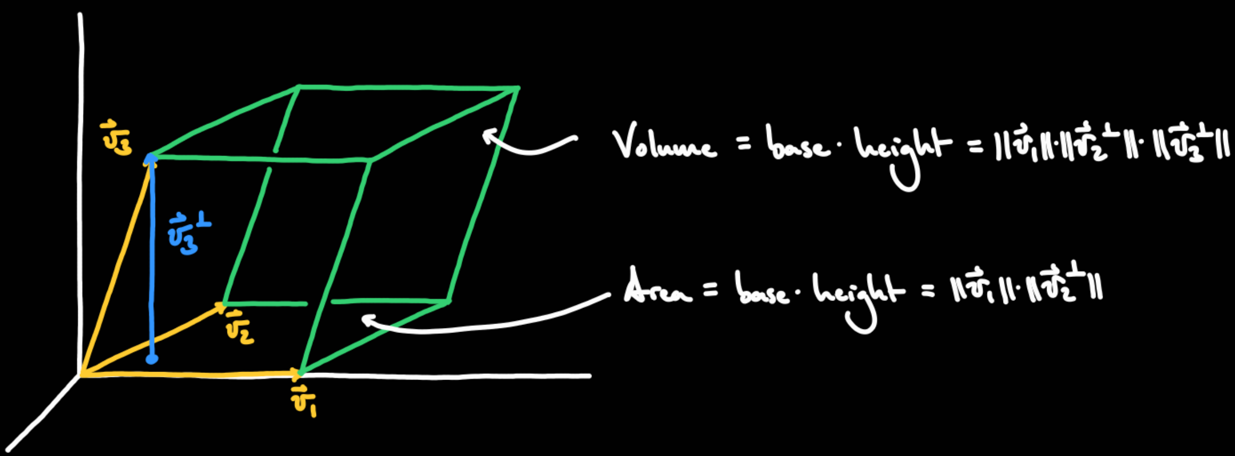
$$|\det(A)| = |\det(QR)| = |\det(Q)| \cdot |\det(R)| = \|\vec{v}_1\| \cdot \|\vec{v}_2^{\perp}\| \cdots \|\vec{v}_n^{\perp}\|$$

where $\vec{v}_i^{\perp}$ is the component of $\vec{v}_i$ perpendicular to $\operatorname{span}(\vec{v}_1, \ldots, \vec{v}_{i-1})$.

Observe that this formula is precisely the generalization of the formula:

$$\text{area} = \text{base} \cdot \text{height}$$

to any dimension. We can illustrate the case $n=3$:



The volume in $\mathbb{R}^n$ of the linearly independent vectors $\vec{v}_1, \ldots, \vec{v}_n$ is $\left| \det \begin{bmatrix} | & & | \\ \vec{v}_1 & \cdots & \vec{v}_n \\ | & & | \end{bmatrix} \right|$.

There is a way of expressing the solutions of a linear system using determinants.

<u>Theorem:</u> (Cramer's rule) Let $A\vec{x} = \vec{b}$ be a linear system. If A is an invertible $n\times n$ matrix then the i-th component of the solution vector is:

$$x_i = \frac{\det(A_{\vec{b},i})}{\det(A)}$$

where $A_{\vec{b},i}$ is the matrix obtained by replacing the i-th column of A by $\vec{b}$.

This is powerful because it allows the computation of inverses of matrices using a close formula. For this, we introduce the <u>classical adjoint</u> of an invertible matrix A, denoted $\mathrm{adj}(A)$, which is the matrix having entries: $(\mathrm{adj}(A))_{ij} = (-1)^{i+j}\cdot \det(A_{ji})$.

<u>Theorem:</u> Let A be an invertible matrix, then $A^{-1} = \dfrac{\mathrm{adj}(A)}{\det(A)}$.

<u>Example:</u> Let $A = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \\ 2 & 1 & 2 \end{bmatrix}$, then to compute A^{-1} we can first compute all the minors of A, and put them in their respective positions in a matrix:

$$\begin{bmatrix} (-1)^{1+1}\det(A_{11}) & (-1)^{1+2}\det(A_{12}) & (-1)^{1+3}\det(A_{13}) \\ (-1)^{2+1}\det(A_{21}) & (-1)^{2+2}\det(A_{22}) & (-1)^{2+3}\det(A_{23}) \\ (-1)^{3+1}\det(A_{31}) & (-1)^{3+2}\det(A_{32}) & (-1)^{3+3}\det(A_{33}) \end{bmatrix} = \begin{bmatrix} 3 & -4 & -1 \\ -1 & -4 & 3 \\ -4 & 8 & -4 \end{bmatrix}$$

we then transpose this matrix:

$$\begin{bmatrix} 3 & -1 & -4 \\ -4 & -4 & 8 \\ -1 & 3 & -4 \end{bmatrix}$$

and we finally divide by the determinant of A:

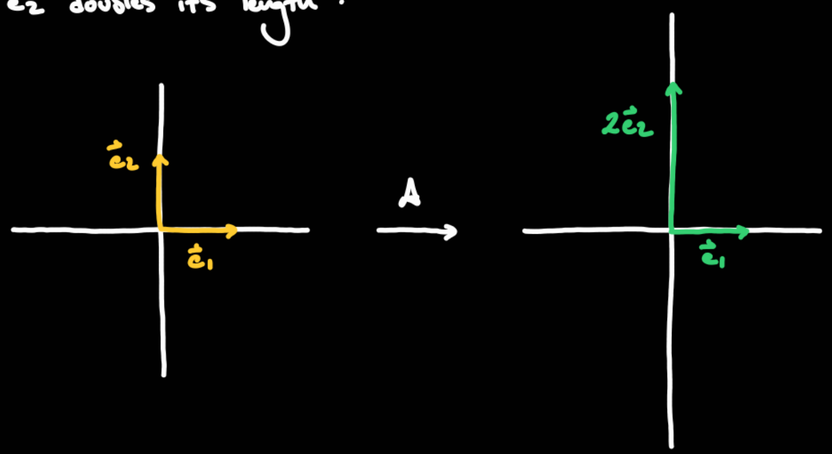
$$A^{-1} = \begin{bmatrix} -3/8 & 1/8 & 1/2 \\ 1/2 & 1/2 & -1 \\ 1/8 & -3/8 & 1/2 \end{bmatrix}$$ as we have already computed.

5. Eigenvalues and eigenvectors (Chapter 7)

Informally, the eigenvectors of a matrix are the "preferred directions" of its associated linear transformation, and the eigenvalues of a matrix are the scaling factors associated to these directions.

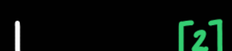
Examples:

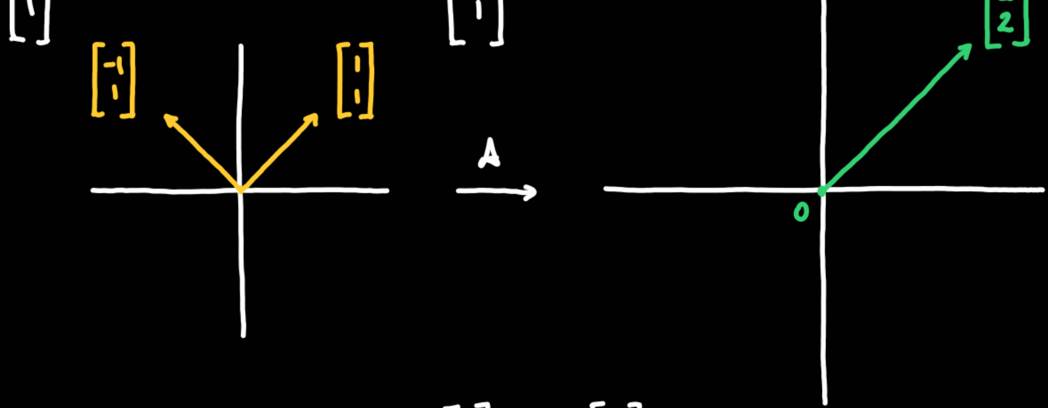
1. Let $A = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$, the associated linear transformation leaves $\vec{e}_1$ untouched, and $\vec{e}_2$ doubles its length:



The "prefered directions" are $\vec{e}_1$ and $\vec{e}_2$, with scaling factors 1 and 2 respectively.

2. Let $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$, the associated linear transformation doubles the length of $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and sends the vector $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$ to zero:





The "preferred directions" are $\begin{bmatrix}1\\1\end{bmatrix}$ and $\begin{bmatrix}-1\\1\end{bmatrix}$, with scaling factors 2 and 0 respectively.

We already knew that, since:

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

is the composition of the projection onto the line $y=x$ conveyed by $\frac{1}{2}\begin{bmatrix}1 & 1\\1 & 1\end{bmatrix}$ with a scaling factor of two conveyed by $\begin{bmatrix}2 & 0\\0 & 2\end{bmatrix}$. We are thus collapsing everything perpendicular to the line $y=x$, and we are doubling everything parallel to the line $y=x$.

3. Let $A = \begin{bmatrix}2/3 & 1/3 & -1/3\\1/3 & 2/3 & 1/3\\-1/3 & 1/3 & 2/3\end{bmatrix}$, the associated linear transformation leaves $\begin{bmatrix}1\\1\\0\end{bmatrix}$ and $\begin{bmatrix}0\\1\\1\end{bmatrix}$ untouched, and sends the vector $\begin{bmatrix}1\\-1\\1\end{bmatrix}$ to zero. The "preferred directions" are $\begin{bmatrix}1\\1\\0\end{bmatrix}$, $\begin{bmatrix}0\\1\\1\end{bmatrix}$, and $\begin{bmatrix}1\\-1\\1\end{bmatrix}$, with scaling factors of 1, 1, and 0 respectively. We already knew that, since A was the matrix associated to the orthogonal projection onto $V = \text{Span}\left(\begin{bmatrix}1\\1\\0\end{bmatrix}, \begin{bmatrix}0\\1\\1\end{bmatrix}\right)$ so it will keep every element in V untouched, and it will collapse everything perpendicular

to it. The line perpendicular to $\vec{v}$ is $L = \operatorname{span}\left(\begin{bmatrix} -1 \\ 1 \end{bmatrix}\right)$, so A collapses $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$.

Working with these "preferred directions" in mind, everything becomes easier.

Examples:

1. We are sending $\vec{e}_1$ and $\vec{e}_2$ to a scalar multiple of themselves, and using $\mathcal{E} = \{\vec{e}_1, \vec{e}_2\}$ as a basis we have that the matrix of the linear transformation given by $A = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$ is $B = \begin{bmatrix} [A(\vec{e}_1)]_{\mathcal{E}} & [A(\vec{e}_2)]_{\mathcal{E}} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$, which is diagonal.

2. We are sending $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$ to a scalar multiple of themselves, and using $\mathcal{B} = \left\{\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix}\right\}$ as a basis we have that the matrix of the linear transformation given by $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ is $B = \begin{bmatrix} \left[A\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right)\right]_{\mathcal{B}} & \left[A\left(\begin{bmatrix} -1 \\ 1 \end{bmatrix}\right)\right]_{\mathcal{B}} \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}$, which is diagonal.

3. We are sending $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$, and $\begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$ to a scalar multiple of themselves, and using $\mathcal{B} = \left\{\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}\right\}$ as a basis we have that the matrix of the linear transformation given by $A = \begin{bmatrix} 2/3 & 1/3 & -1/3 \\ 1/3 & 2/3 & 1/3 \\ -1/3 & 1/3 & 2/3 \end{bmatrix}$ is
$$B = \begin{bmatrix} \left[A\left(\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}\right)\right]_{\mathcal{B}} & \left[A\left(\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}\right)\right]_{\mathcal{B}} & \left[A\left(\begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}\right)\right]_{\mathcal{B}} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$ which is diagonal.

Unfortunally, when we are working over $\mathbb{R}$, there are matrices without "preferred directions".

Example: The matrix $A = \begin{bmatrix} 0 & -1 \\ \cdots \end{bmatrix}$ is encoding a counterclockwise rotation of $\frac{\pi}{2}$. This matrix does not have "preferred directions". In fact, if $\mathcal{B}$ is any basis of $\mathbb{R}^2$, then B the matrix of the counterclockwise rotation of $\frac{\pi}{2}$ in $\mathcal{B}$ is never diagonal, since we are never sending a vector in $\mathbb{R}^2$ to a scalar multiple of itself.

Side note: This existence of matrices without "preferred directions" is a problem arising from the fact that we are working over $\mathbb{R}$. When working over $\mathbb{C}$, every matrix has some "preferred direction".

Let A be an $n\times n$ matrix. A non-zero vector $\vec{v}$ in $\mathbb{R}^n$ is called an eigenvector of A if A sends $\vec{v}$ to a scalar multiple of itself, namely $A\vec{v} = \lambda\vec{v}$ for some λ in $\mathbb{R}$ (which may be zero). The scalar λ is called the eigenvalue associated with the eigenvector $\vec{v}$.

Examples:

1. $A = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$ has eigenvectors $\vec{e}_1$ and $\vec{e}_2$ with eigenvalues 1 and 2 respectively.
2. $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ has eigenvectors $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$ with eigenvalues 2 and 0 respectively.
3. $A = \begin{bmatrix} 2/3 & 1/3 & -1/3 \\ 1/3 & 2/3 & 1/3 \\ -1/3 & 1/3 & 2/3 \end{bmatrix}$ has eigenvectors $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$, and $\begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$ with eigenvalues 1, 1, and 0 respectively.
4. $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ does not have eigenvectors, and thus does not have eigenvalues.

Example: What are the possible (real) eigenvalues of an orthogonal matrix?

Let A be an orthogonal matrix, then A preserves length so $\|A\vec{x}\| = \|\vec{x}\|$ for all $\vec{x}$ in $\mathbb{R}^n$. Thus if $\vec{v}$ is an eigenvector of eigenvalue λ then $A\vec{v} = \lambda\vec{v}$ gives:

$$\|\vec{v}\| = \|A\vec{v}\| = \|\lambda\vec{v}\| = |\lambda|\,\|\vec{v}\| \quad \text{so} \quad |\lambda| = 1 \quad \text{so} \quad \lambda = \pm 1.$$

Thus the eigenvalues of an orthogonal matrix are ± 1.

Theorem: Let A be an $n\times n$ matrix having eigenvectors $\vec{v}_1, \dots, \vec{v}_n$ with eigenvalues $\lambda_1, \dots, \lambda_n$ respectively. If $\mathfrak{B} = \{\vec{v}_1, \dots, \vec{v}_n\}$ is a basis for $\mathbb{R}^n$ then:

$$A \text{ is similar to } B = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} \text{ via } S = \begin{bmatrix} | & & | \\ \vec{v}_1 & \cdots & \vec{v}_n \\ | & & | \end{bmatrix},$$

namely $B = S^{-1}AS$.

The converse is also true.

Theorem: Let A be an $n\times n$ matrix and $\mathfrak{B} = \{\vec{v}_1, \dots, \vec{v}_n\}$ a basis of $\mathbb{R}^n$ such that A is similar to a diagonal matrix $B = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$ via $S = \begin{bmatrix} | & & | \\ \vec{v}_1 & \cdots & \vec{v}_n \\ | & & | \end{bmatrix}$. Then A has eigenvectors $\vec{v}_1, \dots, \vec{v}_n$ with eigenvalues $\lambda_1, \dots, \lambda_n$ respectively.

Let A be an $n\times n$ matrix and λ a real scalar. Then λ is an eigenvalue of A if and only if $\det(A - \lambda I_n) = 0$. This equality is called the <u>characteristic equation</u> of A.

When we see λ as a variable, $\det(A - \lambda I_n)$ is a polynomial of degree n, called the characteristic polynomial of A and denoted by $f_A(\lambda)$.

Example: Find the eigenvalues of $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$. This is equivalent to finding the roots of the characteristic polynomial of A. Now:

$$f_A(\lambda) = \det(A - \lambda I_n) = \det \begin{bmatrix} 1-\lambda & 1 \\ 1 & 1-\lambda \end{bmatrix} = (1-\lambda)^2 - 1 = \lambda^2 + 1 - 2\lambda - 1 = \lambda(\lambda - 2)$$

and thus $f_A(\lambda) = 0$ if and only if $\lambda = 0$ or $\lambda = 2$. The eigenvalues of A are 0 and 2.

Let A be a square matrix of size $n \times n$, the sum of the diagonal entries of A is called the trace of A, denoted $tr(A)$.

Example: Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, then its characteristic polynomial is:

$$f_A(\lambda) = \det \begin{bmatrix} a-\lambda & b \\ c & d-\lambda \end{bmatrix} = (a-\lambda)(d-\lambda) - bc = \lambda^2 - (a+d)\lambda + (ad - bc) =$$

$$= \lambda^2 - tr(A)\lambda + \det(A).$$

Let A be a square matrix and λ_0 an eigenvalue of A. The algebraic multiplicity of λ_0 is the largest positive integer k for which $(\lambda - \lambda_0)^k$ is a factor of $f_A(\lambda)$, namely:

$$f_A(\lambda) = (\lambda - \lambda_0)^k \cdot g(\lambda)$$

for some polynomial $g(\lambda)$ with $g(\lambda_0) \neq 0$.

Example: Find the eigenvalues of $A = \begin{bmatrix} 2/3 & 1/3 & -1/3 \\ 1/3 & 2/3 & 1/3 \\ \end{bmatrix}$, as well as their respective

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

algebraic multiplicities. The characteristic polynomial of A is:

$$f_A(\lambda) = \det \begin{bmatrix} 2/3-\lambda & 1/3 & -1/3 \\ 1/3 & 2/3-\lambda & 1/3 \\ -1/3 & 1/3 & 2/3-\lambda \end{bmatrix} = -\lambda^3 + 2\lambda^2 - \lambda = -\lambda \cdot (\lambda-1)^2$$

which has roots $\lambda = 0$ and $\lambda = 1$. Thus A has eigenvalues 0 and 1 with algebraic multiplicities 1 and 2 respectively.

<u>Theorem:</u> An $n \times n$ matrix has at most n distinct real eigenvalues, counted with multiplicity. If n is odd then the matrix has at least one real eigenvalue. If n is even then the matrix need not have real eigenvalues.

<u>Example:</u> Find the eigenvalues of $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$. The characteristic polynomial of A is:

$$f_A(\lambda) = \det \begin{bmatrix} -\lambda & -1 \\ 1 & -\lambda \end{bmatrix} = \lambda^2 + 1$$, which does not have any real solutions. Thus A does not have real eigenvalues.

<u>Theorem:</u> Let A be an $n \times n$ matrix with eigenvalues $\lambda_1, \dots, \lambda_n$ listed with their algebraic multiplicities. Then:

$$\det(A) = \lambda_1 \cdots \lambda_n \qquad \text{and} \qquad tr(A) = \lambda_1 + \cdots + \lambda_n .$$

Now that we are proficient finding eigenvalues, we need to find their corresponding eigenvectors. These satisfy $A\vec{v} = \lambda \vec{v}$, or equivalently $(A - \lambda I_n)\vec{v} = 0$.

Let A be an $n \times n$ matrix with eigenvalue λ, the kernel of $(A - \lambda I_n)$ is called the

Let A be an $n\times n$ matrix with eigenvalue λ, the kernel of $(A-\lambda I_n)$ is called the

<u>eigenspace</u> of λ, and is denoted by E_λ:

$$E_\lambda = \ker(A-\lambda I_n) = \{\vec{v} \text{ in } \mathbb{R}^n \mid A\vec{v} = \lambda\vec{v}\}.$$

<u>Example:</u> Find the eigenspaces of $A = \begin{bmatrix} 2/3 & 1/3 & -1/3 \\ 1/3 & 2/3 & 1/3 \\ -1/3 & 1/3 & 2/3 \end{bmatrix}$. We know that the eigenvalues of A are $\lambda=0$ and $\lambda=1$. To find $E_0 = \ker(A)$ we have to solve $A\vec{x}=0$:

$$\left[\begin{array}{ccc|c} 2/3 & 1/3 & -1/3 & 0 \\ 1/3 & 2/3 & 1/3 & 0 \\ -1/3 & 1/3 & 2/3 & 0 \end{array}\right] \text{ reduces to } \left[\begin{array}{ccc|c} 1 & 0 & -1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \text{ so } \vec{x} = \begin{bmatrix} t \\ -t \\ t \end{bmatrix} = t\cdot\begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}.$$

Thus $E_0 = \operatorname{span}\left(\begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}\right)$ which is exactly the line perpendicular to $V = \operatorname{span}\left(\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}\right)$.

To find $E_1 = \ker(A-I_n)$ we have to solve $(A-I_n)\vec{x}=0$:

$$\left[\begin{array}{ccc|c} -1/3 & 1/3 & -1/3 & 0 \\ 1/3 & -1/3 & 1/3 & 0 \\ -1/3 & 1/3 & -1/3 & 0 \end{array}\right] \text{ reduces to } \left[\begin{array}{ccc|c} 1 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \text{ so } \vec{x} = \begin{bmatrix} t \\ t+s \\ s \end{bmatrix} = t\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + s\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

Thus $E_1 = \operatorname{span}\left(\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}\right) = V$ the plane onto which A is projecting.

Let A be an $n\times n$ matrix with λ as an eigenvalue. The <u>geometric multiplicity</u> of the eigenvalue λ is the dimension of the eigenspace $E_\lambda = \ker(A-\lambda I_n)$, denoted $\text{gemu}(\lambda)$.

<u>Remark:</u> The geometric multiplicity is easy to find since:

$$\text{gemu}(\lambda) = \dim\left(\ker(A-\lambda I_n)\right) = \text{nullity}(A-\lambda I_n) = n - \text{rank}(A-\lambda I_n).$$

<u>Remark:</u> Over the complex numbers $\mathbb{C}$, we always have as many eigenvalues as the size of

the matrix. This may not be the case over the real numbers. However, this does not mean that we always have as many linearly independent eigenvectors as the size of the matrix, regardless of whether we are working over $\mathbb{C}$ or $\mathbb{R}$.

We now embark on the search for basis of $\mathbb{R}^n$ such that all elements of the basis are eigenvectors of a given matrix.

Let A be an $n \times n$ matrix, an <u>eigenbasis</u> of $\mathbb{R}^n$ is a basis of $\mathbb{R}^n$ consisting of eigenvectors of A.

<u>Example:</u>

1. $A = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$ has eigenbasis $\mathfrak{B} = \{\vec{e}_1, \vec{e}_2\}$.
2. $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ has eigenbasis $\mathfrak{B} = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\}$.
3. $A = \begin{bmatrix} 2/3 & 1/3 & -1/3 \\ 1/3 & 2/3 & 1/3 \\ -1/3 & 1/3 & 2/3 \end{bmatrix}$ has eigenbasis $\mathfrak{B} = \left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \right\}$.
4. $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ does not have an eigenbasis.

<u>Remark:</u> Let A be an $n \times n$ matrix.

(a) If we find a basis of each eigenspace of A and concentenate all these bases, we obtain the eigenvectors $\vec{v}_1, \dots, \vec{v}_s$ with s the sum of the geometric multiplicities of the eigenvalues of A.

(b) The vectors $\vec{v}_1, \dots, \vec{v}_s$ are linearly independent, and thus $s \leq n$.

(c) There is an eigenbasis of A if and only if $s = n$, that is, the sum of the geometric multiplicities add up to the size of the matrix.

<u>Theorem:</u> Let A be an $n \times n$ matrix with n distinct eigenvalues. Then there is an eigenbases of A. We can construct an eigenbasis by finding an eigenvector for each eigenvalue.

We have seen that similar matrices are strongly related. This somewhat extends to eigenbasis.

<u>Theorem:</u> Let A and B be similar matrices. Then:

(a) The characteristic polynomial of A and B coincide: $f_A(\lambda) = f_B(\lambda)$.

(b) The rank and nullity of A and B coincide:

$$\text{rank}(A) = \text{rank}(B) \quad \text{and} \quad \text{nullity}(A) = \text{nullity}(B).$$

(c) The eigenvalues of A and B coincide. The algebraic multiplicities of those eigenvalues are the same for A and B. The geometric multiplicities of those eigenvalues are the same for A and B.

(d) The determinant and the trace of A and B coincide:

$$\det(A) = \det(B) \quad \text{and} \quad \text{tr}(A) = \text{tr}(B).$$

We now illustrate how the geometric and algebraic multiplicities are, in general, different.

Example: Find the eigenvalues and their algebraic and geometric multiplicities of $A = \begin{bmatrix} 8 & -9 \\ 4 & -4 \end{bmatrix}$.

The characteristic polynomial is:

$$f_A(\lambda) = (8-\lambda)(-4-\lambda) + 4\cdot 9 = \lambda^2 - 4\lambda + 4 = (\lambda-2)^2$$

so A has one eigenvalue, of algebraic multiplicity 2. Its geometric multiplicity is:

$\text{gemu}(2) = 2 - \text{rank}(A - 2I_2)$ and since $A - 2I_2 = \begin{bmatrix} 6 & -9 \\ 4 & -6 \end{bmatrix}$

has $\text{rref}(A-2I_2) = \begin{bmatrix} 1 & -3/2 \\ 0 & 0 \end{bmatrix}$ then $\text{rank}(A-2I_2) = 1$ and:

$$\text{gemu}(2) = 2 - \text{rank}(A - 2I_2) = 2 - 1 = 1.$$

Theorem: Let A be a square matrix with eigenvalue λ. Then $\text{gemu}(\lambda) \leq \text{almu}(\lambda)$.

Method to find eigenbasis of matrices.

Let A be an $n\times n$ matrix.

1. Find the eigenvalues of A by soving the characteristic equation: $f_A(\lambda) = 0$.

2. For each eigenvalue λ, find a basis of the eigenspace $E_\lambda = \ker(A - \lambda I_n)$.

3. The matrix A has an eigenbases if and only if the dimensions of the eigenspaces add up to n. In this case, concatenating the bases of the

eigenspaces gives $\vec{v}_1, \dots, \vec{v}_n$ an eigenbases of A. Moreover:

$$S^{-1}AS = B = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} \quad \text{where} \quad S = \begin{bmatrix} | & & | \\ \vec{v}_1 & \cdots & \vec{v}_n \\ | & & | \end{bmatrix}$$

and $\lambda_1, \dots, \lambda_n$ are the eigenvalues of $\vec{v}_1, \dots, \vec{v}_n$.

__Example:__ Find an eigenbasis and a diagonal matrix similar to:

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

The characteristic polynomial is:

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

so A has eigenvalues $\lambda = 1$ and $\lambda = 2$. Since these are different, A will be similar to the matrix $B = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$. To find $E_1 = \ker(A - I_2)$ we solve:

$$(A - I_2)\vec{x} = \vec{0}, \text{ namely } \begin{bmatrix} 0 & 3 \\ 0 & 1 \end{bmatrix} \vec{x} = \vec{0} \quad \text{so} \quad \vec{x} = \begin{bmatrix} t \\ 0 \end{bmatrix} = t \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

To find $E_2 = \ker(A - 2I_2)$ we solve:

$$(A - 2I_2)\vec{x} = \vec{0}, \text{ namely } \begin{bmatrix} -1 & 3 \\ 0 & 0 \end{bmatrix} \vec{x} = \vec{0} \quad \text{so} \quad \vec{x} = \begin{bmatrix} 3t \\ t \end{bmatrix} = t \cdot \begin{bmatrix} 3 \\ 1 \end{bmatrix}.$$

An eigenbasis of A is $\mathfrak{B} = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \end{bmatrix} \right\}$.

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

The characteristic polynomial is:

$$f_A(\lambda) = (1-\lambda)(1-\lambda) = \lambda^2 - 2\lambda + 1$$

so A has eigenvalue $\lambda=1$. We now check the geometric multiplicity of $\lambda=1$.

To find $E_1 = \ker(A - I_2)$ we solve:

$$(A-I_2)\vec{x}=\vec{0}, \text{ namely } \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\vec{x}=\vec{0} \quad \text{so } \vec{x}=\begin{bmatrix} t \\ 0 \end{bmatrix}=t\cdot\begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

Hence E_1 has dimension 1, so $\lambda=1$ has geometric multiplicity 1. Since A is a 2×2 matrix but the sum of its geometric multiplicities is not 2, A does not have any eigenbasis.

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

The characteristic polynomial is:

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

so A has eigenvalues $\lambda=0$ and $\lambda=1$. To find $E_0=\ker(A)$ we solve:

$$A\vec{x}=\vec{0}, \text{ namely } \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}\vec{x}=\vec{0} \quad \text{so } \vec{x}=\begin{bmatrix} -t-s \\ t \\ s \end{bmatrix}=t\cdot\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}+s\cdot\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}.$$

To find $E_1=\ker(A-I_3)$ we solve:

$$(A-I_3)\vec{x}=\vec{0}, \text{ namely } \begin{bmatrix} 0 & 1 & 1 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}\vec{x}=\vec{0} \quad \text{so } \vec{x}=\begin{bmatrix} t \\ 0 \\ 0 \end{bmatrix}=t\cdot\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}.$$

Thus E_0 has dimension 2 so $\lambda=0$ has geometric multiplicity 2, and E_1 has dimension 1 so $\lambda=1$ has geometric multiplicity 1. Since A is a 3×3 matrix and the sum of its geometric multiplicities is 3, then it is similar to

$$B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \text{ and } A \text{ has eigenbasis } \mathfrak{B} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

<u>Example:</u> Find the values a, b, c for which the matrix $A = \begin{bmatrix} 1 & a & b \\ 0 & 0 & c \\ 0 & 0 & 1 \end{bmatrix}$ is diagonalizable.

The characteristic polynomial is:

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

so A has eigenvalues $\lambda = 0$ and $\lambda = 1$. Since each eigenvalue is associated to at least one eigenvector, the eigenspace E_0 will have at least dimension 1. Since the algebraic multiplicity of $\lambda = 0$ is 1, the geometric multiplicity of 1 is exactly 1. To find the geometric multiplicity of $\lambda = 1$ we compute:

$$\text{gemu}(1) = 3 - \text{rank}(A - I_3).$$

Now $A - I_3 = \begin{bmatrix} 0 & a & b \\ 0 & -1 & c \\ 0 & 0 & 0 \end{bmatrix}$. $A - I_3$ has rank 2 when $\begin{bmatrix} a & b \\ -1 & c \end{bmatrix}$ has linearly independent columns, or equivalently when $\begin{bmatrix} a & b \\ -1 & c \end{bmatrix}$ is invertible. $A - I_3$ will never have rank 0 since it has a non-zero entry. $A - I_3$ will have rank 1 when $\begin{bmatrix} a & b \\ -1 & c \end{bmatrix}$ has linearly dependent columns, or equivalently when invertible.

For the matrix A to be diagonalizable we need that the sums of its geometric multiplicities is 3. Since $\operatorname{gemu}(0)=1$, we need $\operatorname{gemu}(1)=2$, and thus we need

$\operatorname{rank}(A-I_3)=1$. This happens if and only if $\begin{bmatrix} a & b \\ -1 & c \end{bmatrix}$ is not invertible, which happens if and only if its determinant is zero.

Thus A is diagonalizable if and only if $ac+b=0$.

<u>Example:</u> Let $A=\begin{bmatrix} 1 & 2 \\ 4 & 3 \end{bmatrix}$, find A^3, A^5, A^{100}.

The characteristic polynomial is $f_A(\lambda)=\lambda^2-4\lambda+5=(\lambda+1)(\lambda-5)$ so $\lambda=-1$ and $\lambda=5$ are the eigenvalues of A. Since they are distinct, A is diagonalizable. We can find that $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ are eigenvectors of eigenvalues -1 and 5 respectively. Now:

$$\begin{bmatrix} 1 & 2 \\ 4 & 3 \end{bmatrix}=A=SBS^{-1}=\begin{bmatrix} -1 & 1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} -1 & 0 \\ 0 & 5 \end{bmatrix}\begin{bmatrix} -2/3 & 1/3 \\ 1/3 & 1/3 \end{bmatrix}$$

So:

$$A^3=(SBS^{-1})(SBS^{-1})(SBS^{-1})=\begin{bmatrix} -1 & 1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} -1 & 0 \\ 0 & 5 \end{bmatrix}^3\begin{bmatrix} -2/3 & 1/3 \\ 1/3 & 1/3 \end{bmatrix},$$

$$A^5=\begin{bmatrix} -1 & 1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} -1 & 0 \\ 0 & 5 \end{bmatrix}^5\begin{bmatrix} -2/3 & 1/3 \\ 1/3 & 1/3 \end{bmatrix}, \text{ and } A^{100}=\begin{bmatrix} -1 & 1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} -1 & 0 \\ 0 & 5 \end{bmatrix}^{100}\begin{bmatrix} -2/3 & 1/3 \\ 1/3 & 1/3 \end{bmatrix}.$$

<u>Remark:</u> We can use these procedures over the complex numbers $\mathbb{C}$.

1. Let a,b be real numbers, b not zero. Then the matrix $A=\begin{bmatrix} a & -b \\ b & a \end{bmatrix}$ can be diagonalized over the complex numbers:

$$\begin{bmatrix} -i/2 & 1/2 \\ i/2 & 1/2 \end{bmatrix}\begin{bmatrix} a & -b \\ b & a \end{bmatrix}\begin{bmatrix} i & -i \\ 1 & 1 \end{bmatrix}=\begin{bmatrix} a+ib & 0 \\ 0 & a-ib \end{bmatrix}$$

so $A = RDR^{-1}$ with $R = \begin{bmatrix} i & -i \\ 1 & 1 \end{bmatrix}$ and $D = \begin{bmatrix} a+ib & 0 \\ 0 & a-ib \end{bmatrix}$.

2. Let A be a real 2×2 matrix with eigenvalues $a+ib$ and $a-ib$ with $b \neq 0$.

   Then A is similar, over the real numbers $\mathbb{R}$, to the matrix $\begin{bmatrix} a & -b \\ b & a \end{bmatrix}$. That is, there is a real 2×2 matrix S such that:

$$A = S \begin{bmatrix} a & -b \\ b & a \end{bmatrix} S^{-1}.$$

All of this relies heavily on the Fundamental Theorem of Algebra:

Theorem: Let $f(x)$ be a polynomial of degree n with coefficients in $\mathbb{C}$. Then $f(x)$ has n roots, counted with multiplicity.

So in particular, an $n\times n$ matrix with complex coefficients has n complex eigenvalues, when counted with multiplicity.

6. Symmetric matrices and quadratic forms. (Chapter 8)

We have been considering when matrices have eigenbasis. We will now be more demanding, and we will consider when matrices have orthonormal eigenbasis. When we achieve this, our eigenbasis will still give us a change of basis matrix S between our original matrix A and a diagonal matrix B: $B = S^{-1}AS$. Since S has orthonormal columns, then $S^{-1} = S^T$ and the above equality becomes: $B = S^T A S$.

We say that a matrix is <u>diagonalizable</u> if it has an eigenbasis.

We say that a matrix is <u>orthogonally diagonalizable</u> if it has an orthonormal eigenbasis.

<u>Theorem:</u> (Spectral Theorem) A matrix is orthogonally diagonalizable if and only if it is symmetric.

The proof is hard! One direction is very easy: if a matrix is orthogonally diagonalizable then it is symmetric. Let A be orthogonally diagonalizable, then $A = SDS^T$ for S orthogonal and D diagonal, so $A^T = (SDS^T)^T = (S^T)^T D^T S^T = SDS^T = A$, so A is symmetric. However, the converse relies on the fact that eigenvectors of a symmetric matrix having different eigenvalues are orthogonal (so, in particular, they are linearly independent), and also on some tools from complex numbers.

<u>Theorem:</u> Let A be a symmetric matrix, $\vec{v}_1$ and $\vec{v}_2$ eigenvectors of A with distinct eigenvalues λ_1 and λ_2. Then $\vec{v}_1 \cdot \vec{v}_2 = 0$.

<u>Proof:</u> We compute $\vec{v}_1^{\,T} A \vec{v}_2$ in two ways:

$$\vec{v}_1^{\,T} A \vec{v}_2 = \vec{v}_1^{\,T}(A\vec{v}_2) = \vec{v}_1^{\,T}(\lambda_2 \vec{v}_2) = \lambda_2 \vec{v}_1^{\,T}\vec{v}_2 = \lambda_2 \cdot (\vec{v}_1 \cdot \vec{v}_2)$$

$$\vec{v}_1^{\,T} A \vec{v}_2 = (\vec{v}_1^{\,T} A)\vec{v}_2 = (\vec{v}_1^{\,T} A^T)\vec{v}_2 = (A\vec{v}_1)^T \vec{v}_2 = (\lambda_1 \vec{v}_1)^T \vec{v}_2 = \lambda_1 \vec{v}_1^{\,T}\vec{v}_2 = \lambda_1 \cdot (\vec{v}_1 \cdot \vec{v}_2)$$

So:

$$\lambda_2 \cdot (\vec{v}_1 \cdot \vec{v}_2) = \lambda_1 \cdot (\vec{v}_1 \cdot \vec{v}_2) \qquad \text{so} \qquad (\lambda_2 - \lambda_1) \cdot (\vec{v}_1 \cdot \vec{v}_2) = 0$$

and thus $\vec{v}_1 \cdot \vec{v}_2 = 0$ since $\lambda_2 \neq \lambda_1$. $\square$.

<u>Example:</u> Find an orthonormal eigenbasis of $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$. We already know that A has an eigenbasis $\mathfrak{B} = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\}$. Since these vectors have different eigenvalues, they are orthogonal, so to make $\mathfrak{B}$ orthonormal we only have to make them of length one.

Thus $\vec{u}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\vec{u}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ form an orthonormal eigenbasis of A.

<u>Example:</u> Find an orthonormal eigenbasis of $A = \begin{bmatrix} 2/3 & 1/3 & -1/3 \\ 1/3 & 2/3 & 1/3 \\ -1/3 & 1/3 & 2/3 \end{bmatrix}$. We know that A has an eigenbasis $\mathfrak{B} = \left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \right\}$, with eigenvalues $1, 1, 0$ respectively. Thus the third vector is orthogonal to the other two because they have different eigenvalues. We can then find an orthonormal basis by making the first two vectors into an orthonormal basis, and dividing the last vector by its length. Setting $\mathfrak{B} = \{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$:

$$\vec{u}_1 = \frac{\vec{v}_1}{\|\vec{v}_1\|} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad \text{spanning the line } L_1 = \mathrm{span}(\vec{u}_1)$$

$$\vec{v}_2^{\parallel} = \mathrm{proj}_{L_1}(\vec{v}_2) = (\vec{v}_2 \cdot \vec{u}_1)\,\vec{u}_1 = \left(\begin{bmatrix} 0 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix} \right) \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$$

$$\vec{v}_2^{\perp} = \vec{v}_2 - \vec{v}_2^{\parallel} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} -1/2 \\ 1/2 \\ 1 \end{bmatrix}$$

$$\vec{u}_2 = \frac{\vec{v}_2}{\|\vec{v}_2^{\perp}\|} = \frac{2}{\sqrt{6}}\begin{bmatrix} -1/2 \\ 1/2 \\ 1 \end{bmatrix} = \frac{1}{\sqrt{6}}\begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix}$$

$$\vec{u}_3 = \frac{\vec{v}_3}{\|\vec{v}_3\|} = \frac{1}{\sqrt{3}}\begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$$

And $\mathfrak{A} = \left\{ \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \frac{1}{\sqrt{6}}\begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix}, \frac{1}{\sqrt{3}}\begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \right\}$ is an orthonormal eigenbasis of A.

<u>Method to find orthonormal eigenbasis of symmetric matrices.</u>

Let A be an $n \times n$ symmetric matrix.

1. Find the eigenvalues of A and a basis for each eigenspace.

2. Find an orthonormal basis for each eigenspace (using the Gram-Schmidt process).

3. Concatenate the basis of the eigenspaces into $\vec{v}_1, \dots, \vec{v}_n$ an orthogonal eigenbasis.

Now:

$$\begin{bmatrix} - \vec{v}_1 - \\ \vdots \\ - \vec{v}_n - \end{bmatrix} A \begin{bmatrix} | & & | \\ \vec{v}_1 & \cdots & \vec{v}_n \\ | & & | \end{bmatrix} = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$$

where $\lambda_1, \dots, \lambda_n$ are the eigenvalues of $\vec{v}_1, \dots, \vec{v}_n$.

A function $q(x_1, \dots, x_n)$ from $\mathbb{R}^n$ to $\mathbb{R}$ is called a <u>quadratic form</u> if it is a linear combination of the products $x_i x_j$ for i and j between 1 and n. A quadratic form can be written as $q(\vec{x}) = \vec{x}^T A \vec{x}$ for a unique symmetric matrix A, called the associated matrix to q.

<u>Example:</u> Consider the function:

$$q(x_1,x_2,x_3)=\frac{2}{3}(x_1^2+x_2^2+x_3^2+x_1x_2-x_1x_3+x_2x_3),$$

is it a quadratic form? Yes, since all the terms are monomials of degree two. Moreover, its associated matrix is the 3×3 matrix with entries:

a_{ii} given by the coefficient of x_i^2,

$a_{ij}=a_{ji}$ given by half the coefficient of x_ix_j for $i\neq j$.

Thus:

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

and indeed:

$$q(\vec{x})=\vec{x}^T A\vec{x}=\begin{bmatrix} x_1 & x_2 & x_3\end{bmatrix}\begin{bmatrix} 2/3 & 1/3 & -1/3 \\ 1/3 & 2/3 & 1/3 \\ -1/3 & 1/3 & 2/3 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3\end{bmatrix}=\frac{2}{3}(x_1^2+x_2^2+x_3^2+x_1x_2-x_1x_3+x_2x_3).$$

Since the associated matrix A of a quadratic form q is symmetric, by the Spectral Theorem we can find an orthonormal eigenbasis $\mathcal{B}=\{\vec{v}_1,\dots,\vec{v}_n\}$ of A. We can then rewrite $\vec{x}=c_1\vec{v}_1+\dots+c_n\vec{v}_n$ in this basis, and now:

$$q(\vec{x})=\vec{x}^T A\vec{x}=(c_1\vec{v}_1+\dots+c_n\vec{v}_n)^T A(c_1\vec{v}_1+\dots+c_n\vec{v}_n)=$$

$$=(c_1\vec{v}_1+\dots+c_n\vec{v}_n)^T(c_1\lambda_1\vec{v}_1+\dots+c_n\lambda_n\vec{v}_n)=(c_1\vec{v}_1+\dots+c_n\vec{v}_n)\cdot(c_1\lambda_1\vec{v}_1+\dots+c_n\lambda_n\vec{v}_n)=$$

$$=\lambda_1c_1^2+\dots+\lambda_nc_n^2.$$

<u>Example:</u> Consider the quadratic form $q(x_1,x_2,x_3)=\frac{2}{3}(x_1^2+x_2^2+x_3^2+x_1x_2-x_1x_3+x_2x_3)$.

Determine whether $x_1 = x_2 = x_3 = 0$ is a maximum or minimum or neither, and whether

it is local or global or neither.

We know that the symmetric matrix associated to q is $A = \begin{bmatrix} 2/3 & 1/3 & -1/3 \\ 1/3 & 2/3 & 1/3 \\ -1/3 & 1/3 & 2/3 \end{bmatrix}$, and

we know that this matrix has eigenvalues 1, 1, 0. We can then rewrite

$q(\vec{x}) = c_1^2 + c_2^2$ which is always positive or zero.

Thus $x_1 = x_2 = x_3 = 0$ is the global minimum of q.

Let $q(\vec{x}) = \vec{x}^T A \vec{x}$ be a quadratic form, with A a symmetric $n \times n$ matrix. We say that A is <u>positive definite</u> if $q(\vec{x})$ is positive for all non-zero $\vec{x}$ in $\mathbb{R}^n$. We say that A is <u>positive semidefinite</u> if $q(\vec{x}) \geq 0$ for all $\vec{x}$ in $\mathbb{R}^n$. We say that A is <u>indefinite</u> if q takes positive and negative values.

<u>Theorem:</u> A symmetric matrix is positive definite if and only if all of its eigenvalues are positive. A symmetric matrix is positive semidefinite if and only if all of its eigenvalues are positive or zero.

<u>Example:</u> Consider the quadratic form $q(x_1, x_2, x_3) = \frac{2}{3}(x_1^2 + x_2^2 + x_3^2 + x_1x_2 - x_1x_3 + x_2x_3)$.

Determine whether the associated symmetric matrix is positive definite, positive semidefinite, or indefinite.

Since $A = \begin{bmatrix} 2/3 & 1/3 & -1/3 \\ 1/3 & 2/3 & 1/3 \\ -1/3 & 1/3 & 2/3 \end{bmatrix}$ has eigenvalues 1, 1, 0, it is positive semidefinite.

We already knew that, since every vector in the line $L = \operatorname{span}\left(\begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}\right)$ is in the kernel of A, and thus is sent to zero.

Let A be a symmetric $n \times n$ matrix. For $i = 1, \dots, n$ set $A^{(i)}$ the $i \times i$ matrix obtained from A by deleting all rows and columns after the i-th ones. The matrices $A^{(i)}$ are called the principal submatrices of A.

Theorem: A symmetric matrix is positive definite if and only if all its principal submatrices have positive determinant.

Example: Consider the quadratic form $q(x_1, x_2, x_3) = \frac{2}{3}(x_1^2 + x_2^2 + x_3^2 + x_1x_2 - x_1x_3 + x_2x_3)$.

Determine whether the associated symmetric matrix is positive definite.

Since $A = \begin{bmatrix} 2/3 & 1/3 & -1/3 \\ 1/3 & 2/3 & 1/3 \\ -1/3 & 1/3 & 2/3 \end{bmatrix}$ satisfies:

$$\det(A^{(1)}) = \det\left(\tfrac{2}{3}\right) = \tfrac{2}{3} > 0,$$

$$\det(A^{(2)}) = \det\begin{bmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} \end{bmatrix} = \frac{1}{3} > 0,$$

$$\det(A^{(3)}) = \det\begin{bmatrix} 2/3 & 1/3 & -1/3 \\ 1/3 & 2/3 & 1/3 \\ -1/3 & 1/3 & 2/3 \end{bmatrix} = 0,$$

and these determinants are not all positive, then A is not positive definite.

As another consequence of the Spectral Theorem we have the existence of orthogonal vectors whose images are also orthogonal, and this will hold for all linear transformations.

<u>Example:</u> Consider the linear transformation given by $L(\vec{x}) = A\vec{x}$ with $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$.

(a) Find an orthonormal basis $\mathbb{B} = \{\vec{v}_1, \vec{v}_2\}$ of $\mathbb{R}^2$ such that $L(\vec{v}_1)$ and $L(\vec{v}_2)$ are orthogonal.

To apply the Spectral Theorem, we consider the symmetric matrix A^TA:

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

which, because of the Spectral Theorem, has an orthonormal eigenbasis. To find it, we begin with an eigenbasis:

$$\vec{v}_1 = \begin{bmatrix} \frac{\sqrt{5}}{2} - \frac{1}{2} \\ 1 \end{bmatrix}, \ \lambda_1 = \frac{3}{2} + \frac{\sqrt{5}}{2} \quad \text{and} \quad \vec{v}_2 = \begin{bmatrix} \frac{-\sqrt{5}}{2} - \frac{1}{2} \\ 1 \end{bmatrix}, \ \lambda_2 = \frac{3}{2} - \frac{\sqrt{5}}{2},$$

and since the eigenvalues are different, we just need to normalize both:

$$\vec{u}_1 = \begin{bmatrix} \sqrt{5-\sqrt{5}}/\sqrt{10} \\ \sqrt{2}/\sqrt{5-\sqrt{5}} \end{bmatrix}, \quad \vec{u}_2 = \begin{bmatrix} -\sqrt{5+\sqrt{5}}/\sqrt{10} \\ \sqrt{2}/\sqrt{5+\sqrt{5}} \end{bmatrix}$$

form an orthonormal eigenbasis of A^TA.

Now $L(\vec{u}_1)$ and $L(\vec{u}_2)$ are orthogonal:

$$L(\vec{u}_1) = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} \sqrt{5-\sqrt{5}}/\sqrt{10} \\ \sqrt{2}/\sqrt{5-\sqrt{5}} \end{bmatrix} = \begin{bmatrix} \sqrt{1+\frac{2}{\sqrt{5}}} \\ \sqrt{2}/\sqrt{5-\sqrt{5}} \end{bmatrix},$$

$$L(\vec{u}_2) = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} -\sqrt{5+\sqrt{5}}/\sqrt{10} \\ \ldots \end{bmatrix} = \begin{bmatrix} -\sqrt{1-\frac{2}{\sqrt{5}}} \\ \ldots \end{bmatrix}.$$

$$\begin{bmatrix} 0 & 1 \end{bmatrix}\begin{bmatrix} \sqrt{2}/\sqrt{5+\sqrt{5}} \end{bmatrix} \quad \begin{bmatrix} \sqrt{2}/\sqrt{5+\sqrt{5}} \end{bmatrix}$$

We could have checked this beforehand:

$$L(\vec{u}_1)\cdot L(\vec{u}_2) = (A\vec{u}_1)\cdot(A\vec{u}_2) = (A\vec{u}_1)^T(A\vec{u}_2) = \vec{u}_1^T A^T A\,\vec{u}_2 = \vec{u}_1^T(A^TA)\,\vec{u}_2 =$$

$$= \vec{u}_1^T \lambda_2 \vec{u}_2 = \lambda_2\, \vec{u}_1^T\vec{u}_2 = \lambda_2(\vec{u}_1\cdot\vec{u}_2) = 0.$$

Here, we do not need that $\vec{u}_1$ and $\vec{u}_2$ are eigenvectors of A.

(b) Find the lengths of $L(\vec{u}_1)$ and $L(\vec{u}_2)$.

We have:

$$\|L(\vec{u}_1)\|^2 = \begin{bmatrix} \sqrt{1+\frac{2}{\sqrt{5}}} \\ \sqrt{2}/\sqrt{5-\sqrt{5}} \end{bmatrix}\cdot\begin{bmatrix} \sqrt{1+\frac{2}{\sqrt{5}}} \\ \sqrt{2}/\sqrt{5-\sqrt{5}} \end{bmatrix} = \frac{3}{2}+\frac{\sqrt{5}}{2} = \lambda_1,$$

$$\|L(\vec{u}_2)\|^2 = \begin{bmatrix} -\sqrt{1-\frac{2}{\sqrt{5}}} \\ \sqrt{2}/\sqrt{5+\sqrt{5}} \end{bmatrix}\cdot\begin{bmatrix} -\sqrt{1-\frac{2}{\sqrt{5}}} \\ \sqrt{2}/\sqrt{5+\sqrt{5}} \end{bmatrix} = \frac{3}{2}-\frac{\sqrt{5}}{2} = \lambda_2.$$

We could also have checked this beforehand:

$$\|L(\vec{u}_1)\|^2 = (A\vec{u}_1)\cdot(A\vec{u}_1) = (A\vec{u}_1)^T(A\vec{u}_1) = \vec{u}_1^T A^T A\,\vec{u}_1 = \lambda_1(\vec{u}_1\cdot\vec{u}_1) = \lambda_1,$$

$$\|L(\vec{u}_2)\|^2 = (A\vec{u}_2)\cdot(A\vec{u}_2) = (A\vec{u}_2)^T(A\vec{u}_2) = \vec{u}_2^T A^T A\,\vec{u}_2 = \lambda_2(\vec{u}_2\cdot\vec{u}_2) = \lambda_2.$$

The <u>singular values</u> of an $n\times m$ matrix A are the square roots of the eigenvalues of the symmetric $m\times m$ matrix A^TA, listed with their algebraic multiplicities. We denote them by $\sigma_1,\dots,\sigma_m$, listed in decreasing order: $\sigma_1 \geq \cdots \geq \sigma_m \geq 0$.

<u>Theorem</u>: Let $L(\vec{x}) = A\vec{x}$ be a linear transformation from $\mathbb{R}^m$ to $\mathbb{R}^n$. Then there exists an orthonormal basis $\vec{v}_1,\dots,\vec{v}_m$ of $\mathbb{R}^m$ such that $L(\vec{v}_1),\dots,L(\vec{v}_m)$ are

orthogonal with lengths the singular values $\sigma_1, \dots, \sigma_m$ of A.

To construct $\vec{v}_1, \dots, \vec{v}_m$ we find an orthonormal eigenbasis for the symmetric matrix A^TA, making sure that the corresponding eigenvalues $\lambda_1, \dots, \lambda_m$ appear in descending order.

<u>Theorem:</u> Let A be an $n\times m$ matrix, $r = \text{rank}(A)$. We can write $A = U\Sigma V^T$ where U is an orthogonal $n\times n$ matrix, V is an orthogonal $m\times m$ matrix, and Σ is an $n\times m$ matrix whose first r diagonal entries are the non-zero singular values $\sigma_1, \dots, \sigma_r$ of A and all other entries are zero. This is called the <u>singular value decomposition</u> of A.

To construct U and V, we first find $\vec{v}_1, \dots, \vec{v}_m$ an orthonormal eigenbasis for the symmetric matrix A^TA and the singular values $\sigma_1, \dots, \sigma_m$ of A. Since $r = \text{rank}(A)$, then $\sigma_1, \dots, \sigma_r$ are non-zero and $\sigma_{r+1}, \dots, \sigma_m$ are zero. Set $\vec{u}_1 = \frac{1}{\sigma_1}A\vec{v}_1, \dots, \vec{u}_r = \frac{1}{\sigma_r}A\vec{v}_r$, and expand this to $\vec{u}_1, \dots, \vec{u}_r, \vec{u}_{r+1}, \dots, \vec{u}_n$ a basis of $\mathbb{R}^n$. Now:

$$U = \begin{bmatrix} | & & | \\ \vec{u}_1 & \cdots & \vec{u}_n \\ | & & | \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \sigma_1 & & & & 0 \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & 0 & \\ 0 & & & & \ddots & 0 \end{bmatrix}, \quad V = \begin{bmatrix} | & & | \\ \vec{v}_1 & \cdots & \vec{v}_m \\ | & & | \end{bmatrix}.$$

<u>Example:</u> Find the singular value decomposition of $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. We already found:

$$\vec{v}_1 = \begin{bmatrix} \sqrt{5-\sqrt{5}}/\sqrt{10} \\ \sqrt{2}/\sqrt{5-\sqrt{5}} \end{bmatrix}, \quad \vec{v}_2 = \begin{bmatrix} -\sqrt{5+\sqrt{5}}/\sqrt{10} \\ \sqrt{2}/\sqrt{5+\sqrt{5}} \end{bmatrix}$$

an orthonormal eigenbasis of A^TA with eigenvalues:

$$\lambda_1 = \frac{3}{2} + \frac{\sqrt5}{2} \text{ and } \lambda_2 = \frac{3}{2} - \frac{\sqrt5}{2}, \text{ so } \sigma_1 = \frac{\sqrt5}{2} + \frac{1}{2} \text{ and } \sigma_2 = \frac{\sqrt5}{2} - \frac{1}{2}.$$

Now:

$$\vec{u}_1 = \frac{1}{\sigma_1} A\vec{v}_1 = \frac{1}{\frac{\sqrt5}{2}+\frac12}\begin{bmatrix}\sqrt{1+\frac{2}{\sqrt5}} \\ \sqrt2/\sqrt{5-\sqrt5}\end{bmatrix} = \begin{bmatrix}\sqrt{5+\sqrt5}/\sqrt{10} \\ \sqrt{5-\sqrt5}/\sqrt{10}\end{bmatrix},$$

$$\vec{u}_2 = \frac{1}{\sigma_2} A\vec{v}_2 = \frac{1}{\frac{\sqrt5}{2}-\frac12}\begin{bmatrix}-\sqrt{1-\frac{2}{\sqrt5}} \\ \sqrt2/\sqrt{5+\sqrt5}\end{bmatrix} = \begin{bmatrix}-\sqrt{5-\sqrt5}/\sqrt{10} \\ \sqrt{5+\sqrt5}/\sqrt{10}\end{bmatrix},$$

so:

$$U = \begin{bmatrix}\sqrt{5+\sqrt5}/\sqrt{10} & -\sqrt{5-\sqrt5}/\sqrt{10} \\ \sqrt{5-\sqrt5}/\sqrt{10} & \sqrt{5+\sqrt5}/\sqrt{10}\end{bmatrix},$$

$$\Sigma = \begin{bmatrix}\frac{\sqrt5}{2}+\frac12 & 0 \\ 0 & \frac{\sqrt5}{2}-\frac12\end{bmatrix},$$

$$V = \begin{bmatrix}\sqrt{5-\sqrt5}/\sqrt{10} & -\sqrt{5+\sqrt5}/\sqrt{10} \\ \sqrt2/\sqrt{5-\sqrt5} & \sqrt2/\sqrt{5+\sqrt5}\end{bmatrix},$$

and indeed:

$$U\Sigma V^T = \begin{bmatrix}\sqrt{5+\sqrt5}/\sqrt{10} & -\sqrt{5-\sqrt5}/\sqrt{10} \\ \sqrt{5-\sqrt5}/\sqrt{10} & \sqrt{5+\sqrt5}/\sqrt{10}\end{bmatrix}\begin{bmatrix}\frac{\sqrt5}{2}+\frac12 & 0 \\ 0 & \frac{\sqrt5}{2}-\frac12\end{bmatrix}\begin{bmatrix}\sqrt{5-\sqrt5}/\sqrt{10} & -\sqrt{5+\sqrt5}/\sqrt{10} \\ \sqrt2/\sqrt{5-\sqrt5} & \sqrt2/\sqrt{5+\sqrt5}\end{bmatrix}^T =$$

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